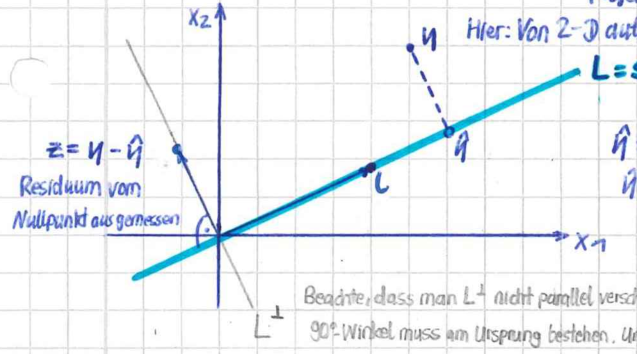
Orthogonale Projektion

Approximation im subspace →

Projektion ist immer eine Art **Dimensionsverlust**, gilt allgemein und nicht nur für orthogonale Proj.
Hier: Von 2-D auf Gerade (1D) projiziert (Unterraum L hat kleinere Dim als Oberraum V - außer triviales $L=V$)



$\hat{y}$: Projektion von y auf L, $\hat{y} \in L$
$\hat{y} = proj_L y$
$z = y - \hat{y}$: Residuum, $z \in L^\perp$

Beachte, dass man $L^\perp$ nicht parallel verschieben darf, sonst nicht mehr orthogonal zu L.
90°-Winkel muss am Ursprung bestehen. Unterraum L und orth. Komplement $L^\perp$ müssen beide durch Ursprung laufen

Andere Def. Orthogonale Projektion als beste Annäherung (Punkt in L zu y)
⇒ Suche Punkt $\in L$ mit minimalem Abstand zu y
Optimierungsproblem: „Approx.-fehler"
$$proj_L y = \underset{l \in L}{\operatorname{argmin}} \|y - l\|$$
äquivalent zu Zerlegungs-Theorem in Skalarproduktraum; ist aber auch außerhalb von VR gültig (eindeutig, jedoch nur auf konvexen Mengen!)

Wollen y (orthogonal) auf Unterraum L projizieren - - - orthogonale Zerlegung von y auf Richtung l [typischerweise ist l ein Einheitsvektor]

⇒ dazu zerlegen wir y in zwei Punkte / Vektoren
- ein Teil (Projektion $\hat{y}$) liegt im Unterraum $L = span\{l\}$, also $\in L$
- $z \perp \hat{y}$
- anderer Teil (Residuum $z = y - \hat{y}$) ist orthogonal zum Unterraum, also $\in L^\perp$

$y = \hat{y} + z$ "orthog. Kompon. von y"

- $z = y - \hat{y}$ ist orthogonal zu L
- Es muss gelten: $\exists c \in \mathbb{R}\; \hat{y} = cl$

Es ist $0 = \langle l, y - \hat{y}\rangle = \langle l, y\rangle - \langle l, \overset{\hat{y}}{cl}\rangle = \langle l, y\rangle - c\langle l, l\rangle$

Also gilt $c = \frac{\langle l, y\rangle}{\langle l, l\rangle}$ und $\hat{y} = proj_L y = \frac{\langle l, y\rangle}{\langle l, l\rangle} \cdot l$ — Falls l Einheitsvektor vereinfacht sich das und der Nenner = 1 fällt weg

z würde durch Proj.matrix auf Nullpunkt abgebildet werden $\hat{=}$ EV zum EW $\lambda = 0$
/ normiert auf Länge 1

Projektionstheorem: Sei L ein endlichdimensionaler Unterraum eines Skalarproduktraums V

Erweiterbar auf unendlich-dimensionale Unter-Räume
Dann muss V aber ein Hilbertraum sein $0\ \forall i \neq j$ orth. basis
$\langle \hat{v}, l_j\rangle = \langle \sum \frac{\langle v, l_i\rangle}{\langle l_i, l_i\rangle} l_i, l_j\rangle = \langle v, l_j\rangle - 0 \ldots - \frac{\langle v, l_j\rangle}{\langle l_j, l_j\rangle}\langle l_j, l_j\rangle - 0 \ldots$

Zerlegung: zwei zueinander orthogonale Vektoren

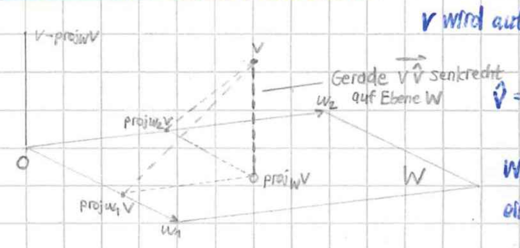
i) Jedes $v \in V$ kann *eindeutig* als $\underline{v = \hat{v} + z}$ dargestellt werden mit $\hat{v} \in L$ und $z\, (= v - \hat{v}) \in L^\perp$ $\langle z, l_j\rangle = \langle v - \hat{v}, l_j\rangle = \langle v, l_j\rangle - \langle \hat{v}, l_j\rangle = \langle v, l_j\rangle - \langle v, l_j\rangle$
$z \perp L$ "⇐" $= 0\ \forall j = 1, \ldots, k$

ii) Für jede Orthogonalbasis $\mathcal{L} = \{l_1, \ldots, l_k\}$ von L gilt: $\hat{v} = proj_L v = \underbrace{\boxed{\frac{\langle v, l_1\rangle}{\langle l_1, l_1\rangle}}}_{c_1} l_1 + \ldots \underbrace{\boxed{\frac{\langle v, l_k\rangle}{\langle l_k, l_k\rangle}}}_{c_k} l_k$ Linearkombination aus $\mathcal{L}$ ⇒ $\hat{v} \in L = span(\mathcal{L})$
Skalare
Jeder VR besitzt Orth. basis
↑ Orthogonalisierungsverfahren wie Gram-Schmidt
Projektion als Summe von Projektionen auf eindimens. UVR $\hat{=}$ einzelne Basisvektoren von L auf Unterraum L mit $dim(L) = k$

Beispiel im $\mathbb{R}^3$: Projektion auf Ebene

w_1 muss orthogonal zu w_2 sein, sonst Formel nicht anwendbar

v wird auf 2D-Ebene $\underline{W = span\{w_1, w_2\}}$ projiziert $\perp$



$$\hat{v} = proj_W v = proj_{w_1} v + proj_{w_2} v = \frac{\langle v, w_1\rangle}{\langle w_1, w_1\rangle} w_1 + \frac{\langle v, w_2\rangle}{\langle w_2, w_2\rangle} w_2$$

Vorteil mit Matrix: geht auch allgemein ohne Input-Vektor
Äquivalent zu Transf. mit Proj. matrix $U(U^TU)^{-1}U$, sofern Spalten(U) Orthog. basis bilden !

Weil $w_1 \perp w_2$ ist $proj_W v$ Summe d. $\perp$ Projektionen auf die einzelnen Unterräume $span\{w_1\}$ und $span\{w_2\}$

Projektionen sind lineare Operationen, die sich in euklidischen Räumen als Matrix-Vektor-Multiplikation darstellen lassen
Proj. matrix allgemein: quadr. Matrix, die (unter Dim. verlust) auf Unterraum projiziert $rang(P) = dim(col(P))$ als Dimension des UVR, auf den projiziert wird $< n$ (fast immer)

$P: \mathbb{R}^n \to \mathbb{R}^n$ bildet auf UVR ab aber NICHT in höhere/niedrigere Dimension ab

quadr. $P \in \mathbb{R}^{n \times n}$ nicht invertierbar, $det(P) = 0$ $col(P)$, in Richtung $ker(P)$ wird Information weggeworfen $\hat{=}$ alle Vektoren $\perp col(P)$

Eigenschaften von Projektionsmatrix × Idempotent, also $P^2 = P$ Logisch: Nach erster Projektion passiert nichts mehr

≠ Orthogonalmatrix!! (nur dann Proj.-matrix auf echten Unterraum)

× Bei orthogonaler Projektion zusätzlich Symmetrie, also $P = P^T$

und U^TU schwerer invertierbar $= P\vec{v}$
ergibt quadr. Projmatrix $P^{n \times n}$, die (L) auf $col(U)$ projiziert $col(P) = col(U)$

Konstruktion v. orthog. Projmatrix   dann aber ≠ Summe orth. Proj. (es reicht sogar lin. unabh. i.s.d. $(U^TU)^{-1}$ ex.)
Wenn $U \in \mathbb{R}^{n \times k}$ mit Spalten gleich Orth. basis für Unterraum $W \subset \mathbb{R}^n$ ⇒ $\boxed{proj_W \vec{v} = U(U^TU)^{-1}U^T\vec{v}}$
$k \leq n$, also nie mehr Spalten als Zeilen (Lin. Abh. ...)   ! U hat $\perp$ Spalten ⇒ U^TU ist Diagonalmatr. also einfach invertier...

siehe Bild oben: $E_0 = span\{z\}$, $E_1 = span\{l\}$

× Reelle Eigenwerte nur 0 oder 1   Sei $\vec{v}$ EV von P, also $Pv = \lambda v$ ⟺ $P(Pv) = \lambda v$ ⇒ $P(\lambda v) = \lambda v$ ⟺ $\lambda(Pv) = \lambda v$ ⇒ $\lambda(\lambda v) = \lambda v$, also $\lambda^2 v = \lambda v$, wobei $\vec{v} \neq \vec{0}$
$P^2 = P$ ⇒ $\lambda^2 = \lambda$ g.d.w. $\lambda \in \{0, 1\}$

$col(P) \oplus ker(P) = \mathbb{R}^n$ & $col(P^T) \oplus ker(P) = \mathbb{R}^n$ ⇏ $col(P) = col(P^T)$ nur falls $P = P^T$ s.o. $p \in col(P)$, $q \in ker(P)$ für Proj. matrix $p = Px$ ist Projektion von x in den Bildraum

□ $\underline{col(P) \oplus ker(P) = \mathbb{R}^n}$ Jedes $x \in \mathbb{R}^n$ lässt sich eindeutig so zerlegen: $x = p + q = Px + (x - Px)$ $q = x - Px$ liegt im Kern von P $P(x - Px) = Px - P^2x \overset{P = P^2}{=} Px - Px = 0$

Veranschaulicht inhärenten Dim. verlust durch Anwendung von P → Ein Teil wird in Zielraum projiziert, ein Teil „verschwindet" im Nullpunkt

Orthogonal. Projektion zudem $col(P) \perp ker(P)$, $p \perp q\ \forall x \in \mathbb{R}^n$
$U(U^TU)^{-1}U^T = UIU^T = UU^T = I_n$ auch wenn Spalten nur orth./lin. unabh. aber muss explizit rechnen
(↗ $U^{n \times n}$ quadratisch, orthonormale Spalten spannen $\mathbb{R}^n$!!)
(mit U orthonormale Spalten $\hat{=}$ orthonorm. basis von W)
Achtung: Wenn $U^{n \times k}$ $\hat{P} = UU^T$ $U(\underbrace{U^TU}_{I_n})^{-1}U^T$

Spezialfall: Projektionsmatrix & Orthogonalmatrix ⇒ $P = I_n$ $\overset{P^2 = P}{}$ $\overset{P^{-1} = P^T \overset{\perp Proj.}{=} P}{}$ $P = P^{-1} = P^2$ ⇒ $P^2 = P \cdot P^{-1} = I_n$ $(P^2 = P)$ ⇒ $P = I_n$
Triviale Projektion auf gesamten Raum, einzige invertierbare Proj. matrix (mit einzigem EW $\lambda = 1$)

[$U^TU = I_n$ dann obwohl U nicht quadratisch, also U keine Orthogonalmatrix (UU^T nicht Skalarpr. selbe Form)]

Beachte: Vektor $\vec{v} \in \mathbb{R}^n$ bleibt von (orthogonaler) Projektion unberührt immer, wenn v bereits im hineinzuprojizierenden Unterraum W liegt.
Hierzu zählt auch der Fall des trivialen Unterraums $W = \mathbb{R}^n$ selbst. Alle $w \in W$ sind EV der Projektionsmatrix zu $\lambda = 1$
Dann ist $proj_W(v) = v$ selbst $z = v - v = \vec{0}$ kein Residuum; $W = \mathbb{R}^n$ ⇒ $col(P) = \mathbb{R}^n$, $rang(P) = n$
$P = I_n$

! Wenn v bereits $\perp W$ ⇒ $\hat{v} = proj_W(v) = \vec{0}$, $z = v - 0 = v$

Orthogonalisierung

Ziel: Wollen aus beliebiger Basis von VR eine Orthogonalbasis machen, ohne den Spann zu verändern also ohne Inform. zu verlieren!
Dies geht immer! Daher wissen wir, dass jeder Vektorraum eine Orthogonalbasis hat.

<u>Verfahren: Gram-Schmidt</u> **Starten mit beliebiger** <u>**Menge** $\{v_1, v_2, \dots, v_n\}$</u> (Linear unabh. Vektoren / Basis) und erhalten am Ende <u>orthogonale **Menge** $\{u_1, \dots, u_n\}$</u> (/ Basis für span $\{v_1, v_2, \dots, v_n\}$)

1. $u_1 = v_1$ (den ersten Vektor behalten wir) **Weil** Startbasis /Startmenge linear unabh. $\Rightarrow \dots \operatorname{span}\{v_1, \dots, v_k\} = \operatorname{span}\{u_1, \dots, u_k\}$, $k = 1, 2, \dots, n$

2. $u_2 = v_2 - \operatorname{proj}_{\operatorname{span}\{u_1\}} v_2$ alle weiteren Basisvektoren ≙ Residuum der Projekt. auf aktuellen Gesamtspann
u_2: Wir ziehen von v_2 das ab, was bereits durch u_1 beschrieben wird, es bleibt übrig orthog. Residuum usw....

3. $u_3 = v_3 - \operatorname{proj}_{\operatorname{span}\{u_1, u_2\}} v_3$ Fall ein v_i linear abh. $\Rightarrow u_i = v_i - v_i = \vec{0}$, zwar ist $\{u_1, \dots, u_n\}$ ⊥ Menge aber Spann wurde "reduziert"! (Verfahren kann weiterlaufen, span ändert sich letztlich nicht)

⋮

Wollen wir nur Orthogonalbasis für span $\{v_1, \dots, v_k\}$, können wir Gram-Schmidt-Verfahren einfach vorzeitig abbrechen. **Jedes Zwischenergebnis** ist eine orthogonale Basis $\{u_1, \dots, u_k\}$ $k = 1, 2, \dots, n$

n. $u_n = v_n - \operatorname{proj}_{\operatorname{span}\{u_1, \dots, u_{n-1}\}}$

Insgesamt ändert sich dadurch der Spann nicht: $u_k = v_k - \sum_{j=1}^{k-1} \operatorname{proj}_{u_j}(v_k)$ ≙ Linearkombination von $v_1, \dots, v_k$; analog kann man v_k mit $u_1, \dots, u_k$ ausdrücken (weil Verfahren invertierbar ist)

Othonormalbasis bekommt man durch Normierung / Standardisierung der resultierenden orthogon. Basisvektoren u_i, also $u_i \cdot \frac{1}{\|u_i\|}$
Manchmal lohnt es sich, (einzelne) u_i zu normalisieren, sodass $\langle u_i, u_i \rangle = 1$ und Berechnung von Projektionen vereinfacht wird, manchmal können dadurch aber auch nervige Brüche Einzug erhalten => dann besser alles erst am Ende normieren

Die Reihenfolge, mit der man $v_1, v_2, \dots$ rauspickt, kann Ergebnis ändern, aber Resultat wird immer orthogonale Menge / Basis sein.

Gram-Schmidt kann genutzt werden, um aus linear abh. Menge abhängige Vektoren rauszufiltern, um am Ende <u>(linear unabh.) orthogonalbasis</u> (für span $\{v_1, \dots, v_n\}$) zu erhalten.
Dazu einfach alle Nullvektoren aus der resultierenden Vektorenmenge des Verfahrens eliminieren.

<u>Anwendung</u> | Linear abhängige Matrix-Spalten orthogonalisieren
z.B. <u>KQ:</u> X orthogonalisieren, um Projektionslösung zu ermöglichen (aber $(X^T X)$ invertierbar machen für indirekte Lösung geht so nicht, weil uns Spalten verloren gehen, also info's)

Weitere Anwendung: Spalten von Matrix (linear unabh.) orthogonal machen, spann-erhaltend

Orthogonale Diagonalisierung

(A quadr. und symmetr.)
Vereinfachung von A zu D ohne Strukturverlust
wesentliche Informationen, "math. Natur" wird bewahrt

quadr.
Matrix $A \in \mathbb{R}^{n\times n}$ orthogonal diagonalisierbar, wenn es Orthogonalmatrix P und Diagonalmatrix D gibt, sodass

$D = P^T A P$ ≙ Wechsel v. A in orthog. EV-Basis $(P^{-1} = P^T)$

$$P^T A P = D \quad \text{bzw.} \quad A = P D P^T \qquad P, P^T, D \in \mathbb{R}^{n\times n}$$

$P^T = P^{-1}$

Zudem Invarianz unter orthogonaler Transformation erhält EW (auch ✓ bei allgem. Diagonalis.), Symmetrie, Länge & ⊥ von Vektoren sowie (positive/negative /...) Definitheit

Orthogonale Diagonalisierung klappt ohne mühselige Berechnung von P^{-1} (P^{-1} gleich P^T) und zugleich kommen uns auch die weiteren schönen Eigenschaften von Orthogonalmatrizen zugute mit dieser Zerlegungsform.

(Symmetr. A auch orth. diagonal., wenn A nicht inv.bar) **!** quadr. Matrix
Wann? A orthogonal diagonalisierbar $\iff$ A symmetrisch

$D = D^T$ für Diagonalmatr. D "$\Leftarrow$" sehr unschöner Beweis

"$\Rightarrow$" A darstellbar als $A = PDP^T = PD^TP^T = (PDP^T)^T = A^T$ ergo A symmetr. □

PDP^T symmetr., auch P^TAP symmetr. & $A \to D$ symmetrie-erhaltend!

Sei A orthogonal diagonalisierbar & inv.bar $\Rightarrow A^{-1}$ orthog. diagon.bar
$A = PDP^T \Rightarrow A^{-1} = (PDP^T)^{-1} = (P^T)^{-1} D^{-1} P^{-1} = P \boxed{D^{-1}} P^T$
P Orth.matrix $(P^T)^{-1} = P$, $P^{-1} = P^T$
D^{-1} existiert, weil aus A inv.bar $\Rightarrow$ kein EW 0 $\quad D^{-1} = \begin{pmatrix} 1/\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_n \end{pmatrix}$

In der Statistik haben wir es häufig mit symmetrischen Matrizen zu tun, weshalb orthogonale Diagonalisierung in diesem Feld so bedeutend ist!

- Kovarianzmatrix (und Abwandlungen davon) ist symmetrische Matrix
- Hauptkomponentenanalyse (PCA) basiert auf orthogonaler Diagonalisierung der Kovarianzmatrix
  Hauptkomponenten in PCA sind ⊥ zueinander

Siehe KQ-Schätzer $\hat{\beta} = (X^TX)^{-1}X^Ty$
- X^TX ist immer symmetrisch $\forall X^{p\times n}$

Für symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ sind alle Eigenräume orthogonal zueinander
darunter alle Diagonalmatrizen(!) $\Rightarrow$ Es existiert eine Orthogonalbasis aus Eigenvektoren (die brauchen wir für Diagonalisierung)

verschiedene EW ≙ verschiedene Eigenräume
↳ bei n verschiedenen Eigenwerten geht (+) Diagonalis. sehr flott!

Seien λ_1, λ_2 verschiedene EW von A mit zugehörigen EV $\vec{v_1}, \vec{v_2}$ aus zwei Eigenräumen $E_{\lambda_1} \neq E_{\lambda_2}$. z.z. $v_1 \perp v_2$ also $\langle v_1, v_2\rangle = v_1^T v_2 = 0$

Starten mit Ausdruck: $\underbrace{\lambda_1 v_1^T v_2}_{(\lambda_1^T = \lambda_1)\ \lambda_1 v_1 = A v_1} = (Av_1)^T v_2 = v_1^T A^T v_2 \underset{A \text{ symm. } A = A^T}{=} v_1^T A v_2 = v_1^T(\lambda_2 v_2) = \lambda_2(v_1^T v_2)$
$\lambda_1 \neq \lambda_2 \Rightarrow v_1^T v_2 = 0$ □

- Außerdem hat symmetrische Matrix netterweise immer nur reelle Eigenwerte
- Zudem folgt aus $A^T = A$, dass $col(A) \perp ker(A)$ • $\langle Ax, y\rangle = (Ax)^T y = x^T A^T y \overset{(A=A^T)}{=} x^T A y = \langle x, Ay\rangle$
  Symmetr. A kann man in Skalarprod. "umherschieben"

Orthogonal diagonalisierbare, also symmetr. Matrizen bilden nur eine kleine Teilmenge aller diagonalis. Matrizen. Die Voraussetzungen sind viel strenger

Schritte zur orthogonalen Diagonalisierung einer symmetrischen Matrix

1. Finde Eigenwerte und zugehörige Eigenräume. Ermittele eine Basis für jeden Eigenraum von A $\quad \dim(E_{\lambda_i}) = 1 \ \forall \lambda_i$
   (entfällt bei n verschiedenen Eigenwerten $\Rightarrow$ alle E_λ = span{einzelnem Eigenvektor} ≙ Basis von E_λ
   Schritt 2. überflüssig bis auf Normierung)
2. Konstruiere für jeden Eigenraum eine Orthonormalbasis mithilfe von Gram-Schmidt-Verfahren und Normierung
   Alle diese (neuen) Basisvekt. ergeben Orthonormalbasis von $\mathbb{R}^n$
3. Schreibe Vektoren dieser Orthonormalbasen in die Spalten von P und passende Eigenwerte (in gleicher Reihenfolge) auf Diagonalen von D

(Orthogonale) Spektralzerlegung von symmetrischer Matrix:

$$A = P\Lambda P^T = \begin{pmatrix} | & & | \\ p_1 & \cdots & p_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} - p_1^T - \\ \vdots \\ - p_n^T - \end{pmatrix} = \sum_{i=1}^{n} \lambda_i p_i p_i^T$$

Siehe Def. von Matrixprodukt $AB = \sum_{i=1}^{n} a_i b^i$

orthonormale EV $p_1, \dots, p_n$ $\quad \Lambda = diag(\lambda_1, \dots, \lambda_n)$

Info aus nur einem Vektor (max. Rang 1)
Summe aus n-Matrizen $\lambda_i \underset{n\times 1}{p_i} \underset{1\times n}{p_i^T} \in \mathbb{R}^{n\times n}$
λ_i als Art Gewichtung, wie wichtig diese Matrix für Gesamtmatrix ist

$\|p_i\| = \sqrt{\langle p_i, p_i\rangle} = \sqrt{p_i^T p_i} = \sqrt{1} = 1 \Rightarrow \vec{x} \mapsto p_i p_i^T \vec{x}$ entspricht orthogonaler Projektion von Vektor $\vec{x}$ auf den Raum span{p_i}
normiert $\langle p_i, p_i\rangle = 1$

Proj.matrix (⊥ Proj.): Skalar 1, normiert
$proj_{span\{p_i\}} = p_i (p_i^T p_i)^{-1} p^T = p_i p_i^T$

alternativ) $\hat{x} = proj_{col(P)}\, x = \frac{\langle x, p_1\rangle}{\langle p_1, p_1\rangle = 1} p_1 + \dots + \frac{\langle x, p_n\rangle}{\langle p_n, p_n\rangle = 1} p_n = p_1\langle p_1, x\rangle + \dots + p_n\langle p_n, x\rangle = p_1 p_1^T x + \dots + p_n p_n^T x = \sum_{i=1}^{n} p_i p_i^T x$

Also insgesamt lässt sich Matrixwirkung als Projektion darstellen: $A\vec{x} = proj_{col(P)}\, \vec{x} = \sum_{i=1}^{n} \lambda_i \cdot proj_{span\{p_i\}} \vec{x}$
und damit als Summe obiger Projektionen

$\vec{x}$ wird **separat** auf p_i-Geraden projiziert, entlang dieser Geraden jeweils um λ_i gestreckt und dann die resultierenden Teil-Vektoren aufsummiert.

Symmetr. Matrix können wir also (in ihrer Wirkung) beschreiben, nur durch Kenntnis ihrer EW und orthogonalen EV!