BESTAPPROXIMIERENDE LÖSUNG VON LINEAREN GLEICHUNGSSYSTEMEN

Wir behandeln das Thema aus dem Vorlesungsskriptum Kapitel 5, Unterkapitel 5, 6 und 7 und erarbeiten uns die dafür benötigten mathematischen Kenntnisse auf unsere eigene Art und Weise, in etwas verkürzter Form. Querverweise auf die entsprechenden Sätze im Vorlesungsskriptum werden gegeben. Damit kann der folgende Stoff auch rein über das Vorlesungsskriptum erarbeitet werden. Dabei nimmt man aber einiges an Zusatzstoff auf sich, den ich bewusst auslasse.

1. Gleichungssysteme

Wir interessieren uns für lineare Gleichunssysteme der Form $A \cdot x = b$ wobei $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$ die aber nicht unbedingt lösbar sind. Dies tritt in der Praxis häufig auf, insbesondere dann, wenn mehr Gleichungen (etwa als Ergebnis von mehrfach wiederholten physikalischen Messprozessen) vorhanden sind als Unbekannte.
(Eine Mehrfachmessung kann der Minimierung von Messfehlern dienen.)

Hilfssatz: Sei $x \in \mathbb{R}^n$ so, dass $A^T \cdot (Ax - b) = 0$, dann gilt für alle $y \in \mathbb{R}^n$: $||Ay - b||^2 \geq ||Ax - b||^2$ und damit auch $||Ay - b|| \geq ||Ax - b||$.
Beweis: Wir halten zunächst fest: Durch Transponieren der Identität $A^T \cdot (Ax - b) = 0$ sieht man, dass $(Ax - b)^T \cdot A = 0$ (wir verwenden hierzu, dass $0^T = 0$,$(A \cdot B)^T = B^T \cdot A^T$ und $(A^T)^T = A$, für entsprechende Matrizen A, B).
Nun verwenden wir die Tatsache, dass sich $||v||^2$ eines Vektors $v \in \mathbb{R}^n$ mit Hilfe einer Matrizenmultiplikation schreiben lässt als $||v||^2 = v^T \cdot v$ (der Zeilenvektor v wir also mit dem Spaltenvektor v multipliziert).
Nun gilt $||Ay - b||^2 = ||(Ay - Ax) + (Ax - b)||^2 = ((Ay - Ax) + (Ax - b))^T \cdot ((Ay - Ax) + (Ax - b))$.
Letzterer Ausdruck wir unter Verwendung der Distributivitätsgesetze ausmultipliziert und wir erhalten: $||Ay - b||^2 = (Ay - Ax)^T (Ay - Ax) + (Ay - Ax)^T (Ax - b) + (Ax - b)^T (Ay - Ax) + (Ax - b)^T (Ax - b)$. Nun gilt $(Ay - Ax)^T (Ax - b) = (y - x)^T A^T (Ax - b) = 0$ laut Voraussetzung und genauso $(Ax - b)^T (Ay - Ax) = (Ax - b)^T A(y - x) = 0$ laut Voraussetzung. Es folgt also insgesamt: $||Ay - b||^2 = (Ay - Ax)^T (Ay - Ax) + (Ax - b)^T (Ax - b) = ||Ay - Ax||^2 + ||Ax - b||^2 \geq ||Ax - b||^2$. Durch Wurzelziehen erhält man dann auch $||Ay - b|| \geq ||Ax - b||$.

Der Hilfssatz sagt in anderen Worten formuliert, dass unser Vektor x welcher der Vorraussetzung $A^T \cdot (Ax - b) = 0$ genügt, jener Vektor ist, sodass Ax von b die geringste Abweichung hat. Etwas mathematischer formuliert bedeutet das, dass $||Ax - b||$ minimal ist.

Definition (Definition 5.27 aus dem Skript): Sei $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Eine bestapproximierende Lösung des Gleichungssystems $A \cdot x = b$ ist ein $x^\star \in \mathbb{R}^n$, sodass $||Ax^\star - b||$ minimal ist, also für alle $y \in \mathbb{R}^n$: $||Ay - b|| \geq ||Ax^\star - b||$.

Wie findet man diesen Vektor $x^\star$ praktisch? Nun, oft gehts ganz einfach.

Satz (Satz 5.28 aus dem Vorlesungsskriptum, etwas anders formuliert): Sei A ein Matrix, so dass $A^T \cdot A$ invertierbar ist. Dann ist die bestapproximierende Lösung $x^\star$ von $A \cdot x = b$ gegeben durch $x^\star = (A^T \cdot A)^{-1} \cdot A^T \cdot b$.
Beweis: Wir brauchen nur den Hilfssatz verwenden und geeignet umformen. Wenn $A^T \cdot (Ax^\star - b) = 0$ gilt, so hat $x^\star$ die gewünschte Eigenschaft. Ausmultiplizieren der Klammer und Umstellen der Gleichung ergibt: $A^T \cdot A \cdot x^\star = A^T \cdot b$. Also folgt durch Invertieren $x^\star = (A^T \cdot A)^{-1} \cdot A^T \cdot b$.

Beachten Sie: Falls A invertierbar und das Gleichungssystem $Ax = b$ damit exakt lösbar ist mit $x = A^{-1} \cdot b$, dann folgt $x^\star = (A^T \cdot A)^{-1} \cdot A^T \cdot b = A^{-1} \cdot (A^T)^{-1} \cdot A^T \cdot b = A^{-1} \cdot b$. $x^\star$ ist dann also tatsächlich die exakte Lösung.
Weiters gilt, falls $Ax = b$ ein Gleichungssystem mit unendlich vielen Lösungen ist, so liefert $x^\star$ nur eine Lösung.
Die Berechnung von $x^\star$ nach obiger Methode macht also nur bei nicht lösbaren Gleichungssystemen einen wirklichen Sinn.
Bleibt noch offen wann für eine Matrix A, die Matrix $A^T \cdot A$ invertierbar ist. Ohne Beweis gilt:

Satz 2 (Lemma 5.17 aus dem Vorlesungsskript): Sind die Spalten einer Matrix A linear unabhängig, so ist $A^T \cdot A$ invertierbar.

Beispiele:
Bestimmung einer bestapproximierenden Geraden durch eine Punktwolke (siehe Vorlesungsskriptum Übungsaufgaben 5.31).

4 Gegenstände sollen durch sechsfaches Messen mit einer Apothekerwaage gewogen werden, um Messfehler zu bekämpfen. Zuerst wiegt man die Gegenstände einzeln, dann die Summe der ersten drei, dann die Summe der letzten drei. Die Messungen ergeben in g: $x_1 = 14.66$, $x_2 = 10.47$, $x_3 = 11.12$, $x_4 = 9.99$, $x_1 + x_2 + x_3 = 36.75$, $x_2 + x_3 + x_4 = 31.41$. Man bestimme die beste Schätzung für die tatsächlichen Massen der Gegenstände.

Lösung: Wir erhalten ein Gleichungssystem $Ax = b$ mit $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$ und $b = \begin{pmatrix} 14.66 \\ 10.47 \\ 11.12 \\ 9.99 \\ 36.75 \\ 31.41 \end{pmatrix}$. Folgende Berechnungen werden am besten mit einem Computeralgebrasystem durchgeführt:

$A^TA = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 3 & 2 & 1 \\ 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix}$, A^TA erweist sich als invertierbar und man berechnet
$x^\star = \begin{pmatrix} 14.855 \\ 10.525 \\ 11.175 \\ 9.85 \end{pmatrix}$.

2. Minimaler Abstand eines Vektors von einem Unterraum

Wir nehmen an wir haben einen Unterraum $U = L(b_1,\dots,b_n)$ des $\mathbb{R}^k$ gegeben, wobei $(b_1,\dots,b_n)$ eine Basis von U ist und $k \in \mathbb{N}$. Weiters sei ein Vektor $b \in \mathbb{R}^k$ gegeben. Gesucht ist der kürzeste Abstand von b zu U und jener Vektor in U der den kürzesten Abstand zu b hat. Wir machen folgende Überlegung:
Wir bilden die Matrix B in deren Spalten die Basisvektoren von U stehen, also $B = (b_1,\dots,b_n)_{Matrix}$. Dann gilt $U = L(b_1,\dots,b_n) = \{B \cdot x | x \in \mathbb{R}^n\}$. U ist also der Spaltenraum der Matrix B. Sei $v \in U$ so, dass v kürzesten Abstand zu b hat. Dann gilt $v = B \cdot x$ für ein geeignetes x und $||v-b|| = ||B \cdot x - b||$ ist minimal. Wir sind also nun in exakt der gleichen Situation wie bei den Gleichungssystemen und suchen $x^\star$ so, dass $x^\star$ bestapproximierende Lösung des Gleichungssystems $||B \cdot x - b||$ ist. Wir wissen, dass $x^\star = (B^T \cdot B)^{-1} \cdot B^T \cdot b$ jener Vektor in $\mathbb{R}^n$ ist, so dass der Abstand zwischen $B \cdot x^\star$ und b minimiert wird. Wir wissen wegen Hilfssatz 2 (Lemma 5.17) auch, dass $(B^T \cdot B)^{-1}$ existiert. Die Spalten der Matrix B sind ja die Basisvektoren von U und damit linear unabhängig.
Der entsprechende Vektor $v \in U$, sodass $||v-b||$ minimal ist, ist also $v = B \cdot x^\star = B \cdot (B^T \cdot B)^{-1} \cdot B^T \cdot b$.
v nennt man den Projektionsvektor von b auf U und $B \cdot (B^T \cdot B)^{-1} \cdot B^T$ die Projektionsmatrix.

Wir haben damit folgenden Satz bewiesen:
Satz 3 (Satz 5.23 aus dem Vorlesungsskriptum in anderer Formulierung):
Sei $U = L(b_1,\dots,b_n)$ ein Unterraum des $\mathbb{R}^k$ wobei $(b_1,\dots,b_n)$ eine Basis von U ist. Sei $b \in \mathbb{R}^k$ beliebig. Jener Vektor $v \in U$ mit kürzestem Abstand zu b, ist gegeben durch $v = B \cdot x^\star = B \cdot (B^T \cdot B)^{-1} \cdot B^T \cdot b$, wobei B die Matrix ist in deren Spalten die Basisvektoren von U stehen.

Bemerkung: Der kürzeste Abstand ist dann gegeben durch $||v-b||$, mit $v = B \cdot x^\star$.

3. Minimaler Abstand eines Vektors von einer linearen Mannigfaltigkeit

Hinweis: Im Vorlesungsskriptum von Herrn Aichinger wird dieses Thema wesentlich umfassender behandelt, nämlich als Abstandsberechnung zwischen 2 linearen Mannigfaltigkeiten.

Sei $U = L(b_1,\dots,b_n)$ ein Unterraum des $\mathbb{R}^k$, wobei $(b_1,\dots,b_n)$ eine Basis von U ist. Sei $p \in \mathbb{R}^k$. Die Menge $p + U := \{p + u | u \in U\}$ heißt eine lineare Mannigfaltigkeit von U. $p + U$ ist im allgemeinen kein Unterraum mehr (ausser $p \in U$, dann ist aber $p+U=U$), da etwa der Nullvektor nicht mehr in $p+U$ liegt. Klassische Beispiele für lineare Mannigfaltigkeiten sind Geraden im $\mathbb{R}^2$ bzw. Ebenen im $\mathbb{R}^3$ die nicht durch den Koordinatenursprung gehen. Solche Geraden bzw. Ebenen kann man sich vorstellen als Geraden bzw. Ebenen die durch Verschiebung um einen Vektor p einer Geraden / Ebene durch den Ursprung entstehen. Im Falle dass eine Gerade bzw. Ebene in parametrisierter Form angegeben ist, ist diese Tatsache leicht zu sehen.

Beispiele: $g : X = \begin{pmatrix} 4 \\ 3 \end{pmatrix} + \lambda \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ist eine Gerade in $\mathbb{R}^2$ die nicht durch den Nullpunkt geht. Sie lässt sich darstellen als, $g = p+U$ wobei $p = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ und $U = L(\begin{pmatrix} 1 \\ 1 \end{pmatrix})$, die Gerade durch Null mit Richtungsvektor $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
$e : X = \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix} + \lambda \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \mu \cdot \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$ ist eine Ebene in $\mathbb{R}^3$ die nicht durch den Nullpunkt geht. Sie lässt sich darstellen als, $e = p+U$ wobei $p = \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}$ und $U = L(\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix})$, die Ebene durch Null mit Richtungsvektoren $\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$.

Wir suchen also nun den kürzesten Abstand eines Vektors b von der linearen Mannigfaltigkeit $p+U$ und den Vektor $v \in p+U$ der diesen kürzesten Abstand zu b hat. Die Situation ist wieder ganz ähnlich zu vorher, wir fassen U wieder als Spaltenraum der Matrix B auf in deren Spalten die Basisvektoren aus U stehen. Sei $v \in p+U$ so, dass v kürzesten Abstand zu b hat. Dann gilt $v = p + B \cdot x$ für ein geeignetes x und $||v-b|| = ||p+B\cdot x - b||$ ist minimal. Anstatt $||p+B\cdot x-b||$ schreiben wir nun $||B\cdot x-(b-p)||$. Wir sind also nun wieder in exakt der gleichen Situation wie bei den Gleichungssystemen. Wir wissen, dass $x^\star = (B^T \cdot B)^{-1} \cdot B^T \cdot (b-p)$ jener Vektor in $\mathbb{R}^n$ ist, so dass der Abstand zwischen $B \cdot x^\star$ und $(b-p)$ minimal ist. Wir wissen wegen Hilfssatz 2 (Lemma 5.17) auch, dass $(B^T \cdot B)^{-1}$ existiert. Die Spalten der Matrix B sind ja die Basisvektoren von U und damit linear unabhängig.
Wir erhalten also nun, dass $p + B\cdot x^\star = p + B \cdot (B^T \cdot B)^{-1} \cdot B^T \cdot (b-p)$ der gesuchte Vektor in $p+U$ mit kürzestem Abstand zu b ist. Der kürzeste Abstand selbst ist $||p+B\cdot x^\star - b||$. Falls $p = \vec{0}$ erhalten wir natürlich wieder die Formel aus Satz 3.

Satz 4 (Seite 95 bis 97 aus dem Vorlesungsskriptum in anderer Formulierung): Sei $U = L(b_1, \ldots, b_n)$ ein Unterraum des $\mathbb{R}^k$ wobei $(b_1, \ldots, b_n)$ eine Basis von U ist. Seien $b, p \in \mathbb{R}^k$ beliebig. Jener Vektor $v \in p+U$ mit kürzestem Abstand zu b, ist gegeben durch $v = p + B \cdot (B^T \cdot B)^{-1} \cdot B^T \cdot (b-p)$.

Bemerkung: Man kann zeigen, dass es sich bei den kürzesten Abständen im Falle dass $p+U$ eine Gerade im $\mathbb{R}^2$ bzw. eine Ebene im $\mathbb{R}^3$ ist um die altbekannten und in der Vorlesung bereits hergeleiteten Normalabstände handelt. Der Nachweis dafür entfällt allerdings.