Anhang B: Pseudoinverse, Projektor–Matrix und pseudoinverse Kopplungsmatrix

Die Pseudoinverse A^Ψ einer Matrix A kann als verallgemeinerte Inverse des Matrizenprodukts aufgefaßt werden. Ein einfacher Zugang erfolgt über die Singular Value Decomposition, welche eine explizite Konstruktion der Pseudoinversen liefert. Die aus der Pseudoinversen abgeleitete Projektormatrix besitzt für die Analyse der Dynamik Neuronaler Netzwerke günstige analytische Eigenschaften.

B.1 Definition und Konstruktion der Pseudoinversen

B.1.1 Definition der Pseudoinversen

Für beliebige Matrizen A kann die verallgemeinerte Inverse oder Pseudoinverse A^Ψ durch folgende vier Gleichungen definiert werden [Pe55]:

$$AA^\Psi A = A \tag{B.1}$$
$$A^\Psi AA^\Psi = A^\Psi \tag{B.2}$$
$$(AA^\Psi)^T = AA^\Psi \tag{B.3}$$
$$(A^\Psi A)^T = A^\Psi A \tag{B.4}$$

Die Gleichungen (B.1) bis (B.4) legen, bei gegebener Matrix A, deren Pseudoinverse A^Ψ eindeutig fest.

Mit Hilfe der Pseudoinversen A^Ψ ist es möglich, *immer* eine Lösung (oder zumindest eine wohldefinierte Näherungslösung) x_0 des linearen Gleichungssystems

$$Ax = b \tag{B.5}$$

zu finden, unabhängig davon, ob das Gleichungssystem lösbar, über– oder unterbestimmt ist. Der Lösungsvektor (oder die Näherungslösung) ist immer durch

$$x_0 = A^\Psi b \tag{B.6}$$

gegeben. Für den Fall, daß das Gleichungsystem (B.5) eindeutig lösbar ist, reduziert sich die Pseudoinverse A^Ψ auf die normale Inverse A^{-1} der Matrix A; wir haben dann, wie üblich, $x_0 = A^{-1}b$. Ist das Gleichungsystem (B.5) hingegen unterbestimmt, so liefert $x_0 = A^\Psi b$ aus der Menge der Lösungen den Lösungsvektor mit der kleinsten Norm $|x_0|$; falls das Gleichungsystem überbestimmt ist, minimiert die Näherungslösung x_0 den Fehler $f = |Ax_0 - b|$.

B.1.2 Singular Value Decomposition (SVD)

Die Eigenschaften der Pseudoinversen bezüglich der Lösung linearer Gleichungssysteme lassen sich am elegantesten durch die Anwendung der Singular Value Decomposition (SVD) zeigen [Pr86]. Die SVD ist eine Zerlegung einer vorgegebenen Matrix A in drei spezielle Matrizen. Jede $n \times m$–Matrix A kann nämlich als Produkt der folgenden drei Matrizen geschrieben werden:

a) einer $m \times n$–Matrix U, mit orthogonalen Spalten:

$$\sum_i u_{ik} u_{il} = \delta_{kl} \ ,$$

b) einer $n \times n$–Diagonal–Matrix

$$W = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \ ,$$

wobei die Ordnung der nichtsingulären Diagonal–Matrix D gleich dem Rang der Matrix A ist,

c) und einer $n \times n$–Orthogonalmatrix V.

Wir haben damit die Dekomposition

$$A = UWV^T = U \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V^T$$

der Matrix A vorliegen; dies definiert die SVD. Hierbei gilt für die Matrizen U und V, daß

$$U^T U = V^T V = V V^T = \mathbb{1} \, .$$

Es sei angemerkt, daß die SVD nicht eindeutig ist. Bei gegebener SVD können beliebige Permutationen der Spalten der Matrizen U und V durchgeführt werden, falls die entsprechenden Diagonalelemente von W mitpermutiert werden. Ferner können noch zwischen den Spalten der Matrizen U und V beliebige Linearkombinationen gebildet werden, wenn die dazu korrespondierenden Elemente von W untereinander identisch sind.

Liegt nun aber eine SVD der Matrix A vor, so kann deren Pseudoinverse durch

$$\begin{aligned} A^\Psi &= V \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T \\ &\equiv V W^{-1} U^T \end{aligned} \tag{B.7}$$

dargestellt werden. Wie man leicht nachrechnet, erfüllt die so definierte Matrix A^Ψ die obigen Definitionsgleichungen der Pseudoinversen. Die Nichteindeutigkeit der SVD spielt hierbei keine Rolle, da nur eine Matrix, nämlich die Pseudoinverse A^Ψ, alle Definitionsgleichungen (B.1) bis (B.4) erfüllt.

B.1.3 Eigenschaften der Pseudoinversen

Ausgehend von einer SVD–Darstellung der Matrizen A und A^Ψ lassen sich die optimalen Eigenschaften der Pseuodinversen zur Lösung linearer Gleichungssysteme zeigen. Betrachten wir dazu die durch die Matrix A induzierte, lineare Abbildung

$$x \in \mathbb{R}^n \rightarrow y \in \mathbb{R}^m \quad : \quad y = Ax = (UWV^T)x \ .$$

Offensichtlich spannen die Spalten der Matrix U, deren korrespondierende Elemente w_{ii} der Diagonalmatrix W ungleich null sind, den Bildbereich der linearen Abbildung, $\{y | y = Ax; x \in \mathbb{R}^n\}$, auf. Die Zeilen i der Matrix V^T (oder die Spalten von V) mit korrespondierendem $w_{ii} = 0$ spannen hingegen den Nullraum $\{x | Ax = 0\}$ der linearen Abbildung auf. Die SVD konstruiert also explizit orthogonale Basen für den Bildbereich und den Nullraum der linearen Abbildung A.

Bei der Lösung des Gleichungsystems (B.5) ergeben sich damit zwei verschiedene Möglichkeiten. Ist der Vektor b ein Element des Bildbereiches von A, dann ist das Gleichungssystem lösbar. Eine der möglicherweise vielen Lösungen ist $x_0 = A^\Psi b$. Die anderen Lösungen erhält man aus $x = x_0 + \tilde{x}$, wobei $\tilde{x}$ ein beliebiges Element des Nullraums von A ist. Die $\tilde{x}$ sind also Linearkombinationen der Spalten der Matrix V, deren korrespondierendes Element $w_{ii} = 0$ ist. Der durch die Pseudoinverse gegebene Lösungsvektor $x_0 = A^\Psi$ ist aber unter allen Lösungen dadurch ausgezeichnet, daß er die kleinste Norm $|x_0|$ besitzt.

Beweis: Betrachten wir $|x_0 + \tilde{x}|$, wobei $\tilde{x}$ ein Element des Nullraums der linearen Abbildung sei. Dann gilt

$$\begin{aligned} |x_0 + \tilde{x}| &= |VW^{-1}U^T b + \tilde{x}| \\ &= |V(W^{-1}U^T b + V^T \tilde{x})| \\ &= |W^{-1}U^T b + V^T \tilde{x}| \end{aligned}$$

Der Vektor $W^{-1}U^T b + V^T \tilde{x}$ hat nun Komponenten

- vom ersten Summanden nur dort, wo $w_{ii} \neq 0$ ist, denn W^{-1} war ja eine etwas saloppe Schreibweise für
$$W^{-1} = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$
(vergl. (B.7)).
- vom zweiten Summanden nur dort, wo $w_{ii} = 0$ ist, denn $\tilde{x}$ war ja aus dem Nullraum der linearen Abbildung.

Deshalb können, egal wie $\tilde{x}$ gewählt wird, die Komponenten des ersten Summanden, und damit die Norm von $|x_0 + \tilde{x}|$ nicht reduziert werden. Die minimale Norm ergibt sich für $\tilde{x} = 0$. $\lhd$

Der andere Fall liegt vor, falls der Vektor b kein Element des Bildbereiches von A ist. Dann kann Gleichung (B.5) eigentlich nicht gelöst werden. Der Vektor $x_0 = A^\Psi b$ ist dann aber eine optimale Näherungslösung und minimiert den Fehler $f = |Ax - b|$.

Beweis: Es sei $x = x_0 + x'$; $b' = Ax'$. Wir finden damit für den Fehler $|Ax - b|$, daß

$$\begin{aligned} |Ax - b| &= |Ax_0 - b + b'| \\ &= |(UWV^T)(VW^{-1}U^T)b - b + b'| \\ &= |(UWW^{-1}U^T - 1)b + b'| \\ &= |U(WW^{-1} - 1)U^T b + UU^T b'| \\ &= |(WW^{-1} - 1)U^T b + U^T b'| \end{aligned}$$

Nun ist aber $(WW^{-1} - 1)$ eine Diagonalmatrix, deren Komponenten nur dort ungleich null sind, wo $w_{ii} = 0$ ist. Da $b' = Ax'$ ein Element des Bildbereiches von A ist, kann $U^T b'$ nur dort Komponenten haben, wo $w_{ii} \neq 0$ ist. Das Minimum wird also für $b' = 0$ erzielt. $\lhd$

B.2 Die Projektor–Matrix

Die Eigenschaften der Pseudoinversen bei der Lösung linearer Gleichungssysteme können vorteilhaft bei der Erzeugung von Kopplungsmatrizen für Autoassoziative Neuronale Netzwerke ausgenutzt werden. Die Kopplungsmatrix eines Neuronalen Netzes soll p binäre Mustervektoren $\xi_i^\mu = \pm 1$ als dynamisch stabile Netzwerk–Konfigurationen eingespeichern. Dies ist äquivalent der Forderung

$$\mathrm{sign}\Big(\sum_j J_{ij}\xi_j^\mu\Big) = \xi_i^\mu \ ,$$

welche für alle Muster ξ_i^μ erfüllt sein muß. Schärfer ist die Bedingung

$$\sum_j J_{ij}\xi_j^\mu = \xi_i^\mu \ ,$$

die wir, mit der aus den p Muster–Vektoren ξ_i^μ zusammengesetzten Matrix A

$$A = \left(\xi_i^1, \xi_i^2, \ldots, \xi_i^p\right) ; \qquad i = 1, \ldots, N \ ,$$

auch als Matrizengleichung

$$JA = A \tag{B.8}$$

schreiben können. Diese Gleichung hat unendlich viele Lösungen, die mit der zur Matrix A gehörenden Pseudoinversen A^Ψ als

$$J = AA^\Psi + B(\mathbb{1} - AA^\Psi)$$

bei beliebiger $n \times n$–Matrix B, dargestellt werden kann [Ko84, Pe86].

Für Anwendungen im Bereich Neuronaler Netzwerke wird normalerweise die Lösung von (B.8) mit verschwindender B–Matrix genommen, also die Matrix

$$\mathcal{P} = AA^\Psi \ .$$

Dies ist, wie man leicht nachrechnet, der Projektionsoperator in den Raum der Mustervektoren. Diese Matrix hat unter allen Lösungen von (B.8) die minimalste Norm und sorgt damit für maximale Stabilitäten der Muster (vergl. Anhang C). Außerdem ist sie eine symmetrische Matrix. Die pseudoinverse Kopplungsmatrix unterscheidet sich von der Projektormatrix lediglich durch eine genullte Diagonale:

$$J_{ij} = \begin{cases} \mathcal{P}_{ij} & \text{für } i \neq j \\ 0 & \text{für } i = j \end{cases} \ .$$

Die Projektormatrix kann für linear unabhängige Muster direkt durch die Formel

$$\mathcal{P}_{ij} = N^{-1} \sum_{\mu,\nu} \xi_i^\mu C_{\mu\nu}^{-1} \xi_j^\nu \tag{B.9}$$

berechnet werden. Dabei ist $C_{\mu\nu}$ die Korrelationsmatrix der Muster

$$C_{\mu\nu} = N^{-1} \sum_i \xi_i^\mu \xi_i^\nu \ .$$

Formel (B.9) ist numerisch weniger aufwendig als die Anwendung der SVD und bietet gleichzeitig einen analytischen Zugang zur Projektormatrix. Zur numerischen Berechnung der Projektormatrix können aber noch schnellere Verfahren angewendet werden (siehe Anhang D), die auch eine Invertierung der Korrelationsmatrix überflüssig machen. Wichtig ist, und zwar für eine Implementation als paralleler Algorithmus, daß auch eine lokale Lernregel zur Berechnung der Projektormatrix existiert [Di87].

B.2.1 Eigenschaften der Projektor–Matrix und der pseudoinversen Kopplungsmatrix

Für binäre Zufallsmuster, also für $\xi_i^\mu = \pm 1$, gestattet Formel (B.9) einige für die Analytik nützliche Eigenschaften der Projektor–Matrix $\mathcal{P}_{ij}$ und der pseudoinversen Kopplungsmatrix J_{ij} herzuleiten.

Zunächst erhält man für die Diagonalelemente von $\mathcal{P}$, daß

$$\sum_i \mathcal{P}_{ii} = N^{-1} \sum_{i\mu\nu} C^{-1}_{\mu\nu} \xi_i^\mu \xi_i^\nu = N\alpha \; .$$

Hieraus folgt, daß im Limes $N \to \infty$

$$\overline{\mathcal{P}_{ii}} = \alpha \tag{B.10}$$

gilt, mit Flukuationen der Ordnung $\mathcal{O}(\frac{1}{\sqrt{N}})$. Damit erhalten wir für die internen Felder der Muster bei der pseudoinversen Kopplungsmatrix

$$\sum_j J_{ij} \xi_j^\mu = (1 - \alpha) \xi_i^\mu \; . \tag{B.11}$$

Mit (B.10) erhalten wir weiter, wiederum im Limes $N \to \infty$,

$$\begin{aligned} \sum_i J_{ij} J_{ik} &= \sum_i \mathcal{P}_{ij}(1 - \delta_{ij}) \mathcal{P}_{ik}(1 - \delta_{ik}) \\ &= \mathcal{P}_{jk}(1 - \mathcal{P}_{jj} - \mathcal{P}_{kk}) + \mathcal{P}_{kk}^2 \delta_{jk} \\ &= \mathcal{P}_{jk}(1 - 2\alpha) + \alpha^2 \delta_{jk} \; . \end{aligned} \tag{B.12}$$

Hierbei wurde noch die Projektoreigenschaft von $\mathcal{P}$ und die Symmetrie dieser Matrix ausgenutzt. Aus (B.12) erhält man zunächst für $j = k$ die Formel

$$\sum_j J_{ij}^2 = \alpha(1 - \alpha) \; .$$

Als weitere wichtige Formel folgt

$$\begin{aligned} \sum_{ijk} J_{ij} J_{ik} S_j(t) S_k(\tau) &= \alpha(1 - \alpha) \sum_k S_k(t) S_k(\tau) \\ &\quad + (1 - 2\alpha) \sum_k h_k(t) S_k(\tau) \; . \end{aligned} \tag{B.13}$$

Damit erhalten wir beispielsweise für $\alpha = \frac{1}{2}$, daß

$$\overline{h_i(t) h_i(\tau)} = \frac{1}{4} \overline{S_i(t) S_i(\tau)}$$

gilt, woraus, wiederum für diese Speicherdichte, die einfache Formel

$$\overline{h_i^2(t)} = \frac{1}{4}$$

folgt (siehe auch Abbildung 4.4 auf Seite 66).

Die Iterationsgleichungen der DPD

Mit Gleichung (B.11) folgt in einem beliebigen Systemzustand $S_i(t)$ für den Mittelwert der Projektion der internen Felder auf einen der Mustervektoren die Gleichung

$$\begin{aligned}\overline{\xi_i^\mu h_i(t)} &= N^{-1}\sum_{i,j} J_{ij}\xi_i^\mu S_j(t)\\ &= (1-\alpha)N^{-1}\sum_j \xi_j^\mu S_j(t)\\ &= (1-\alpha)m_\mu(t)\ .\end{aligned}$$

Mit $\xi_i^1 \overset{!}{=} 1$ folgt sofort die Iterationsgleichung (4.22) der DPD auf Seite 62.

Zur Herleitung der zweiten Iterationsgleichung der DPD für die Varianz der Feldverteilung, $\overline{\Delta h_i^2(t)}$, benutzen wir Formel (B.13). Wir finden zunächst

$$\begin{aligned}\overline{h_i^2(t+1)} &= \alpha(1-\alpha)+(1-2\alpha)\,\overline{h_i(t+1)\,\mathrm{sign}(h_i(t))}\\ &\approx \alpha(1-\alpha)+(1-2\alpha)\,\overline{h_i(t+1)}\cdot\overline{\mathrm{sign}(h_i(t))}\\ &\quad +(1-2\alpha)\underbrace{\frac{\overline{\Delta h_i(t+1)\,\Delta h_i(t)}}{\overline{\Delta h_i^2(t)}}\Big\{\overline{|h_i(t)|}-\overline{\mathrm{sign}(h_i(t))}\cdot\overline{h_i(t)}\Big\}}_{=\,a_i(t+1)}\end{aligned}$$

mit der Näherung

$$\overline{\Delta a(x)\,\Delta b(y)} \approx \frac{\overline{\Delta x\,\Delta y}}{\overline{\Delta x^2}\;\overline{\Delta y^2}}\cdot\Big\{\overline{a(x)\,\Delta x}\cdot\overline{b(y)\,\Delta y}\Big\}\ .$$

Dies ist eine Entwicklung der linken Seite nach kleinen Korrelationen $\overline{\Delta x\,\Delta y}$ für schwach korrelierte Zufallsvariable x und y.

Mit

$$\begin{aligned}\overline{\Delta h_i(t+1)\,\Delta h_i(t)} &= \overline{h_i(t+1)\,h_i(t)}-\overline{h_i(t+1)}\cdot\overline{h_i(t)}\\ &= (1-2\alpha)\,\overline{|h_i(t)|}-(1-\alpha)^2 m(t+1)m(t)\\ &\quad +\alpha(1-\alpha)\,\overline{\mathrm{sign}\big(h_i(t)\big)\,\mathrm{sign}\big(h_i(t-1)\big)}\ ,\end{aligned}$$

und

$$\begin{aligned}\overline{\mathrm{sign}\big(h_i(t)\big)\,\mathrm{sign}\big(h_i(t-1)\big)} &\approx m(t+1)\,m(t)\\ &\quad +\frac{a_i(t)}{\overline{\Delta h_i^2(t)}}\Big\{\overline{|h_i(t)|}-(1-\alpha)m(t+1)m(t)\Big\}\end{aligned}$$

erhält man nach schließlich unter Verwendung der DPD–Gleichung

$$\overline{|h_i(t)|}-(1-\alpha)m(t+1)m(t) = (1-m_t^2)d_{t+1}$$

die Rekursionsgleichung (4.23) auf Seite 62.