Zusammenfassung der Großübung am 23. November

Normalengleichung

In der Vorlesung und den Tutorien behandelt man den Fall, dass die Systemmatrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ überbestimmt ist, also $n > m$ und $\mathrm{Rang}(\mathbf{A}) = m$ (voller Spaltenrang). In diesem Fall ist $\mathbf{A}^\mathsf{T}\mathbf{A}$ positiv definit, damit regulär und das LGS $\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{x} = \mathbf{A}^\mathsf{T}\mathbf{b}$ hat immer genau eine Lösung.

- Die Normalengleichung hat immer mindestens eine Lösung, egal wie $\mathbf{A}$ und $\mathbf{b}$ aussehen. Das kann man an der Herleitung dieser sehen: Man fordert, dass $\mathbf{x}$ so gewählt wird, dass $\hat{\mathbf{b}} = \mathbf{A}\mathbf{x}$ möglichst nah an $\mathbf{b}$ ist. Egal wie $\mathbf{A}$ und $\mathbf{b}$ aussehen, dieser „nächste" Vektor (oder Punkt) existiert in jedem Fall und ist eindeutig.
- Was nicht unbedingt eindeutig ist, ist wie man diesen nächsten Punkt $\hat{\mathbf{b}}$ mit den Spalten von $\mathbf{A}$ erreicht. Wir erinnern uns, dass $\mathrm{Kern}(\mathbf{A}^\mathsf{T}\mathbf{A}) = \mathrm{Kern}(\mathbf{A})$. Falls $\mathbf{A}$ also einen nicht-trivialen Kern hat, hat $\hat{\mathbf{b}}$ unendlich viele Darstellungen bezüglich der Spalten von $\mathbf{A}$ und damit die Normalengleichung unendlich viele Lösungen.
- Ergo: Falls $\mathrm{Kern}(\mathbf{A}) = \{\mathbf{0}\}$, so hat $\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{x} = \mathbf{A}^\mathsf{T}\mathbf{b}$, egal ob $\mathbf{A}\mathbf{x} = \mathbf{b}$ keine oder genau eine Lösung hat, immer genau eine Lösung. Falls $\mathrm{Kern}(\mathbf{A}) \neq \{\mathbf{0}\}$, so hat die Normalengleichung unendlich viele Lösungen.

Orthogonale Projektion

Zunächst einige Definitionen:

- Eine Matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ heißt Projektion, falls sie idempotent ist, d.h. $\mathbf{P}^2 = \mathbf{P}$.
- Eine Projektion $\mathbf{P}$ heißt orthogonale Projektion, falls sie zusätzlich symmetrisch ist: $\mathbf{P}^\mathsf{T} = \mathbf{P}$ (Achtung: Orthogonale Projektionen sind im Allgemeinen nicht orthogonal, unter anderem, da sie nicht regulär sein müssen! Der Begriff „orthogonal" bezieht sich darauf, dass ihr Kern und ihr Bild das orthogonale Komplement des jeweils anderen sind. Dazu unten mehr.).

Aus der Normalengleichung kann man sich die orthogonale Projektion auf den Spaltenraum einer Matrix $\mathbf{A}$ definieren, falls ihre Spalten linear unabhängig sind:

$$\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{x} = \mathbf{A}^\mathsf{T}\mathbf{b} \Leftrightarrow \mathbf{x} = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\mathbf{b} \Leftrightarrow \underbrace{\mathbf{A}\mathbf{x}}_{\mathbf{P_A}\mathbf{b}} = \underbrace{\mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}}_{\mathbf{P_A}}\mathbf{b}.$$

Eine Studentin fragte, ob das nicht einfach die Identitätsmatrix ist. Schließlich könnte man das so umformen:

$$\mathbf{P_A} = \mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T} \stackrel{?}{=} \underbrace{\mathbf{A}\mathbf{A}^{-1}}_{\mathbf{I}}\underbrace{\mathbf{A}^{-\mathsf{T}}\mathbf{A}^\mathsf{T}}_{\mathbf{I}} = \mathbf{I}.$$

Ein guter Punkt, der ein häufiges Missverständnis enthält: Die Matrix $\mathbf{A}$ muss nicht regulär sein. Falls sie regulär ist, ihre Spalten den gesamten $\mathbb{R}^n$ aufspannen, dann ist die orthogonale Projektion tatsächlich die Identität. Dies ergibt auch Sinn: Da jeder Punkt im Spann dieser Matrix ist, wird er auf sich selbst projiziert. Bevor wir dies vertiefen, wollen wir ein kleines Rechenbeispiel durchführen:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 0 & 0 \end{bmatrix}$$

Nun wollen wir die zugehörige orthogonale Projektion $\mathbf{P_A}$ berechnen.

$$\begin{aligned}
\mathbf{P_A} &= \begin{bmatrix}1 & 3\\ 2 & 4\\ 0 & 0\end{bmatrix} \left(\begin{bmatrix}1 & 2 & 0\\ 3 & 4 & 0\end{bmatrix} \begin{bmatrix}1 & 3\\ 2 & 4\\ 0 & 0\end{bmatrix} \right)^{-1} \begin{bmatrix}1 & 2 & 0\\ 3 & 4 & 0\end{bmatrix}\\
&= \begin{bmatrix}1 & 3\\ 2 & 4\\ 0 & 0\end{bmatrix} \begin{bmatrix}5 & 11\\ 11 & 25\end{bmatrix}^{-1} \begin{bmatrix}1 & 2 & 0\\ 3 & 4 & 0\end{bmatrix}\\
&= \frac{1}{4} \begin{bmatrix}1 & 3\\ 2 & 4\\ 0 & 0\end{bmatrix} \begin{bmatrix}25 & -11\\ -11 & 4\end{bmatrix} \begin{bmatrix}1 & 2 & 0\\ 3 & 4 & 0\end{bmatrix}\\
&= \frac{1}{4} \begin{bmatrix}-8 & 4\\ 6 & -2\end{bmatrix} \begin{bmatrix}1 & 2 & 0\\ 3 & 4 & 0\end{bmatrix}\\
&= \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix}
\end{aligned}$$

Das Ergebnis ist nicht gerade überraschend, da die Spalten von $\mathbf{A}$ die x-y-Ebene aufspannen. Die Projektion erfolgt dann dadurch, dass man die z-Koordinate auf Null setzt. Aber Achtung: Dass die Projektionsmatrix so ähnlich wie die Identität aussieht, ist purer Zufall, bzw. kommt dadurch zustande, dass der Unterraum die x-y-Ebene ist. Man überzeuge sich davon, dass dies nicht immer so sein muss, indem man $\mathbf{P_B}$ für

$$\mathbf{B} = \begin{bmatrix}1 & 1\\ 1 & 0\\ 0 & 1\end{bmatrix}$$

berechnet.

Nun zeigen wir, dass $\mathbf{P_A}$ eine orthogonale Projektion ist. Sei $\mathbf{A} \in \mathbb{R}^{n\times k}$ mit $k \leq n$ und $\mathrm{Rang}(\mathbf{A}) = k$ eine Matrix, deren Spalten linear unabhängig sind und eine Basis eines linearen Unterraumes des $\mathbb{R}^n$ bilden. Es gilt:

- Idempotent: $\mathbf{P_A^2} = \mathbf{A}\cancel{(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{A}}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top = \mathbf{P_A}$
- Symmetrisch: $\mathbf{P_A^\top} = (\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top)^\top = {\mathbf{A}^\top}^\top(\mathbf{A}^\top\mathbf{A})^{-\top}\mathbf{A}^\top = \mathbf{A}((\mathbf{A}^\top\mathbf{A})^\top)^{-1}\mathbf{A}^\top$
  $= \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top = \mathbf{P_A}$ (wir wissen ja, dass $\mathbf{A}^\top\mathbf{A}$ symmetrisch ist).

Zunächst ein paar Eigenschaften von allgemeinen Projektionen $\mathbf{P}$:

- Sie haben nur die Eigenwerte 0 und 1. Sei $\mathbf{v}$ ein Eigenvektor von $\mathbf{P}$ zum Eigenwert λ. Es gilt:

  $$\lambda\mathbf{v} = \mathbf{P}\mathbf{v} = \mathbf{P}^2\mathbf{v} = \lambda\mathbf{P}\mathbf{v} = \lambda^2\mathbf{v}$$

  Aus $\lambda^2 = \lambda$ folgt, dass λ nur Eins oder Null sein darf.
- Punkte in ihrem Bild sind Fixpunkte: Falls $\mathbf{x} \in \mathrm{Bild}(\mathbf{P})$, dann ist $\mathbf{P}\mathbf{x} = \mathbf{x}$. Dies folgt unmittelbar daraus, dass $\mathbf{P}^2 = \mathbf{P}$. Fixpunkte einer linearen Abbildung sind übrigens Eigenvektoren zum Eigenwert $\lambda = 1$.

Und was passiert bei der orthogonalen Projektion eines Punktes $\mathbf{x}$, der nicht im Bild von $\mathbf{P}$ liegt? Sie werden auf den Punkt $\mathbf{y} = \mathbf{P}\mathbf{x}$ im Bild von $\mathbf{P}$ projiziert, der möglichst nah (bzgl. euklidischer Norm) zu $\mathbf{x}$ ist. Dies gilt aber nur dann, wenn $\mathbf{P}$ zusätzlich symmetrisch ist.

Ab hier wird es zu fortgeschritten

Wir schauen uns zum Schluss zwei Eigenschaften zur orthogonalen Projektion an:

- Der Kern von $\mathbf{P_A}$ ist das orthogonale Komplement vom Spaltenraum von $\mathbf{A}$. Das lässt sich sowohl zeichnerisch als auch rechnerisch überprüfen. Das orthogonale Komplement $V^\perp$ eines linearen Unterraumes $V \subseteq \mathbb{R}^n$ ist ebenfalls ein linearer Unterraum, der alle Vektoren enthält, die zu allen Vektoren in V orthogonal sind:

  $$V^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^\top\mathbf{y} = 0,\ \forall \mathbf{y} \in V\}. \tag{1}$$

Falls V die Dimension k hat, hat $V^\perp$ die Dimension $n-k$ (dies entspricht der Kodimension von V). Beispiele:

- Das orthogonale Komplement einer Geraden im $\mathbb{R}^2$ ist eine Gerade.
- Das orthogonale Komplement einer Ebene im $\mathbb{R}^3$ ist eine Gerade.
- Das orthogonale Komplement einer Geraden im $\mathbb{R}^3$ ist eine Ebene.
- Das orthogonale Komplement des $\mathbb{R}^3$ im $\mathbb{R}^3$ ist der Nullvektor.

Die Projektionsmatrix $\mathbf{P}_\mathbf{A}^\perp$ auf das orthogonale Komplement des Bildes von $\mathbf{A}$ hat eine zu schöne Darstellung, um sie euch vorzuenthalten. Es gilt $\mathbf{P}_\mathbf{A}^\perp = \mathbf{I} - \mathbf{P}_\mathbf{A}$. Dies lässt sich zum Beispiel dadurch begründen, dass die Matrix $[\mathbf{A}, \mathbf{A}^\perp]$ regulär ist, wobei $\mathbf{A}^\perp$ in ihren Spalten eine Basis des orthogonalen Komplements von $\mathrm{Bild}(\mathbf{A})$ enthält. Diese Formel ist insofern interessant, als dass man in diesem Kurs die Addition von Matrizen nur selten zu Gesicht bekommt.

Für die Interessierten: Diese Formel lässt sich auch mit der Eigenzerlegung von $\mathbf{P}_\mathbf{A}$ begründen, und hat eine interessante Verbindung zur Pseudoinversen $\mathbf{A}^+$.

Ein Anwendungsfall dafür ist bekannt aus dem Gram-Schmidt-Verfahren zum orthogonalisieren einer Basis. Zuerst verinnerlichen wir uns folgende Vereinfachung: Falls die Spalten von $\mathbf{A}$ orthonormal sind, also $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$, dann lässt sich $\mathbf{P}_\mathbf{A}$ vereinfachen zu $\mathbf{P}_\mathbf{A} = \mathbf{A}\mathbf{A}^\mathsf{T}$. Die orthogonale Projektion für $\mathrm{Bild}(\mathbf{A})^\perp$ ist $\mathbf{P}_\mathbf{A}^\perp = \mathbf{I} - \mathbf{A}\mathbf{A}^\mathsf{T}$.

Man möchte zum Beispiel die Spalten von $\mathbf{B}$ (oben) orthogonalisieren. Zunächst normiert man die erste Spalte:

$$\mathbf{b}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Nun berechnet man die orthogonale Projektion auf das orthogonale Komplement des Vektors:

$$\mathbf{P}_{\mathbf{b}_1}^\perp = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Nun projiziert man die zweite Spalte in das orthogonale Komplement und normiert das Ergebnis:

$$\tilde{\mathbf{b}_2} = \begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}, \quad \mathbf{b}_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

Damit ist $[\mathbf{b}_1, \mathbf{b}_2]$ eine Matrix mit orthonormalen Spalten, die dasselbe Bild wie $\mathbf{B}$ hat.

- Die orthogonale Projektion ist unabhängig davon, wie man die Basis wählt. Sei $\mathbf{B} \in \mathbb{R}^{n \times k}$ eine Matrix, deren Spalten denselben Spaltenraum wie $\mathbf{A}$ aufspannen, d.h. es gibt eine reguläre Matrix $\mathbf{C} \in \mathbb{R}^{k \times k}$, sodass $\mathbf{B} = \mathbf{A}\mathbf{C}$. Es gilt:

$$\begin{aligned} \mathbf{P}_\mathbf{B} &= \mathbf{B}(\mathbf{B}^\mathsf{T}\mathbf{B})^{-1}\mathbf{B}^\mathsf{T} \\ &= \mathbf{A}\mathbf{C}((\mathbf{A}\mathbf{C})^\mathsf{T}\mathbf{A}\mathbf{C})^{-1}(\mathbf{A}\mathbf{C})^\mathsf{T} \\ &= \mathbf{A}\mathbf{C}(\mathbf{C}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{C})^{-1}\mathbf{C}^\mathsf{T}\mathbf{A}^\mathsf{T} \\ &= \mathbf{A}\mathbf{C}\mathbf{C}^{-1}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{C}^{-\mathsf{T}}\mathbf{C}^\mathsf{T}\mathbf{A}^\mathsf{T} \\ &= \mathbf{A}(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T} \\ &= \mathbf{P}_\mathbf{A} \end{aligned}$$

Dies lässt sich auch anekdotisch überprüfen. Man berechnet $\mathbf{P}_\mathbf{D}$ für

$$\mathbf{D} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

und vergleicht es mit der von uns berechneten $\mathbf{P}_\mathbf{A}$ oben. Es ist offensichtlich, dass die Spalten von $\mathbf{D}$ ebenfalls die x-y-Ebene aufspannen.