

Aspekte der linearen Minimax-Schätzung

Dissertation

zur

Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

im Fachbereich Mathematik/Informatik
der Universität Kassel

vorgelegt von

Stefan Heilmann

aus

Köln

Kassel, den 27. Januar 2004

Als Dissertation vom Fachbereich Mathematik/Informatik
angenommen am: 11.Mai 2004

Erster Gutachter: Prof. Dr. H. Drygas
Zweiter Gutachter: Prof. Dr. H. Läuter

Tag der mündlichen Prüfung: 22. Juni 2004

Inhaltsverzeichnis

1	Einleitung	5
2	Minimax-Schätzung bei ellipsoider Vorinformation	11
2.1	Schätzproblem	11
2.2	Minimax-Schätzung	15
3	Bayes'scher Ansatz	21
3.1	Bayes-Verfahren	22
3.2	Zweite Moment-Matrix	24
3.3	Minimax-Schätzer als Bayes-Verfahren	26
3.4	Ungünstigste a priori Verteilung	29
4	Spektrale Methoden	33
4.1	Spektralgleichungen	34
4.2	Der Fall $j = 1$	42
4.3	Der Fall $j = k$	43

4.4	Der Fall $j = k - 1$	45
5	Darstellung von Hoffmann und Lauter	51
6	Gemeinsamer Eigenvektor	57
6.1	Falluntersuchung	58
6.2	Orthogonal zum maximalen Eigenraum	66
7	Im maximalen Eigenraum	69
7.1	Eintrage der Matrizen	70
7.2	Entwicklung von Kriterien	82
7.3	Notwendigkeit der Kriterien	93
7.4	Ausblick	110
8	Darstellung der Losung	113
8.1	Algebraische Grundlagen	114
8.2	Keine Losung durch Radikale	118
9	Fixpunktiteration	125
9.1	Differentiation von Matrix-Funktionen	125
9.2	Fixpunktgleichung	128
9.3	Simulation	133
10	Zusammenfassung und Ausblick	139

A	Matrizenrechnung	143
B	Minimax-Theorem	149
C	Konvexe Analysis	153
D	Mathematica Programme	155
D.1	1. Simulation	155
D.2	2. Simulation	158
D.3	Beispiel	162
	Abbildungsverzeichnis	165
	Literaturverzeichnis	167

Kapitel 1

Einleitung

Die Untersuchung von Einflussgrößen x_1, \dots, x_k auf eine Zielgröße y ist in der Statistik eine bedeutende Anwendung. Kann der Zusammenhang zwischen Einflussgrößen und Zielgröße annähernd durch eine lineare Funktion

$$y = \beta_1 x_1 + \dots + \beta_k x_k$$

beschrieben werden, so findet das lineare Regressionsmodell Anwendung. Durch Schätzung der Koeffizienten (Parameter) β_1, \dots, β_k mittels n -facher Beobachtung der Zielgröße und zugehörigen Einflussgrößen kann die Wirkung der Einflussgrößen auf die Zielgröße ermittelt werden. Da der Zusammenhang aber nur annähernd linear ist, wird die Beobachtungsgleichung durch

$$y = X\beta + \varepsilon$$

gegeben. Dabei ist y der Vektor der Beobachtungen der Zielgröße, in der Matrix X stehen jeweils in den Spalten die Beobachtungen der entsprechenden Einflussgrößen und ε ist ein Fehlervektor.

Für den Fall, dass keine Vorinformationen über den Parametervektor β bekannt sind, kann dieser mit der Methode der kleinsten Quadrate optimal geschätzt werden. Anders sieht es dabei aus, wenn man die Information besitzt, dass die Koeffizienten aus einem Ellipsoid stammen. In dieser Situation ist die Kleinste-Quadrate-Schätzung nicht optimal. Daher kommt hier der Minimax-Schätzer zum Einsatz. Dieser Einsatz ist dadurch gerechtfertigt, weil er zulässig ist, das heißt es gibt keinen gleichmäßig besseren Schätzer, und weil er außerdem die beste Verbesserung des Kleinsten-Quadrate-Schätzers ist.

Die Bestimmung und Darstellung des Minimax-Schätzers ist bisher nur selten in geschlossener Form möglich. Das heißt der Minimax-Schätzer kann in Abhängigkeit der Matrix X aus der Beobachtungsgleichung und der Gewichtsmatrix der Verlustfunktion bzw. durch Funktionen dieser Matrizen oder deren Einträge angegeben werden. Dies liegt daran, dass zur Bestimmung des Minimax-Schätzers eine auf der Menge der Matrizen definierte Funktion minimiert werden muss. Diese Funktion ist aber im Allgemeinen nicht differenzierbar. Durch die Übertragung auf andere Zugänge (Spektraler Ansatz, Bayes'scher Ansatz, Darstellung von Hoffmann und Läuter) konnten äquivalente Kriterien zur Bestimmung von Minimax-Schätzern angegeben werden. Aber bei diesen Zugängen hängt der Minimax-Schätzer von einer Funktion von sich selbst ab oder es sind zu seiner Bestimmung nichtlineare Eigenwertprobleme zu lösen. Dennoch ist es dadurch gelungen in einigen Situationen Minimax-Schätzer in geschlossener Form anzugeben.

In dieser Arbeit geht es darum die bisherigen Vorgehensweisen und Resultate so zusammenzustellen, dass man daraus Wege zur Bestimmung von Minimax-Schätzern in noch ungelösten Fällen öffnen kann. Durch die häufig unbefriedigende Darstellung von Minimax-Schätzer ist man auch an einer Verschärfung

der bereits bestehenden Kriterien bzw. an neuen Kriterien interessiert. Durch diese Schwierigkeiten der Darstellung der Minimax-Schätzer spielen natürlich auch einfach zu handhabende numerische Verfahren eine Rolle. Diese können sowohl im Zusammenhang mit den bestehenden bzw. verschärften Kriterien als auch in der generellen Situation von Interesse sein.

In Kapitel zwei wird die Parameterschätzung im linearen Modell bei ellipsoider Vorinformation als Schätzproblem im Rahmen der statistischen Entscheidungstheorie formuliert. Es folgt die Definition sowie die Rechtfertigung der Minimax-Schätzung als optimales Schätzverfahren.

Das dritte Kapitel gibt einen Überblick über Bayes-Verfahren im linearen Modell und erläutert den vorhandenen Zusammenhang zwischen Bayes-Verfahren und Minimax-Schätzung. Dadurch wird die Bestimmung des Minimax-Schätzers auf ein Minimierungsproblem einer differenzierbaren Funktion mit Nebenbedingung reduziert. Doch bei Anwendung der Lagrange-Multiplikatorenregel stößt man auf ein nichtlineares Eigenwertproblem.

Mit Hilfe der konvexen Analysis wird in Kapitel vier der spektrale Ansatz vorgestellt. Es gibt aber noch weitere Herangehensweisen an den spektralen Ansatz. Dieser transferiert das Minimierungsproblem, das aus der Definition der Minimax-Schätzung resultiert, in Gleichungen, die von den Eigenwerten und Eigenvektoren einer Funktion des Minimax-Schätzers abhängen. Aus diesem Ansatz heraus war es aber möglich für einige Fälle den Minimax-Schätzer in geschlossener Form darzustellen. Ein besonderes Augenmerk liegt dabei auf der Vielfachheit des maximalen Eigenwertes der Funktion des Minimax-Schätzers. Mit Kenntnis dieser Vielfachheit kann entschieden werden, ob ein Fall, in dem es geschlossene Lösungen gibt, vorliegt. Ein Bezug zum Bayes'schen Ansatz kann auch aufgezeigt werden.

Das Kapitel fünf handelt von der Darstellung von Hoffmann und Läuter, die in den 70er Jahren das Minimierungsproblem auf das Lösen einer Matrixgleichung mit Nebenbedingungen gebracht haben. Diese Matrixgleichung ist ebenfalls ein nichtlineares Eigenwertproblem. Obwohl diese Darstellung historisch gesehen der älteste Zugang ist, wird sie hier aus dem in Kapitel vier beschriebenen spektralen Ansatz hergeleitet.

Ausgehend von einem Beispiel aus ([8], S. 9ff) wird im sechsten Kapitel ein spezielles Modell mit drei Einflussgrößen definiert und mit Methoden des spektralen Ansatzes strukturiert. Anhand der Kriterien aus Kapitel vier können ungelöste Fälle ausgemacht werden.

In Kapitel sieben wird ein ungelöster Fall einer näheren Untersuchung unterzogen. Durch Verwendung eines Satzes aus Kapitel vier sowie die Ausnutzung der Strukturierung aus Kapitel sechs kann die Bestimmung des Minimax-Schätzers auf die Bestimmung einer Nullstelle einer reellwertigen Funktion zurückgeführt werden. Desweiteren kann ein von der Matrix X aus der Beobachtungsgleichung und der Gewichtsmatrix unabhängiges Intervall, das die Nullstelle überdeckt, angegeben werden. Mit der Intervallschachtelung ist es dann möglich bei gegebener Genauigkeit die Nullstelle numerisch zu berechnen.

Das Kapitel acht beschäftigt sich mit der Frage, ob der Minimax-Schätzer für das Beispiel aus ([8], S. 9ff) in Abhängigkeit der Einträge der Matrix X und der Gewichtsmatrix der Verlustfunktion dargestellt werden kann. Doch wird sich zeigen, dass dieses nicht möglich ist.

In Kapitel neun kann ausgehend von der Darstellung von Hoffmann und Läuter eine Fixpunktgleichung formuliert werden. In einer Simulation werden für viele Beispiele mittels Fixpunktiteration oftmals Ergebnisse erhalten, die zur Berechnung des Minimax-Schätzers verwendet werden können. Die Kon-

vergenz wie auch der Erhalt des gewünschten Fixpunktes ist bis jetzt nicht bewiesen. Wenn die triviale Lösung zur Berechnung des Minimax-Schätzers verwendet werden kann, ist die lokale Konvergenz gesichert.

Kapitel zehn fasst die Ergebnisse zusammen und gibt einen Ausblick.

Kapitel 2

Minimax-Schätzung bei ellipsoider Vorinformation

Man betrachte ein lineares Regressionsmodell mit ellipsoider Vorinformation. Mit Hilfe der statistischen Entscheidungstheorie kann die Minimax-Schätzung als sinnvolles und optimales Schätzverfahren hergeleitet werden. Eine weitere Berechtigung erfährt die Minimax-Schätzung durch Bezug zur Spieltheorie.

2.1 Schätzproblem

Die Formulierung des Schätzproblems erfolgt hier auf die gleiche Weise, wie die beim linearen Modell ohne Vorinformation (vgl. [31], S. 9). Dort entstammt der Parametervektor dem \mathbb{R}^k . Sind jedoch Vorinformationen über den Parametervektor bekannt, so kann man seinen Bereich dementsprechend einschränken. Die Parametervektoren werden in einem Ellipsoid liegen.

Definition 2.1 Das Schätzproblem im linearen Regressionsmodell bei ellipsoider Vorinformation ist gegeben durch

(1) das statistische lineare Modell

$$y = X\beta + \varepsilon$$

mit den n -dimensionalen Zufallsvektoren y und ε . Es gelte $E\varepsilon = 0$, $\text{Cov } \varepsilon = I_n$. Dabei sind E der Erwartungswert, Cov die Kovarianzmatrix und I_n die $(n \times n)$ -Einheitsmatrix. Weiter seien $X \in \mathbb{R}^{n \times k}$, $k < n$, $\beta \in \mathcal{E} \subset \mathbb{R}^k$ und $\mathcal{E} = \{b \in \mathbb{R}^k | b'b \leq 1\}$,

(2) den Entscheidungsraum \mathbb{R}^k ,

(3) die Menge der (nichtrandomisierten) linearen Schätzfunktionen

$$\Delta = \{\delta : \mathbb{R}^n \rightarrow \mathbb{R}^k | \delta \text{ linear}\}$$

und (4) die Gauß'sche Verlustfunktion $s : \mathcal{E} \times \Delta \rightarrow [0, \infty)$ mit

$$s(\beta, \delta) = (\beta - \delta)' B' B (\beta - \delta), \quad (2.1)$$

wobei $B \in \mathbb{R}^{m \times k}$ mit $\text{rg } B = m$ die Gewichtsmatrix der Verlustfunktion ist.

Mit M' werde das Transponierte einer Matrix M bezeichnet.

Das verallgemeinerte lineare Modell mit $\text{Cov } \varepsilon = \sigma^2 V$, $\mathcal{E} = \{b \in \mathbb{R}^k | (b - b_0)' T (b - b_0) \leq \tau\}$, $V \in \mathbb{R}^{n \times n}$ positiv definit, $T \in \mathbb{R}^{k \times k}$ positiv definit, $b_0 \in \mathbb{R}^k$, $\sigma^2 > 0$ und $\tau > 0$ kann durch die Transformationen

$$y \rightarrow \sigma^{-1} V^{-1/2} (y - X b_0),$$

$$T \rightarrow \tau^{-1} T,$$

$$\beta \rightarrow T^{1/2} (\beta - b_0)$$

und

$$X \rightarrow \sigma^{-1} V^{-1/2} X T^{-1/2}$$

auf das Modell aus Definition 2.1 (1) zurückgeführt werden (vgl. [6], S. 17).

Desweiteren kann $B'B$ ohne Einschränkung als Diagonalmatrix angenommen werden. Da $B'B$ nicht negativ definit (n.n.d.) ist, besitzt es die Spektraldarstellung $B'B = D\Gamma D'$, wobei D eine orthogonale Matrix und Γ eine Diagonalmatrix ist. Mit $\beta_1 = D'\beta$ und $\delta_1 = D'\delta$ erhält man das Modell

$$y = X\beta_1 + \varepsilon$$

mit $\beta_1'\beta_1 = \beta'DD'\beta = \beta'\beta \leq 1$ und der Verlustfunktion

$$s(\beta_1, \delta_1) = (\beta_1 - \delta_1)'\Gamma(\beta_1 - \delta_1).$$

Die Menge Δ aus Definition 2.1 (3) kann alternativ durch

$$\Delta = \{\delta : \mathbb{R}^n \rightarrow \mathbb{R}^k \mid \delta = Cy, C \in \mathbb{R}^{k \times n}\} \quad (2.2)$$

dargestellt werden. Dadurch wird dann (2.1) zu

$$s(\beta, C) = (\beta - Cy)'\beta'\beta(\beta - Cy).$$

Es ergibt sich nun als Risikofunktion

$$R(\beta, C) = \beta'(CX - I_k)'\beta'\beta(CX - I_k)\beta + \text{tr}(BC)'\beta'\beta(BC). \quad (2.3)$$

Ein optimales Schätzverfahren sollte die Risikofunktion R über \mathcal{E} minimieren. Dies ist jedoch nach ([31], S. 53) nicht möglich. Aber man kann zuerst alle linearen Schätzer, für die es keinen gleichmäßig besseren Schätzer gibt, in Betracht ziehen, bevor man weitere Einschränkungen vollzieht.

Definition 2.2 Sei \mathcal{D} eine Menge von Schätzern und Θ ein Parameterraum. Dann heißt $\delta_1 \in \mathcal{D}$ besser als $\delta_2 \in \mathcal{D}$ bezüglich der Risikofunktion R , wenn gilt:

- (1) $\forall \vartheta \in \Theta \quad R(\vartheta, \delta_1) \leq R(\vartheta, \delta_2)$
 (2) $\exists \vartheta_0 \in \Theta \quad R(\vartheta_0, \delta_1) < R(\vartheta_0, \delta_2)$.

Ein Schätzer $\delta \in \mathcal{D}$ heißt zulässig, falls es keinen anderen Schätzer aus \mathcal{D} gibt, der besser als δ bezüglich R ist.

Hoffmann [16] hat für das Schätzproblem aus Definition 2.1 mit $\text{rg } X = k$ und $\text{rg } B = k$ Bedingungen für die Zulässigkeit eines Schätzers gegeben. In einer Variante von Schipp ([27], S. 64) ergibt sich folgender Satz.

Satz 2.3 In Definition 2.1 mit $\text{rg } X = k$ und $\text{rg } B = k$ ist $\delta \in \Delta$ mit $\delta = Cy$ genau dann zulässig, wenn gilt:

- (1) $C = LX'$,
 (2) $(X'X)^{\frac{1}{2}}L(X'X)^{\frac{1}{2}} \quad n.n.d.$,
 (3) $(X'X)^{-1} - L \quad \text{positiv definit (p.d)}$,
 (4) $\text{tr} (X'X)^{-1}B'B [(I_k - LX'X)^{-1} - I_k] \leq 1$.

BEWEIS:

vgl. [16], S. 435, Theorem 2. □

Der Kleinste-Quadrate-Schätzer $\hat{\beta}_{KQ} = (X'X)^{-1}X'y$ (vgl. [27], S. 65) ist im linearen Regressionsmodell ohne Vorinformation und $\text{rg } X = k$ der gleichmäßig beste erwartungstreue Schätzer. Jedoch im Modell aus Definition 2.1 ist die Bedingung (3) aus Satz 2.3 nicht erfüllt. Es ist $L = (X'X)^{-1}$ und $(X'X)^{-1} - (X'X)^{-1} = 0$ ist nicht positiv definit. Damit ist hier für dieses Modell der Schätzer $\hat{\beta}_{KQ}$ nicht zulässig. Das heißt, es gibt einen Schätzer aus Δ , der besser als $\hat{\beta}_{KQ}$ ist. Mit dem Minimax-Schätzer kann nun ein zulässiger und besserer Schätzer gefunden werden.

2.2 Minimax-Schätzung

Mit Hilfe der Spieltheorie ist es möglich einen Zugang zur Lösung eines statistischen Entscheidungsproblems herzustellen. Dabei betrachtet man ein Zweipersonen-Nullsummenspiel zwischen Natur als Spieler 1 und Statistiker als Spieler 2 mit erwarteter Auszahlung $R(\beta, \delta)$. Die Natur versucht ihren eigenen Gewinn zu maximieren; der Statistiker will seinen Schaden minimieren. Das heißt, die Natur will $R(\beta, \delta)$ bezüglich β maximieren und der Statistiker will $R(\beta, \delta)$ bezüglich δ minimieren. Da im Allgemeinen beiden Spielern die Strategie des Gegenspielers nicht bekannt ist, nimmt ein Spieler die für ihn ungünstigste Strategie des Gegenspielers an und wählt dann dazu die Strategie, die seinen Gewinn maximiert. In diesem Falle gibt R dann den Mindestgewinn der Natur und den Höchstverlust des Statistikers an. Stimmen diese beiden Werte überein, so kann man von einer optimalen Lösung des Spiels sprechen. Aus der Sicht des Statistikers liegt ein Minimax-Verfahren vor, da er versucht den maximalen Gewinn der Natur zu minimieren (vgl. [18], Vol. 5, S.511; [31], S.75ff.).

Im entscheidungstheoretischen Sinn heißt das, dass durch die Annahme des maximalen Risikos die Risikofunktion von einer funktionswertigen auf eine reellwertige Funktion überführt werden kann. Somit besteht die Möglichkeit ein Minimum zu finden.

Definition 2.4 Der Schätzer $\hat{\beta}_M \in \Delta$ mit $\hat{\beta}_M = \hat{C}_M y$, $\hat{C}_M \in \mathbb{R}^{k \times n}$, heißt linearer Minimax-Schätzer (MILE), wenn gilt:

$$\sup_{\beta \in \mathcal{E}} R(\hat{C}_M, \beta) = \inf_{C \in \mathbb{R}^{k \times n}} \sup_{\beta \in \mathcal{E}} R(C, \beta).$$

Hoffmann [17] konnte 1979 zeigen, dass der MILE existiert und für $\text{rg } X = k$ und $\text{rg } B = k$ eindeutig und zulässig ist. Weiterhin konnte dort gezeigt wer-

den, dass der MILE die beste lineare Verbesserung des Kleinsten-Quadrat-Schätzers ist. Unter der besten linearen Verbesserung versteht man den Schätzer, der die minimale Differenz über \mathcal{E} maximiert. Also gilt hier für $\delta \in \Delta$

$$\sup_{\delta \in \Delta} \inf_{\beta \in \mathcal{E}} \left[R(\hat{\beta}_{KQ}, \beta) - R(\delta, \beta) \right] = \inf_{\beta \in \mathcal{E}} \left[R(\hat{\beta}_{KQ}, \beta) - R(\hat{\beta}_M, \beta) \right]$$

(vgl. [27], S. 67).

Selbst wenn man die Menge der nichtrandomisierten linearen Schätzfunktionen Δ auf die Menge der randomisierten linearen Schätzfunktionen $\tilde{\Delta}$ erweitert, kann das Minimax-Risiko nicht verringert werden.

Definition 2.5 *Eine Wahrscheinlichkeitsverteilung $\tilde{\delta}$ auf der Menge der nichtrandomisierten linearen Schätzfunktionen Δ heißt randomisierte lineare Schätzfunktion. Die Menge aller randomisierten linearen Schätzfunktionen werde mit $\tilde{\Delta}$ bezeichnet.*

Sei Z eine Zufallsvariable, die gemäß einem $\tilde{\delta} \in \tilde{\Delta}$ verteilt ist und Werte in Δ annimmt. Dann ist durch

$$\tilde{s}(\beta, \tilde{\delta}) = \int s(\beta, Z) d\tilde{\delta}$$

die Verlustfunktion bei Gebrauch von randomisierten linearen Schätzfunktionen gegeben (vgl. [10], S. 23f). Im Schätzproblem aus Definition 2.1 ist der Entscheidungsraum $\mathcal{E} \subset \mathbb{R}^k$ die Einheitskugel und damit auch konvex. Die Gauß'sche Verlustfunktion s ist in δ eine konvexe Funktion, für die $\lim_{\|\delta\| \rightarrow \infty} s(\beta, \delta) = \infty$ gilt. Nach Theorem 1 aus Ferguson ([10], S. 78) gibt es für jede randomisierte Schätzfunktion $\tilde{\delta} \in \tilde{\Delta}$ eine nichtrandomisierte Schätzfunktion $\delta \in \Delta$, die für jedes $\beta \in \mathcal{E}$ keinen größeren Verlust hat. Deswegen reicht es aus nur nichtrandomisierte lineare Schätzfunktionen zu betrachten.

Zur Bestimmung des MILE muss nun das maximale Risiko

$$r(C) = \sup_{\beta \in \mathcal{E}} R(C, \beta)$$

minimiert werden. Es ergibt sich mit Lemma A.2

$$\begin{aligned} & \sup_{\beta \in \mathcal{E}} R(C, \beta) \\ &= \sup_{\beta \in \mathcal{E}} \beta' (CX - I_k)' B' B (CX - I_k) \beta + \text{tr} (BC)' BC \\ &= \lambda_{\max} [(CX - I_k)' B' B (CX - I_k)] + \text{tr} (BC)' BC. \end{aligned} \quad (2.4)$$

Dabei bezeichne $\lambda_{\max}(M)$ den maximalen Eigenwert der Matrix M . Bei der Minimierung der Funktion r bezüglich C besteht das Problem, dass r im Allgemeinen nicht differenzierbar ist. Nur für den Fall $m = 1$ ist das Differenzieren möglich. Für diesen Fall haben Kuks und Ol'man [19] bereits 1972 den Minimax-Schätzer bestimmt. Ist man aber interessiert Gewichtsmatrizen B mit $\text{rg } B > 1$ zuzulassen, so kann man die Dimension des Definitionsbereichs von r von $k \cdot n$ auf $k \cdot k$ verringern. Dazu verwendet man die Transformation

$$D = CX. \quad (2.5)$$

Damit ist $D \in \mathbb{R}^{k \times k}$. Die Matrizen C , die (2.5) erfüllen, haben die Form

$$C = DX^+ + R(I_n - XX^+),$$

wobei $R \in \mathbb{R}^{k \times n}$ beliebig und X^+ die Moore-Penrose-Inverse von X ist (vgl. Definition A.1). Betrachtet man nun die Funktion $g : \mathbb{R}^{k \times k} \rightarrow \mathbb{R}$ mit

$$g(D) = \lambda_{\max} [(D - I_k)' B' B (D - I_k)] + \text{tr} (X^+)' D' B' B D X^+, \quad (2.6)$$

so minimiert $\hat{D} = \hat{C}X$ stets g , falls r sein Minimum in \hat{C} annimmt. Andererseits wird f in $\hat{C} = \hat{D}X^+$ minimal, wenn g von \hat{D} minimiert wird (vgl. [29]). Dadurch wird das Minimierungsproblem der Funktion r auf die Funktion g übertragen.

Ist nun die Gewichtsmatrix B regulär, so kann die Substitution

$$Z = B(CX - I) \quad (2.7)$$

durchgeführt werden. Löst man (2.7) nach C auf, so ergibt sich als Lösung

$$C = B^{-1}(Z + B)X^+ + R(I - XX^+)$$

mit $R \in \mathbb{R}^{k \times n}$ beliebig. Durch Zulässigkeitsüberlegungen im nächsten Lemma besteht ein eindeutiger Zusammenhang zwischen C und Z .

Lemma 2.6 *Sei $C_1 = B^{-1}(Z + B)X^+ + R(I - XX^+)$, $R \in \mathbb{R}^{k \times n}$ und $R(I - XX^+) \neq 0$. Weiter sei $C_0 = B^{-1}(Z + B)X^+$. Dann gilt für alle $\beta \in \mathcal{E}$:*

$$R(\beta, C_1) > R(\beta, C_0).$$

BEWEIS:

Es gilt:

- (1) $C_1 = C_0 + R(I - XX^+)$,
- (2) $B(C_1X - I) = B[C_0X + R(I - XX^+)X - I] = B(C_0X - I)$,
- (3) $\text{tr } C_0' B' B R (I - XX^+) = \text{tr } (X^+)' (Z + B)' (B^{-1})' B' B R (I - XX^+)$
 $= \text{tr } (Z + B)' B R (I - XX^+) (X^+)' = 0$,
- (4) $\text{tr } (I - XX^+) R' B' B R (I - XX^+) > 0$, da wegen der Regularität von B auch $B R (I - XX^+) \neq 0$ ist.

Dann ist

$$\begin{aligned} R(\beta, C_1) &= \beta' (C_1X - I)' B' B (C_1X - I) \beta + \text{tr } C_1' B' B C_1 \\ &= \beta' (C_0X - I)' B' B (C_0X - I) \beta + \text{tr } C_0' B' B C_0 \\ &\quad + 2 \underbrace{\text{tr } C_0' B' B R (I - XX^+)}_{=0 \text{ (3)}} \\ &\quad + \underbrace{\text{tr } (I - XX^+) R' B' B R (I - XX^+)}_{>0 \text{ (4)}} \end{aligned}$$

$$\begin{aligned}
&> \beta'(C_0X - I)'B'B(C_0X - I)\beta + \text{tr } C_0'B'BC_0 \\
&= R(\beta, C_0).
\end{aligned}$$

(vgl. [6], S. 22)

□

Sei nun

$$f(Z) = \lambda_{\max}(Z'Z) + \text{tr}(X^+)'(Z + B)'(Z + B)X^+.$$

Dann ist der Minimax-Schätzer $\hat{\beta}_M = \hat{C}_M y$ durch

$$\hat{C}_M = B^{-1}(\hat{Z}_M + B)X^+ \quad (2.8)$$

gegeben, falls

$$f(\hat{Z}_M) = \min_{Z \in \mathbb{R}^{k \times k}} f(Z)$$

gilt.

Aber auch mit der Dimensionsreduktion des Definitionsbereichs des Minimax-Risikos r bleibt das Problem der Nichtdifferenzierbarkeit bestehen. Im Folgenden werden nun drei Wege (Bayes'scher Ansatz, spektrale Methoden und Hoffmann-Läuter-Gleichung) beschrieben, wie dieses Minimierungsproblem auf äquivalente Probleme übertragen werden kann.

Kapitel 3

Bayes'scher Ansatz

Eine weitere Möglichkeit, einen Zugang zu den Minimax-Schätzern im linearen Regressionsmodell mit Vorinformation zu finden, ist mittels der Bayes-Verfahren möglich. Um ein solches Bayes-Verfahren zu gewinnen wird zuerst der Parameterraum mit einer Wahrscheinlichkeitsverteilung, der a priori Verteilung, ausgestattet. Mit Hilfe dieser a priori Verteilung kann die Risikofunktion auf eine reellwertige Funktion überführt und durch Minimierung dieser ein Bayes-Verfahren gewonnen werden. Es wird sich herausstellen, dass für das Schätzproblem aus Definition 2.1 das Bayes-Verfahren zu der a priori Verteilung mit dem größten Risiko der Minimax-Schätzer ist.

Wie schon bei den Minimax-Schätzern ist auch bei den Bayes-Verfahren ein Bezug zur Spieltheorie möglich. Jedoch ist es hier der Natur erlaubt auch gemischte Strategien einer bekannten Verteilung bezüglich β zu spielen. Der Statistiker wählt eine Strategie um die erwartete Auszahlung $E_{\beta}R(\beta, \delta)$ zu minimieren. Diese optimale Strategie heißt Bayes-Strategie zur a priori Verteilung bezüglich β . Gibt es eine a priori Verteilung, wobei die Bayes-Strategie des Statistikers die höchste erwartete Auszahlung einbringt, und

eine Minimax-Strategie des Statistikers derart, dass die beiden Auszahlungen übereinstimmen, so ist die Minimax-Strategie auch Bayes-Strategie zur eben beschriebenen a priori Verteilung (vgl [31], S. 76ff).

Dieses Kapitel orientiert sich hauptsächlich an den Ergebnissen aus Pilz [24]. Jedoch wird auf Verwendung Bayes'scher Versuchsplanung verzichtet.

3.1 Bayes-Verfahren

Es werde nun ein parametrisches statistisches Modell mit Parameterraum Θ , Menge der Entscheidungsfunktionen \mathcal{D} , Verlustfunktion $s : \Theta \times \mathcal{D} \rightarrow [0, \infty)$ und der daraus resultierenden Risikofunktion R betrachtet. Für die Bestimmung von Bayes-Verfahren ist die a priori Verteilung, eine Verteilung auf dem Parameterraum Θ , von Bedeutung.

Definition 3.1 (A Priori Verteilung) *Sei Π ein Wahrscheinlichkeitsmaß auf Θ mit einer geeigneten σ -Algebra über Θ . Dann heißt Π eine a priori Verteilung auf Θ . Mit \mathcal{T} werde die Menge aller a priori Verteilungen auf Θ bezeichnet.*

Mit Hilfe der a priori Verteilung ist es nun möglich das Bayes-Risiko anzugeben.

Definition 3.2 (Bayes-Risiko) *Sei $\delta \in \mathcal{D}$. Für $\Pi \in \mathcal{T}$ heißt*

$$\varrho(\Pi, \delta) = \int_{\Theta} R(\vartheta, \delta) d\Pi \tag{3.1}$$

Bayes-Risiko von δ bezüglich der a priori Verteilung Π und

$$\inf_{\delta \in \mathcal{D}} \varrho(\Pi, \delta)$$

minimales Bayes-Risiko bezüglich Π .

Von Interesse für eine Schätzung von β ist diejenige Schätzfunktion, die (3.1) minimiert.

Definition 3.3 (Bayes-Verfahren) Ein $\delta_\Pi \in \mathcal{D}$ heißt Bayes-Verfahren zu Π , wenn gilt:

$$\varrho(\Pi, \delta_\Pi) = \inf_{\delta \in \mathcal{D}} \varrho(\Pi, \delta).$$

Die Verwendung von Bayes-Verfahren setzt ein ausreichendes Vorwissen über die Verteilung der Parameter voraus. Ist dies nicht der Fall, kann überdies der Bayes'sche Ansatz bei der Existenz einer ungünstigsten a priori Verteilung zur Anwendung kommen.

Definition 3.4 (Ungünstigste a priori Verteilung) Ein $\Pi^* \in \mathcal{T}$ heißt ungünstigste a priori Verteilung, falls gilt:

$$\inf_{\delta \in \mathcal{D}} \varrho(\Pi^*, \delta) = \sup_{\Pi \in \mathcal{T}} \inf_{\delta \in \mathcal{D}} \varrho(\Pi, \delta).$$

Die ungünstigste a priori Verteilung ist die a priori Verteilung, dessen Bayes-Verfahren das größte Bayes-Risiko besitzt. Es ist nun möglich einen Zusammenhang zwischen Bayes- und Minimax-Verfahren herzustellen. Existiert neben der ungünstigsten a priori Verteilung ein Minimax-Verfahren δ_M , d.h. gibt es eine Lösung δ_M von

$$\sup_{\vartheta \in \Theta} R(\vartheta, \delta_M) = \inf_{\delta \in \mathcal{D}} \sup_{\vartheta \in \Theta} R(\vartheta, \delta)$$

und besitzt das Bayes-Risiko ϱ in (Π^*, δ_M) einen Sattelpunkt, so ist das Minimax-Verfahren ein Bayes-Verfahren zur ungünstigsten a priori Verteilung (vgl. [31], Satz 1.78). Im Abschnitt 3.3 wird auf den Zusammenhang zwischen Minimax-Schätzern und Bayes-Schätzern speziell im linearen Regressionsmodell bei Vorinformation nochmals eingegangen.

3.2 Zweite Moment-Matrix

Es werde wiederum das Schätzproblem aus Definition 2.1 betrachtet und damit auch die Risikofunktion R aus (2.3). Weiter werde mit \mathcal{T} die Menge aller a priori Verteilungen auf \mathcal{E} und durch

$$\mathcal{M} = \left\{ M_{\Pi} \in \mathbb{R}^{k \times k} \mid M_{\Pi} = \int_{\mathcal{E}} \beta \beta' d\Pi, \Pi \in \mathcal{T} \right\} \quad (3.2)$$

die Menge der zweiten Moment-Matrizen der a priori Verteilung bezeichnet. Dieses Integral existiert, da bezüglich der Norm $\|A\| = (\text{tr } A'A)^{\frac{1}{2}}$ die Abschätzung

$$\left\| \int_{\mathcal{E}} \beta \beta' d\Pi \right\| \leq \int_{\mathcal{E}} \|\beta \beta'\| d\Pi \leq \int_{\mathcal{E}} d\Pi = 1$$

gilt. Damit ergibt sich das Bayes-Risiko zur a priori Verteilung $\Pi \in \mathcal{T}$ zu

$$\varrho(\Pi, C) = \text{tr } B'B \left[CC' + (CX - I_k) \int_{\mathcal{E}} \beta \beta' d\Pi (CX - I_k)' \right]$$

(siehe [24], S. 212). Das Bayes-Risiko hängt damit von der zweiten Moment-Matrix $M_{\Pi} \in \mathcal{M}$ der a priori Verteilung Π ab und besitzt die Darstellung

$$\varrho(M_{\Pi}, C) = \text{tr } B'B [CC' + (CX - I_k)M_{\Pi}(CX - I_k)'].$$

Durch Minimierung des Bayes-Risikos ϱ bezüglich C erhält man den Bayes-Schätzer $\hat{\beta}_{\Pi} = \hat{C}_{\Pi}y$ zur a priori Verteilung Π . Er hat nach ([24], Lemma 5.3) mit Lemma A.3 die Form

$$\begin{aligned} \hat{C}_{\Pi} &= (B'B)^+ B' B M_{\Pi} X' (I_n + X M_{\Pi} X')^{-1} \\ &\quad + [I_k - (B'B)^+ B'B] R \\ &= B'(BB')^{-1} B M_{\Pi} X' (I_n + X M_{\Pi} X')^{-1} \\ &\quad + [I_k - B'(BB')^{-1} B] R \end{aligned} \quad (3.3)$$

für $R \in \mathbb{R}^{k \times n}$. Für den Fall $\text{rg } B = k$ ergibt sich

$$\hat{C}_\Pi = M_\Pi X' (I_n + X M_\Pi X')^{-1}.$$

Lemma 3.5 *Das Bayes-Risiko für den Bayes-Schätzer \hat{C}_Π aus (3.3) ist:*

$$\varrho(M_\Pi, \hat{C}_\Pi) = \text{tr} \left[B' B M_\Pi - B' B M_\Pi X' (I_n + X M_\Pi X')^{-1} X M_\Pi \right]. \quad (3.4)$$

BEWEIS:

Definiere $L = (I_n + X M_\Pi X')^{-1}$ und $\hat{C}_\Pi = B'(BB')^{-1}BC_0 + [I_k - B'(BB')^{-1}B]R$ mit $C_0 = M_\Pi X' L$.

Dann gilt:

$$\begin{aligned} (1) \quad & \text{tr } B' B \hat{C}_\Pi \hat{C}_\Pi' = \text{tr } B' B C_0 C_0' \\ & = \text{tr } C_0 \{ C_0' B' (BB')^{-1} B + R' [I_k - B' (BB')^{-1} B] B' B \} = \text{tr } B' B C_0 C_0', \\ (2) \quad & \text{tr } B' B (\hat{C}_\Pi X - I_k) M_\Pi (\hat{C}_\Pi X - I_k)' \\ & = \text{tr } B' B (C_0 X - I_k) M_\Pi [X' C_0 + X' R' (I_k - B' (BB')^{-1} B) - I_k] \\ & = \text{tr } B' B (C_0 X - I_k) M_\Pi (C_0 X - I_k)'. \end{aligned}$$

Mit (1) und (2) folgt dann

$$\begin{aligned} \varrho(M_\Pi, \hat{C}_\Pi) &= \text{tr } B' B [C_0 C_0' + (C_0 X - I_k) M_\Pi (C_0 X - I_k)'] \\ &= \text{tr } B' B [C_0 C_0' + C_0 X M_\Pi X' C_0' - C_0 X M_\Pi - M_\Pi X' C_0' + M_\Pi] \\ &= \text{tr } B' B M_\Pi X' L^2 X M_\Pi + \text{tr } B' B M_\Pi X' L X M_\Pi X' L X M_\Pi \\ &\quad - 2 \text{tr } B' B M_\Pi X' L X M_\Pi + \text{tr } B' B M_\Pi \\ &= \text{tr } B' B M_\Pi X' L L^{-1} L X M_\Pi - 2 \text{tr } B' B M_\Pi X' L X M_\Pi + \text{tr } B' B M_\Pi \\ &= \text{tr } B' B M_\Pi - \text{tr } B' B M_\Pi X' L X M_\Pi. \end{aligned}$$

(vgl. [24], Theorem 5.3)

□

3.3 Minimax-Schätzer als Bayes-Verfahren

In diesem Abschnitt wird der Zusammenhang zwischen Bayes-Verfahren und Minimax-Schätzung im linearen Regressionsmodell bei nichtlinearer Vorinformation untersucht. Dazu werden Aussagen über die Menge der zweiten Moment-Matrizen der a priori Verteilungen benötigt um die Existenz einer ungünstigsten a priori Verteilung sicherzustellen.

Lemma 3.6 *Sei \mathcal{T} die Menge aller a priori Verteilungen über \mathcal{E} und \mathcal{M} aus (3.2). Dann gilt:*

$$\mathcal{M} = \{M_{\Pi} \in \mathbb{R}^{k \times k} \mid M \text{ n.n.d, } \text{tr } M_{\Pi} \leq 1\}.$$

BEWEIS:

Sei M eine zweite Moment-Matrix gemäß (3.2). Dann ist M offensichtlich nicht negativ definit. Für $M \in \mathbb{R}^{k \times k}$ ist $\text{tr } M$ in M linear und es folgt

$$\text{tr} \int_{\mathcal{E}} \beta \beta' d\Pi = \text{tr } E\beta \beta' = E(\text{tr } \beta \beta') = E\beta' \beta \leq 1.$$

vgl. auch ([24], S. 155f) □

Lemma 3.7 *Die Menge \mathcal{M} aus Lemma 3.6 ist kompakt bezüglich $\|M\| = (\text{tr } M^2)^{\frac{1}{2}}$ und konvex.*

BEWEIS:

Kompaktheit:

Sei $(M_n : n \in \mathbb{N}) \subset \mathcal{M}$ eine Folge von Matrizen. Da die Matrix M_n n.n.d ist, besitzt sie k reellwertige Eigenwerte $\lambda_1, \dots, \lambda_k \geq 0$. Es gilt:

$$\|M_n\|^2 = \sum_{i=1}^k \lambda_i^2 \leq \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j = \left(\sum_{i=1}^k \lambda_i \right)^2 \leq 1,$$

da $\text{tr } M \leq 1$. Wegen der Beschränktheit von M_n und der Vollständigkeit des $\mathbb{R}^{k \times k}$ existiert eine konvergente Teilfolge $(M_{n_\nu} : \nu \in \mathbb{N})$. Auf Grund der Stetigkeit der Spur und der quadratischen Formen ist $\lim_{\nu \rightarrow \infty} M_{n_\nu} \in \mathcal{M}$.

Konvexität:

Sei $\alpha \in [0, 1]$, $M_1, M_2 \in \mathcal{M}$. Dann gilt für alle $x \in \mathbb{R}^k$:

$$x' [\alpha M_1 + (1 - \alpha) M_2] x = \alpha x' M_1 x + (1 - \alpha) x' M_2 x \geq 0$$

und

$$\text{tr} [\alpha M_1 + (1 - \alpha) M_2] = \alpha \text{tr} M_1 + (1 - \alpha) \text{tr} M_2 \leq \alpha + 1 - \alpha = 1.$$

□

Lemma 3.8 *Sei das Schätzproblem aus Definition 2.1 gegeben und \mathcal{T} die Menge aller a priori Verteilungen über \mathcal{E} . Dann existiert eine ungünstigste a priori Verteilung $\Pi_0 \in \mathcal{T}$.*

BEWEIS:

Wegen der Kompaktheit von \mathcal{M} und der Stetigkeit von $\varrho(M_\Pi, \hat{C}_\Pi)$ bezüglich M_Π gilt

$$\sup_{M_\Pi \in \mathcal{M}} \varrho(M_\Pi, \hat{C}_\Pi) < \infty$$

und es existiert ein $M_{\Pi_0} \in \mathcal{M}$ mit

$$\varrho(M_{\Pi_0}, \hat{C}_{\Pi_0}) = \sup_{M_\Pi \in \mathcal{M}} \varrho(M_\Pi, \hat{C}_\Pi). \quad (3.5)$$

Dabei ist M_{Π_0} die zweite Moment-Matrix der a priori Verteilung $\Pi_0 \in \mathcal{T}$ und wegen (3.5) ist Π_0 die ungünstigste a priori Verteilung. □

Nun lässt sich (3.5) in der Form

$$\inf_{C \in \mathbb{R}^{k \times k}} \varrho(M_{\Pi_0}, C) = \sup_{M_\Pi \in \mathcal{M}} \inf_{C \in \mathbb{R}^{k \times k}} \varrho(M_\Pi, C)$$

darstellen.

Lemma 3.9 *Es gilt:*

$$\inf_{C \in \mathbb{R}^{k \times k}} \sup_{M_\Pi \in \mathcal{M}} \varrho(M_\Pi, C) = \sup_{M_\Pi \in \mathcal{M}} \inf_{C \in \mathbb{R}^{k \times k}} \varrho(M_\Pi, C). \quad (3.6)$$

BEWEIS:

Da wegen Lemma 3.7 \mathcal{M} kompakt und konvex ist, $\varrho(M_\Pi, C)$ bezüglich M_Π stetig und konkav und $\varrho(M_\Pi, C)$ bezüglich C konvex ist, kann für die Vertauschung von Infimum und Supremum das Minimax-Theorem aus Lemma B.5 angewendet werden. \square

Durch die Vertauschung von Supremum und Infimum des Bayes-Risikos und Anwendung der Formel (15.4) aus ([24], S. 292)

$$\sup_{\beta \in \mathcal{E}} R(\beta, C) = \sup_{M_\Pi \in \mathcal{M}} \varrho(M_\Pi, C),$$

ergibt sich (3.6) zu

$$\inf_{C \in \mathbb{R}^{k \times k}} \sup_{\beta \in \mathcal{E}} R(\beta, C) = \sup_{M_\Pi \in \mathcal{M}} \inf_{C \in \mathbb{R}^{k \times k}} \varrho(M_\Pi, C)$$

und weiter zu

$$\inf_{C \in \mathbb{R}^{k \times k}} r(C) = \sup_{M_\Pi \in \mathcal{M}} \varrho(M_\Pi, \hat{C}_\Pi).$$

Damit kann folgendes Lemma formuliert werden.

Lemma 3.10 *Der Schätzer $\hat{\beta}_M = \hat{C}_M y$ für das Schätzproblem aus Definition 2.1 ist der MILE, genau dann wenn er Bayes-Schätzer zur ungünstigsten a priori Verteilung Π_0 aus Lemma 3.8 ist.*

BEWEIS:

siehe ([24], Lemma 15.3) \square

Es ist nun eine Darstellung des MILE gefunden. Diese hängt aber von der zweiten Moment-Matrix der ungünstigsten a priori Verteilung ab.

3.4 Ungünstigste a priori Verteilung

Der Vorteil des Bayes'schen Ansatzes ist, dass das Minimierungsproblem einer im Allgemeinen nicht differenzierbaren Funktion auf ein Minimierungsproblem einer stetig differenzierbaren Funktion übertragen werden kann. Aber bei dem neuen Minimierungsproblem handelt es sich um eins mit Nebenbedingungen. Es sind die Bedingungen, dass die Matrizen nicht negativ definit sind und ihre Spur nicht größer als eins ist, also solche, die nicht adhoc abzuhandeln sind. Desweiteren lässt sich der Minimax-Schätzer auch nicht immer eindeutig über die zweite Moment-Matrix identifizieren, da die ungünstigste a priori Verteilung nicht notwendigerweise eindeutig sein muss (vgl. [24], S. 214).

Wegen Lemma 3.9, (3.3) und Lemma A.3 kann der Minimax-Schätzer $\hat{\beta}_M = \hat{C}_M y$ durch

$$\hat{C}_M = B'(BB')^{-1}BM_0X'(I_n + XM_0X')^{-1} + [I_k - B'(BB')^{-1}B]R$$

für $R \in \mathbb{R}^{k \times k}$ und M_0 die zweite Moment-Matrix der ungünstigsten a priori Verteilung angegeben werden. Für $\text{rg } B = k$ vereinfacht sich die Darstellung des Minimax-Schätzers zu

$$\hat{C}_M = M_0X'(I_n + XM_0X')^{-1}. \quad (3.7)$$

Gilt neben $\text{rg } B = k$ auch $\text{rg } X = k$, so ist die zweite Moment-Matrix M_0 der ungünstigsten a priori Verteilung eindeutig bestimmt (vgl. [24], S. 214).

Zur Bestimmung der zweiten Moment-Matrix der ungünstigsten a priori Verteilung M_0 muss das Maximierungsproblem

$$\sup_{M_{\Pi} \in \mathcal{M}} \varrho(M_{\Pi}, \hat{C}_{\Pi}) \quad (3.8)$$

gelöst werden. Dieses Supremum existiert auf Grund der Kompaktheit von

\mathcal{M} und der Stetigkeit von ϱ in M_{Π} . Im folgenden Satz wird herausgestellt, dass es ausreicht das Supremum auf einer echten Teilmenge von \mathcal{M} zu suchen. Dazu definiere man gemäß (3.4) die Funktion $f : \mathcal{M} \rightarrow \mathbb{R}$ durch

$$f(M) = \operatorname{tr} B'BM - B'BMX'(I_n + XMX')^{-1}XM.$$

Durch Verwendung von Lemma A.2 erhält man

$$\begin{aligned} f(M) &= \operatorname{tr} B'BM - B'BMX'(I_n + XMX')^{-1}XM \\ &= \operatorname{tr} B'BM[I_k - X'(I_n + XMX')^{-1}XM] \\ &= \operatorname{tr} B'BM(I_k - X'XM)^{-1} \\ &= \operatorname{tr} B'B(I_k - MX'X)^{-1}M. \end{aligned} \tag{3.9}$$

Satz 3.11 Sei $\mathcal{M} = \{Mn.n.d \mid \operatorname{tr} M \leq 1\} \subset \mathbb{R}^{k \times k}$ und $\mathcal{M}_0 = \{Mn.n.d \mid \operatorname{tr} M = 1\} \subset \mathcal{M}$. Die Funktion $f : \mathcal{M} \rightarrow \mathbb{R}$ ist definiert durch

$$f(M) = \operatorname{tr} B'B(I_k - MX'X)^{-1}M.$$

Dann existiert zu jedem $M \in \mathcal{M}$ mit $\operatorname{tr} M < 1$ ein $M_0 \in \mathcal{M}_0$ mit

$$f(M) \leq f(M_0).$$

BEWEIS:

Sei $M \in \mathcal{M}$ mit $\operatorname{tr} M = m < 1$. Definiere $M_0 = \frac{1}{m}M$. Dann ist $M_0 \in \mathcal{M}_0$. Es werde $L = I_k + MX'X$, $L_m = mI_k + MX'X$ und $K = LL_m$ gesetzt. Es folgt, dass L und L_m und damit auch K regulär sind. Weiter gilt

$$LL_m = mI_k + (1+m)MX'X + MX'XMX'X = L_mL.$$

Wegen Lemma A.6 folgt

$$L_m^{-1} - L^{-1} = K^{-1}(L - L_m) = (1-m)K^{-1}.$$

Da MK' nicht negativ definit ist, erhält man

$$K^{-1}M = K^{-1}MK'(K')^{-1} \quad \text{n.n.d..}$$

Aus

$$\begin{aligned} f(M_0) - f(M) &= \text{tr } B'B(I + M_0X'X)^{-1}M_0 - \text{tr } B'B(I + MX'X)^{-1}M \\ &= \text{tr } B \left[(mI + MX'X)^{-1} - (I + MX'X)^{-1} \right] MB' \\ &= \text{tr } B(L_m^{-1} - L^{-1})MB' \\ &= \underbrace{(1-m)}_{>0} \text{tr } B \underbrace{K^{-1}M}_{\text{n.n.d.}} B' \geq 0 \end{aligned}$$

folgt die Behauptung. \square

Es verändert sich nun das Maximierungsproblem zu

$$\sup_{M \in \mathcal{M}_0} f(M).$$

Dabei ist f wie in (3.9). Eine notwendige Bedingung für ein Maximum im eben beschriebenen Minimierungsproblem wird im nächsten Satz gegeben.

Satz 3.12 *Die Matrix M_0 sei im Schätzproblem aus Definition 2.1 mit $\text{rg } B = k$ die zweite Moment-Matrix zur ungünstigsten a priori Verteilung. Dann erfüllt M_0 für ein $\alpha > 0$ die Gleichung*

$$M_0^* M_0 = \alpha M_0.$$

Dabei ist $M_0^* = (I_k + X'X M_0)^{-1} B'B (I_k + M_0 X'X)^{-1}$.

BEWEIS:

siehe ([7], Theorem 3.3) \square

In Kapitel 4 kann schließlich eine hinreichende Bedingung für die zweite Moment-Matrix zur ungünstigsten a priori Verteilung angegeben werden.

Kapitel 4

Spektrale Methoden

Anfang und Mitte der 90-er Jahre wurde ein weiterer Zugang zur Minimax-Schätzung im linearen Regressionsmodell bei ellipsoider Vorinformation gefunden. Die Bestimmung des Minimax-Schätzers wurde auf das Lösen einer Matrixgleichung geführt. Jedoch hängt diese Gleichung von dem maximalen Eigenwert und dazugehörigen Eigenvektoren der Matrix $B(CX - I_k)[B(CX - I_k)]'$ ab. Besondere Aufmerksamkeit liegt auf der Vielfachheit des maximalen Eigenwerts. Den Beweis der Notwendigkeit einer Spektralgleichung erbrachte 1996 Girko [13] unter Verwendung der Eigenwerttheorie zufälliger Matrizen und Störungstheorie zufälliger Operatoren. Die andere Richtung konnte 1996 von Drygas [6] bewiesen werden. Von ihm stammen weitere Beweismethoden der Spektralgleichung (vgl. [3], [5]). Durch Verwendung konvexer Analysis erbrachten Christopeit und Helmes [2] ebenfalls einen Beweis. In diesem Kapitel wird mit einer Verallgemeinerung dieses Beweises die Gültigkeit der Spektralgleichung aufgezeigt. Anschließend wird noch kurz auf die Verbindung zum Bayes'schen Ansatz aus Kapitel 3 eingegangen. Diese Verbindung konnte 1996 durch Drygas und Pilz [7] hergestellt werden. Eine besondere

Betrachtung liegt auf einem zu den Spektralgleichungen äquivalenten Gleichungssystem, das in Drygas ([6], Theorem 2.3) zu finden ist. Bei gegebenen Vielfachheiten des maximalen Eigenwerts von $B(CX - I_k)[B(CX - I_k)]'$ können aus diesem Gleichungssystem viele Resultate erhalten werden.

4.1 Spektralgleichungen

Es werde das lineare Regressionsmodell aus Definition 2.1 betrachtet. Die Vorgehensweise zur Bestimmung einer Spektralgleichung wird mit Hilfe von konvexer Analysis durchgeführt und orientiert sich im Wesentlichen an Christopeit und Helmes [2].

Zur Bestimmung des MILE muss die Funktion r aus (2.4) bezüglich C minimiert werden. Definiere dazu

$$u(C) = \lambda_{\max} [(CX - I_k)'B'B(CX - I_k)]$$

und

$$v(C) = \text{tr}(BC)'BC.$$

Die Funktionen u und v sind konvex. Es gilt

$$r(C) = u(C) + v(C).$$

Damit ist gemäß Lemma C.2 das Subdifferential von r durch

$$\partial r(C) = \partial u(C) + \partial v(C) \tag{4.1}$$

gegeben. Da v stetig partiell differenzierbar ist, erhält man mit Lemma A.8 und Lemma C.3

$$\partial v(C) = \{2B'BC\}.$$

Um das Subdifferential von u zu bestimmen betrachte man $I = \{x \in \mathbb{R}^k \mid \|x\| = 1\}$ und

$$u_x(C) = x'(CX - I_k)'B'B(CX - I_k)x$$

für $x \in I$. Die Funktionen u_x sind konvex und stetig partiell differenzierbar in C . Mit

$$u_x(C) = \text{tr}(CX - I_k)'B'B(CX - I_k)xx',$$

Lemma A.8 und Lemma C.3 erhält man

$$\partial u_x(C) = \{2B'B(CX - I_k)xx'X'\}$$

für $x \in I$. Dann sind

$$u(C) = \sup_{x \in I} u_x(C) < \infty$$

und

$$I(C) = \{x \in \mathbb{R}^k \mid (CX - I_k)'B'B(CX - I_k)x = \lambda_1^2 x, \|x\| = 1\}$$

die aktive Indexmenge. Dabei ist $\lambda_1^2 = \lambda_{\max}[(CX - I_k)'B'B(CX - I_k)]$. Aus Lemma C.4 erhält man

$$\partial u(C) = \text{conv} \{2B'B(CX - I_k)xx'X' \mid x \in I(C)\}.$$

Lemma 4.1 *Sei $E(\lambda_1^2)$ der Eigenraum zu λ_1^2 und $\Phi \in \partial u(C)$. Dann besitzt Φ die Darstellung*

$$\Phi = 2B'B(CX - I_k) \left(\sum_{i=1}^j p_i v_i v_i' \right) X'.$$

Dabei ist j die Vielfachheit von λ_1^2 , $\{v_1, \dots, v_j\}$ bilden ein Orthonormalbasis von $E(\lambda_1^2)$, und es ist $p_i > 0$ für $i = 1, \dots, j$ und $\sum_{i=1}^j p_i = 1$.

BEWEIS:

a) Definiere $\mathcal{L} = \{xx' \mid x \in E(\lambda_1^2)\}$. Durch $x = \sum_{i=1}^j \alpha_i v_i$ besitzt xx' die Darstellung

$$xx' = \sum_{i=1}^j \alpha_i^2 v_i v_i' + \sum_{i=1}^{j-1} \sum_{k=i+1}^j \alpha_i \alpha_k (v_i v_k' + v_k v_i')$$

und dadurch gilt

$$\text{span}(\mathcal{L}) = \text{span}(\{v_i v_i', v_i v_k' + v_k v_i' \mid i \in \{1, \dots, j\}, k \in \{i+1, \dots, j\}\}).$$

Weiter stehen alle Elemente aus $\{v_i v_i', v_i v_k' + v_k v_i' \mid i \in \{1, \dots, j\}, k \in \{i+1, \dots, j\}\}$ paarweise senkrecht bezüglich des Skalarproduktes $(A, B) = \text{tr } A'B$ für $A, B \in \mathbb{R}^{k \times k}$ aufeinander. Damit erhält man, dass $\{v_i v_i', v_i v_k' + v_k v_i' \mid i \in \{1, \dots, j\}, k \in \{i+1, \dots, j\}\}$ eine Basis von $\text{span}(\mathcal{L})$ ist und somit

$$\dim(\text{span}(\mathcal{L})) = \frac{j(j+1)}{2} = r.$$

b) Seien nun für $i \in \{1, \dots, r+1\}$ Zahlen $q_i \geq 0$ mit $\sum_{i=1}^{r+1} q_i = 1$ und Vektoren $x_i \in E(\lambda_1^2)$ mit $\|x_i\| = 1$ gegeben. Desweiteren sei v ein normierter Eigenvektor von $\sum_{i=1}^{r+1} q_i x_i x_i'$ zum Eigenwert p . Es ist $p \geq 0$. Mit

$$\begin{aligned} & \left(\sum_{i=1}^{r+1} q_i x_i x_i' \right) v = p v \\ \Rightarrow & \left(\sum_{i=1}^{r+1} q_i (CX - I_k)' B' B (CX - I_k) x_i x_i' \right) v \\ & = p (CX - I_k)' B' B (CX - I_k) v \\ \Rightarrow & \lambda_1^2 \left(\sum_{i=1}^{r+1} q_i x_i x_i' \right) v = p (CX - I_k)' B' B (CX - I_k) v \\ \Rightarrow & p \lambda_1^2 v = p (CX - I_k)' B' B (CX - I_k) v \end{aligned}$$

folgt, dass v auch Eigenvektor zu $(CX - I_k)' B' B (CX - I_k)$ ist, falls $p > 0$ ist. Das heißt, man findet eine Orthonormalbasis $\{v_1, \dots, v_j\}$ von $E(\lambda_1^2)$ und

für $i = 1, \dots, j$ Zahlen $p_i > 0$ derart, dass

$$\sum_{i=1}^{r+1} q_i x_i x_i' = \sum_{i=1}^j p_i v_i v_i'$$

gilt. Man erhält weiter

$$\sum_{i=1}^j p_i = \text{tr} \left(\sum_{i=1}^j p_i v_i v_i' \right) = \text{tr} \left(\sum_{i=1}^{r+1} q_i x_i x_i' \right) = 1.$$

c) Ein $\Phi \in \partial u(D)$ besitzt gemäß a) für $q_i \geq 0$ mit $\sum_{i=1}^{r+1} q_i = 1$ und $x_i \in E(\lambda_1^2)$ mit $\|x_i\| = 1$ ($i \in \{1, \dots, r+1\}$) die Darstellung

$$\Phi = 2B'B(CX - I_k) \left(\sum_{i=1}^{r+1} q_i x_i x_i' \right) X'.$$

Durch Anwendung von b) folgt die Behauptung. \square

Mit $V_1 = (v_1 \dots v_j)$ und $P = \text{diag}\{p_1, \dots, p_j\}$ wird die Darstellung von Φ aus Lemma 4.1 zu

$$\Phi = 2B'B(CX - I_k)V_1 P V_1' X'. \quad (4.2)$$

Satz 4.2 (Rechte Spektralgleichung) *Im Schätzproblem aus Definition 2.1 ist $\hat{\beta}_M = \hat{C}_M y$ ein MILE, wenn für eine orthogonale Matrix $V_1 \in \mathbb{R}^{k \times j}$ mit $[B(\hat{C}_M X - I_k)]' B(\hat{C}_M X - I_k) V_1 = \lambda_1^2 V_1$, $\lambda_1^2 = \lambda_{\max} \left\{ [B(\hat{C}_M X - I_k)]' B(\hat{C}_M X - I_k) \right\}$ und für eine Diagonalmatrix $P \in \mathbb{R}^{j \times j}$ mit $P = \text{diag}\{p_1, \dots, p_j\}$, $p_1, \dots, p_j > 0$ und $\sum_{i=1}^j p_i = 1$ die Gleichung*

$$B(\hat{C}_M X - I_k)V_1 P V_1' X' + B\hat{C}_M = 0 \quad (4.3)$$

erfüllt ist.

BEWEIS:

Dass $\hat{\beta}_M = \hat{C}_M y$ ein MILE ist, ist äquivalent dazu, dass die Funktion r

aus (2.4) minimiert wird. Nach Satz C.5 ist dies äquivalent zu $0 \in \partial r(\hat{C}_M)$. Wegen (4.2) und Lemma 4.1 gilt dann

$$\begin{aligned} & 2B'B(\hat{C}_M X - I_k)V_1PV_1'X' + 2B'B\hat{C}_M = 0 \\ \Leftrightarrow & B(\hat{C}_M X - I_k)V_1PV_1'X' + B\hat{C}_M = 0 \end{aligned}$$

und (4.3) ist erfüllt. \square

Die Spalten von V_1 sind die Eigenvektoren von $\left[B(\hat{C}_M X - I_k)\right]' B(\hat{C}_M X - I_k)$ zum maximalen Eigenwert, der mit Vielfachheit j auftritt. Es ist nun möglich den MILE mit Hilfe der Eigenvektoren von $B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k)\right]'$ zum maximalen Eigenwert anzugeben.

Satz 4.3 (Linke Spektralgleichung) *Im Schätzproblem aus Definition 2.1 ist $\hat{\beta}_M = \hat{C}_M y$ ein MILE, wenn für eine orthogonale Matrix $U_1 \in \mathbb{R}^{k \times j}$ mit $B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k)\right]' U_1 = \lambda_1^2 U_1$, $\lambda_1^2 = \lambda_{\max} \left\{ B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k)\right]' \right\}$ und für eine Diagonalmatrix $P \in \mathbb{R}^{j \times j}$ mit $P = \text{diag}\{p_1, \dots, p_j\}$, $p_1, \dots, p_j > 0$ und $\sum_{i=1}^j p_i = 1$ die Gleichung*

$$U_1 P U_1' B(\hat{C}_M X - I_k) X' + B\hat{C}_M = 0$$

erfüllt ist.

BEWEIS:

Sei V_1 und P wie in Satz 4.2. Die Singulärwertzerlegung von $B(\hat{C}_M X - I_k)$ hat die Form

$$B(\hat{C}_M X - I_k) = \lambda_1 U_1 V_1' + U_2 \Lambda_2 V_2'. \quad (4.4)$$

Dabei ist $\lambda_1 I_{k-j} - \Lambda_2$ positiv definit. Damit ergibt sich

$$\begin{aligned} & B(\hat{C}_M X - I_k) V_1 P V_1' \\ &= (\lambda_1 U_1 V_1' + U_2 \Lambda_2 V_2') V_1 P V_1' \end{aligned}$$

$$\begin{aligned}
&= \lambda_1 U_1 P V_1' \\
&= \lambda_1 U_1 P U_1' U_1 V_1' + U_1 P U_1' U_2 \Lambda_2 V_2' \\
&= U_1 P U_1' B (\hat{C}_M X - I_k).
\end{aligned}$$

Aus dieser Identität folgt die Behauptung. \square

Mit der Singulärwertzerlegung aus (4.4) kann durch

$$\lambda_1 U_1 P V_1 X' + B \hat{C}_M = 0$$

eine weitere Spektralgleichung angegeben werden. Sie wird als gemischte Spektralgleichung bezeichnet und U_1 , V_1 und P sind wie in Satz 4.2 und Satz 4.3. Die Darstellung des MILE lautet demnach

$$B \hat{C}_M y = -\lambda_1 U_1 P V_1' X' y.$$

Für $\text{rg } B = k$ ist nach ([7], Theorem 3.1) die rechte Spektralgleichung äquivalent zu

$$B(\hat{C}_M X - I_k) V_1 P V_1' X' X + B(\hat{C}_M X - I_k) + B = 0.$$

Daraus folgt dann sofort die Äquivalenz von

$$U_1 P U_1' B(\hat{C}_M X - I_k) X' X + B(\hat{C}_M X - I_k) + B = 0$$

zur linken Spektralgleichung.

Drygas ([7], S. 46f) konnte zeigen, dass ein Bayes-Schätzer $\hat{C}_\Pi = M_\Pi X' (I_n + X M_\Pi X')^{-1}$ genau dann die rechte Spektralgleichung erfüllt, wenn die zugehörige zweite Moment-Matrix M_Π die Gleichung

$$M_\Pi X' X = V_1 P V_1' X' X$$

erfüllt. Daraus konnte dann ebenfalls in ([7], S. 51) gefolgert werden, dass

für die zweite Moment-Matrix M_0 der ungünstigsten a priori Verteilung die Bedingung

$$\begin{aligned} & (I_k + X'X M_0)^{-1} B' B (I_k + M_0 X' X)^{-1} M_0 \\ & = \lambda_{\max} [(I_k + X'X M_0)^{-1} B' B (I_k + M_0 X' X)^{-1}] M_0 \end{aligned}$$

gelten muss. Dies ist demnach auch die hinreichende Bedingung im Minimierungsproblem aus Satz 3.12. Das heißt aber, dass zur Bestimmung der zweiten Moment-Matrix M_0 zur ungünstigsten a priori Verteilung aus dem Minimierungsproblem mit Nebenbedingung einer stetig differenzierbaren Funktion ein nichtlineares Eigenwertproblem geworden ist. Damit stellt der Bayes'sche Ansatz keine Vereinfachung zur Bestimmung des MILE zu den Spektralgleichungen dar. Deshalb wird die Betrachtung der Spektralgleichungen fortgesetzt.

Unter Verwendung der Singulärwertzerlegung (4.4) und $\text{rg } B = k$ ist die gemischte Spektralgleichung durch

$$\lambda_1 (U_1 P V_1' X' X + U_1 V_1') + U_2 \Lambda_2 V_2' + B = 0. \quad (4.5)$$

gegeben.

In [6] können Aussagen über die Darstellung des MILE und das Minimax-Risiko des MILE gemacht werden:

- (1) $\hat{C}_M y = -\lambda_1 B^{-1} U_1 P V_1' X' y$
- (2) $r(\hat{C}_M) = -\lambda_1 \text{tr } B' U_1 P V_1'$
- (3) $r(\hat{C}_M) \geq \lambda_{\min} [B(I + X'X)^{-1} B']$.

Die Spektralgleichungen werden in ([6], Theorem 2.3) in ein System aus Gleichungen und Ungleichungen überführt.

Satz 4.4 *Im Schätzproblem aus Definition 2.1 mit $\text{rg } B = k$ ist $\hat{C}_M y$ ein MILE für β genau dann, wenn für $j \leq k$, $u_i \in \mathbb{R}^k$, $p_i \in \mathbb{R}$ mit $i \in \{1, 2, \dots, j\}$ und $\lambda_1 > 0$ die Bedingungen*

- (1) $\{u_1, \dots, u_j\}$ bilden ein Orthonormalsystem,
- (2) $\{v_1, \dots, v_j\}$ bilden ein Orthonormalsystem, wobei $v_i = \frac{1}{\lambda_1} (I_k + p_i X'X)^{-1} B' u_i$ mit $i \in \{1, 2, \dots, j\}$ sind,
- (3) $p_i \geq 0$ und $\sum_{i=1}^j p_i = 1$,
- (4) $\lambda_1^2 \geq \lambda_{\max} \left[(I_k - \sum_{i=1}^j u_i u_i') B B' (I_k - \sum_{i=1}^j u_i u_i') \right]$,
- (5) $(I_k - \sum_{i=1}^j u_i u_i') B (I_k + p_l X'X)^{-1} B' u_l = 0$, $l = 1, \dots, j$

erfüllt sind. Der MILE ist dann durch

$$\hat{C}_M y = B^{-1} \sum_{i=1}^j p_i u_i u_i' B (I_k + p_i X'X)^{-1} X' y$$

gegeben.

BEWEIS:

siehe ([6], S.27ff) □

Wie in Satz 4.3 gibt j die Vielfachheit des maximalen Eigenwerts von $B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k) \right]'$ an und u_1, \dots, u_j sind dessen Eigenvektoren. In vielen Fällen ist die praktische Berechnung mit diesem System nicht möglich. Falls aber bekannt ist, dass $B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k) \right]'$ einen einfachen maximalen Eigenwert besitzt, reduziert sich (5) zu einem Eigenwertproblem. Tritt der maximale Eigenwert mit Vielfachheit k auf, so sind (4) und (5) sofort erfüllt. Daher werden diese Situationen weitergehend untersucht. Im weiteren Verlauf wird davon gesprochen, dass der Fall $j = q$, $q \in \{1, \dots, k\}$ vorliegt, wenn die Matrix $B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k) \right]'$ einen q -fachen maximalen Eigenwert besitzt.

4.2 Der Fall $j = 1$

Wie eben angedeutet reduziert sich (5) aus Satz 4.4 für den Fall $j = 1$ zu einem Eigenwertproblem und es ist $p_1 = 1$. Damit kann dann u_1 als normierter Eigenwert von $B(I_k + X'X)^{-1}B'$ mit zugehörigem Eigenwert ϱ berechnet werden. Somit ist (1) aus Satz 4.4 erfüllt. Falls weiter

$$u_1'(BB')^{-1}u_1\varrho^2 \geq \lambda_{\max}[(I_k - u_1u_1')BB'(I_k - u_1u_1')] \quad (4.6)$$

gilt, sind auch (2) und (4) erfüllt. Der MILE besitzt dann die Darstellung

$$\hat{C}_M = \varrho(B^{-1}u_1)(B^{-1}u_1)X'y.$$

Erfüllen alle Eigenwerte von $B(I_k + X'X)^{-1}B'$ die Bedingung (4.6) nicht, so kann der Fall $j = 1$ nicht vorkommen (vgl. [6], S. 29). Ist also $B(I_k + X'X)^{-1}B'$ eine $(k \times k)$ -Matrix müssten bis zu k Eigenwertprobleme gelöst und k Überprüfungen der Bedingung (4.6) stattfinden. Es kann jedoch gezeigt werden, dass es ausreicht nur den maximalen Eigenwert von $B(I_k + X'X)^{-1}B'$ zu untersuchen.

Satz 4.5 *Im Schätzproblem aus Definition 2.1 mit $\text{rg } B = \text{rg } X = k$ liegt der Fall $j = 1$ genau dann vor, wenn für $\varrho = \lambda_{\max}[B(I_k + X'X)^{-1}B']$ mit zugehörigem normierten Eigenvektor u gilt:*

$$(1) \quad u'(BB')^{-1}u\varrho^2 \geq \lambda_{\max}[(I_k - uu')BB'(I_k - uu')],$$

$$(2) \quad \varrho \text{ ist einfach.}$$

BEWEIS:

„ \Rightarrow “

Es gilt der Fall $j = 1$, d.h. ein Eigenwert ϱ von $B(I_k + X'X)^{-1}B'$ mit zu-

gehörigem normierten Eigenvektor u erfüllt die Ungleichung

$$u'(BB')^{-1}u \varrho^2 \geq \lambda_{\max} [(I - uu')BB'(I - uu')]. \quad (4.7)$$

Mit $\varrho_2, \dots, \varrho_k$ werden die restlichen Eigenwerte von $B(I_k + X'X)^{-1}B'$ bezeichnet. Die dazugehörigen Eigenvektoren sind dann u_2, \dots, u_k , derart dass $\{u, u_2, \dots, u_k\}$ ein Orthonormalsystem ergibt. Sei $i \in \{2, \dots, m\}$ beliebig.

$$\begin{aligned} \varrho &= u'(BB')^{-1}u \varrho^2 \frac{1}{u'(BB')^{-1}u \varrho} \\ &\stackrel{\text{Lemma A.5 (b),(4.7)}}{>} \lambda_{\max} [(I - uu')BB'(I - uu')] \\ &\geq u'_i(I - uu')BB'(I - uu')u_i \\ &\stackrel{u_i \perp u}{=} u'_i BB' u_i \\ &\stackrel{\text{Lemma A.5 (a)}}{>} \varrho_i \end{aligned}$$

Also gilt für alle $i \in \{2, \dots, k\}$ stets $\varrho > \varrho_i$.

„ \Leftarrow “

$\varrho = \lambda_{\max} [B(I_k + X'X)^{-1}B']$ mit zugehörigem normierten Eigenvektor u erfüllt (4.6). \square

Es kann durch Kenntnis der Matrizen X und B entschieden werden, ob der Fall $j = 1$ vorliegt oder nicht. Desweiteren gibt es eine Darstellung des MILE, die nur von den Matrizen X und B sowie ihren Eigenwerten und -vektoren abhängt. Aber die Darstellung hängt nicht mehr von den Eigenwerten und -vektoren des MILE selbst ab.

4.3 Der Fall $j = k$

Die Entscheidung über das Vorliegen des Falls $j = k$ kann mit Hilfe der Kenntnis von X und B beantwortet werden. Außerdem kann der Minimax-

Schätzer in Abhängigkeit dieser Matrizen geschlossen angegeben werden. Aus den Bedingungen (1), (2) und (3) aus Satz 4.4 kann dann folgender Satz formuliert werden.

Satz 4.6 *Im Schätzproblem aus Definition 2.1 mit $\text{rg } B = \text{rg } X = k$ liegt der Fall $j = k$ genau dann vor, wenn*

$$H_0 = -(X'X)^{-1} + \frac{1}{\lambda_1} [(X'X)^{-1}B'B(X'X)^{-1}]^{\frac{1}{2}}$$

nicht negativ definit ist. Dabei ist

$$\lambda_1 = \frac{\text{tr} [(X'X)^{-1}B'B(X'X)^{-1}]^{\frac{1}{2}}}{1 + \text{tr} (X'X)^{-1}}.$$

BEWEIS:

siehe ([6], S. 30ff) □

Korollar 4.7 *Eine äquivalente Bedingung dafür, dass H_0 aus Satz 4.6 nicht negativ definit ist, ist durch*

$$\lambda_{\min} \left\{ (X'X)^{\frac{1}{2}} [(X'X)^{-1}B'B(X'X)^{-1}]^{\frac{1}{2}} (X'X)^{\frac{1}{2}} \right\} \geq \lambda_1$$

gegeben. Dabei ist λ_1 wie in Satz 4.6.

BEWEIS:

siehe ([6], S. 33) □

Für den Fall $j = k$ hat die gemischte Spektralgleichung die Darstellung

$$\lambda_1 U_1 P V_1' X' X + \lambda_1 U_1 V_1' + B = 0 \quad (4.8)$$

und es gilt nach Drygas([8], S. 4)

$$V_1 P V_1' = H_0 \quad (4.9)$$

sowie

$$U_1 V_1' = -B(X'X)^{-1} [(X'X)^{-1}B'B(X'X)^{-1}]^{-\frac{1}{2}}. \quad (4.10)$$

Der MILE $\hat{C}_M y$ hat die Form

$$\hat{C}_M y = \left\{ I_k - \lambda_1 (X'X)^{-1} [(X'X)^{-1} B' B (X'X)^{-1}]^{-\frac{1}{2}} \right\} (X'X)^{-1} X' y$$

(vgl. [6], S. 33).

4.4 Der Fall $j = k - 1$

Drygas [8] hat aufgezeigt, wie man den linearen Minimax-Schätzer im Fall $k = 3, j = 2$ bestimmen kann. Die Vorgehensweise aus ([8], S. 35ff) kann auf den Fall $j = k - 1$ verallgemeinert werden. Jedoch ist hier die Kenntnis, dass dieser Fall vorliegt, entscheidend.

Für den Fall $j = k - 1$ wird die gemischte Spektralgleichung (4.5) zu

$$\lambda_1 (U_1 P V_1' X' X + U_1 V_1') + \lambda_2 u_k v_k' + B = 0. \quad (4.11)$$

Im folgenden Satz können nun Kriterien angegeben werden um die unbekannt Matrizen und Zahlen zu bestimmen.

Satz 4.8 *Im Schätzproblem aus Definition 2.1 mit $\text{rg } B = \text{rg } X = k$ liege der Fall $j = k - 1$ vor. Seien $\{a_1, \dots, a_{k-1}\}$ und $\{b_1, \dots, b_{k-1}\}$ Orthonormalsysteme im \mathbb{R}^k , $X_1 = X(b_1 \cdots b_{k-1})$ und $B_1 = (a_1 \cdots a_{k-1})' B (b_1 \cdots b_{k-1})$. Falls ein $u \in \mathbb{R}^k$ mit $\|u\| = 1$ existiert, so dass $\{a_1, \dots, a_{k-1}, u\}$ und $\{b_1, \dots, b_{k-1}, v\}$ mit $v = -\frac{B'u}{\|B'u\|}$ zu Orthonormalbasen des \mathbb{R}^k werden, ist*

$$\begin{aligned} \hat{\beta}_M &= \hat{C}_M y \\ &= B^{-1} [\lambda_1 (a_1 \cdots a_{k-1}) G_1 (b_1 \cdots b_{k-1})' + (I - uu') B] (X'X)^{-1} X' y \end{aligned}$$

der MILE genau dann, wenn

(1) $\lambda_1^2 \geq u'BB'u$ mit

$$\lambda_1 = \frac{\text{tr} [(X_1'X_1)^{-1}B_1'B_1(X_1'X_1)^{-1}]^{\frac{1}{2}}}{1 + \text{tr} (X_1'X_1)^{-1}}, \quad (4.12)$$

(2) $H_1 = -(X_1'X_1)^{-1} + \frac{1}{\lambda_1} [(X_1'X_1)^{-1}B_1'B_1(X_1'X_1)^{-1}]^{\frac{1}{2}}$ ist nicht negativ definit,

(3) $\lambda_1 W X'X B'u + BB'u - uBB'u = 0$ für

$$W = (a_1 \cdots a_{k-1}) G_1 H_1 (b_1 \cdots b_{k-1})'$$

mit

$$G_1 = -B_1(X_1'X_1)^{-1} [(X_1'X_1)^{-1}B_1'B_1(X_1'X_1)^{-1}]^{-\frac{1}{2}}$$

erfüllt sind.

BEWEIS:

Dass $\hat{C}_M y$ der MILE ist, ist äquivalent zur Gültigkeit von (4.11).

„ \Rightarrow “

Die Matrix der normierten Eigenvektoren u_1, \dots, u_{k-1} des maximalen Eigenwerts von $B(\hat{C}_M X - I_k) [B(\hat{C}_M X - I_k)]'$ kann durch

$$U_1 = (u_1 \cdots u_{k-1}) = (a_1 \cdots a_{k-1}) R_1$$

angegeben werden. Dabei ist $R_1 \in \mathbb{R}^{(k-1) \times (k-1)}$ eine entsprechende orthogonale Matrix und $\{a_1, \dots, a_{k-1}\}$ ein Orthonormalsystem. Ähnlich verfährt man für die Matrix der normierten Eigenvektoren v_1, \dots, v_{k-1} zum maximalen Eigenwert von $[B(\hat{C}_M X - I_k)]' B(\hat{C}_M X - I_k)$ durch

$$V_1 = (v_1 \cdots v_{k-1}) = (b_1 \cdots b_{k-1}) R_2$$

mit einer entsprechenden orthogonalen Matrix $R_2 \in \mathbb{R}^{(k-1) \times (k-1)}$ und einem

Orthonormalsystem $\{b_1, \dots, b_{k-1}\}$. Multipliziert man (4.11) von links mit u'_k , so ergibt sich

$$u'_k B + \lambda_2 v'_k = 0$$

und daraus

$$v_k = -\frac{1}{\lambda_2} B' u_k. \quad (4.13)$$

Da $\|v_k\| = 1$ sein soll, ist $v_k = -\frac{B' u_k}{\|B' u_k\|}$ und somit

$$\lambda_2 = \sqrt{u'_k B B' u_k}. \quad (4.14)$$

Außerdem steht $v_k = -\frac{B' u_k}{\|B' u_k\|}$ senkrecht auf b_i ($i = 1, \dots, k - 1$). Nach Einsetzen von (4.13) und (4.14) in (4.11) und der Multiplikation von links mit $(a_1 \cdots a_{k-1})'$ und von rechts mit $(b_1 \cdots b_{k-1})$ erhält man

$$\lambda_1 R_1 P R'_2 X'_1 X_1 + \lambda_1 R_1 R'_2 + B_1 = 0. \quad (4.15)$$

Dies ist eine Spektralgleichung für den Fall $j = k$, nur dass hierbei die Matrizen $X'_1 X_1$ und B_1 aus dem $\mathbb{R}^{(k-1) \times (k-1)}$ stammen. Wegen Satz 4.6 und (4.14) folgt (1) und (2). Aus (4.15) folgt mit (4.9) und (4.10) $R_2 P R'_2 = H_1$ und $R_1 R'_2 = G_1$. Durch

$$\begin{aligned} & \lambda_1 W X' X B' u_k + B B' u_k - u'_k B B' u_k u_k \\ &= \lambda_1 (a_1 \cdots a_{k-1}) G_1 H_1 (b_1 \cdots b_{k-1})' X' X B' u_k \\ & \quad + B B' u_k - u'_k B B' u_k u_k \\ &= \lambda_1 (a_1 \cdots a_{k-1}) R_1 P R'_2 (b_1 \cdots b_{k-1}) X' X B' u_k \\ & \quad + B B' u_k - u'_k B B' u_k u_k \\ &= \lambda_1 U_1 P V'_1 X' X B' u_k + (B B' u_k - u'_k B B' u_k u_k) \\ & \stackrel{(4.13)}{=} -\lambda_2 \lambda_1 U_1 P V'_1 X' X v_k - \lambda_2 B v_k - \lambda_2^2 u_k \\ & \stackrel{(4.11)}{=} -\lambda_2 (-\lambda_1 U_1 V'_1 - B - \lambda_2 u_k v'_k) v_k - \lambda_2 B v_k - \lambda_2^2 u_k = 0 \end{aligned}$$

ist auch (3) erfüllt.

„ \Leftarrow “

Aus (2) und (4.12) folgt mit Satz 4.6, dass orthogonale Matrizen $Q_1, Q_2 \in \mathbb{R}^{(k-1) \times (k-1)}$ und eine positiv definite Diagonalmatrix $P \in \mathbb{R}^{(k-1) \times (k-1)}$ mit $\text{tr } P = 1$ existieren, für die

$$\lambda_1 Q_1 P Q_2' X_1' X_1 + \lambda_1 Q_1 Q_2' + B_1 = 0$$

gilt. Dabei ist wegen (4.9) und (4.10) $Q_1 P Q_2' = G_1 H_1$ und $Q_1 Q_2' = G_1$. Daraus folgt

$$\begin{aligned} \lambda_1 Q_1 P Q_2' (b_1 \cdots b_{k-1})' X' X (b_1 \cdots b_{k-1}) + \lambda_1 Q_1 Q_2' \\ + (a_1 \cdots a_{k-1})' B (b_1 \cdots b_{k-1}) = 0. \end{aligned}$$

Definiere $U_1 = (a_1 \cdots a_{k-1}) Q_1$ und $V_1 = (b_1 \cdots b_{k-1}) Q_2$. Es ist $U_1' U_1 = V_1' V_1 = I_{k-1}$ und $U_1' u = V_1' v = 0$. Dann gilt nach Multiplikation von links mit $(a_1 \cdots a_{k-1})$ und von rechts mit $(b_1 \cdots b_{k-1})'$

$$\begin{aligned} \lambda_1 U_1 P V_1' X' X (b_1 \cdots b_{k-1}) (b_1 \cdots b_{k-1})' + \lambda_1 U_1 V_1' \\ + (a_1 \cdots a_{k-1}) (a_1 \cdots a_{k-1})' B (b_1 \cdots b_{k-1}) (b_1 \cdots b_{k-1})' = 0. \end{aligned}$$

Da $\{a_1, \dots, a_{k-1}, u\}$ und $\{b_1, \dots, b_{k-1}, v\}$ jeweils eine Orthonormalbasis des \mathbb{R}^k bilden, gilt:

$$(a_1 \cdots a_{k-1}) (a_1 \cdots a_{k-1})' = I_k - uu'$$

bzw.

$$(b_1 \cdots b_{k-1}) (b_1 \cdots b_{k-1})' = I_k - vv'$$

Damit gilt

$$\lambda_1 U_1 P V_1' X' X (I_k - vv') + \lambda_1 U_1 V_1' + (I_k - uu') B (I_k - vv') = 0.$$

Daraus folgt

$$\begin{aligned} & \lambda_1 U_1 P V_1 X' X + \lambda_1 U_1 V_1' + (I_k - uu')B \\ & - [\lambda_1 U_1 P V_1 X' X + (I_k - uu')B] v v' = 0. \end{aligned} \quad (4.16)$$

Mit $v = -\frac{B'u}{\|B'u\|}$ und

$$U_1 P V_1' = (a_1 \cdots a_{k-1}) Q_1 P Q_2' (b_1 \cdots b_{k-1})' = W$$

gilt

$$\begin{aligned} & - [\lambda_1 W X' X + (I_k - uu')B] v v' \\ & = \frac{1}{\|B'u\|} [\lambda_1 W X' X B'u + (B B'u - u' B B' u u)] v' \\ & \stackrel{(3)}{=} 0 \end{aligned}$$

Mit Hilfe von $v = -\frac{B'u}{\|B'u\|}$ ergibt sich (4.16) zu

$$\lambda_1 (U_1 P V_1' X' X + U_1 V_1') + \lambda_2 u v' + B = 0.$$

Dabei ist $\lambda_2 = \sqrt{u B B' u}$. Wegen (1) handelt es sich hier um die gemischte Spektralgleichung für den Fall $j = k - 1$.

Durch die Singulärwertzerlegung von $B(\hat{C}_M X - I_k)$ gilt

$$\begin{aligned} B \hat{C}_M X &= \lambda_1 U_1 V_1' + \lambda_2 u v' + B \\ &\stackrel{(4.13)}{=} \lambda_1 U_1 V_1' - uu' B + B \\ &= \lambda_1 U_1 V_1' + (I - uu') B \\ &= \lambda_1 (a_1 \cdots a_{k-1}) G_1 (b_1 \cdots b_{k-1})' + (I - uu') B. \end{aligned}$$

Der MILE hat deshalb die Form

$$\hat{C}_M y = B^{-1} [\lambda_1 (a_1 \cdots a_{k-1}) G_1 (b_1 \cdots b_{k-1})' + (I - uu') B] (X' X)^{-1} X' y.$$

□

Es bleiben jedoch weiterhin die Probleme bestehen, dass man Kenntnis über den vorliegenden Fall und über den Eigenvektor zum kleinsten Eigenwert von $B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k) \right]'$ haben muss.

Kapitel 5

Darstellung von Hoffmann und Läuter

In den 70er Jahren haben Läuter [21] 1975 und Hoffmann [17] 1979 das Minimierungsproblem (2.4) auf eine Matrixgleichung mit Nebenbedingungen übertragen können. Drygas und Pilz [7] 1996 sowie Drygas und Läuter [4] 1994 konnten mit Hilfe der Bayes-Verfahren den Zusammenhang zwischen spektralen Methoden und der Darstellung von Hoffmann und Läuter herstellen. Hier ist die Möglichkeit gegeben die Gültigkeit der von Hoffmann und Läuter entwickelten Gleichung über die in Kapitel 4 beschriebenen spektralen Methoden herzuleiten.

Man gehe hier wiederum von den Voraussetzungen $\text{rg } X = \text{rg } B = k$ aus. Desweiteren werde $S = X'X$ und $F = S^{-1}B'BS^{-1}$ definiert. Damit sind S und F positiv definit.

Lemma 5.1 Seien $S, F \in \mathbb{R}^{k \times k}$ p.d., $A, M \in \mathbb{R}^{k \times k}$ n.n.d und $M = \sqrt{a}(F + A)^{\frac{1}{2}} - S^{-1}$ mit $a > 0$. Dann gilt

$$\begin{aligned} (M + S^{-1})^{-1}F(M + S^{-1})^{-1}M &= \frac{1}{a}M \\ \Leftrightarrow \frac{1}{\sqrt{a}}(F + A)^{-\frac{1}{2}}A &= SA. \end{aligned}$$

BEWEIS:

siehe ([4], S. 21) □

Satz 5.2 (Hoffmann-Lauter-Gleichung) Der Schatzer $\hat{\beta}_M = \hat{C}_M y$ ist der MILE fur β im Schatzproblem aus Definition 2.1 mit $\text{rg } X = k$ und $\text{rg } B = k$ genau dann, wenn eine reelle Zahl $a > 0$ und eine Matrix $A \in \mathbb{R}^{k \times k}$ n.n.d. existieren, fur die gelten:

- (1) $M_0 = \sqrt{a}(F + A)^{\frac{1}{2}} - S^{-1}$ n.n.d,
- (2) $\frac{1}{\sqrt{a}}(F + A)^{-\frac{1}{2}}A = SA$,
- (3) $\sqrt{a} \text{tr}(F + A)^{\frac{1}{2}} = 1 + \text{tr } S^{-1}$.

Der MILE ist dann durch $\hat{\beta}_M = \hat{C}_M y$ mit

$$\hat{C}_M = M_0 X' (I_n + X M_0 X')^{-1} y$$

gegeben.

BEWEIS:

Aquivalent zu $\hat{C}_M y$ MILE ist, dass $\hat{C}_M y$ die Bedingungen aus Satz 4.2 erfullt. Da $\text{rg } B = k$ gilt, ist der MILE durch $\hat{Z}_M = B(\hat{C}_M X - I)$ gema (2.7) eindeutig bestimmt. Es ist die Spektralgleichung

$$\hat{Z}_M V_1 P V_1' X' X + \hat{Z}_M + B = 0$$

erfullt.

„ \Rightarrow “

Definiere $M = V_1 P V_1'$ mit V_1 und P aus Satz 4.2. Dann gilt M n.n.d, $\text{tr } M = 1$,

$$\hat{Z}_M = -B(I_k + MS)^{-1}$$

und schließlich

$$\hat{Z}'_M \hat{Z}_M = (M + S^{-1})^{-1} F (M + S^{-1})^{-1}.$$

Da $\hat{Z}'_M \hat{Z}_M V_1 = \lambda_1^2 V_1$ ist, folgt

$$\hat{Z}'_M \hat{Z}_M M = \lambda_1^2 M. \quad (5.1)$$

Definiere $a = \frac{1}{\lambda_1^2}$ und $A = \frac{1}{a}(M + S^{-1})^2 - F$. Dann ist $a > 0$ und A n.n.d., da

$$\begin{aligned} A &= \frac{1}{a}(M + S^{-1})^2 - F \\ &= (M + S^{-1})[\lambda_1^2 I_k - (M + S^{-1})^{-1} F (M + S^{-1})^{-1}](M + S^{-1}) \\ &= (M + S^{-1})(\lambda_1^2 I_k - \hat{Z}'_M \hat{Z}_M)(M + S^{-1}). \end{aligned}$$

Weiter ist $M_0 = \sqrt{a}(F + A)^{\frac{1}{2}} - S^{-1} = M$ n.n.d und auf Grund von $\text{tr } M_0 = \text{tr } M = 1$ gilt $\sqrt{a} \text{tr } (F + A)^{\frac{1}{2}} = 1 + \text{tr } S^{-1}$. Damit sind (1) und (3) erfüllt.

Wegen $M = M_0$ und (5.1) gilt

$$(M_0 + S^{-1})^{-1} F (M_0 + S^{-1})^{-1} M_0 = \lambda_1^2 M_0$$

und durch Lemma 5.1 dann auch (2).

Die Darstellung des MILE folgt sofort aus (4.3) und (2.8).

„ \Leftarrow “

Aus der Darstellung des MILE

$$\hat{C}_M = M_0 X' (I_n + X M_0 X')^{-1}$$

folgt mit Lemma A.4 und (1)

$$\begin{aligned}
\hat{Z}'_M \hat{Z}_M &= (M_0 + S^{-1})^{-1} F (M_0 + S^{-1})^{-1} \\
&= \frac{1}{a} (F + A)^{-\frac{1}{2}} F (F + A)^{-\frac{1}{2}} \\
&= \frac{1}{a} I_k - \frac{1}{a} (F + A)^{-\frac{1}{2}} A (F + A)^{-\frac{1}{2}}.
\end{aligned} \tag{5.2}$$

Fur alle $x \in \mathbb{R}^k$ mit $\|x\| = 1$ gilt

$$x' \hat{Z}'_M \hat{Z}_M x \leq \frac{1}{a}, \tag{5.3}$$

da $\frac{1}{a} (F + A)^{-\frac{1}{2}} A (F + A)^{-\frac{1}{2}}$ n.n.d.. Desweiteren ist A singularur. Denn ware A regulur, ware (2) aquivalent zu

$$(F + A)^{\frac{1}{2}} = \frac{1}{\sqrt{a}} S^{-1}.$$

Durch Einsetzen in (3) ergabe sich der Widerspruch

$$\text{tr } S^{-1} = 1 + \text{tr } S^{-1}.$$

Dadurch ist mit (5.2)

$$\begin{aligned}
&\det \left(\frac{1}{a} I_k - \hat{Z}'_M \hat{Z}_M \right) \\
&= \det \frac{1}{a} (F + A)^{-\frac{1}{2}} A (F + A)^{-\frac{1}{2}} = 0.
\end{aligned}$$

Daraus folgt, dass $\frac{1}{a}$ ein Eigenwert von $\hat{Z}'_M \hat{Z}_M$ ist und mit (5.3)

$$\lambda_{\max}(\hat{Z}'_M \hat{Z}_M) = \frac{1}{a}.$$

Durch (2) und Lemma 5.1 gilt

$$\hat{Z}'_M \hat{Z}_M M_0 = \lambda_{\max}(\hat{Z}'_M \hat{Z}_M) M_0$$

und mit (3) folgt $\text{tr } M_0 = 1$. Damit kann $M_0 = V_1 P V_1'$ dargestellt werden, wobei $P \in \mathbb{R}^{j \times j}$ eine positiv definite Diagonalmatrix mit $\text{tr } P = 1$ ist und die

Spalten von V_1 aus den normierten Eigenvektoren von $\hat{Z}'_M \hat{Z}_M$ zum maximalen Eigenwert bestehen. Mit

$$\begin{aligned} & \hat{Z}_M M_0 X' X + \hat{Z}_M + B \\ = & -B(I_k + M_0 X' X)^{-1} M_0 X' X - B(I_k + M_0 X' X)^{-1} + B \\ = & B [I_k - (I_k + M_0 X' X)^{-1} (I_k + M_0 X' X)] = 0 \end{aligned}$$

ist die rechte Spektralgleichung erfüllt. \square

Läuter und Hoffmann verwendeten in ihren Beweisen aus den Jahren 1975 und 1979 die Minimax-Risikofunktion bzw. Funktionen mit gleichen Minimierungseigenschaften.

Direkt aus dem Beweis von Satz 5.2 ist zu erkennen, dass M_0 aus Satz 5.2 (1) die Form

$$M_0 = V_1 P V_1'$$

besitzt, wobei V_1 die j Eigenvektoren zum maximalen Eigenwert von $\hat{Z}'_M \hat{Z}_M$ beinhaltet und P eine Diagonalmatrix mit positiven Einträgen ist (siehe Satz 4.2). Dadurch ist $\text{rg } M_0 = j$ und man kann an M_0 erkennen, in welchem Fall man sich befindet. Bezüglich des Bayes'schen Zugangs stimmt die Matrix M_0 mit der zweiten Moment-Matrix der ungünstigsten a priori Verteilung überein (vgl. [4], Theorem 6). Auch zwischen der Matrix A und der Spektraldarstellung des Minimax-Schätzers kann ein Zusammenhang aufgedeckt werden. Sei für $(\lambda_1 I_{k-j} - \Lambda_2)$ p.d.

$$\hat{Z}_M = \lambda_1 U_1 V_1' + U_2 \Lambda_2 V_2'$$

die Singulärwertzerlegung von \hat{Z}_M , so ist nach (vgl. [9], S. 312)

$$A = \begin{cases} S^{-1} V_2 (\lambda_1^2 I_{k-j} - \Lambda_2^2) V_2' S^{-1} & , \text{ Fall } j < k \text{ liegt vor} \\ 0 & , \text{ Fall } j = k \text{ liegt vor.} \end{cases} \quad (5.4)$$

Dadurch kann sofort auf $\operatorname{rg} A = k - j$ geschlossen werden, das heißt, man kann bei Kenntnis von A auf den vorliegenden Fall schließen.

Zur Bestimmung von A muss aber die Hoffmann-Läuter-Gleichung aus Satz 5.2 gelöst werden, das heißt, man muss eine nichtlineare Matrixgleichung mit Nebenbedingung lösen.

Kapitel 6

Gemeinsamer Eigenvektor

Drygas untersuchte in ([8], S. 9ff) das lineare Regressionsmodell bei ellipsoider Vorinformation mit

$$X'X = \begin{pmatrix} \frac{15}{44} & \frac{9}{11} & 0 \\ \frac{9}{11} & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

und für $\mu > 0$

$$B = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \mu \end{pmatrix}.$$

Das Vorliegen der Fälle $j = 1$ bzw. $j = 3$ konnte in Abhängigkeit von μ angegeben werden. Für $\mu > 2$ liegt der Fall $j = 1$ vor und für etwa $0.51048 \leq \mu \leq 0.87525$ liegt der Fall $j = 3$ vor. Da für $\mu < 0.51048$ und $0.87525 < \mu < 2$ die Fälle $j = 1$ und $j = 3$ ausgeschlossen werden können, muss hier der Fall $j = 2$ vorliegen. Aber dennoch konnte im Fall $j = 2$ nur für $\mu < 0.51048$ der Minimax-Schätzer in geschlossener Form angegeben werden.

In diesem Beispiel fällt auf, dass $X'X$ und B den gemeinsamen Eigenvektor $(0 \ 0 \ 1)'$ besitzen. Im Folgenden soll eine Verallgemeinerung dieser Situation stattfinden, jedoch wird die Anzahl der Einflussgrößen von drei beibehalten. Es ist dann möglich alle Fälle zu identifizieren. Außerdem wird eine Strukturierung dieses Problems in Abhängigkeit zum Eigenwert des Eigenvektors $(0 \ 0 \ 1)'$ der Gewichtsmatrix aus der Verlustfunktion angestrebt.

6.1 Falluntersuchung

Wie schon in Kapitel 2.1 angemerkt wurde, kann die Gewichtsmatrix B aus Definition 2.1 ohne Einschränkung als Diagonalmatrix verwendet werden. Da die Einträge der Diagonalen einer positiv definiten Diagonalmatrix positiv sind, kann durch entsprechende Division ein Eintrag auf 1 gebracht werden. Dies verändert zwar das Risiko, aber die Stelle, an der die Risikofunktion aus (2.4) minimiert wird, bleibt unverändert.

Definition 6.1 *Das Schätzproblem aus Definition 2.1 mit*

$$X'X = \begin{pmatrix} s_{11} & s_{12} & 0 \\ s_{12} & s_{22} & 0 \\ 0 & 0 & s_3 \end{pmatrix}, \quad (6.1)$$

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & \mu \end{pmatrix}, \quad (6.2)$$

$s_{11}, s_{22}, s_3 > 0$, $0 < s_{12}^2 < s_{11} s_{22}$, $b \in (0, 1)$ und $\mu > 0$ heißt dreidimensionales lineares Regressionsmodell mit ellipsoider Vorinformation bei einem gemeinsamen Eigenvektor.

Man beachte, dass $X'X$ und $B'B$ positiv definit sind. Zuerst werden die Bedingungen für das Vorliegen des Falls $j = 1$ ermittelt. Dazu definiere man

$$S = X'X,$$

$$\tilde{S} = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} \quad (6.3)$$

und

$$\tilde{B} = \begin{pmatrix} 1 & 0 \\ 0 & b \end{pmatrix}. \quad (6.4)$$

Lemma 6.2 *Seien die Voraussetzungen aus Definition 6.1 erfüllt. Dann gilt:*

(a)

$$\lambda_{\max} [(I_3 - uu')BB'(I_3 - uu')] = 1,$$

falls $u = (0 \ 0 \ 1)'$,

(b)

$$\lambda_{\max} [(I_3 - uu')BB'(I_3 - uu')] \\ = \max \left\{ \lambda_{\max} [(I_2 - \tilde{u}\tilde{u}')\tilde{B}\tilde{B}'(I_2 - \tilde{u}\tilde{u}')], \mu^2 \right\},$$

falls $u = \begin{pmatrix} \tilde{u} \\ 0 \end{pmatrix}$, $\tilde{u} \in \mathbb{R}^2$ und $\|\tilde{u}\| = 1$.

BEWEIS:

(a)

$$[(I_3 - uu')BB'(I_3 - uu')] \\ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & b^2 & 0 \\ 0 & 0 & \mu^2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Daraus folgt sofort die Behauptung.

(b) Es ist

$$(I_3 - uu') = \begin{pmatrix} I_2 - \tilde{u}\tilde{u}' & 0 \\ 0 & 1 \end{pmatrix}.$$

Mit

$$\begin{aligned} & [(I_3 - uu')BB'(I_3 - uu')] \\ &= \begin{pmatrix} (I_2 - \tilde{u}\tilde{u}')\tilde{B}\tilde{B}'(I_2 - \tilde{u}\tilde{u}') & 0 \\ 0 & \mu^2 \end{pmatrix} \end{aligned}$$

folgt die Behauptung. \square

Nun lässt sich mit Hilfe von Satz 4.5 angeben, für welche μ der Fall $j = 1$ vorliegt.

Satz 6.3 *Es seien die Voraussetzungen aus Definition 6.1 gegeben. Der Fall $j=1$ liegt genau dann vor, wenn*

(1)

$$\mu \geq 1 + s_3$$

oder

(2)

$$\mu \leq \sqrt{\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u}} \lambda_{\max} [\tilde{B}(I + \tilde{S})^{-1}\tilde{B}]$$

und

$$\begin{aligned} & \tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u} \lambda_{\max}^2 [\tilde{B}(I + \tilde{S})^{-1}\tilde{B}] \\ & \geq \lambda_{\max}(I_2 - \tilde{u}\tilde{u}')\tilde{B}\tilde{B}'(I_2 - \tilde{u}\tilde{u}'). \end{aligned}$$

Dabei ist \tilde{u} der normierte Eigenvektor zum maximalen Eigenwert von $\tilde{B}(I + \tilde{S})^{-1}\tilde{B}$.

BEWEIS:

„ \Rightarrow “

Wenn der Fall $j = 1$ vorliegt, gilt gemäß Satz 4.5 für $\varrho = \lambda_{\max}[B(I_3 + X'X)^{-1}B']$ mit zugehörigem normierten Eigenvektor u

$$(1) \quad u'(BB')^{-1}u\varrho^2 \geq \lambda_{\max}[(I_3 - uu')(BB')(I_3 - uu')],$$

(2) ϱ ist einfach.

Wegen

$$B(I_3 + X'X)^{-1}B' = \begin{pmatrix} \tilde{B}(I_2 + \tilde{S})^{-1}\tilde{B} & 0 \\ 0 & \frac{\mu^2}{1+s_3} \end{pmatrix}$$

und (2) ist

$$\varrho = \begin{cases} \frac{\mu^2}{1+s_3} & , \quad \frac{\mu^2}{1+s_3} > \tilde{\varrho} \\ \tilde{\varrho} & , \quad \frac{\mu^2}{1+s_3} < \tilde{\varrho} \end{cases}$$

mit $\tilde{\varrho} = \lambda_{\max}[\tilde{B}(I_2 + \tilde{S})^{-1}\tilde{B}]$.

Für $\varrho = \frac{\mu^2}{1+s_3}$ ist der zugehörige normierte Eigenvektor $u = (0 \ 0 \ 1)'$. Daraus folgt mit (1) und Lemma 6.2 (a)

$$u'(BB')^{-1}u\varrho^2 = \frac{\mu^2}{(1+s_3)^2} \geq 1$$

und damit auch $\mu \geq 1 + s_3$.

Für den Fall $\varrho = \tilde{\varrho}$ hat der zugehörige normierte Eigenvektor die Form $u = \begin{pmatrix} \tilde{u} \\ 0 \end{pmatrix}$ mit $\tilde{u} \in \mathbb{R}^2$ und $\|\tilde{u}\| = 1$. Durch (1) und Lemma 6.2 (b) gilt dann:

$$\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u}\tilde{\varrho}^2 \geq \max\left\{\lambda_{\max}\left[(I_2 - \tilde{u}\tilde{u}')\tilde{B}\tilde{B}'(I_2 - \tilde{u}\tilde{u}')\right], \mu^2\right\}.$$

Daraus folgt sofort

$$\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u}\tilde{\varrho}^2 \geq \lambda_{\max}\left[(I_2 - \tilde{u}\tilde{u}')\tilde{B}\tilde{B}'(I_2 - \tilde{u}\tilde{u}')\right]$$

und

$$\mu \leq \sqrt{\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u}} \tilde{\varrho}.$$

„ \Leftarrow “

Sei $\mu \geq 1 + s_3$. Dann ist

$$\frac{\mu^2}{1 + s_3} \geq 1 + s_3 > 1 = \lambda_{\max}(\tilde{B}'\tilde{B}) \stackrel{\text{Lemma A.5 (a)}}{>} \tilde{\varrho}.$$

Damit ist $\varrho = \frac{\mu^2}{1+s_3}$ einfach und $u = (0 \ 0 \ 1)'$ der zugehörige normierte Eigenvektor. Durch

$$u'(BB')^{-1}u \varrho^2 = \frac{\mu^2}{(1 + s_3)^2} \geq 1 = \lambda_{\max}[(I_3 - uu')(BB')(I_3 - uu')]$$

wird (1) erfüllt.

Es gelte $\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u} \tilde{\varrho}^2 \geq \lambda_{\max}[(I_2 - \tilde{u}\tilde{u}')\tilde{B}'\tilde{B}(I_2 - \tilde{u}\tilde{u}')] \tilde{\varrho}^2$ und $\mu \leq \sqrt{\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u}} \tilde{\varrho}$. Daraus folgt

$$\frac{\mu^2}{1 + s_3} \leq \frac{\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u} \tilde{\varrho}^2}{1 + s_3} < \tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u} \tilde{\varrho}^2 \stackrel{\text{Lemma A.5 (b)}}{<} \tilde{\varrho}.$$

Dadurch ist $\varrho = \tilde{\varrho}$ einfach. Mit dem zugehörigen Eigenvektor $u = \begin{pmatrix} \tilde{u} \\ 0 \end{pmatrix}$ ist

(1) erfüllt, denn es gilt

$$\begin{aligned} u'(BB')^{-1}u \varrho^2 &= \tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u} \tilde{\varrho}^2 \\ &\geq \max \left\{ \lambda_{\max}[(I_2 - \tilde{u}\tilde{u}')\tilde{B}'\tilde{B}(I_2 - \tilde{u}\tilde{u}')] , \mu^2 \right\} \\ &= \lambda_{\max}[(I_3 - uu')(BB')(I_3 - uu')]. \end{aligned}$$

□

Für $\mu \geq 1 + s_3$ ist der MILE durch

$$\hat{C}_M y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{1+s_3} \end{pmatrix} X' y$$

gegeben. In der Situation

$$\mu \leq \sqrt{\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u}} \tilde{\varrho}$$

und

$$\tilde{u}'(\tilde{B}'\tilde{B})^{-1}\tilde{u} \lambda_{\max}^2 \left[\tilde{B}(I + \tilde{S})^{-1}\tilde{B} \right] \geq \lambda_{\max}(I_2 - \tilde{u}\tilde{u}')\tilde{B}\tilde{B}'(I_2 - \tilde{u}\tilde{u}')$$

hat der MILE die Darstellung

$$\hat{C}_M y = \begin{pmatrix} \tilde{\varrho}(\tilde{B}^{-1}\tilde{u})'\tilde{B}^{-1}\tilde{u} & 0 \\ 0 & 0 \end{pmatrix} X'y.$$

Für die Betrachtung des Falls $j = 3$ werden zuerst

$$F = S^{-1}B'BS^{-1}$$

und

$$\kappa = \lambda_{\min} \left(S^{\frac{1}{2}}F^{\frac{1}{2}}S^{\frac{1}{2}} \right)$$

definiert. Entsprechend zu (6.3) und (6.4) schreibt man

$$\tilde{F} = \tilde{S}^{-1}\tilde{B}'\tilde{B}\tilde{S}^{-1} \tag{6.5}$$

und

$$\tilde{\kappa} = \lambda_{\min} \left(\tilde{S}^{\frac{1}{2}}\tilde{F}^{\frac{1}{2}}\tilde{S}^{\frac{1}{2}} \right). \tag{6.6}$$

Daraus erhält man

$$F = \begin{pmatrix} \tilde{F} & 0 \\ 0 & \frac{\mu^2}{s_3^2} \end{pmatrix}$$

und

$$S^{\frac{1}{2}}F^{\frac{1}{2}}S^{\frac{1}{2}} = \begin{pmatrix} \tilde{S}^{\frac{1}{2}}\tilde{F}^{\frac{1}{2}}\tilde{S}^{\frac{1}{2}} & 0 \\ 0 & \mu \end{pmatrix}.$$

So ergibt sich $\kappa = \min\{\tilde{\kappa}, \mu\}$.

Satz 6.4 Die Voraussetzungen aus Definition 6.1 seien erfüllt. Dann liegt der Fall $j = 3$ genau dann vor, wenn $\tilde{\kappa} \geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}$ und

$$\mu \in \left[\frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}, \tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}} \right]$$

sind.

BEWEIS:

Das Eintreten des Falls $j = 3$ ist nach Korollar 4.7 äquivalent zu $\kappa \geq \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}}$

und es gilt

$$\begin{aligned} \kappa &\geq \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} \\ \Leftrightarrow \min\{\tilde{\kappa}, \mu\} &\geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}} + \frac{\mu}{s_3}}{1 + \operatorname{tr} \tilde{S}^{-1} + \frac{1}{s_3}} \\ \Leftrightarrow \tilde{\kappa} &\geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}} + \frac{\mu}{s_3}}{1 + \operatorname{tr} \tilde{S}^{-1} + \frac{1}{s_3}} \quad \wedge \quad \mu \geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}} + \frac{\mu}{s_3}}{1 + \operatorname{tr} \tilde{S}^{-1} + \frac{1}{s_3}} \\ \Leftrightarrow \mu &\leq \tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}} \\ &\wedge \quad \mu \geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}} \\ \Leftrightarrow \mu &\in \left[\frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}, \tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}} \right]. \end{aligned}$$

Damit das Intervall nicht leer ist, muss gelten

$$\begin{aligned} \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}} &\leq \tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}} \\ \Leftrightarrow \tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) &\geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1})}{1 + \operatorname{tr} \tilde{S}^{-1}} \\ \Leftrightarrow \tilde{\kappa} &\geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}. \end{aligned}$$

□

Können die Fälle $j = 1$ und $j = 3$ ausgeschlossen werden, so befindet man sich im Fall $j = 2$. In Satz 6.3 (2) und Satz 6.4 sind von μ unabhängi-

ge Bedingungen gegeben, wann dieser Fall überhaupt eintreten kann. Im nachfolgenden Lemma wird gezeigt, dass sich diese Bedingungen gegenseitig ausschließen.

Lemma 6.5 *Die Voraussetzungen aus Definition 6.1 seien gegeben. Dann sind*

$$\tilde{\kappa} < \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}} \quad (6.7)$$

und

$$\tilde{u}(\tilde{B}'\tilde{B})^{-1}\tilde{u} \tilde{\varrho}^2 \geq \lambda_{\max} \left[(I_2 - \tilde{u}\tilde{u}')\tilde{B}\tilde{B}'(I_2 - \tilde{u}\tilde{u}') \right] \quad (6.8)$$

äquivalent.

BEWEIS:

Betrachtet man das Schätzproblem aus Definition 2.1 mit dem Modell $y = \tilde{X}\beta + \varepsilon$ und \tilde{B} als Gewichtsmatrix der Verlustfunktion sowie \tilde{X} derart, dass $\tilde{S} = \tilde{X}'\tilde{X}$ ist, so hat man ein Modell mit zwei Einflussgrößen. Die Bedingung (6.8) ist äquivalent zum Vorliegen des Falls $j = 1$. Damit kann der noch übrigbleibende Fall $j = 2$ nicht eintreten und (6.7) ist äquivalent zur Umkehrung, dass der Fall $j = 2$ eintritt. \square

Dies führt nun zu zwei unterschiedlichen Strukturen dieses Problems. Damit erhält man folgende Übersicht.

Struktur I	$\tilde{\kappa} \geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}$
j=1	$\mu \geq 1 + s_3$
j=2	$\mu \in \left(\tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}}, 1 + s_3 \right)$
j=3	$\mu \in \left[\frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}, \tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}} \right]$
j=2	$\mu \in \left(0, \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}} \right)$

Struktur II	$\tilde{\kappa} < \frac{\text{tr } \tilde{F}^{\frac{1}{2}}}{1 + \text{tr } \tilde{S}^{-1}}$
j=1	$\mu \geq 1 + s_3$
j=2	$\mu \in \left(\sqrt{\tilde{u}(\tilde{B}'\tilde{B})^{-1}\tilde{u}\tilde{q}}, 1 + s_3 \right)$
j=1	$\mu \in \left(0, \sqrt{\tilde{u}(\tilde{B}'\tilde{B})^{-1}\tilde{u}\tilde{q}} \right]$

Anhand der Einträge der Matrizen ist es jetzt möglich das Vorliegen eines Falls zu erkennen. Weiter weiß man, dass μ , also der Eigenwert zum Eigenvektor $(0 \ 0 \ 1)'$ der Gewichtsmatrix B , eine bedeutende Rolle spielt. Aber die Entscheidung, welche Struktur diesem Problem zu Grunde liegt, hängt nur von den Matrizen \tilde{B} und \tilde{S} ab. In Kapitel 7.1 kann dann sogar gezeigt werden, dass das Vorliegen einer Struktur durch b in Abhängigkeit der Einträge aus \tilde{S} angegeben werden kann.

6.2 Orthogonal zum maximalen Eigenraum

Wie im letzten Abschnitt gesehen kann der Fall $j = 2$ in unterschiedlichen Situationen auftreten. Für die Situation aus Struktur I mit $\mu < \frac{\text{tr } \tilde{F}^{\frac{1}{2}}}{1 + \text{tr } \tilde{S}^{-1}}$ konnte bereits durch Drygas ([9], S. 313ff) eine Lösung dieses Problems gefunden werden. Dabei fand die Darstellung von Hoffmann und Läuter Anwendung. Es ist aber auch möglich diesen Sachverhalt mittels Satz 4.8 zu zeigen. Dazu nehme man an, dass $u_3 = (0 \ 0 \ 1)'$ derjenige Vektor ist, der senkrecht auf dem Eigenraum des maximalen Eigenwerts von $\left[B(\hat{C}_M X - I_3) \right]' B(\hat{C}_M X - I_3)$ steht. Dabei ist $\hat{C}_M y$ der MILE. Nun kann man $\{(1 \ 0 \ 0)', (0 \ 1 \ 0)'\}$ als Orthonormalbasis des Eigenraums des maximalen Eigenwerts von $\left[B(\hat{C}_M X - I_3) \right]' B(\hat{C}_M X - I_3)$ wählen. Damit ergibt sich gemäß Satz 4.8 der Vektor $v_3 = (0 \ 0 \ -1)'$. Die Vektoren $(1 \ 0 \ 0)', (0 \ 1 \ 0)'$ und

v_3 bilden eine weitere Orthonormalbasis des \mathbb{R}^3 . Man erhält

$$X_1'X_1 = \begin{pmatrix} I_2 & 0 \end{pmatrix} \begin{pmatrix} \tilde{S} & 0 \\ 0 & s_3 \end{pmatrix} \begin{pmatrix} I_2 \\ 0 \end{pmatrix} = \tilde{S}$$

und

$$B_1 = \begin{pmatrix} I_2 & 0 \end{pmatrix} \begin{pmatrix} \tilde{B} & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} I_2 \\ 0 \end{pmatrix} = \tilde{B}.$$

Die Überprüfung der drei Bedingungen aus Satz 4.8 ergibt folgendes:

(1)

$$\begin{aligned} \lambda_1^2 &= \left\{ \frac{\text{tr} [(X_1'X_1)^{-1}B_1'B_1(X_1'X_1)^{-1}]^{\frac{1}{2}}}{1 + \text{tr} (X_1'X_1)^{-1}} \right\}^2 \\ &= \left(\frac{\text{tr} \tilde{F}^{\frac{1}{2}}}{1 + \text{tr} \tilde{S}^{-1}} \right)^2 > \mu^2 = u_3' B' B u_3. \end{aligned}$$

(2)

$$\begin{aligned} \tilde{\kappa} &\geq \frac{\text{tr} \tilde{F}^{\frac{1}{2}}}{1 + \text{tr} \tilde{S}^{-1}} \\ &\Leftrightarrow \tilde{S}^{\frac{1}{2}} \tilde{F}^{\frac{1}{2}} \tilde{S}^{\frac{1}{2}} - \frac{\text{tr} \tilde{F}^{\frac{1}{2}}}{1 + \text{tr} \tilde{S}^{-1}} I_2 \quad \text{n.n.d} \\ &\Leftrightarrow \frac{1}{\lambda_1} \tilde{S}^{\frac{1}{2}} \tilde{F}^{\frac{1}{2}} \tilde{S}^{\frac{1}{2}} - I_2 \quad \text{n.n.d} \\ &\Leftrightarrow \frac{1}{\lambda_1} \tilde{F}^{\frac{1}{2}} - \tilde{S}^{-1} \quad \text{n.n.d.} \end{aligned}$$

(3) Es folgt durch

$$G_1 = -\tilde{B}\tilde{S}^{-1}\tilde{F}^{-\frac{1}{2}}$$

und

$$W = \begin{pmatrix} G_1 H_1 & 0 \\ 0 & 0 \end{pmatrix},$$

dass

$$\lambda_1 W X' X B' u_3 = 0$$

ist. Mit

$$\begin{pmatrix} \tilde{B}' \tilde{B} & 0 \\ 0 & \mu^2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \mu^2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

ist auch diese Bedingung erfüllt.

Damit ist gezeigt, dass Satz 4.8 angewendet werden kann. Mit einer entsprechenden Singulärwertzerlegung ist es möglich den MILE $\hat{\beta}_M$ zu berechnen.

Kapitel 7

Im maximalen Eigenraum

Im Kapitel 6.2 spielt der gemeinsame Eigenvektor $(0 \ 0 \ 1)'$ von $X'X$ und $B'B$ eine bedeutende Rolle. Dieser gemeinsame Eigenvektor ist ebenfalls ein normierter Eigenvektor zu einem Eigenwert von $B(\hat{C}_M X - I_3) \left[B(\hat{C}_M X - I_3) \right]'$ und steht senkrecht auf dem Eigenraum zum maximalen Eigenwert von $B(\hat{C}_M X - I_3) \left[B(\hat{C}_M X - I_3) \right]'$. Für die noch ungelösten Fälle $j = 2$ aus Struktur I und Struktur II, die in Kapitel 6.1 beschrieben werden, liegt nahe, dass dieser gemeinsame Eigenvektor ein Eigenvektor zu $B(\hat{C}_M X - I_3) \left[B(\hat{C}_M X - I_3) \right]'$ ist. Jedoch wird er diesmal Eigenvektor zum maximalen Eigenwert sein. Unter der Annahme, dass für $\mu \in \left(\tilde{\kappa} (1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}}, 1 + s_3 \right)$ der Vektor $(0 \ 0 \ 1)'$ senkrecht auf dem Eigenraum zum maximalen Eigenwert von $B(\hat{C}_M X - I_3) \left[B(\hat{C}_M X - I_3) \right]'$ stehen würde, erhält man analog zu Kapitel 6.2, dass $X'_1 X_1 = \tilde{S}$ und $B_1 = \tilde{B}$ sind. Es gilt dann

$$\begin{aligned}\lambda_1^2 &= \left\{ \frac{\operatorname{tr} [(X_1' X_1)^{-1} B_1' B_1 (X_1' X_1)^{-1}]^{\frac{1}{2}}}{1 + \operatorname{tr} (X_1' X_1)^{-1}} \right\}^2 \\ &= \left(\frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}} \right)^2 < \mu^2 = u_3' B' B u_3.\end{aligned}$$

Dies führt sofort zu einer Verletzung von (1) aus Satz 4.8.

Im Folgenden wird sich auf die Struktur I und $s_{12} > 0$ beschränkt um die eben angeführte Vermutung, dass $(0 \ 0 \ 1)'$ im Eigenraum zum maximalen Eigenwert von $\left[B(\hat{C}_M X - I_3) \right]' B(\hat{C}_M X - I_3)$ liegt, zu bestätigen. Außerdem werden die Kriterien aus Satz 4.8 zur Bestimmung des MILE vereinfacht. Abschließend wird noch ein kurzer Ausblick auf die Situation im Fall $j = 2$ ohne gemeinsamen Eigenvektor getätigt.

7.1 Einträge der Matrizen

Die Bestätigung der Vermutung, dass $(0 \ 0 \ 1)'$ im Eigenraum zum maximalen Eigenwert von $B(\hat{C}_M X - I_3) \left[B(\hat{C}_M X - I_3) \right]'$ liegt und die Vereinfachung der Kriterien aus Satz 4.8 wird durch Terme, in denen die Einträge der Matrizen \tilde{S} und \tilde{B} aus (6.3) und (6.4) vorkommen, durchgeführt. In diesem Abschnitt werden Ausdrücke, die in Kapitel 6 bisher durch Matrizen bzw. Funktionen von Matrizen dargestellt wurden, in Abhängigkeit von den Einträgen dieser Matrizen angeben. Weiterhin werden für die nächsten Abschnitte wichtige Funktionen, die von den Einträgen abhängen, bereitgestellt.

Lemma 7.1 *Für \tilde{S} , \tilde{B} und \tilde{F} aus (6.3), (6.4) und (6.5) gilt:*

(a)

$$\operatorname{tr} \tilde{S}^{-1} = \frac{s_{11} + s_{22}}{s_{11} s_{22} - s_{12}^2},$$

(b)

$$\operatorname{tr} \tilde{F} = \frac{b^2 s_{11}^2 + (1 + b^2) s_{12}^2 + s_{22}^2}{(s_{11} s_{22} - s_{12}^2)^2},$$

(c)

$$\operatorname{tr} \tilde{F}^{\frac{1}{2}} = \frac{\sqrt{(b s_{11} + s_{22})^2 + (1 - b)^2 s_{12}^2}}{s_{11} s_{22} - s_{12}^2}.$$

BEWEIS:

(a) Es ist

$$\tilde{S}^{-1} = \begin{pmatrix} \frac{s_{22}}{s_{11} s_{22} - s_{12}^2} & \frac{-s_{12}}{s_{11} s_{22} - s_{12}^2} \\ \frac{-s_{12}}{s_{11} s_{22} - s_{12}^2} & \frac{s_{11}}{s_{11} s_{22} - s_{12}^2} \end{pmatrix}$$

und damit

$$\operatorname{tr} \tilde{S}^{-1} = \frac{s_{11} + s_{22}}{s_{11} s_{22} - s_{12}^2}.$$

(b) Aus

$$\begin{aligned} \tilde{F} &= \tilde{S}^{-1} \tilde{B}' \tilde{B} \tilde{S}^{-1} \\ &= \begin{pmatrix} \frac{s_{22}}{s_{11} s_{22} - s_{12}^2} & \frac{-s_{12}}{s_{11} s_{22} - s_{12}^2} \\ \frac{-s_{12}}{s_{11} s_{22} - s_{12}^2} & \frac{s_{11}}{s_{11} s_{22} - s_{12}^2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & b^2 \end{pmatrix} \begin{pmatrix} \frac{s_{22}}{s_{11} s_{22} - s_{12}^2} & \frac{-s_{12}}{s_{11} s_{22} - s_{12}^2} \\ \frac{-s_{12}}{s_{11} s_{22} - s_{12}^2} & \frac{s_{11}}{s_{11} s_{22} - s_{12}^2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{s_{22}^2 + b^2 s_{12}^2}{(s_{11} s_{22} - s_{12}^2)^2} & -\frac{s_{12} s_{22} + b^2 s_{11} s_{12}}{(s_{11} s_{22} - s_{12}^2)^2} \\ -\frac{s_{12} s_{22} + b^2 s_{11} s_{12}}{(s_{11} s_{22} - s_{12}^2)^2} & \frac{s_{12}^2 + b^2 s_{11}^2}{(s_{11} s_{22} - s_{12}^2)^2} \end{pmatrix} \end{aligned}$$

folgt

$$\operatorname{tr} \tilde{F} = \frac{b^2 s_{11}^2 + (1 + b^2) s_{12}^2 + s_{22}^2}{(s_{11} s_{22} - s_{12}^2)^2}.$$

(c) Seien l_1 und l_2 die Eigenwerte von \tilde{F} . Da \tilde{F} positiv definit ist, sind $l_1, l_2 > 0$. Damit gilt dann

$$\begin{aligned} \left(\operatorname{tr} \tilde{F}^{\frac{1}{2}}\right)^2 &= \left(\sqrt{l_1} + \sqrt{l_2}\right)^2 = l_1 + l_2 + 2\sqrt{l_1 l_2} \\ &= \operatorname{tr} \tilde{F} + 2 \det \tilde{F}^{\frac{1}{2}}. \end{aligned} \tag{7.1}$$

Daraus folgt

$$\operatorname{tr} \tilde{F}^{\frac{1}{2}} = \sqrt{\operatorname{tr} \tilde{F} + 2 \det \tilde{F}^{\frac{1}{2}}}$$

und mit

$$\begin{aligned} \det \tilde{F}^{\frac{1}{2}} &= \sqrt{\det \tilde{S}^{-1} \tilde{B}' \tilde{B} \tilde{S}^{-1}} \\ &= \frac{\det \tilde{B}}{\det \tilde{S}} = \frac{b}{s_{11} s_{22} - s_{12}^2} \end{aligned} \quad (7.2)$$

erhält man

$$\begin{aligned} \operatorname{tr} \tilde{F}^{\frac{1}{2}} &= \sqrt{\frac{b^2 s_{11}^2 + (1+b^2) s_{12}^2 + s_{22}^2}{(s_{11} s_{22} - s_{12}^2)^2}} + \frac{2b}{s_{11} s_{22} - s_{12}^2} \\ &= \frac{\sqrt{b^2 s_{11}^2 + 2b s_{11} s_{22} + s_{22}^2 + (1-2b+b^2) s_{12}^2}}{s_{11} s_{22} - s_{12}^2} \\ &= \frac{\sqrt{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}}{s_{11} s_{22} - s_{12}^2}. \end{aligned}$$

□

Lemma 7.2 *Definiere $\tilde{M} = \tilde{S}^{\frac{1}{2}} \tilde{F}^{\frac{1}{2}} \tilde{S}^{\frac{1}{2}}$. Dann ist*

$$\operatorname{tr} \tilde{M} = \frac{(1+b)(b s_{11} + s_{22})}{\sqrt{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}}.$$

BEWEIS:

Sei nun

$$\tilde{F} = \begin{pmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{pmatrix}$$

und

$$\tilde{F}^{\frac{1}{2}} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}.$$

Dann ergeben sich folgende Gleichungen

$$\tilde{F}^{\frac{1}{2}} \tilde{F}^{\frac{1}{2}} = \begin{pmatrix} h_{11}^2 + h_{12}^2 & h_{12}(h_{11} + h_{22}) \\ h_{12}(h_{11} + h_{22}) & h_{22}^2 + h_{12}^2 \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{pmatrix} = \tilde{F}.$$

Durch Auflösen nach h_{12} , h_{11} und h_{22} erhält man

$$h_{12} = \frac{f_{12}}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}},$$

$$h_{11} = \sqrt{f_{11} - \frac{f_{12}^2}{(\operatorname{tr} \tilde{F}^{\frac{1}{2}})^2}}$$

und

$$h_{22} = \sqrt{f_{22} - \frac{f_{12}^2}{(\operatorname{tr} \tilde{F}^{\frac{1}{2}})^2}}.$$

Da $\tilde{F}^{\frac{1}{2}}$ positiv definit ist, müssen $h_{11}, h_{22} > 0$ sein. Weiter ist

$$\begin{aligned} h_{11} &= \frac{1}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}} \sqrt{(\operatorname{tr} \tilde{F}^{\frac{1}{2}})^2 f_{11} - f_{12}^2} \\ &\stackrel{(7.1)}{=} \frac{1}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}} \sqrt{f_{11}^2 + f_{11} f_{22} + 2 f_{11} \det \tilde{F}^{\frac{1}{2}} - f_{12}^2} \\ &= \frac{1}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}} \sqrt{f_{11}^2 + 2 f_{11} \det \tilde{F}^{\frac{1}{2}} + \det \tilde{F}} \\ &= \frac{1}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}} (f_{11} + \det \tilde{F}^{\frac{1}{2}}). \end{aligned}$$

Analog erhält man für

$$h_{22} = \frac{1}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}} (f_{22} + \det \tilde{F}^{\frac{1}{2}}).$$

Damit ergibt sich

$$\begin{aligned} \operatorname{tr} \tilde{M} &= \operatorname{tr} \tilde{S}^{\frac{1}{2}} \tilde{F}^{\frac{1}{2}} \tilde{S}^{\frac{1}{2}} = \operatorname{tr} \tilde{S} \tilde{F}^{\frac{1}{2}} \\ &= s_{11} h_{11} + s_{22} h_{22} + 2 s_{12} h_{12} \\ &\stackrel{(7.2)}{=} \frac{s_{11} f_{11} + s_{22} f_{22} + b \frac{s_{11} + s_{22}}{s_{11} s_{22} - s_{12}^2} + 2 s_{12} f_{12}}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}} \\ &= \frac{\frac{s_{11} s_{22}^2 + b^2 s_{11} s_{12}^2 + s_{12}^2 s_{22} + b^2 s_{11}^2 s_{22} - 2 s_{12}^2 s_{22} - 2 b^2 s_{11} s_{12}^2}{(s_{11} s_{22} - s_{12}^2)^2} + b \frac{s_{11} + s_{22}}{s_{11} s_{22} - s_{12}^2}}{\operatorname{tr} \tilde{F}^{\frac{1}{2}}} \\ &= \frac{\frac{s_{11} s_{22}^2 - b^2 s_{11} s_{12}^2 - s_{12}^2 s_{22} + b^2 s_{11}^2 s_{22}}{s_{11} s_{22} - s_{12}^2} + b (s_{11} + s_{22})}{(s_{11} s_{22} - s_{12}^2) \operatorname{tr} \tilde{F}^{\frac{1}{2}}} \\ &= \frac{b^2 s_{11} + s_{22} + b s_{11} + b s_{22}}{\sqrt{(b s_{11} + s_{22})^2 + (1 - b)^2 s_{12}^2}} \end{aligned}$$

$$= \frac{(1+b)(b s_{11} + s_{22})}{\sqrt{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}}.$$

□

Nun ist es möglich $\tilde{\kappa}$ aus (6.6) ebenfalls über die Einträge von \tilde{B} und \tilde{S} auszudrücken sowie Terme, in denen $\tilde{\kappa}$ vorkommt, zu vereinfachen.

Lemma 7.3 *Sei $\tilde{\kappa}$ wie in (6.6). Dann gilt:*

(a)

$$\tilde{\kappa} = \frac{(1+b)(b s_{11} + s_{22}) - (1-b)\sqrt{(b s_{11} + s_{22})^2 - 4 b s_{12}^2}}{2\sqrt{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}},$$

(b)

$$\tilde{\kappa}^2 = \tilde{\kappa} \operatorname{tr} \tilde{M} - b,$$

(c)

$$(\tilde{\kappa}^2 - b^2)(1 - \tilde{\kappa}^2) = \tilde{\kappa}^2 \frac{(1-b^2)^2 s_{12}^2}{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}.$$

BEWEIS:

(a) Es ist $\tilde{M} \in \mathbb{R}^{2 \times 2}$ und daher $\det(\lambda I_2 - \tilde{M}) = \lambda^2 - \operatorname{tr} \tilde{M} \lambda + \det \tilde{M}$. Somit hat $\tilde{\kappa}$ die Darstellung

$$\tilde{\kappa} = \lambda_{\min}(\tilde{M}) = \frac{1}{2} \operatorname{tr} \tilde{M} - \frac{1}{2} \sqrt{(\operatorname{tr} \tilde{M})^2 - 4 \det \tilde{M}}. \quad (7.3)$$

Weiter ist

$$\begin{aligned} \det \tilde{M} &= \det \tilde{S}^{\frac{1}{2}} \tilde{F}^{\frac{1}{2}} \tilde{S}^{\frac{1}{2}} = \det \tilde{S} \tilde{F}^{\frac{1}{2}} \\ &\stackrel{(7.3)}{=} \det \tilde{S} \frac{\det \tilde{B}}{\det \tilde{S}} = b. \end{aligned}$$

Mit Lemma 7.2 und (7.3) ergibt sich

$$\begin{aligned}
\tilde{\kappa} &= \frac{(1+b)(b s_{11} + s_{22})}{2\sqrt{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}} \\
&\quad - \frac{1}{2} \sqrt{\frac{(1+b)^2 (b s_{11} + s_{22})^2 - 4b [(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2]}{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}} \\
&= \frac{(1+b)(b s_{11} + s_{22}) - (1-b)\sqrt{(b s_{11} + s_{22})^2 - 4b s_{12}^2}}{2\sqrt{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}}.
\end{aligned}$$

(b) Mit Anwendung von (7.3) gilt:

$$\begin{aligned}
\tilde{\kappa}^2 &= \left(\frac{1}{2} \operatorname{tr} \tilde{M} - \frac{1}{2} \sqrt{(\operatorname{tr} \tilde{M})^2 - 4b} \right)^2 \\
&= \frac{1}{4} (\operatorname{tr} \tilde{M})^2 - \frac{1}{2} \operatorname{tr} \tilde{M} \sqrt{(\operatorname{tr} \tilde{M})^2 - 4b} + \frac{1}{4} (\operatorname{tr} \tilde{M})^2 - b \\
&= \left(\frac{1}{2} \operatorname{tr} \tilde{M} - \frac{1}{2} \sqrt{(\operatorname{tr} \tilde{M})^2 - 4b} \right) \operatorname{tr} \tilde{M} - b \\
&= \tilde{\kappa} \operatorname{tr} \tilde{M} - b.
\end{aligned}$$

(c) Mit Anwendung von Lemma 7.3 (b) und Lemma 7.2 ergibt sich

$$\begin{aligned}
&(\tilde{\kappa}^2 - b^2)(1 - \tilde{\kappa}^2) \\
&= (\tilde{\kappa} \operatorname{tr} \tilde{M} - b - b^2)(1 - \tilde{\kappa} \operatorname{tr} \tilde{M} + b) \\
&= \tilde{\kappa} \operatorname{tr} \tilde{M} - \tilde{\kappa}^2 (\operatorname{tr} \tilde{M})^2 + 2b \tilde{\kappa} \operatorname{tr} \tilde{M} - b - 2b^2 + b^2 \tilde{\kappa} \operatorname{tr} \tilde{M} - b^3 \\
&= \tilde{\kappa} \operatorname{tr} \tilde{M} (1+b)^2 - b(1+b)^2 - \tilde{\kappa}^2 (\operatorname{tr} \tilde{M})^2 \\
&= \tilde{\kappa}^2 \left[(1+b)^2 - (\operatorname{tr} \tilde{M})^2 \right] \\
&= \tilde{\kappa}^2 \frac{(1+b)^2 [(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2] - (1+b)^2 (b s_{11} + s_{22})^2}{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2} \\
&= \tilde{\kappa}^2 \frac{(1-b^2)^2 s_{12}^2}{(b s_{11} + s_{22})^2 + (1-b)^2 s_{12}^2}.
\end{aligned}$$

□

Im Kapitel 6.1 wurde gezeigt, dass die Struktur I genau dann vorliegt, wenn

$$\tilde{\kappa} \geq \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}$$

ist. Diese Bedingung kann aber auch für b in Abhängigkeit der Einträge der Matrix \tilde{S} angegeben werden. Zur Vorbereitung dienen die folgenden Lemmata. Dazu sei

$$\tau = \sqrt{(b s_{11} + s_{22})^2 + (1 - b)^2 s_{12}^2}. \quad (7.4)$$

Lemma 7.4 *Seien $s_{11}, s_{22} > 0$, $s_{12}^2 < s_{11} s_{22}$ und $d = s_{11} + s_{22} + s_{11} s_{22} - s_{12}^2$.*

Dann besitzt die Funktion $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$\Psi(b) = (1 + s_{11})b^2 - (2 + d)b + (1 + s_{22})$$

genau die beiden Nullstellen

$$b_0 = \frac{2 + d - \sqrt{d^2 - 4 s_{12}^2}}{2(1 + s_{11})} \quad (7.5)$$

und

$$b_1 = \frac{2 + d + \sqrt{d^2 - 4 s_{12}^2}}{2(1 + s_{11})}.$$

Dabei ist $b_0 \in (0, 1)$ und $b_1 > 1$.

BEWEIS:

Die Darstellung der Nullstellen folgt aus der Formel für quadratische Gleichungen. Es gilt

$$d^2 - 4s_{12}^2 > (s_{11} + s_{22})^2 - 4s_{11}s_{22} = (s_{11} - s_{22})^2 \geq 0.$$

Weiter ist

$$\begin{aligned} b_0 &= \frac{2 + d - \sqrt{d^2 - 4 s_{12}^2}}{2(1 + s_{11})} < \frac{2 + s_{11} + s_{22} - |s_{11} - s_{22}|}{2(1 + s_{11})} \\ &= \frac{2[1 + \min(s_{11}, s_{22})]}{2(1 + s_{11})} \leq 1, \end{aligned}$$

$$b_0 = \frac{2 + d - \sqrt{d^2 - 4 s_{12}^2}}{2(1 + s_{11})} > \frac{2}{(1 + s_{11})} > 0,$$

da $d > 0$, und

$$\begin{aligned} b_1 &= \frac{2 + d + \sqrt{d^2 - 4 s_{12}^2}}{2(1 + s_{11})} > \frac{2 + s_{11} + s_{22} + |s_{11} - s_{22}|}{2(1 + s_{11})} \\ &= \frac{2[1 + \max(s_{11}, s_{22})]}{2(1 + s_{11})} \geq 1. \end{aligned}$$

□

Lemma 7.5 Seien $s_{11}, s_{22} > 0$, $s_{12}^2 < s_{11} s_{22}$, $d = s_{11} + s_{22} + s_{11} s_{22} - s_{12}^2$ und $\Xi : [0, 1] \rightarrow \mathbb{R}$ durch

$$\begin{aligned} \Xi(b) &= (1 + b)(b s_{11} + s_{22})d \\ &\quad - (1 - b)d \sqrt{(b s_{11} + s_{22})^2 - 4 b s_{12}^2} - 2 \tau^2 \end{aligned}$$

gegeben. Dann gilt

$$\Xi(b) = \begin{cases} > 0 & , \quad b > b_0 \\ = 0 & , \quad b = b_0 \\ < 0 & , \quad b < b_0. \end{cases}$$

für b_0 aus (7.5).

BEWEIS:

Es ist

$$\begin{aligned} \Xi(b) &= 0 \\ \Leftrightarrow (1 + b)(b s_{11} + s_{22})d - 2 \tau^2 &= (1 - b)d \sqrt{(b s_{11} + s_{22})^2 - 4 b s_{12}^2} \\ \Rightarrow (1 + b)^2 (b s_{11} + s_{22})^2 d^2 - 4(1 + b)(b s_{11} + s_{22})d \tau^2 + 4 \tau^4 \\ &= (1 - b)^2 d^2 (b s_{11} + s_{22})^2 - 4 b (1 - b)^2 d^2 s_{12}^2 \\ \Leftrightarrow 4 b d^2 \tau^2 - 4(1 + b)(b s_{11} + s_{22})d \tau^2 + 4 \tau^4 &= 0 \end{aligned}$$

$$\begin{aligned}
&\Leftrightarrow b^2 (s_{11}^2 - d s_{11} + s_{12}^2) + b [d^2 - d (s_{11} + s_{22}) + 2 (s_{11} s_{22} - s_{12}^2)] \\
&\quad + s_{22}^2 - d s_{22} + s_{12}^2 = 0 \\
&\Leftrightarrow -(s_{11} s_{22} - s_{12}^2) (1 + s_{11}) b^2 + (s_{11} s_{22} - s_{12}^2) (2 + d) b \\
&\quad - (s_{11} s_{22} - s_{12}^2) (1 + s_{22}) = 0 \\
&\Leftrightarrow \Psi(b) = 0.
\end{aligned}$$

Dabei ist Ψ aus Lemma 7.4. Somit ist b_0 die einzige mögliche Nullstelle von Ξ . Die Funktion Ξ ist auf $[0, 1]$ stetig. Es gilt

$$\Xi(0) = d s_{22} - d s_{22} - 2 (s_{22}^2 + s_{12}^2) < 0$$

und

$$\begin{aligned}
\Xi(1) &= 2 (s_{11} + s_{22}) d - 2 (s_{11} + s_{22})^2 \\
&= 2 (s_{11} + s_{22}) (s_{11} s_{22} - s_{12}^2) > 0.
\end{aligned}$$

Damit erhält man für Ξ

$$\Xi(b) = \begin{cases} > 0 & , \quad b > b_0 \\ = 0 & , \quad b = b_0 \\ < 0 & , \quad b < b_0. \end{cases}$$

□

Nachdem jetzt alle notwendigen Lemmata behandelt wurden, kann der folgende Satz formuliert werden.

Satz 7.6 *Man betrachte das dreidimensionale Regressionsproblem bei einem gemeinsamen Eigenvektor aus Definition 6.1. Dann liegt die Struktur I genau dann vor, wenn $b \geq b_0$ mit b_0 aus (7.5) ist.*

BEWEIS:

Die Struktur I liegt genau dann vor, wenn $\tilde{\kappa} \geq \frac{\text{tr } \tilde{F}^{\frac{1}{2}}}{1 + \text{tr } \tilde{S}^{-1}}$ ist.

$$\begin{aligned}
\tilde{\kappa} \geq \frac{\text{tr } \tilde{F}^{\frac{1}{2}}}{1 + \text{tr } \tilde{S}^{-1}} &\stackrel{\text{Lemma 7.1}}{\Leftrightarrow} \tilde{\kappa} d \geq \tau \\
&\stackrel{\text{Lemma 7.3 (a)}}{\Leftrightarrow} (1+b)(bs_{11} + s_{22})d \\
&\quad - (1-b)d\sqrt{(bs_{11} + s_{22})^2 - 4bs_{12}^2} \\
&\geq 2\tau^2 \\
&\Leftrightarrow \Xi(b) \geq 0 \\
&\stackrel{\text{Lemma 7.5}}{\Leftrightarrow} b \geq b_0
\end{aligned}$$

□

Es werden noch zwei Lemmata mit Blick auf den Inhalt von Kapitel 7.3 angegeben.

Lemma 7.7 *Seien $s_{11}, s_{22} > 0$, $s_{12}^2 < s_{11}s_{22}$, $d = s_{11} + s_{22} + s_{11}s_{22} - s_{12}^2$ und Ψ aus Lemma 7.4. Dann gelten folgende Identitäten:*

(a)

$$\begin{aligned}
&[b^2(1 + s_{11}) - (1 + s_{22})]^2 - [b^2(1 + s_{11}) + (1 + s_{22}) - 2b]^2 \\
&= 4b(\Psi(b) - bs_{12}^2),
\end{aligned}$$

(b)

$$\begin{aligned}
&(1-b)[b^2(1 + s_{11}) - (1 + s_{22})](bs_{11} + s_{22}) \\
&\quad + [b^2(1 + s_{11}) + (1 + s_{22}) - 2b]^2 + b(1-b)^2s_{12}^2 \\
&= b(bs_{11} + s_{22})^2 + (1-b)^2\Psi(b).
\end{aligned}$$

BEWEIS:

Es gilt:

(a)

$$\begin{aligned}
& [b^2(1+s_{11}) - (1+s_{22})]^2 - [b^2(1+s_{11}) + (1+s_{22}) - 2b]^2 \\
&= -4b^2(1+s_{11})(1+s_{22}) + 4b^3(1+s_{11}) + 4b(1+s_{22}) - 4b^2 \\
&= 4b[b^2(1+s_{11}) - b(2+s_{11}+s_{22}+s_{11}s_{22}) + 1+s_{22}] \\
&= 4b(\Psi(b) - bs_{12}^2),
\end{aligned}$$

(b)

$$\begin{aligned}
& (1-b)[b^2(1+s_{11}) - (1+s_{22})](bs_{11}+s_{22}) \\
&\quad + [b^2(1+s_{11}) + (1+s_{22}) - 2b]^2 + b(1-b)^2s_{12}^2 \\
&= b(bs_{11}+s_{22})^2 + (bs_{11}+s_{22})[b^2+b-b^3(1+s_{11}) - (1+s_{22})] \\
&\quad + b^4(1+s_{11})^2 + (1+s_{22})^2 + 4b^2 + 2b^2(1+s_{11})(1+s_{22}) \\
&\quad - 4b^3(1+s_{11}) - 4b(1+s_{22}) + b(1-b)^2s_{12}^2 \\
&= b(bs_{11}+s_{22})^2 \\
&\quad + (1+s_{11})[-b^3(bs_{11}+s_{22}) + b^4(1+s_{11}) - 4b^3] \\
&\quad + (1+s_{22})[-(bs_{11}+s_{22}) + (1+s_{22}) - 4b] \\
&\quad + 2b^2(1+s_{11})(1+s_{22}) \\
&\quad + 4b^2 + b(1+b)(bs_{11}+s_{22}) + b(1-b)^2s_{12}^2 \\
&= b(bs_{11}+s_{22})^2 + (1+s_{11})(b^4 - 3b^3) + (1+s_{22})(1-3b) \\
&\quad + (1+s_{11})(1+s_{22})(-b^3 + 2b^2 - b) + 4b^2 \\
&\quad + (b^3 + b^2)s_{11} + (b^2 + b)s_{22} + b(1-b)^2s_{12}^2 \\
&= b(bs_{11}+s_{22})^2 + b^2(1-b)^2s_{11} + (1-b)^2s_{22} - b(1-b)^2(1+d) \\
&\quad + b^2(1-b^2) - b(1-b)^2 + (1-b)^2 \\
&= b(bs_{11}+s_{22})^2 + (1-b)^2\Psi(b).
\end{aligned}$$

□

Im nachfolgenden Lemma spielen die Nullstellen der Funktion Ψ aus Lemma 7.4 eine zentrale Rolle.

Lemma 7.8 *Seien $s_{11}, s_{22} > 0$, $s_{12}^2 < s_{11} s_{22}$, $d = s_{11} + s_{22} + s_{11} s_{22} - s_{12}^2$ und $\Phi : [0, 1] \rightarrow \mathbb{R}$ durch*

$$\begin{aligned} \Phi(b) &= [b^2(1 + s_{11}) - (1 + s_{22})](b s_{11} + s_{22}) + 2b(1 - b)s_{12}^2 \\ &\quad + [b^2(1 + s_{11}) + (1 + s_{22}) - 2b] \sqrt{(b s_{11} + s_{22})^2 - 4b s_{12}^2} \end{aligned}$$

gegeben. Dann ist für $b \geq b_0$ mit b_0 aus (7.5) stets $\Phi(b) \geq 0$.

BEWEIS:

Es gilt mit Lemma 7.7

$$\begin{aligned} \Phi(b) &= 0 \\ \Leftrightarrow & [b^2(1 + s_{11}) - (1 + s_{22})](b s_{11} + s_{22}) + 2b(1 - b)s_{12}^2 \\ &= -[b^2(1 + s_{11}) + (1 + s_{22}) - 2b] \sqrt{(b s_{11} + s_{22})^2 - 4b s_{12}^2} \\ \Rightarrow & [b^2(1 + s_{11}) - (1 + s_{22})]^2 (b s_{11} + s_{22})^2 + 4b^2(1 - b)^2 s_{12}^4 \\ &+ 4b(1 - b)[b^2(1 + s_{11}) - (1 + s_{22})](b s_{11} + s_{22}) s_{12}^2 \\ &= [b^2(1 + s_{11}) + (1 + s_{22}) - 2b]^2 [(b s_{11} + s_{22})^2 - 4b s_{12}^2] \\ \Leftrightarrow & (b s_{11} + s_{22})^2 \\ &\cdot [b^2(1 + s_{11}) - (1 + s_{22})]^2 - [b^2(1 + s_{11}) + (1 + s_{22}) - 2b]^2 \\ &+ 4b \{ (1 - b)[b^2(1 + s_{11}) - (1 + s_{22})](b s_{11} + s_{22}) \\ &+ [b^2(1 + s_{11}) + (1 + s_{22}) - 2b]^2 + b(1 - b)^2 s_{12}^2 \} s_{12}^2 = 0 \\ \Leftrightarrow & (b s_{11} + s_{22})^2 \Psi(b) - b(b s_{11} + s_{22})^2 s_{12}^2 \\ &+ b(b s_{11} + s_{22})^2 s_{12}^2 + (1 - b)^2 \Psi(b) s_{12}^2 = 0 \\ \Leftrightarrow & \tau^2 \Psi(b) = 0 \\ \Leftrightarrow & \Psi(b) = 0. \end{aligned}$$

Damit ist b_0 die einzige mögliche Nullstelle von Φ . Wegen der Stetigkeit von Φ auf $[0, 1]$ und

$$\begin{aligned}\Phi(1) &= (s_{11} + s_{22})(s_{11} - s_{22}) + (s_{11} + s_{22})\sqrt{(s_{11} + s_{22})^2 - 4s_{12}^2} \\ &= (s_{11} + s_{22})\left[(s_{11} - s_{22}) + \sqrt{(s_{11} - s_{22})^2 + 4(s_{11}s_{22} - s_{12}^2)}\right] \\ &> (s_{11} + s_{22})(s_{11} - s_{22} + |s_{11} - s_{22}|) \geq 0\end{aligned}$$

folgt die Behauptung. \square

7.2 Entwicklung von Kriterien

Es wird nun ein Fall $j = 2$ aus Struktur I aus Kapitel 6.1 betrachtet. Dies ist derjenige, bei dem

$$\mu \in \left((\tilde{\kappa}(1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}}, 1 + s_3) \right)$$

ist. Von nun an gelten folgende Voraussetzungen.

Voraussetzung 7.9 *Man betrachte das Schätzproblem aus Definition 6.1 und zusätzlich seien $s_{12} > 0$ und*

$$\mu \in \left((\tilde{\kappa}(1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}}, 1 + s_3) \right).$$

Durch Anwendung von Satz 4.8 sollen Kriterien zur Bestimmung des MILE für diese spezielle Situation gefunden werden.

Man nehme nun an, dass der Vektor $(0 \ 0 \ 1)'$ im Eigenraum zum maximalen Eigenwert von $B(\hat{C}_M X - I_3) \left[B(\hat{C}_M X - I_3) \right]'$ liegt. Aus dieser Annahme heraus kann $a_2 = (0 \ 0 \ 1)'$ gewählt werden. Mit $a_1 = (-y \ x \ 0)'$ und $u = (x \ y \ 0)$, wobei $x, y \in [-1, 1]$ und $x^2 + y^2 = 1$ gelten, ist $\{a_1, a_2, u\}$ eine

Orthonormalbasis des \mathbb{R}^3 . Es ergibt sich mit

$$\lambda_2 = \sqrt{x^2 + b^2 y^2}, \quad (7.6)$$

dass

$$v = \begin{pmatrix} -\frac{x}{\lambda_2} \\ -\frac{by}{\lambda_2} \\ 0 \end{pmatrix}$$

ist. Mit $b_1 = (\frac{by}{\lambda_2} - \frac{x}{\lambda_2} 0)'$ und $b_2 = (0 0 1)'$ wird $\{b_1, b_2, v\}$ zu einer Orthonormalbasis des \mathbb{R}^3 , da

$$\begin{aligned} b_1' b_1 &= \frac{b^2 y^2 + x^2}{\lambda_2^2} \stackrel{(7.6)}{=} 1, \\ b_2' b_2 &= 1, \\ v' v &= \frac{x^2 + b^2 y^2}{\lambda_2^2} \stackrel{(7.6)}{=} 1, \\ b_1' b_2 &= 0, \\ b_1' v &= -\frac{bxy}{\lambda_2^2} + \frac{bxy}{\lambda_2^2} = 0, \\ b_2' v &= 0. \end{aligned}$$

Definiere

$$U_1 = (a_1 \ a_2) = \begin{pmatrix} y & 0 \\ -x & 0 \\ 0 & 1 \end{pmatrix}$$

und

$$V_1 = (b_1 \ b_2) = \begin{pmatrix} \frac{by}{\lambda_2} & 0 \\ -\frac{x}{\lambda_2} & 0 \\ 0 & 1 \end{pmatrix}.$$

Lemma 7.10 *Es gelte die Voraussetzung 7.9. Dann ist*

$$\lambda_1 = \frac{\mu q + b s_3 \lambda_2}{q(1 + s_3) + s_3 \lambda_2^2}$$

und

$$H_1 = \frac{1}{\lambda_1} \begin{pmatrix} \lambda_2 \frac{b-\lambda_1 \lambda_2}{q} & 0 \\ 0 & \frac{\mu-\lambda_1}{s_3} \end{pmatrix}$$

mit

$$q = b^2 s_{11} y^2 - 2b s_{12} x y + s_{22} x^2.$$

BEWEIS:

Es ist

$$\begin{aligned} X_1' X_1 &= V_1' S V_1 \\ &= \begin{pmatrix} \frac{by}{\lambda_2} & -\frac{x}{\lambda_2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} & 0 \\ s_{12} & s_{22} & 0 \\ 0 & 0 & s_3 \end{pmatrix} \begin{pmatrix} \frac{by}{\lambda_2} & 0 \\ -\frac{x}{\lambda_2} & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{by}{\lambda_2} s_{11} - \frac{x}{\lambda_2} s_{12} & \frac{by}{\lambda_2} s_{12} - \frac{x}{\lambda_2} s_{22} & 0 \\ 0 & 0 & s_3 \end{pmatrix} \begin{pmatrix} \frac{by}{\lambda_2} & 0 \\ -\frac{x}{\lambda_2} & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{q}{\lambda_2^2} & 0 \\ 0 & s_3 \end{pmatrix} \end{aligned}$$

und

$$\begin{aligned} B_1 &= U_1' B V_1 \\ &= \begin{pmatrix} y & -x & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & \mu \end{pmatrix} \begin{pmatrix} \frac{by}{\lambda_2} & 0 \\ -\frac{x}{\lambda_2} & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} y & -bx & 0 \\ 0 & 0 & \mu \end{pmatrix} \begin{pmatrix} \frac{by}{\lambda_2} & 0 \\ -\frac{x}{\lambda_2} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{b}{\lambda_2} & 0 \\ 0 & \mu \end{pmatrix}. \end{aligned}$$

Daraus erhält man

$$\begin{aligned}
 F_1 &= (X_1' X_1)^{-1} B_1' B_1 (X_1' X_1)^{-1} \\
 &= \begin{pmatrix} \frac{\lambda_2^2}{q} & 0 \\ 0 & \frac{1}{s_3} \end{pmatrix} \begin{pmatrix} \frac{b^2}{\lambda_2^2} & 0 \\ 0 & \mu^2 \end{pmatrix} \begin{pmatrix} \frac{\lambda_2^2}{q} & 0 \\ 0 & \frac{1}{s_3} \end{pmatrix} \\
 &= \begin{pmatrix} \frac{b^2 \lambda_2^2}{q^2} & 0 \\ 0 & \frac{\mu^2}{s_3} \end{pmatrix}
 \end{aligned}$$

und

$$F_1^{\frac{1}{2}} = \begin{pmatrix} \frac{b \lambda_2}{q} & 0 \\ 0 & \frac{\mu}{s_3} \end{pmatrix}.$$

Nun ist

$$\begin{aligned}
 \lambda_1 &= \frac{\text{tr } F_1^{\frac{1}{2}}}{1 + \text{tr } (X_1' X_1)^{-1}} \\
 &= \frac{\frac{b \lambda_2}{q} + \frac{\mu}{s_3}}{1 + \frac{\lambda_2^2}{q} + \frac{1}{s_3}} \\
 &= \frac{\mu q + b s_3 \lambda_2}{q(1 + s_3) + s_3 \lambda_2^2}
 \end{aligned}$$

und

$$\begin{aligned}
 H_1 &= -(X_1' X_1)^{-1} + \frac{1}{\lambda_1} F_1^{\frac{1}{2}} \\
 &= - \begin{pmatrix} \frac{\lambda_2^2}{q} & 0 \\ 0 & \frac{1}{s_3} \end{pmatrix} + \frac{1}{\lambda_1} \begin{pmatrix} \frac{b \lambda_2}{q} & 0 \\ 0 & \frac{\mu}{s_3} \end{pmatrix} \\
 &= \frac{1}{\lambda_1} \begin{pmatrix} \lambda_2 \frac{b - \lambda_1 \lambda_2}{q} & 0 \\ 0 & \frac{\mu - \lambda_1}{s_3} \end{pmatrix}.
 \end{aligned}$$

□

Es bleibt noch zu bemerken, dass stets $q > 0$ gilt, da q eine quadratische Form einer positiv definiten Matrix ist. Man beachte außerdem, dass q, λ_1 und λ_2 von den Einträgen x und y aus u abhängen. Desweiteren stellt dieses

Lemma für die Voraussetzung 7.9 die Bedingungen (1) und (2) aus Satz 4.8 dar. Damit eine Lösung von (3) aus Satz 4.8 zur Berechnung des MILE dient, muss $\lambda_1 \geq \lambda_2$ und H_1 nicht negativ definit, also hier $b - \lambda_1 \lambda_2 \geq 0$ und $\mu - \lambda_1 \geq 0$ gelten.

Das Hauptinteresse liegt aber darin eine Lösung der Gleichung (3) aus Satz 4.8 zu finden, weil diese zur Berechnung des MILE beiträgt. Die Lösung sollte in Abhängigkeit von den Einträgen aus u gegeben sein.

Durch

$$\begin{aligned} G_1 &= B_1(X_1'X_1)^{-1} [(X_1'X_1)^{-1}B_1'B_1(X_1'X_1)^{-1}]^{-\frac{1}{2}} \\ &= B_1(X_1'X_1)^{-1}F_1^{-\frac{1}{2}} \\ &= \begin{pmatrix} \frac{b}{\lambda_2} & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} \frac{\lambda_2^2}{q} & 0 \\ 0 & \frac{1}{s_3} \end{pmatrix} \begin{pmatrix} \frac{q}{b\lambda_2} & 0 \\ 0 & \frac{s_3}{\mu} \end{pmatrix} = I_2 \end{aligned}$$

erhält man

$$\begin{aligned} W &= U_1G_1H_1V_1' \\ &= -\frac{1}{\lambda_1} \begin{pmatrix} y & 0 \\ -x & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_2 \frac{b-\lambda_1\lambda_2}{q} & 0 \\ 0 & \frac{\mu-\lambda_1}{s_3} \end{pmatrix} \begin{pmatrix} \frac{by}{\lambda_2} & -\frac{x}{\lambda_2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= -\frac{1}{\lambda_1} \begin{pmatrix} \frac{b^2-b\lambda_1\lambda_2}{q}y^2 & -\frac{b-\lambda_1\lambda_2}{q}xy & 0 \\ -\frac{b^2-b\lambda_1\lambda_2}{q}xy & \frac{b-\lambda_1\lambda_2}{q}x^2 & 0 \\ 0 & 0 & \frac{\mu-\lambda_1}{s_3} \end{pmatrix} \\ &= \frac{1}{\lambda_1} \begin{pmatrix} \frac{\lambda_1\lambda_2-b}{q}by^2 & -\frac{\lambda_1\lambda_2-b}{q}xy & 0 \\ -\frac{\lambda_1\lambda_2-b}{q}bxy & \frac{\lambda_1\lambda_2-b}{q}x^2 & 0 \\ 0 & 0 & \frac{\mu-\lambda_1}{s_3} \end{pmatrix}. \end{aligned}$$

Wegen

$$\begin{aligned}
 & X'XB'u_3 \\
 &= \begin{pmatrix} s_{11} & s_{12} & 0 \\ s_{12} & s_{22} & 0 \\ 0 & 0 & s_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & \mu \end{pmatrix} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \\
 &= \begin{pmatrix} s_{11}x + b s_{12}y \\ s_{12}x + b s_{22}y \\ 0 \end{pmatrix}
 \end{aligned}$$

ergibt sich mit

$$p = b(s_{11} - s_{22})xy + b^2 s_{12}y^2 - s_{12}x^2$$

die Umformung

$$\begin{aligned}
 & \lambda_1 W X' X B' u_3 \\
 &= \begin{pmatrix} \frac{\lambda_1 \lambda_2 - b}{q} b y^2 & -\frac{\lambda_1 \lambda_2 - b}{q} x y & 0 \\ -\frac{\lambda_1 \lambda_2 - b}{q} b x y & \frac{\lambda_1 \lambda_2 - b}{q} x^2 & 0 \\ 0 & 0 & \frac{\mu - \lambda_1}{s_3} \end{pmatrix} \begin{pmatrix} s_{11}x + b s_{12}y \\ s_{12}x + b s_{22}y \\ 0 \end{pmatrix} \\
 &= \frac{(\lambda_1 \lambda_2 - b)p}{q} \begin{pmatrix} y \\ -x \\ 0 \end{pmatrix}.
 \end{aligned}$$

Da

$$B' B u_3 - \lambda_2^2 u_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b^2 & 0 \\ 0 & 0 & \mu^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} - \lambda_2^2 \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} (1 - \lambda_2^2)x \\ (b^2 - \lambda_2^2)y \\ 0 \end{pmatrix}$$

ist, wird (3) aus Satz 4.8 zu

$$\begin{pmatrix} \frac{(\lambda_1 \lambda_2 - b)p}{q} y + (1 - \lambda_2^2) x \\ -\frac{(\lambda_1 \lambda_2 - b)p}{q} x + (b^2 - \lambda_2^2) y \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (7.7)$$

Offensichtlich erfüllt $-u$ alle Bedingungen aus Satz 4.8, falls u diese Bedingungen bereits erfüllt. Somit kann o.B.d.A. $y \geq 0$ angenommen werden. Außerdem kann man die Lösungen $x = 0$ oder $y = 0$ ausschließen, was im folgenden Lemma gezeigt wird.

Lemma 7.11 *Unter der Voraussetzung 7.9 gibt es für $x = 0$ oder $y = 0$ keine Lösungen, die alle Bedingungen aus Satz 4.8 erfüllen.*

BEWEIS:

Sei $x = 0$, dann gilt $y = 1$, $\lambda_2 = b$, $q = b^2 s_{11}$ und $p = b^2 s_{12} \neq 0$. Wegen $\mu < 1 + s_3$ ist $\lambda_1 = \frac{b^2 s_{11} \mu + b^2 s_3}{b^2 s_{11}(1+s_3) + b^2 s_3} < 1$ und deshalb $\lambda_1 \lambda_2 - b \neq 0$. Die Gleichung (7.7) ist nicht erfüllt.

Sei nun $y = 0$, dann ist $x = 1$ oder $x = -1$. Weiter gilt $\lambda_2 = 1$ und $q = s_{22}$. Wegen $\lambda_1 = \frac{s_{22} \mu + b s_3}{s_{22}(1+s_3) + s_3} < 1 = \lambda_2$ ist die Bedingung (1) aus Satz 4.8 nicht erfüllt. \square

Durch Verwendung von $\lambda_2^2 = x^2 + b^2 y^2$ (siehe (7.6)) und $x^2 + y^2 = 1$ erhält man $1 - \lambda_2^2 = (1 - b^2)y^2$ und $b^2 - \lambda_2^2 = -(1 - b^2)x^2$. Die Gleichung (7.7) ist dadurch äquivalent zu

$$\begin{pmatrix} (\lambda_1 \lambda_2 - b) p y + (1 - b^2) q x y^2 \\ -(\lambda_1 \lambda_2 - b) p x - (1 - b^2) q x^2 y \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

und wegen $x \neq 0$ und $y \neq 0$ ebenfalls äquivalent zu

$$\begin{pmatrix} (\lambda_1 \lambda_2 - b)p + (1 - b^2)qxy \\ -(\lambda_1 \lambda_2 - b)p - (1 - b^2)qxy \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Damit wird die Bedingung (3) aus Satz 4.8 auf Lösungen des Gleichungssystems

$$(1) \quad (\lambda_1 \lambda_2 - b)p + (1 - b^2)qxy = 0 \quad (7.8)$$

$$(2) \quad p = b(s_{11} - s_{22})xy + b^2 s_{12}y^2 - s_{12}x^2$$

$$(3) \quad q = b^2 s_{11}y^2 - 2b s_{12}xy + s_{22}x^2$$

$$(4) \quad \lambda_1 = \frac{\mu q + b s_3 \lambda_2}{q(1 + s_3) + s_3 \lambda_2^2}$$

$$(5) \quad \lambda_2 = \sqrt{x^2 + b^2 y^2}$$

$$(6) \quad y = \sqrt{1 - x^2}$$

geführt.

Lemma 7.12 *Es gelte die Voraussetzung 7.9 und λ_1 sei wie in Lemma 7.10.*

Dann gilt:

$$\lambda_1 \lambda_2 - b = \frac{\mu \lambda_2 - b(1 + s_3)}{q(1 + s_3) + s_3 \lambda_2^2} q.$$

BEWEIS:

$$\begin{aligned} \lambda_1 \lambda_2 - b &= \frac{\mu q \lambda_2 + b s_3 \lambda_2^2 - b q(1 + s_3) - b s_3 \lambda_2^2}{q(1 + s_3) + s_3 \lambda_2^2} \\ &= \frac{\mu \lambda_2 - b(1 + s_3)}{q(1 + s_3) + s_3 \lambda_2^2} q. \end{aligned}$$

□

Durch entsprechende Einsetzungen kann nun das Gleichungssystem (7.8) auf genau eine Gleichung in der Variablen x gebracht werden.

Satz 7.13 *Unter der Voraussetzung 7.9 ist eine Lösung des Gleichungssystems (7.8) eine Nullstelle der Funktion $f: [-1, 1] \rightarrow \mathbb{R}$ mit*

$$\begin{aligned} f(x) = & 2b(1-b^2)s_{12}(1+s_3)x^4 + (1-b^2)(\alpha-\beta)x^3\sqrt{1-x^2} \\ & -b(1-3b^2)s_{12}(1+s_3)x^2 \\ & -(1+b^2)s_{12}\mu x^2\sqrt{b^2+(1-b^2)x^2} \\ & +b^2(\alpha-\beta)x\sqrt{1-x^2} \\ & +b(s_{11}-s_{22})\mu x\sqrt{1-x^2}\sqrt{b^2+(1-b^2)x^2} \\ & +b^2s_{12}\mu\sqrt{b^2+(1-b^2)x^2} - b^3s_{12}(1+s_3). \end{aligned}$$

Dabei sind $\alpha = s_{22} + s_{22}s_3 + s_3$ und $\beta = b^2(s_{11} + s_{11}s_3 + s_3)$.

BEWEIS:

a) Durch Einsetzen von (7.8,(3)) und (7.8,(5)) erhält man

$$q(1+s_3) + s_3\lambda_2^2 = \alpha x^2 - 2bs_{12}(1+s_3)xy + \beta y^2$$

mit $\alpha = s_{22} + s_{22}s_3 + s_3$ und $\beta = b^2(s_{11} + s_{11}s_3 + s_3)$. Verwendet man zusätzlich $y^2 = 1 - x^2$, so ergibt sich schließlich

$$q(1+s_3) + s_3\lambda_2^2 = (\alpha - \beta)x^2 - 2bs_{12}(1+s_3)xy + \beta.$$

b) Man erhält durch Verwendung von (7.8,(2))

$$\begin{aligned} & [\mu\lambda_2 - b(1+s_3)]p \\ = & b(s_{11}-s_{22})\mu xy\lambda_2 + s_{12}\mu\lambda_2(b^2y^2 - x^2) \\ & -b^2(1+s_3)(s_{11}-s_{22})xy - b(1+s_3)s_{12}(b^2y^2 - x^2) \\ \stackrel{y^2=1-x^2}{=} & b(s_{11}-s_{22})\mu xy\lambda_2 - b^2(1+s_3)(s_{11}-s_{22})xy \\ & +b^2s_{12}\mu\lambda_2 - (1+b^2)s_{12}\mu x^2\lambda_2 \\ & -b^3(1+s_3)s_{12} + b(1+b^2)(1+s_3)s_{12}x^2. \end{aligned}$$

c) Unter Verwendung von Lemma 7.12 und $q(1 + s_3) + s_3 \lambda_2^2 > 0$ ist Gleichung (1) aus (7.8) äquivalent zu

$$[\mu \lambda_2 - b(1 + s_3)] p + (1 - b^2) [q(1 + s_3) + s_3 \lambda_2^2] x y = 0.$$

Durch Anwendung von a), b) und $y^2 = 1 - x^2$ ergibt sich

$$\begin{aligned} & [\mu \lambda_2 - b(1 + s_3)] p + (1 - b^2) [q(1 + s_3) + s_3 \lambda_2^2] x y \\ &= b(s_{11} - s_{22}) \mu x y \lambda_2 - b^2(1 + s_3)(s_{11} - s_{22}) x y \\ &\quad + b^2 s_{12} \mu \lambda_2 - (1 + b^2) s_{12} \mu x^2 \lambda_2 \\ &\quad - b^3(1 + s_3) s_{12} + b(1 + b^2)(1 + s_3) s_{12} x^2 \\ &\quad + (1 - b^2)(\alpha - \beta) x^3 y - 2b(1 - b^2) s_{12}(1 + s_3) x^2(1 - x^2) \\ &\quad + (1 - b^2) \beta x y \\ &= 2b(1 - b^2) s_{12}(1 + s_3) x^4 + (1 - b^2)(\alpha - \beta) x^3 y \\ &\quad - b(1 - 3b^2) s_{12}(1 + s_3) x^2 \\ &\quad - (1 + b^2) s_{12} \mu x^2 \lambda_2 \\ &\quad + b^2(\alpha - \beta) x y + b(s_{11} - s_{22}) \mu x y \lambda_2 \\ &\quad + b^2 s_{12} \mu \lambda_2 - b^3 s_{12}(1 + s_3). \end{aligned}$$

Wird noch $y = \sqrt{1 - x^2}$ und $\lambda_2 = \sqrt{b^2 + (1 - b^2)x^2}$ gesetzt, erhält man die gewünschte Funktion f . \square

Damit ist gezeigt, dass unter der Voraussetzung 7.9 und der Annahme, dass $(0 \ 0 \ 1)$ im Eigenraum zum maximalen Eigenwert von $\left[B(\hat{C}_M X - I_k) \right]' B(\hat{C}_M X - I_k)$ liegt, eine Lösung von (3) aus Satz 4.8 eine Nullstelle der Funktion f aus Satz 7.13 ist und umgekehrt. Im nachfolgenden Korollar werden nochmals die wichtigsten Erkenntnisse dieses Kapitels zusammengefasst.

Korollar 7.14 *Es gelte die Voraussetzung 7.9. Seien*

$$U_1 = \begin{pmatrix} \sqrt{1-x^{*2}} & 0 \\ -x^* & 0 \\ 0 & 1 \end{pmatrix},$$

$$V_1 = \begin{pmatrix} \frac{b\sqrt{1-x^{*2}}}{\lambda_2^*} & 0 \\ \frac{-x^*}{\lambda_2^*} & 0 \\ 0 & 1 \end{pmatrix},$$

$$u_3 = \begin{pmatrix} x^* \\ \sqrt{1-x^{*2}} \\ 0 \end{pmatrix},$$

$$\lambda_1^* = \frac{\mu [(s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1-x^{*2}} + b^2 s_{11}] + b s_3 \lambda_2^*}{[(s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1-x^{*2}} + b^2 s_{11}] (1 + s_3) + s_3 \lambda_2^{*2}}$$

und

$$\lambda_2^* = \sqrt{b^2 + (1 - b^2)x^{*2}}.$$

Dann ist

$$\hat{\beta}_M = B^{-1} [\lambda_1 U_1 V_1' + (I_k - u_3 u_3') B] (X'X)^{-1} X'y$$

der MILE, falls $x^* \in (-1, 1) \setminus \{0\}$ die Bedingungen

(1) x^* ist Nullstelle von f aus Satz 7.13,

(2) $\lambda_1^* \geq \lambda_2^*$,

(3) $b - \lambda_1^* \lambda_2^* \geq 0$,

(4) $\mu \geq \lambda_1^*$

erfüllt.

BEWEIS:

Aus (1) folgt mit Satz 7.13, Gleichungssystem (7.8) und Lemma 7.11, dass die Bedingung (3) aus Satz 4.8 erfüllt ist. Wegen (2), (3) und (4) sind auf Grund von Lemma 7.10 die Bedingungen (1) und (2) aus Satz 4.8 erfüllt. Mit Satz 4.8 folgt dann die Behauptung. \square

Nun stellt sich die Frage, ob so ein x^* aus Korollar 7.14 überhaupt existiert. Der nächste Abschnitt wird sich mit dieser Frage beschäftigen.

7.3 Notwendigkeit der Kriterien

Aus Kapitel 7.2 weiß man, dass, wenn ein x^* mit den Bedingungen aus Korollar 7.14 existiert, der MILE in Abhängigkeit von x^* angegeben werden kann. Aber die Existenz solch eines x^* ist noch nicht gesichert. In diesem Abschnitt wird die Existenz eines x^* aus Korollar 7.14 bewiesen. Auch ist es möglich zu zeigen, dass x^* aus einem Intervall $I \subset (0, \frac{1}{\sqrt{2}})$ stammt.

Zuerst wird mittels Nullstellensatz die Existenz einer Nullstelle der Funktion f aus Satz 7.13 auf I gezeigt. Danach wird die Gültigkeit von (2), (3) und (4) aus Korollar 7.14 für die Nullstellen von f auf I bestätigt. Dafür werden vorab ein paar Abschätzungen und Identitäten benötigt.

Lemma 7.15 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9:*

$$b - [(1 - b)^2 + (b^2 s_{11} + s_{22})] \frac{\tilde{\kappa}}{\tau} \geq 0.$$

BEWEIS:

Da die Voraussetzung 7.9 gilt, ist nach Satz 7.6 $b \geq b_0$ mit b_0 aus (7.5).

Weiter gilt mit Lemma 7.3 und Lemma 7.8, dass

$$\begin{aligned}
& b - [(1-b)^2 + (b^2 s_{11} + s_{22})] \frac{\tilde{\kappa}}{\tau} \\
&= \frac{2b(b s_{11} + s_{22})^2 + 2b(1-b)^2 s_{12}^2}{2\tau^2} \\
&\quad - \frac{[(1-b)^2 + (b^2 s_{11} + s_{22})]}{2\tau^2} \\
&\quad \cdot \left[(1+b)(b s_{11} + s_{22}) - (1-b) \sqrt{(b s_{11} + s_{22})^2 - 4b s_{12}^2} \right] \\
&= \frac{b s_{11} + s_{22}}{2\tau^2} [(1-b)(b^2 s_{11} - s_{22}) - (1-b)(1-b^2)] \\
&\quad + \frac{2b(1-b)^2 s_{12}^2}{2\tau^2} \\
&\quad + \frac{1-b}{2\tau^2} [(1-b)^2 + (b^2 s_{11} + s_{22})] \sqrt{(b s_{11} + s_{22})^2 - 4b s_{12}^2} \\
&= \frac{1-b}{2\tau^2} \Phi(b) \geq 0
\end{aligned}$$

ist. □

Korollar 7.16 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9:*

$$b - (b^2 s_{11} + s_{22}) \frac{\tilde{\kappa}}{\tau} \geq 0.$$

BEWEIS:

Folgt mit $(1-b)^2 > 0$ sofort aus Lemma 7.15. □

Lemma 7.17 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9:*

$$\begin{aligned}
& -(1+b^2) s_{12} \tilde{\kappa} \frac{\tilde{\kappa}^2 - b^2}{1-b^2} + b(s_{11} - s_{22}) s_{12} \frac{\tilde{\kappa}^2}{\tau} + b^2 s_{12} \tilde{\kappa} \\
&= \frac{(1+b) s_{12}}{1-b} \tilde{\kappa} \left[b - \frac{\tilde{\kappa}}{\tau} (b^2 s_{11} + s_{22}) \right].
\end{aligned}$$

BEWEIS:

Mit Anwendung von Lemma 7.2 und Lemma 7.3 (b) folgt:

$$\begin{aligned}
& -(1+b^2) s_{12} \tilde{\kappa} \frac{\tilde{\kappa}^2 - b^2}{1-b^2} + b(s_{11} - s_{22}) s_{12} \frac{\tilde{\kappa}^2}{\tau} + b^2 s_{12} \tilde{\kappa} \\
&= \frac{s_{12}}{1-b^2} \tilde{\kappa} \left[2b^2 - \tilde{\kappa}^2 (1+b^2) + b(1-b^2) (s_{11} - s_{22}) \frac{\tilde{\kappa}}{\tau} \right] \\
&= \frac{(1+b) s_{12}}{1-b^2} \tilde{\kappa} \\
&\quad \left[b(1+b) - (1+b^2) (b s_{11} + s_{22}) \frac{\tilde{\kappa}}{\tau} + b(1-b) (s_{11} - s_{22}) \frac{\tilde{\kappa}}{\tau} \right] \\
&= \frac{(1+b) s_{12}}{1-b} \tilde{\kappa} \left[b - \frac{\tilde{\kappa}}{\tau} (b^2 s_{11} + s_{22}) \right].
\end{aligned}$$

□

Lemma 7.18 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9:*

(a)

$$2 \frac{(\tilde{\kappa}^2 - b^2)^2}{1-b^2} - (1-3b^2) \frac{\tilde{\kappa}^2 - b^2}{1-b^2} - b^2 = \frac{2\tilde{\kappa}^2 - 1 - b^2}{1-b^2} \tilde{\kappa}^2,$$

(b)

$$\begin{aligned}
& -(1+b^2) \frac{\tilde{\kappa}^2 - b^2}{1-b^2} + b(s_{11} - s_{22}) \frac{\tilde{\kappa}}{\tau} + b^2 \\
&= \frac{1+b}{1-b} \left[b - (b^2 s_{11} + s_{22}) \frac{\tilde{\kappa}}{\tau} \right].
\end{aligned}$$

BEWEIS:

Es ist:

(a)

$$\begin{aligned}
& 2 \frac{(\tilde{\kappa}^2 - b^2)^2}{1-b^2} - (1-3b^2) \frac{\tilde{\kappa}^2 - b^2}{1-b^2} - b^2 \\
&= \frac{1}{1-b^2} (2\tilde{\kappa}^4 - b^2 \tilde{\kappa}^2 - \tilde{\kappa}^2) \\
&= \frac{2\tilde{\kappa}^2 - 1 - b^2}{1-b^2} \tilde{\kappa}^2,
\end{aligned}$$

(b)

$$\begin{aligned}
& -(1+b^2)\frac{\tilde{\kappa}^2 - b^2}{1-b^2} + b(s_{11} - s_{22})\frac{\tilde{\kappa}}{\tau} + b^2 \\
&= -(1+b^2)\frac{(b s_{11} + s_{22})\frac{\tilde{\kappa}}{\tau} - b}{1-b} + b(s_{11} - s_{22})\frac{\tilde{\kappa}}{\tau} + b^2 \\
&= \frac{1}{1-b} \left[b(1+b) - (b^3 s_{11} + b^2 s_{11} + s_{22} + b s_{22})\frac{\tilde{\kappa}}{\tau} \right] \\
&= \frac{1+b}{1-b} \left[b - (b^2 s_{11} + s_{22})\frac{\tilde{\kappa}}{\tau} \right].
\end{aligned}$$

□

Lemma 7.19 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9:*

(a)

$$\begin{aligned}
& b(2\tilde{\kappa}^2 - 1 - b^2) + b(1+b)^2 - (1+b)^2(b^2 s_{11} + s_{22})\frac{\tilde{\kappa}}{\tau} \\
& -(1-b^2)(b^2 s_{11} - s_{22})\frac{\tilde{\kappa}}{\tau} = 0,
\end{aligned}$$

(b)

$$[b\tau - (b^2 s_{11} + s_{22})\tilde{\kappa}] [(s_{11} + s_{22})\tilde{\kappa} - \tau] = -(1-b)^2(s_{11}s_{22} - s_{12}^2)\tilde{\kappa}^2.$$

BEWEIS:

Durch Verwendung von Lemma 7.3 (b) ergibt sich:

(a)

$$\begin{aligned}
& b(2\tilde{\kappa}^2 - 1 - b^2) + b(1+b)^2 - (1+b)^2(b^2 s_{11} + s_{22})\frac{\tilde{\kappa}}{\tau} \\
& -(1-b^2)(b^2 s_{11} - s_{22})\frac{\tilde{\kappa}}{\tau} \\
&= 2b(1+b)(b s_{11} + s_{22})\frac{\tilde{\kappa}}{\tau} - [2b^2(s_{11} + s_{22}) + 2b(b^2 s_{11} + s_{22})]\frac{\tilde{\kappa}}{\tau} \\
&= 0,
\end{aligned}$$

(b)

$$\begin{aligned}
& [b\tau - (b^2 s_{11} + s_{22})\tilde{\kappa}] [(s_{11} + s_{22})\tilde{\kappa} - \tau] \\
&= b(s_{11} + s_{22})\tau\tilde{\kappa} - (b^2 s_{11} + s_{22})(s_{11} + s_{22})\tilde{\kappa}^2 - b\tau^2 + (b^2 s_{11} + s_{22})\tau\tilde{\kappa} \\
&= (1+b)(b s_{11} + s_{22})\tau\tilde{\kappa} - [(b s_{11} + s_{22})^2 + (1-b)^2 s_{11} s_{22}] \tilde{\kappa}^2 - b\tau^2 \\
&= (1+b)(b s_{11} + s_{22})\tau\tilde{\kappa} - [\tau^2 + (1-b)^2(s_{11} s_{22} - s_{12}^2)] \tilde{\kappa}^2 - b\tau^2 \\
&= -(1-b)^2(s_{11} s_{22} - s_{12}^2) \tilde{\kappa}^2.
\end{aligned}$$

□

Korollar 7.20 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9*

$$\frac{\tilde{\kappa}}{\tau} < \frac{1}{s_{11} + s_{22}}.$$

BEWEIS:

Da wegen Korollar 7.16 $b\tau - (b^2 s_{11} + s_{22})\tilde{\kappa} > 0$ und $-(1-b^2)(s_{11} s_{22} - s_{12}^2)\tilde{\kappa}^2 < 0$ sind, muss mit Lemma 7.19 (b) auch $(s_{11} + s_{22})\tilde{\kappa} - \tau < 0$ sein und die Behauptung ist erfüllt. □

Mit Hilfe der vorangegangenen Lemmata kann nun im nächsten Satz die Existenz einer Nullstelle der Funktion f aus Satz 7.13 auf einem gegebenen Intervall sichergestellt werden.

Satz 7.21 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor besitzt unter der Voraussetzung 7.9 die Funktion f aus Satz 7.13 eine Nullstelle im Intervall $(0, \sqrt{\frac{\tilde{\kappa}^2 - b^2}{1 - b^2}})$.*

BEWEIS:

a) Es gilt

$$f(0) = b^3 s_{12} \mu - b^3 s_{12} (1 + s_3) = b^3 s_{12} [\mu - (1 + s_3)] < 0.$$

b) Definiere

$$x_r := \sqrt{\frac{\tilde{\kappa}^2 - b^2}{1 - b^2}} \quad (7.9)$$

und

$$\mu_0 := \tilde{\kappa}(1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}}.$$

Dann ist

$$\begin{aligned} \sqrt{1 - x_r^2} &= \sqrt{\frac{1 - \tilde{\kappa}^2}{1 - b^2}}, \\ \sqrt{b^2 + (1 - b^2)x_r^2} &= \tilde{\kappa} \end{aligned}$$

und mit Lemma 7.3 (c)

$$x_r \sqrt{1 - x_r^2} = \frac{\sqrt{\tilde{\kappa}^2 - b^2} \sqrt{1 - \tilde{\kappa}^2}}{1 - b^2} = s_{12} \frac{\tilde{\kappa}}{\tau}.$$

Dabei ist τ wie in (7.4). Somit ist weiter

$$\begin{aligned} f(x_r) &= 2b s_{12} (1 + s_3) \frac{(\tilde{\kappa}^2 - b^2)^2}{1 - b^2} + (\alpha - \beta) s_{12} (\tilde{\kappa}^2 - b^2) \frac{\tilde{\kappa}}{\tau} \\ &\quad - b(1 - 3b^2) s_{12} (1 + s_3) \frac{\tilde{\kappa}^2 - b^2}{1 - b^2} \\ &\quad - (1 + b^2) s_{12} \mu \frac{\tilde{\kappa}^2 - b^2}{1 - b^2} \tilde{\kappa} + b^2 (\alpha - \beta) s_{12} \frac{\tilde{\kappa}}{\tau} \\ &\quad + b(s_{11} - s_{22}) s_{12} \mu \frac{\tilde{\kappa}^2}{\tau} + b^2 s_{12} \mu \tilde{\kappa} - b^3 s_{12} (1 + s_3) \\ &\stackrel{\text{Lemma 7.18}}{=} \frac{s_{12}}{1 - b^2} \tilde{\kappa}^2 [b(1 + s_3)(2\tilde{\kappa}^2 - 1 - b^2)] \\ &\quad + (\alpha - \beta) s_{12} \frac{\tilde{\kappa}^3}{\tau} \\ &\quad + \frac{(1 + b) s_{12}}{1 - b} \tilde{\kappa} \left[b - \frac{\tilde{\kappa}}{\tau} (b^2 s_{11} + s_{22}) \right] \mu \\ &\stackrel{\text{Lemma 7.15}}{>} \frac{s_{12} (1 + s_3)}{1 - b^2} \tilde{\kappa}^2 \\ &\quad \cdot \left[b(2\tilde{\kappa}^2 - 1 - b^2) - (1 - b^2)(b^2 s_{11} - s_{22}) \frac{\tilde{\kappa}}{\tau} \right] \\ &\quad + (1 - b^2) s_{12} s_3 \frac{\tilde{\kappa}^3}{\tau} \end{aligned}$$

$$\begin{aligned}
& + \frac{(1+b)s_{12}}{1-b} \tilde{\kappa} \left[b - \frac{\tilde{\kappa}}{\tau} (b^2 s_{11} + s_{22}) \right] \mu_0 \\
= & \frac{s_{12}(1+s_3)}{1-b^2} \tilde{\kappa}^2 \\
& \cdot \left[b(2\tilde{\kappa}^2 - 1 - b^2) - (1-b^2)(b^2 s_{11} - s_{22}) \frac{\tilde{\kappa}}{\tau} \right] \\
& + (1-b^2) s_{12} s_3 \frac{\tilde{\kappa}^3}{\tau} \\
& + \frac{(1+b)s_{12}}{1-b} \tilde{\kappa} \left[b - \frac{\tilde{\kappa}}{\tau} (b^2 s_{11} + s_{22}) \right] \\
& \cdot \left[\tilde{\kappa}(1+s_3) + (\tilde{\kappa} \operatorname{tr} \tilde{S}^{-1} - \tilde{F}^{\frac{1}{2}}) s_3 \right] \\
\stackrel{\text{Lemma 7.1}}{=} & \frac{s_{12}(1+s_3)}{1-b^2} \tilde{\kappa}^2 \\
& \cdot \left[b(2\tilde{\kappa}^2 - 1 - b^2) + b(1+b)^2 \right. \\
& \left. - (1+b)^2 (b^2 s_{11} + s_{22}) \frac{\tilde{\kappa}}{\tau} - (1-b^2)(b^2 s_{11} - s_{22}) \frac{\tilde{\kappa}}{\tau} \right] \\
& + (1-b^2) s_{12} s_3 \frac{\tilde{\kappa}^3}{\tau} \\
& + \frac{(1+b)s_{12}s_3}{(1-b)(s_{11}s_{22} - s_{12}^2)} \frac{\tilde{\kappa}}{\tau} \\
& \cdot \left[b\tau - \tilde{\kappa}(b^2 s_{11} + s_{22}) \right] \left[\tilde{\kappa}(s_{11} + s_{22}) - \tau \right] \\
\stackrel{\text{Lemma 7.19}}{=} & (1-b^2) s_{12} s_3 \frac{\tilde{\kappa}^3}{\tau} - (1-b^2) s_{12} s_3 \frac{\tilde{\kappa}^3}{\tau} = 0
\end{aligned}$$

und schließlich $f(x_r) > 0$.

Durch die Stetigkeit von f auf dem Intervall $[-1, 1]$ folgt aus a) und b) mit dem Nullstellensatz, dass ein $x^* \in (0, x_r)$ mit $f(x^*) = 0$ existiert. \square

Die obere Grenze des Intervalls $(0, x_r)$ mit x_r aus (7.9) ist unabhängig von den Einträgen der Matrizen $X'X$ und B beschränkt.

Lemma 7.22 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9:*

$$x_r < \frac{1}{\sqrt{2}}.$$

BEWEIS:

Man betrachte den Fall $b s_{11} - s_{22} \geq 0$. Dann gilt mit Lemma 7.3 (b) und Korollar 7.20

$$\begin{aligned}\tilde{\kappa}^2 - b^2 &= (1+b)(b s_{11} + s_{22}) \frac{\tilde{\kappa}}{\tau} - b(1+b) \\ &< (1-b^2) \frac{s_{22}}{s_{11} + s_{22}} \\ &\leq (1-b^2) \frac{b s_{22}}{s_{22} + b s_{22}} < \frac{1}{2}(1-b^2).\end{aligned}$$

Im Fall $b s_{11} - s_{22} < 0$ erhält man ebenfalls mit Lemma 7.3 (b) und Korollar 7.16

$$\begin{aligned}\tilde{\kappa}^2 - b^2 &= (1+b)(b s_{11} + s_{22}) \frac{\tilde{\kappa}}{\tau} - b(1+b) \\ &< b(1+b) \frac{(b s_{11} + s_{22})}{(b^2 s_{11} + s_{22})} - b(1+b) \\ &= b^2(1-b^2) \frac{s_{11}}{b^2 s_{11} + s_{22}} \\ &\leq (1-b^2) \frac{b}{1+b} < \frac{1}{2}(1-b^2).\end{aligned}$$

Aus beiden Fällen folgt nun

$$x_r = \sqrt{\frac{\tilde{\kappa}^2 - b^2}{1 - b^2}} < \frac{1}{\sqrt{2}}.$$

□

Für große μ , also $\mu \in (\frac{b}{\tilde{\kappa}}(1+s_3), 1+s_3)$, kann das Intervall, in dem die Nullstelle liegt, noch verkleinert werden. Definiere

$$x_s := \frac{b}{\mu} \sqrt{\frac{(1+s_3)^2 - \mu^2}{1-b^2}}. \quad (7.10)$$

Dann ist $x_s > 0$ und für $\mu > \frac{b}{\tilde{\kappa}}(1+s_3)$ gilt

$$x_s < \frac{\tilde{\kappa}}{1+s_3} \sqrt{\frac{(1+s_3)^2 - \frac{b^2}{\tilde{\kappa}^2}(1+s_3)^2}{1-b^2}} = \sqrt{\frac{\tilde{\kappa}^2 - b^2}{1-b^2}} = x_r.$$

Für $\mu > \frac{b}{\kappa}(1 + s_3)$ existiert schon im verkleinerten Intervall $(0, x_s)$ mit x_s aus (7.10) eine Nullstelle. Dies zeigt der folgende Satz.

Satz 7.23 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor besitzt unter der Voraussetzung 7.9 und $\mu > \frac{b}{\kappa}(1 + s_3)$ die Funktion f aus Satz 7.13 eine Nullstelle im Intervall $(0, x_s)$. Dabei ist x_s wie in (7.10).*

BEWEIS:

a) Es ist wegen des Beweises von Satz 7.21 $f(0) < 0$.

b) Es gilt

$$\sqrt{1 - x_s^2} = \frac{1}{\mu} \sqrt{\frac{\mu^2 - b^2(1 + s_3)^2}{1 - b^2}}$$

und

$$\sqrt{b^2 + (1 - b^2)x_s^2} = \frac{b(1 + s_3)}{\mu}.$$

Durch Definition von

$$w := \sqrt{(1 + s_3)^2 - \mu^2} \sqrt{\mu^2 - b^2(1 + s_3)^2}$$

ist dann

$$\begin{aligned} f(x_s) &= 2b s_{12}(1 + s_3) \frac{b^4 [(1 + s_3)^2 - \mu^2]^2}{\mu^4 (1 - b^2)} \\ &\quad + \frac{b^3}{\mu^4} (\alpha - \beta) \frac{(1 + s_3)^2 - \mu^2}{1 - b^2} w \\ &\quad - \frac{b^3}{\mu^2} (1 - 3b^2) s_{12}(1 + s_3) \frac{(1 + s_3)^2 - \mu^2}{1 - b^2} \\ &\quad - \frac{b^3}{\mu^2} (1 + b^2) s_{12}(1 + s_3) \frac{(1 + s_3)^2 - \mu^2}{1 - b^2} \\ &\quad + \frac{b^3}{\mu^2} (\alpha - \beta) \frac{w}{1 - b^2} + \frac{b^3}{\mu^2} (s_{11} - s_{22}) (1 + s_3) \frac{w}{1 - b^2} \\ &\quad + b^3 s_{12}(1 + s_3) - b^3 s_{12}(1 + s_3) \end{aligned}$$

$$\begin{aligned}
&= \frac{b^3}{\mu^2} s_{12} (1 + s_3) \frac{(1 + s_3)^2 - \mu^2}{1 - b^2} \\
&\quad \cdot \left[2 b^2 \frac{(1 + s_3)^2 - \mu^2}{\mu^2} - 2(1 - b^2) \right] \\
&\quad + \frac{b^3}{\mu^2} \frac{w}{1 - b^2} \\
&\quad \cdot \left[(\alpha - \beta) \frac{(1 + s_3)^2 - \mu^2}{\mu^2} + (\alpha - \beta) + (s_{11} - s_{22})(1 + s_3) \right] \\
&= \frac{b^3}{\mu^4} (1 + s_3) \frac{w}{1 - b^2} \\
&\quad \cdot \left[-2 s_{12} w + (\alpha - \beta)(1 + s_3) + \mu^2 (s_{11} - s_{22}) \right].
\end{aligned}$$

Da $\frac{b^3}{\mu^4} (1 + s_3) \frac{w}{1 - b^2} > 0$ gilt, ist nur noch die Betrachtung von $-2 s_{12} w + (\alpha - \beta)(1 + s_3) + \mu^2 (s_{11} - s_{22})$ notwendig. Aus

$$\begin{aligned}
&(\alpha - \beta)(1 + s_3) + \mu^2 (s_{11} - s_{22}) \\
&= (1 + s_3)^2 (s_{22} - b^2 s_{11}) + (1 - b^2)(1 + s_3) s_3 + \mu^2 (s_{11} - s_{22}) \\
&= [(1 + s_3)^2 - \mu^2] s_{22} + [\mu^2 - b^2 (1 + s_3)^2] s_{11} + (1 - b^2)(1 + s_3) s_3
\end{aligned}$$

folgt

$$\begin{aligned}
&-2 s_{12} w + (\alpha - \beta)(1 + s_3) + \mu^2 (s_{11} - s_{22}) \\
&= -2 s_{12} w + [(1 + s_3)^2 - \mu^2] s_{22} + [\mu^2 - b^2 (1 + s_3)] s_{11} \\
&\quad + (1 - b^2)(1 + s_3) s_3 \\
&= \left[\sqrt{(1 + s_3)^2 - \mu^2} \sqrt{s_{22}} - \sqrt{\mu^2 - b^2 (1 + s_3)} \sqrt{s_{11}} \right]^2 \\
&\quad + 2(\sqrt{s_{11} s_{22}} - s_{12}) w + (1 - b^2)(1 + s_3) s_3 > 0
\end{aligned}$$

und somit $f(x_s) > 0$.

Da f auf $[-1, 1]$ stetig ist, existiert mit a) und b) durch Verwendung des Nullstellensatz ein $x^* \in (0, x_s)$ mit $f(x^*) = 0$. \square

Aus den beiden vorangegangenen Sätzen erhält man nun, dass f aus Satz 7.13 auf dem Intervall $(0, \min\{x_r, x_s\})$ mit x_r und x_s aus (7.9) bzw. (7.10) eine

Nullstelle besitzt. Damit eine dieser Nullstellen zur Berechnung des Minimax-Schätzers verwendet werden kann, ist das Verhalten dieser Nullstellen auf dem Intervall $(0, \min\{x_r, x_s\})$ unter den Bedingungen (2), (3) und (4) von Korollar 7.14 von Interesse. Diese Bedingungen sind, dass für eine Nullstelle x^* von f stets $\lambda_1 \geq \lambda_2$ mit λ_1 aus Lemma 7.10 gilt, sowie die Matrix H_1 aus Lemma 7.10 nicht negativ definit ist. Das heißt, es sind $b - \lambda_1 \lambda_2 \geq 0$ und $\mu - \lambda_1 \geq 0$.

Für die Bedingung (2) aus Korollar 7.14 wird sich zeigen, dass die Verkleinerung des Intervalls $(0, x_r)$ für große μ sinnvoll war.

Satz 7.24 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor gilt unter der Voraussetzung 7.9 für alle $x \in (0, x_s)$ mit x_s aus (7.10)*

$$b - \lambda_1 \lambda_2 \geq 0.$$

BEWEIS:

Es ist mit Lemma 7.12

$$b - \lambda_1 \lambda_2 = \frac{b(1 + s_3) - \mu \lambda_2}{q(1 + s_3) + s_3 \lambda_2^2} q.$$

Für $x < x_s$ gilt

$$\lambda_2 = \sqrt{b^2 + (1 - b^2)x^2} < \frac{b(1 + s_3)}{\mu}$$

und damit auch

$$b - \lambda_1 \lambda_2 > \frac{b(1 + s_3) - \mu \frac{b(1+s_3)}{\mu}}{q(1 + s_3) + s_3 \lambda_2^2} q = 0.$$

□

Dadurch, dass für alle Elemente auf dem Intervall die Ungleichung erfüllt ist, wird sie insbesondere auch für die Nullstellen von f erfüllt.

Um die Gültigkeit der Bedingungen (3) und (4) aus Korollar 7.14 nachzuprüfen benötigt man vorab ein paar Identitäten und Abschätzungen.

Lemma 7.25 *Sei die Voraussetzung 7.9 erfüllt und $t : (-1, 1) \rightarrow \mathbb{R}$ durch $t(x) = (1 - b) s_{12} x^2 + b s_{12} - (b s_{11} + s_{22}) x \sqrt{1 - x^2}$ gegeben. Dann gilt für alle $x \in (0, x_r)$ mit x_r aus (7.9) stets $t(x) > 0$.*

BEWEIS:

Die Funktion t ist auf $(-1, 1)$ stetig differenzierbar und die Ableitung lautet

$$t'(x) = 2(1 - b) s_{12} x - (b s_{11} + s_{22}) \frac{1 - 2x^2}{\sqrt{1 - x^2}}.$$

Da wegen Lemma 7.22 $x_r^2 < \frac{1}{2}$ ist, erhält man für $x \in (0, x_r)$ mit Lemma 7.3

$$\begin{aligned} t'(x) &\leq t'(x_r) \\ &= \frac{1}{\sqrt{1 - x_r^2}} \\ &\quad \cdot \left[2(1 - b) s_{12} x_r \sqrt{1 - x_r^2} - (b s_{11} + s_{22}) (1 - 2x_r^2) \right] \\ &= \frac{1}{\sqrt{1 - x_r^2}} \\ &\quad \cdot \left[2(1 - b) s_{12}^2 \frac{\tilde{\kappa}}{\tau} - (b s_{11} + s_{22}) \frac{1 + b^2 - 2\tilde{\kappa}^2}{1 - b^2} \right] \\ &= \frac{1}{\sqrt{1 - x_r^2}} \\ &\quad \cdot \left[2(1 - b) s_{12}^2 \frac{\tilde{\kappa}}{\tau} - (b s_{11} + s_{22}) \frac{1 + b}{1 - b} + 2 \frac{(b s_{11} + s_{22})^2}{1 - b} \frac{\tilde{\kappa}}{\tau} \right] \\ &= \frac{1}{(1 - b) \sqrt{1 - x_r^2}} \left[2\tau^2 \frac{\tilde{\kappa}}{\tau} - (1 + b) (b s_{11} + s_{22}) \right] \\ &= -\frac{1}{\sqrt{1 - x_r^2}} \sqrt{(b s_{11} + s_{22})^2 - 4b s_{12}^2} < 0. \end{aligned}$$

Somit ist t streng monoton fallend und man erhält für $x \in (0, x_r)$ mit Lemma 7.3

$$\begin{aligned}
t(x) &> t(x_r) \\
&= (b s_{11} + s_{22}) s_{12} \frac{\tilde{\kappa}}{\tau} - b s_{12} + b s_{12} - (b s_{11} + s_{22}) s_{12} \frac{\tilde{\kappa}}{\tau} = 0.
\end{aligned}$$

□

Lemma 7.26 Seien $s_{11}, s_{22} > 0$, $s_{12}^2 < s_{11} s_{22}$, $b \in (0, 1)$,

$$\lambda_2 = \sqrt{b^2 + (1 - b^2)x^2},$$

$$p = b(s_{11} - s_{22})x\sqrt{1 - x^2} + b^2 s_{12} - (1 + b^2)s_{12}x^2,$$

und

$$q = (s_{22} - b^2 s_{11})x^2 - 2b s_{12}x\sqrt{1 - x^2} + b^2 s_{11}.$$

Dann gilt:

$$\begin{aligned}
&bp - (1 - b^2)qx\sqrt{1 - x^2} \\
&= \lambda_2^2 \left[b s_{12} (1 - 2x^2) - (s_{22} - b^2 s_{11})x\sqrt{1 - x^2} \right].
\end{aligned}$$

BEWEIS:

Es gilt

$$\begin{aligned}
&bp - (1 - b^2)qx\sqrt{1 - x^2} \\
&= b^2(s_{11} - s_{22})x\sqrt{1 - x^2} + b^3 s_{12} - b(1 + b^2)s_{12}x^2 \\
&\quad - (1 - b^2)(s_{22} - b^2 s_{11})x^3\sqrt{1 - x^2} + 2b(1 - b^2)s_{12}x^2 \\
&\quad - 2b(1 - b^2)s_{12}x^4 - b^2(1 - b^2)s_{11}x\sqrt{1 - x^2} \\
&= -2b(1 - b^2)s_{12}x^4 - (1 - b^2)(s_{22} - b^2 s_{11})x^3\sqrt{1 - x^2} \\
&\quad + b(1 - b^2)s_{12}x^2 - 2b^3 s_{12}x^2 \\
&\quad + b^2(b^2 s_{11} - s_{22})x\sqrt{1 - x^2} + b^3 s_{12} \\
&= \lambda_2^2 \left[b s_{12} (1 - 2x^2) - (s_{22} - b^2 s_{11})x\sqrt{1 - x^2} \right].
\end{aligned}$$

□

Nun kann die Richtigkeit der Ungleichung $\lambda_1 \geq \lambda_2$ für die Nullstellen $x^* \in (0, x_r)$ der Funktion f gezeigt werden.

Satz 7.27 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor sei unter der Voraussetzung 7.9 $x^* \in (0, x_r)$ mit x_r aus (7.9) eine Nullstelle der Funktion f aus Satz 7.13. Dann gilt für*

$$\lambda_1^* = \frac{\mu [(s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1 - x^{*2}} + b^2 s_{11}] + b s_3 \lambda_2^*}{[(s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1 - x^{*2}} + b^2 s_{11}] (1 + s_3) + s_3 \lambda_2^{*2}}$$

und

$$\lambda_2^* = \sqrt{b^2 + (1 - b^2)x^{*2}}$$

stets

$$\lambda_1^* \geq \lambda_2^*.$$

BEWEIS:

Sei $q^* = (s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1 - x^{*2}} + b^2 s_{11}$ und $p^* = b(s_{11} - s_{22})x^* \sqrt{1 - x^{*2}} + b^2 s_{12} - (1 + b^2) s_{12} x^{*2}$. Da x^* eine Nullstelle von f ist, hält x^* das Gleichungssystem (7.8) und es gilt

$$\lambda_1^* \lambda_2^* = b - \frac{(1 - b^2) q^* x^* \sqrt{1 - x^{*2}}}{p^*}.$$

Auf Grund von Satz 7.24 ist $\lambda_1^* \lambda_2^* - b < 0$ und es ist $(1 - b^2) q^* x^* \sqrt{1 - x^{*2}} > 0$.

Dadurch muss $p^* > 0$ sein. Es folgt nun mit Lemma 7.26

$$\begin{aligned} & \lambda_1^* - \lambda_2^* \\ &= \frac{1}{\lambda_2^*} (\lambda_1^* \lambda_2^* - \lambda_2^{*2}) \\ &= \frac{1}{p^* \lambda_2^*} (b p^* - (1 - b^2) q^* x^* \sqrt{1 - x^{*2}} - \lambda_2^{*2} p^*) \\ &= \frac{\lambda_2^*}{p^*} \left[b s_{12} (1 - 2x^{*2}) - (s_{22} - b^2 s_{11}) x^* \sqrt{1 - x^{*2}} \right. \\ & \quad \left. - b (s_{11} - s_{22}) x^* \sqrt{1 - x^{*2}} - b^2 s_{12} + (1 + b^2) s_{12} x^{*2} \right] \end{aligned}$$

$$= \frac{(1-b)\lambda_2^*}{p^*} t(x^*) > 0.$$

Dabei ist t wie in Lemma 7.25. □

Die Bedingung (3) aus Korollar 7.14 ist damit für jede Nullstelle von f auf dem Intervall $(0, x_r)$ erfüllt.

Lemma 7.28 *Sei die Voraussetzung 7.9 erfüllt und $w : (-1, 1) \rightarrow \mathbb{R}$ durch $w(x) = b s_{12}(1 - 2x^2) - (s_{22} - b^2 s_{11} + 1 - b^2)x\sqrt{1-x^2}$ gegeben. Dann gilt für alle $x \in (0, x_r)$ mit x_r aus (7.9) stets $w(x) > 0$.*

BEWEIS:

1.Fall: Sei $s_{22} - b^2 s_{11} + 1 - b^2 \leq 0$. Da wegen Lemma 7.18 $x_r^2 < \frac{1}{2}$ folgt sofort $w(x) > 0$.

2.Fall: Sei $s_{22} - b^2 s_{11} + 1 - b^2 > 0$. Für $x \in (0, x_r)$ ist $x\sqrt{1-x^2}$ streng monoton wachsend. Damit gilt mit Lemma 7.3:

$$\begin{aligned} w(x) &> w(x_r) \\ &= b s_{12} \left[1 - \frac{2(\tilde{\kappa}^2 - b^2)}{1 - b^2} \right] - s_{12} (s_{22} - b^2 s_{11} + 1 - b^2) \frac{\tilde{\kappa}}{\tau} \\ &= \frac{b s_{12}}{1 - b^2} \left[(1 + b)^2 - 2(1 + b)(b s_{11} + s_{22}) \frac{\tilde{\kappa}}{\tau} \right] \\ &\quad - s_{12} (s_{22} - b^2 s_{11} + 1 - b^2) \frac{\tilde{\kappa}}{\tau} \\ &= \frac{b(1+b)s_{12}}{1-b} - \frac{s_{12}(1+b)(b^2 s_{11} + s_{22}) + (1+b)(1-b)^2 s_{12}}{1-b} \frac{\tilde{\kappa}}{\tau} \\ &= \frac{(1+b)s_{12}}{1-b} \left\{ b - [b^2 s_{11} + s_{22} + (1-b)^2] \frac{\tilde{\kappa}}{\tau} \right\} \\ &\geq 0 \quad \text{nach Lemma 7.15.} \end{aligned}$$

□

Auf Grund der vorangegangenen Lemmata ist es nun möglich die Richtigkeit

der Ungleichung $\mu \geq \lambda_1$ für die Nullstellen $x^* \in (0, x_r)$ der Funktion f zu beweisen.

Satz 7.29 *Im linearen Modell bei ellipsoider Vorinformation und bei einem gemeinsamen Eigenvektor sei unter der Voraussetzung 7.9 $x^* \in (0, x_r)$ mit x_r aus (7.9) eine Nullstelle der Funktion f aus Satz 7.13. Dann gilt für*

$$\lambda_1^* = \frac{\mu [(s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1 - x^{*2}} + b^2 s_{11}] + b s_3 \lambda_2^*}{[(s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1 - x^{*2}} + b^2 s_{11}] (1 + s_3) + s_3 \lambda_2^{*2}}$$

stets

$$\mu \geq \lambda_1^*.$$

BEWEIS:

Sei $q^* = (s_{22} - b^2 s_{11})x^{*2} - 2b s_{12} x^* \sqrt{1 - x^{*2}} + b^2 s_{11}$ und $p^* = b(s_{11} - s_{22})x^* \sqrt{1 - x^{*2}} + b^2 s_{12} - (1 + b^2) s_{12} x^{*2}$. Da x^* eine Nullstelle von f ist, hält x^* das Gleichungssystem (7.8) und es gilt

$$\lambda_1^* \lambda_2^* = b - \frac{(1 - b^2) q^* x^* \sqrt{1 - x^{*2}}}{p^*}.$$

Auf Grund von Satz 7.24 ist $\lambda_1^* \lambda_2^* - b < 0$ und es ist $(1 - b^2) q^* x^* \sqrt{1 - x^{*2}} > 0$.

Dadurch muss $p^* > 0$ sein. Desweiteren ist

$$\mu = \frac{\lambda_1^* [q^* (1 + s_3) + s_3 \lambda_2^{*2}] - b s_3 \lambda_2^*}{q^*}.$$

Für $x^* \in (0, x_r)$ folgt mit Lemma 7.26:

$$\begin{aligned} \mu - \lambda_1^* &= \frac{1}{\lambda_2^*} (\mu \lambda_2^* - \lambda_1^* \lambda_2^*) \\ &= \frac{1}{\lambda_2^*} \left\{ \frac{[q^* (1 + s_3) + s_3 \lambda_2^{*2} - q^*] \lambda_1^* \lambda_2^* - b s_3 \lambda_2^{*2}}{q^*} \right\} \\ &= \frac{s_3}{\lambda_2^*} \left\{ \frac{(q^* + \lambda_2^{*2}) \left[b - \frac{(1 - b^2) q^* x^* \sqrt{1 - x^{*2}}}{p^*} \right] - b \lambda_2^{*2}}{q^*} \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{s_3}{\lambda_2^* p^*} \left[b p^* - (1 - b^2) (q^* + \lambda_2^{*2}) x^* \sqrt{1 - x^{*2}} \right] \\
&= \frac{s_3 \lambda_2^*}{p^*} \\
&\quad \cdot \left[b s_{12} (1 - 2 x^{*2}) - (s_{22} - b^2 s_{11}) x^* \sqrt{1 - x^{*2}} - (1 - b^2) x^* \sqrt{1 - x^{*2}} \right] \\
&= \frac{s_3 \lambda_2^*}{p^*} w(x^*) > 0 \quad \text{nach Lemma 7.28.}
\end{aligned}$$

□

Damit ist nun ebenfalls gezeigt, dass jede Nullstelle von f auf dem Intervall $(0, x_r)$ die Bedingung (4) aus Korollar 7.14 hält. Da alle Nullstellen die Ungleichungen aus Satz 7.27 und Satz 7.29 halten, sind diese Aussagen auch für die Nullstellen auf $(0, x_s)$ mit x_s aus (7.10) für $\mu > \frac{b}{\kappa} (1 + s_3)$ richtig.

Zusammengefasst erhält man aus Satz 7.24, Satz 7.27 und Satz 7.29, dass jede Nullstelle auf dem Intervall $(0, \min(x_r, x_s))$ alle drei Bedingungen erfüllt. Dadurch ist jede Nullstelle zur Berechnung des MILE gemäß Korollar 7.14 geeignet. Da jedoch der MILE eindeutig ist, kann die Funktion f aus Satz 7.13 auf dem Intervall $(0, \min(x_r, x_s))$ nur genau eine Nullstelle besitzen. Damit wurde das Problem der Bestimmung des MILE unter der Voraussetzung 7.9 auf das Berechnen von einer Nullstelle auf einem vorgegebenen Intervall einer reellwertigen Funktion reduziert.

In Kapitel 8 wird sich mit der Darstellung dieser Nullstellen anhand des Beispiels aus Kapitel 6 beschäftigt. Aber für eine numerische Bestimmung der Nullstelle sind die Intervallschachtelung und Regula falsi geeignet. Das liegt daran, dass f für alle Einträge der Matrizen S und B , die die Voraussetzung 7.9 halten, auf $(0, x_r)$ stetig und $f(0) < 0$ sowie $f(x_r) > 0$ gilt. Wegen der Abschätzung $x_r < \frac{1}{\sqrt{2}}$ kann bei der Intervallschachtelung eine maximale Iterationsanzahl für eine gegebene Genauigkeit von l Stellen bestimmt werden. Sei $x_{(k)}$ der Intervallmittelpunkt nach k Intervallhalbierungen und x^*

die Nullstelle, so gilt die Abschätzung

$$|x_{(k)} - x^*| \leq \frac{1}{2^{k+\frac{3}{2}}}.$$

Bei einer Genauigkeit von l Stellen muss man maximal $[(l+1) \frac{\log 10}{\log 2} - \frac{3}{2}] + 1$ Iterationen durchführen. Mit $\lceil \cdot \rceil$ werde die Gaußklammer bezeichnet. Zum Beispiel benötigt man bei einer Genauigkeit von 10 Stellen maximal 36 Iterationen.

7.4 Ausblick

Es stellt sich nun die Frage, wie der MILE bestimmt wird, wenn X und B keinen gemeinsamen Eigenvektor besitzen, also wenn

$$X'X = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{pmatrix}$$

mit $s_{13} \neq 0$ oder $s_{23} \neq 0$ gilt, und der Fall $j = 2$ eingetreten ist. Eine Strukturierung der Fälle wie in Kapitel 6.1 ist ohne weiteres nicht mehr möglich, da die Abhängigkeiten der Einträge der Matrizen beim Fall $j = 2$ durch mehrere und komplexere Gleichungen und Ungleichungen beschrieben werden müssten. Ungeachtet dessen besteht trotzdem ein Interesse (3) aus Satz 4.8 zu lösen, denn beim Eintreten des Falls $j = 2$ ist eine Lösung dieses Gleichungssystems zur Bestimmung des MILE geeignet. Es wird nun kurz eine Idee gegeben, wie man dieses Gleichungssystem aufstellen könnte. Dazu sei $u = (x \ y \ z)'$ mit $x^2 + y^2 + z^2 = 1$. Man erhält $\lambda_2 = x^2 + b^2 y^2 + \mu^2 z^2$ und $v = -\frac{1}{\sqrt{\lambda_2}} (x \ by \ \mu z)'$. Zur Bestimmung der Orthonormalbasen $\{a_1, a_2, u\}$ und $\{b_1, b_2, v\}$ kann man sich die Struktur der Helmert-Matrix von der Ordnung 3

aus ([25], S. 13)

$$\mathcal{H}_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{pmatrix}$$

zu Nutzen machen. Diese Matrix \mathcal{H}_3 ist eine orthogonale Matrix. Betrachtet man für $z \neq 0$ und $z \neq 1$ die Matrix

$$\mathcal{U} = \begin{pmatrix} x & y & x \\ y & -x & y \\ z & 0 & -\frac{x^2+y^2}{z} \end{pmatrix},$$

so gilt wegen

$$x^2 + y^2 + \frac{(x^2 + y^2)^2}{z^2} = \frac{x^2 + y^2}{z^2} (x^2 + y^2 + z^2) = \frac{x^2 + y^2}{z^2},$$

dass

$$\mathcal{U}'\mathcal{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & x^2 + y^2 & 0 \\ 0 & 0 & \frac{x^2+y^2}{z^2} \end{pmatrix}$$

ist. Mit $a_1 = \frac{1}{\sqrt{x^2+y^2}} (y \ -x \ 0)'$ und $a_2 = \frac{1}{\sqrt{x^2+y^2}} (xz \ yz \ -(x^2 + y^2))'$ ist durch $\{a_1, a_2, u\}$ eine Orthonormalbasis des \mathbb{R}^3 gegeben. Ähnlich verfährt man bei der Bestimmung von b_1 und b_2 und erhält $b_1 = \frac{1}{\sqrt{x^2+b^2y^2}} (by \ -x \ 0)'$ und $b_2 = \frac{1}{\lambda_2 \sqrt{x^2+y^2}} (\mu xz \ b\mu yz \ -(x^2 + b^2y^2))'$. Dann ist $\{b_1, b_2, v\}$ eine Orthonormalbasis des \mathbb{R}^3 . Es ist nun möglich $X_1, B_1, \lambda_1, H_1, G_1$ und W aus Satz 4.8 in Abhängigkeit von x und y anzugeben und ein zu (3) aus Satz 4.8 äquivalentes Gleichungssystem zu formulieren. Falls $z = 0$ ist, kann man mit den Orthonormalbasen aus Kapitel 7.2 starten. Die Orthonormalbasen $\{(1 \ 0 \ 0)', (0 \ 1 \ 0)', (0 \ 0 \ 1)'\}$ bzw. $\{(1 \ 0 \ 0)', (0 \ 1 \ 0)', (0 \ 0 \ -1)'\}$ können bei $z = 1$ zum Einsatz kommen. Bei mehr als drei Einflussgrößen ($k \geq 4$) kann

die Orthonormalbasis des \mathbb{R}^k aus der Struktur der Helmert-Matrix von der Ordnung k aus ([25], S. 13)

$$\mathcal{H}_k = \begin{pmatrix} \frac{1}{\sqrt{k}} & \frac{1}{\sqrt{1 \cdot 2}} & \frac{1}{\sqrt{2 \cdot 3}} & \cdots & \cdots & \frac{1}{\sqrt{(k-1) \cdot k}} \\ \frac{1}{\sqrt{k}} & -\frac{1}{\sqrt{1 \cdot 2}} & \frac{1}{\sqrt{2 \cdot 3}} & \cdots & \cdots & \frac{1}{\sqrt{(k-1) \cdot k}} \\ \frac{1}{\sqrt{k}} & 0 & -\frac{2}{\sqrt{2 \cdot 3}} & \cdots & \cdots & \frac{1}{\sqrt{(k-1) \cdot k}} \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \frac{1}{\sqrt{(k-1) \cdot k}} \\ \frac{1}{\sqrt{k}} & 0 & 0 & \cdots & \cdots & -\frac{k-1}{\sqrt{(k-1) \cdot k}} \end{pmatrix}$$

hergeleitet werden. Gibt man sich den Vektor $u = (\alpha_1 \ \alpha_2 \ \cdots \ \alpha_k)$ mit $\sum_{i=1}^k \alpha_i^2 = 1$, $\alpha_1^2 + \alpha_2^2 \neq 0$ und $\alpha_i \neq 0$ für $i = 3, \dots, k$ vor, so erhält man mit

$$\mathcal{U}_k = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_1 & \alpha_1 & \cdots & \cdots & \alpha_1 \\ \alpha_2 & -\alpha_1 & \alpha_2 & \alpha_2 & \cdots & \cdots & \alpha_2 \\ \alpha_3 & 0 & -\frac{\alpha_1^2 + \alpha_2^2}{\alpha_3} & \alpha_3 & \cdots & \cdots & \alpha_3 \\ \alpha_4 & 0 & 0 & -\frac{\sum_{i=1}^3 \alpha_i^2}{\alpha_4} & \cdots & \cdots & \alpha_4 \\ \vdots & \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \alpha_{k-1} \\ \alpha_k & 0 & 0 & 0 & \cdots & 0 & -\frac{\sum_{i=1}^{k-1} \alpha_i^2}{\alpha_k} \end{pmatrix}$$

eine Matrix, bei der alle Spalten paarweise senkrecht aufeinander stehen. Somit stehen die Spalten als Vektoren einer Orthonormalbasis nach entsprechender Normierung zur Verfügung. Jedoch ist hier zu bedenken, dass es bis jetzt keine Kriterien, mit denen man den Fall $j = k - 1$ für $k \geq 4$ aufspüren kann, gibt.

Kapitel 8

Darstellung der Lösung

In diesem Kapitel geht es darum, ob der MILE im linearen Regressionsmodell bei ellipsoider Vorinformation, gemeinsamem Eigenvektor und Gültigkeit der Voraussetzung 7.9 in einer geschlossenen Form angegeben werden kann, also hier durch die Einträge der Matrizen $X'X$ und B aus (6.1) und (6.2). In Kapitel 7 wurde gezeigt, dass sich die Bestimmung des MILE auf das Auffinden einer Nullstelle einer nicht rationalen Funktion reduziert. Diese Funktion f aus Satz 7.13 enthält zwei verschiedene Wurzelterme bzw. das Produkt dieser Wurzelterme. Daher ist es möglich diese Funktion durch entsprechende Umformungen und zweifaches Quadrieren in ein Polynom g zu transformieren, so dass gilt:

$$\forall x \in (-1, 1) \quad f(x) = 0 \quad \Rightarrow \quad g(x) = 0.$$

Es stellt sich nun die Frage, ob das Polynom g durch Radikale auflösbar ist und dadurch dann die entsprechende Nullstelle von f und folglich auch der MILE mittels der Einträge der Matrizen $X'X$ und B aus (6.1) und (6.2) angegeben werden kann. Hier wird nun das Beispiel aus Drygas([8], S. 9ff),

welches auch am Anfang von Kapitel 6 beschrieben ist, betrachtet. Doch um diese Frage zu beantworten werden erst einmal die notwendigen algebraischen Grundlagen dazu bereit gestellt.

8.1 Algebraische Grundlagen

Es gilt nun zu entscheiden, wann ein bestimmtes Polynom durch Radikale auflösbar ist und wann nicht. Dies ist nicht sofort an der Darstellung des Polynoms zu erkennen, aber die Galois-Theorie bietet eine Möglichkeit zur Erkennung, ob ein Polynom durch Radikale auflösbar ist oder nicht. Durch Eigenschaften der Galois-Gruppe des Polynoms kann die Entscheidung der Auflösbarkeit gefällt werden. Deswegen werden erst einmal die benötigten Definitionen und Sätze aus der Gruppentheorie bereitgestellt.

Dieser Abschnitt orientiert sich hauptsächlich an den Ergebnissen aus Bosch [1] und Kunz [20].

Definition 8.1 *Sei G eine endliche Gruppe. Die Ordnung von G ist gleich der Anzahl ihrer Elemente, also $\text{ord}(G) = |G|$.*

Die zentrale Eigenschaft für die Entscheidung, ob ein Polynom durch Radikale aufgelöst werden kann, ist die Auflösbarkeit einer Gruppe.

Definition 8.2 *Sei G eine Gruppe mit Verknüpfung „ \cdot “. Eine Untergruppe $U \subset G$ heißt Normalteiler von G , wenn $\{a \cdot u | u \in U\} = \{u \cdot a | u \in U\}$ für alle $a \in G$ gilt.*

Für eine Gruppe G mit Normalteiler U bezeichnet die Restklassengruppe G/U die Menge $\{\{a \cdot u | u \in U\}, a \in G\}$ mit entsprechender Ver-

knüpfung. Diese Gruppe G/U existiert und ist bis auf Isomorphie eindeutig (vgl. [20], S. 131f).

Definition 8.3 *Eine Gruppe G mit neutralem Element e heißt auflösbar, wenn es eine Kette*

$$\{e\} = U_0 \subset U_1 \subset \cdots \subset U_{m-1} \subset U_m = G$$

von Untergruppen U_i ($i = 0, \dots, m$) gibt, für die gilt:

- (1) U_{i-1} ist Normalteiler von U_i ($i = 1, \dots, m$),
- (2) U_i/U_{i-1} ist abelsch für alle $i \in \{1, \dots, m\}$.

Eine wichtige Rolle in der Galois-Theorie spielt die symmetrische Gruppe \mathcal{S}_n . Diese Gruppe besteht aus den bijektiven Selbstabbildungen auf der Menge $\{1, \dots, n\}$ mit $n \in \mathbb{N}$. Es gilt $\text{ord}(\mathcal{S}_n) = n!$.

Satz 8.4 *Sei $n \in \mathbb{N}$ und \mathcal{S}_n die symmetrische Gruppe. Für $n \geq 5$ ist \mathcal{S}_n nicht auflösbar.*

BEWEIS:

vgl. ([1], S. 253) □

Für die Definition bzw. Konstruktion der Galois-Gruppe werde ein Polynom g vom Grad n ($n \in \mathbb{N}$) mit Koeffizienten aus den rationalen Zahlen betrachtet. Die Menge der Polynome mit Koeffizienten aus einem Körper K werde mit $K[X]$ bezeichnet.

Definition 8.5 *Ein Polynom $g \in K[X]$ heißt irreduzibel, wenn es keine nicht konstanten Polynome $p, q \in K[X]$ mit $g = pq$ gibt.*

Eine Überprüfung der Irreduzibilität ist mit dem Reduktionskriterium (vgl. [1], S. 66) möglich. Sei dazu h ein Polynom mit Koeffizienten in \mathbb{Z} . Es gebe eine Primzahl p , die nicht den höchsten Koeffizienten des Polynoms

h teilt. Ist das Polynom h_p , welches durch Überführung der Koeffizienten des Polynoms h von \mathbb{Z} nach $\mathbb{Z}/p\mathbb{Z}$ entsteht, in $\mathbb{Z}/p\mathbb{Z}$ irreduzibel, so ist h in \mathbb{Q} irreduzibel.

Aus dem Fundamentalsatz der Algebra ist bekannt, dass ein Polynom g vom Grad n in den komplexen Zahlen genau n Nullstellen x_1, \dots, x_n besitzt und dort in Linearfaktoren zerfällt. Jedoch gibt es Teilkörper von \mathbb{C} , die auch diese Eigenschaften inne haben. Dies führt zu folgender Definition.

Definition 8.6 Sei $g \in \mathbb{Q}[X]$ ein nicht konstantes Polynom. Ein Erweiterungskörper L von \mathbb{Q} heißt Zerfällungskörper von g , falls gilt:

- (1) g zerfällt über L vollständig in Linearfaktoren,
- (2) L entsteht aus \mathbb{Q} durch Adjunktion der Nullstellen von g .

Zu jedem nicht konstanten Polynom existiert ein bis auf Isomorphie eindeutiger Zerfällungskörper. Der Grad dieser Körpererweiterung wird durch die Vektorraumdimension von L über \mathbb{Q} angegeben. Man schreibt für den Grad der Erweiterung $[L : \mathbb{Q}]$.

Ein Zerfällungskörper L kann durch die Menge der Automorphismen auf L , die \mathbb{Q} elementweise festlassen, in Zusammenhang gebracht werden.

Definition 8.7 Sei K ein Körper und L eine Körpererweiterung von K . Dann heißt eine Abbildung $\sigma : L \rightarrow L$ mit

- (1) $\sigma(a + b) = \sigma(a) + \sigma(b)$ für alle $a, b \in L$,
- (2) $\sigma(a \cdot b) = \sigma(a) \cdot \sigma(b)$, für alle $a, b \in L$,
- (3) $\sigma|_K = \text{id}$,
- (4) σ ist bijektiv

K-Automorphismus von *L*.

Die *K*-Automorphismen einer Körpererweiterung *L* von *K* bilden eine Gruppe.

Definition 8.8 Sei $g \in \mathbb{Q}[X]$ ein nicht konstantes irreduzibles Polynom vom Grad n und *L* sein Zerfällungskörper. Dann heißt $G(g) = \{\sigma : L \rightarrow L \mid \sigma \text{ } \mathbb{Q}\text{-Automorphismus}\}$ mit der Komposition von Abbildungen als Verknüpfung die Galois-Gruppe von g .

Für die Ordnung der Galois-Gruppe gilt:

$$\text{ord}[G(g)] = [L : \mathbb{Q}] \leq n!.$$

Seien $x_1, \dots, x_n \in L$ die Nullstellen von g . Jedes $\sigma \in G(g)$ ist durch $\sigma(x_i)$ ($i = 1, \dots, n$) bereits eindeutig festgelegt und $\sigma(x_i)$ ist wiederum eine Nullstelle der Funktion g , d.h. $\sigma \in G(g)$ permutiert die Nullstellen von g und ist dadurch eindeutig festgelegt. Damit kann $G(g)$ als Untergruppe der symmetrischen Gruppe \mathcal{S}_n aufgefasst werden (vgl. [20], S. 114f).

Der nächste Satz zeigt, wie man unter gewissen Voraussetzungen anhand der Anzahl der reellen Nullstellen eines Polynoms die Galois-Gruppe erkennen kann.

Satz 8.9 Sei p eine Primzahl und $g \in \mathbb{Q}[X]$ ein irreduzibles Polynom vom Grad p . Es habe g genau zwei nicht reelle Nullstellen. Dann ist \mathcal{S}_p die Galois-Gruppe von g .

BEWEIS:

vgl. ([30], S. 140)

□

Der Zusammenhang zwischen der Auflösbarkeit eines Polynoms durch Radi-

kale und der Galois-Gruppe dieses Polynoms wird im folgenden Satz formuliert.

Satz 8.10 *Sei $g \in \mathbb{Q}[X]$ ein irreduzibles Polynom. Das Polynom g ist genau dann durch Radikale auflösbar, wenn seine Galois-Gruppe $G(g)$ auflösbar ist.*

BEWEIS:

vgl. ([30], S. 139) □

8.2 Keine Lösung durch Radikale

Es werde nun das Beispiel aus Drygas([8], S. 9ff) mit $\mu = 1$ betrachtet. In diesem Abschnitt soll nun die Frage beantwortet werden, ob der MILE in Abhängigkeit der Einträge der Matrizen $X'X$ und B aus (6.1) und (6.2) angegeben werden kann. Mit den Resultaten aus dem Kapitel 7 ist es nun möglich diese Frage mit „nein“ zu beantworten. Es sind

$$X'X = \begin{pmatrix} 3 & \frac{9}{11} & 0 \\ \frac{9}{11} & \frac{15}{44} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

und

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Für dieses Beispiel ist die Voraussetzung 7.9 erfüllt, da mit $s_{11} = \frac{15}{44}$, $s_{22} = 3$, $s_{12} = \frac{9}{11} > 0$, $s_3 = 1$, $b = \frac{1}{2}$ und $\mu = 1$ gemäß Lemma 7.1 und Lemma 7.3 (a)

$$\tilde{\kappa} = \sqrt{\frac{5}{17}} = \frac{596 \sqrt{85}}{10132} \geq \frac{561 \sqrt{85}}{10132} = \frac{33 \sqrt{85}}{596} = \frac{\operatorname{tr} \tilde{F}^{\frac{1}{2}}}{1 + \operatorname{tr} \tilde{S}^{-1}}$$

gilt. Somit liegt die Struktur I aus Kapitel 6.1 vor. Es ist

$$\tilde{\kappa}(1 + s_3 + s_3 \operatorname{tr} \tilde{S}^{-1}) - s_3 \operatorname{tr} \tilde{F}^{\frac{1}{2}} = \frac{92\sqrt{85}}{969} < 1 = \mu$$

und

$$\mu = 1 < 2 = 1 + s_3.$$

Zur Bestimmung des MILE ist die Nullstelle der Funktion f aus Satz 7.13 im Intervall $(0, x_r)$ mit x_r aus (7.9) von Interesse. Man erhält

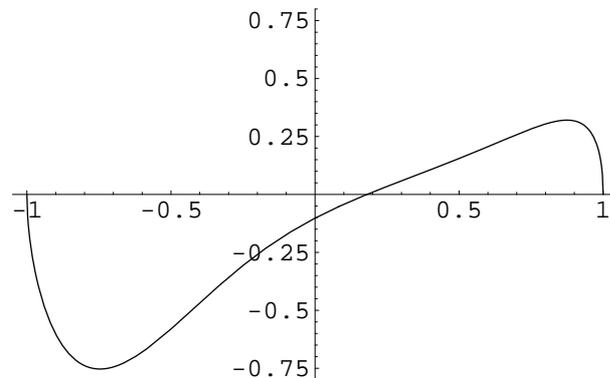
$$x_r = \frac{1}{\sqrt{17}}$$

und f (siehe Abbildung 8.1) durch

$$\begin{aligned} f(x) &= \frac{27}{22}x^4 - \frac{9}{176}x^3\sqrt{1-x^2} - \frac{9}{44}x^2 - \frac{45}{88}x^2\sqrt{1+3x^2} \\ &\quad - \frac{3}{176}x\sqrt{1-x^2} + \frac{117}{176}x\sqrt{1-x^2}\sqrt{1+3x^2} \\ &\quad + \frac{9}{88}\sqrt{1+3x^2} - \frac{9}{44} \\ &= \frac{3}{176}\left(72x^4 - 3x^3\sqrt{1-x^2} - 12x^2 - 30x^2\sqrt{1+3x^2}\right. \\ &\quad \left.- x\sqrt{1-x^2} + 39x\sqrt{1-x^2}\sqrt{1+3x^2}\right. \\ &\quad \left.+ 6\sqrt{1+3x^2} - 12\right) \\ &= \frac{3\sqrt{1+3x^2}}{176}\left(24x^2\sqrt{1+3x^2} - 12\sqrt{1+3x^2}\right. \\ &\quad \left.- x\sqrt{1-x^2}\sqrt{1+3x^2} + 39x\sqrt{1-x^2} - 30x^2 + 6\right). \end{aligned}$$

Weiter gilt:

$$\begin{aligned} f(x) &= 0 \\ \Leftrightarrow & 12(2x^2 - 1)\sqrt{1+3x^2} + 39x\sqrt{1-x^2} \\ &= 6(5x^2 - 1) + x\sqrt{1-x^2}\sqrt{1+3x^2}. \end{aligned}$$

Abbildung 8.1: Graph von f

Durch Quadrieren folgt

$$\begin{aligned}
 & 144(4x^4 - 4x^2 + 1)(1 + 3x^2) + 1521x^2(1 - x^2) \\
 & + 936x(2x^2 - 1)\sqrt{1 - x^2}\sqrt{1 + 3x^2} \\
 & = 36(25x^4 - 10x^2 + 1) + x^2(1 - x^2)(1 + 3x^2) \\
 & + 12x(5x^2 - 1)\sqrt{1 - x^2}\sqrt{1 + 3x^2} \\
 \Leftrightarrow & 36(48x^6 - 57x^4 + 6x^2 + 3) - (1 - x^2)(3x^4 - 1520x^2) \\
 & = 12x\sqrt{1 - x^2}\sqrt{1 + 3x^2}(-156x^2 + 78 + 5x^2 - 1) \\
 \Leftrightarrow & (1 - x^2)[108(16x^4 - 3x^2 - 1) + 3x^4 - 1520x^2] \\
 & = 12x\sqrt{1 - x^2}\sqrt{1 + 3x^2}(151x^2 - 77) \\
 \Leftrightarrow & \sqrt{1 - x^2}(1731x^4 - 1844x^2 - 108) \\
 & = 12x\sqrt{1 + 3x^2}(151x^2 - 77).
 \end{aligned}$$

Durch nochmaliges Quadrieren folgt

$$\begin{aligned}
 & (1 - x^2)(2996361x^8 - 6383921x^6 + 3026440x^4 \\
 & + 398304x^2 + 11664) \\
 & = (432x^4 + 144x^2)(22801x^4 - 23254x^2 + 5929) \\
 \Leftrightarrow & -2996361x^{10} + 9380289x^8 - 9410368x^6
 \end{aligned}$$

$$\begin{aligned}
& 2628136 x^4 - 386640 x^2 + 11664 \\
& = 9850032 x^8 - 6762384 x^6 - 787248 x^4 + 853776 x^2 \\
\Leftrightarrow & 2996361 x^{10} + 469743 x^8 + 2647984 x^6 - 3415384 x^4 \\
& + 467136 x^2 - 11664 = 0.
\end{aligned}$$

Nach Substitution von $z = x^2$ ist die gesuchte Nullstelle von f eine Nullstelle der Funktion g (siehe Abbildung 8.2) mit

$$\begin{aligned}
g(z) = & 2996361 z^5 + 469743 z^4 + 2647984 z^3 - 3415384 z^2 \\
& + 467136 z - 11664.
\end{aligned}$$

auf dem Intervall $(0, \frac{1}{17})$.

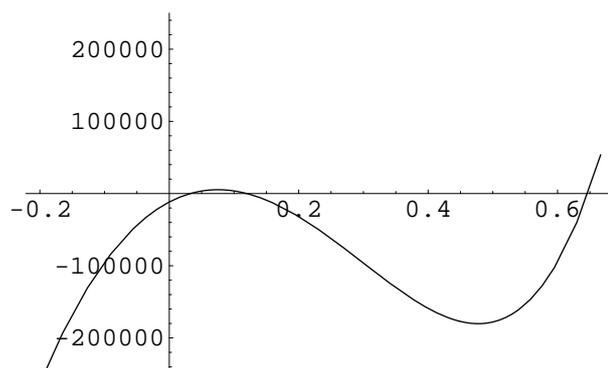


Abbildung 8.2: Graph von g

Die Überprüfung der Irreduzibilität von g wird mit dem Reduktionskriterium durchgeführt. Da 5 nicht 2996361 teilt, erhält man in $\mathbb{Z}/5\mathbb{Z}$

$$g_5(z) = z^5 + 3z^4 + 4z^3 + z^2 + z + 1.$$

Durch Einsetzen der Zahlen $0, \dots, 4 \in \mathbb{Z}/5\mathbb{Z}$ in g_5 ergibt sich durch $g_5(0) = 1$, $g_5(1) = 1$, $g_5(2) = 4$, $g_5(3) = 2$ und $g_5(4) = 4$, dass g_5 in $\mathbb{Z}/5\mathbb{Z}$ keine Nullstellen und daher auch keine linearen Faktoren besitzt. Wegen der

Division mit Rest lässt sich g_5 in der Form

$$g_5(z) = q_5(z)(z^2 + sz + t) + r_5(z)$$

schreiben. Dabei sind q_5 , $(z^2 + sz + t)$ und r_5 Polynome in $\mathbb{Z}/5\mathbb{Z}$ und der Grad von r_5 ist kleiner oder gleich 1. Führt man diese Division mit Rest für alle $s, t \in \mathbb{Z}/5\mathbb{Z}$, $t \neq 0$, durch, ergeben sich in Abbildung 8.3 aufgelisteten Polynome r_5 .

$s \setminus t$	1	2	3	4
0	$3 + 3z$	$1 + 2z$	$3z$	z
1	$3 + 2z$	$2 + 4z$	$3 + 3z$	$1 + 4z$
2	$3 + 4z$	$3 + 3z$	$1 + 4z$	$2 + 2z$
3	$4 + 2z$	$1 + 2z$	$2 + 4z$	$2 + 3z$
4	$2 + 3z$	$3 + 3z$	4	$4z$

Abbildung 8.3: Restpolynome von g_5

Da dabei keine Nullpolynome vorkommen, kann g_5 keine quadratischen Faktoren enthalten. Somit ist gezeigt, dass g_5 in $\mathbb{Z}/5\mathbb{Z}$ irreduzibel ist. Mit dem Reduktionskriterium ist dann ebenfalls g in \mathbb{Q} irreduzibel.

Als nächstes wird die Anzahl der reellen Nullstellen von g betrachtet. Dazu wird die zweite Ableitung bestimmt. Man erhält

$$g''(z) = 59927220 z^3 + 5636916 z^2 + 15887904 z - 6830768.$$

Mit dem Softwarepaket Mathematica 4.1 wird eine Nullstelle z^* von g'' in der Form

$$z^* = -\frac{-2179459053279 + 313162 \gamma^{\frac{1}{3}} + \gamma^{\frac{2}{3}}}{9987870 \gamma^{\frac{1}{3}}} + i \frac{(2179459053279 + \gamma^{\frac{2}{3}})}{3329290 \sqrt{3} \gamma^{\frac{1}{3}}}$$

mit

$$\gamma = 7611944195664680569 + 6658580 \sqrt{1540354492802913263625061}$$

berechnet. Offensichtlich ist $z^* \in \mathbb{C} \setminus \mathbb{R}$. Dadurch ist das konjugiert komplexe von z^* , welches von z^* verschieden ist, auch eine nicht reelle Nullstelle von g'' und g'' besitzt genau zwei nicht reelle Nullstellen. Daraus folgt, dass g mindestens zwei nicht reelle, also höchstens drei reelle Nullstellen besitzt. Wegen $g(0) = -11664$, $g(\frac{1}{10}) = \frac{362068191}{100000}$, $g(\frac{1}{2}) = -\frac{5694361}{32}$ und $g(1) = 3154176$ besitzt nach dem Nullstellensatz g mindestens drei reelle Nullstellen. Insgesamt erhält man, dass g genau drei reelle und genau zwei nicht reelle Nullstellen besitzt. Mit Satz 8.9 ist die Galois-Gruppe von g die Gruppe \mathcal{S}_5 . Wegen Satz 8.4 und Satz 8.10 kann g nicht durch Radikale aufgelöst werden.

Abschließend kann bemerkt werden, dass für dieses Beispiel der MILE nicht in Abhängigkeit der Einträge der Matrizen $X'X$ und B angegeben werden kann. Man ist also auf eine numerische Lösung wie am Ende von Kapitel 7 beschrieben angewiesen.

Kapitel 9

Fixpunktiteration

In Kapitel 5 wurde durch die Darstellung von Hoffmann und Lauter eine Moglichkeit aufgezeigt den MILE zu gewinnen. Dies geschieht durch das Losen einer Matrixgleichung mit Nebenbedingung. Diese Matrixgleichung kann in die Form $F(Y) = Y$ gebracht werden. Durch eine Simulation von vielen Beispielen werden fast immer die gewunschten Losungen erhalten. Aber es ist bisher nur fur den Fall $j = k$ eine Konvergenz in einer Umgebung des gesuchten Fixpunktes gesichert. Um diese lokale Konvergenz zu uberprufen benotigt man die Ableitung einer matrixwertigen Funktion.

9.1 Differentiation von Matrix-Funktionen

Es werden nun Funktionen vom Typ $F : \mathbb{R}^{n \times q} \rightarrow \mathbb{R}^{m \times p}$ betrachtet.

Definition 9.1 Sei $D \subset \mathbb{R}^{n \times q}$ offen und fur $r > 0$, $Y_0 \in D$ sei $B(Y_0, r) = \{Y \in D \mid \|Y - Y_0\| < r\}$ mit $\|Y\| = (\text{tr } Y'Y)^{\frac{1}{2}}$. Weiter sei $U \in \mathbb{R}^{n \times q}$ mit $\|U\| < r$ derart, dass $Y_0 + U \in B(Y_0, r)$ ist. Die Funktion $F : D \rightarrow \mathbb{R}^{m \times p}$

heißt differenzierbar in Y_0 , falls eine Matrix $A(Y_0) \in \mathbb{R}^{mp \times nq}$, die nur von Y_0 abhängt, existiert, so dass

$$\text{vec } F(Y_0 + U) = \text{vec } F(Y_0) + A(Y_0) \text{vec } U + \text{vec } R_{Y_0}(U)$$

für alle U mit $\|U\| < r$ und

$$\lim_{U \rightarrow \infty} \frac{R_{Y_0}(U)}{\|U\|} = 0$$

gilt. Die Matrix $A(Y_0)$ heißt die erste Ableitung von F in Y_0 und $\text{vec } dF = \text{vec } dF(Y_0, U) = A(Y_0) \text{vec } U$ heißt das erste Differential.

Ist F für jedes $Y_0 \in D$ differenzierbar, so heißt F differenzierbar auf D .

In der Definition 9.1 ist vec der Vec-Operator (siehe Definition A.9). Die Differenzierbarkeit der Funktion F kann auf die Differenzierbarkeit von vektorwertigen Funktionen zurückgeführt werden. Mit Hilfe des Vec-Operators gibt es einen Isomorphismus zwischen $\mathbb{R}^{n \times q}$ und \mathbb{R}^{nq} bzw. $\mathbb{R}^{m \times p}$ und \mathbb{R}^{mp} . Dadurch kann dann an Stelle von F die Funktion $f : \text{vec } D \rightarrow \mathbb{R}^{mp}$, wobei $\text{vec } D = \{y \in \mathbb{R}^{nq} \mid y = \text{vec } Y, Y \in D\}$ ist, mit $f(\text{vec } Y) = \text{vec } F(Y)$ betrachtet werden (vgl. [22], S. 95). Die Jacobi-Matrix von $\mathbf{DF}(Y)$ ist durch

$$\mathbf{DF}(Y) = \mathbf{Df}(\text{vec } Y)$$

gegeben. Aus dem Identifikationssatz für Matrix-Funktionen erhält man, dass $\text{vec } dF = A(Y) \text{vec } U$ genau dann für alle U gilt, wenn

$$\mathbf{DF}(Y) = A(Y)$$

ist (siehe [22] S. 96). Die Rechenregeln für vektorwertige Funktionen können somit auf matrixwertige Funktionen übertragen werden.

Lemma 9.2 (Kettenregel) Sei $D \subset \mathbb{R}^{n \times q}$ offen, $F : D \rightarrow \mathbb{R}^{m \times p}$ differenzierbar und sei $G : F(\mathbb{R}^{m \times p}) \rightarrow \mathbb{R}^{r \times s}$ im Inneren von $F(\mathbb{R}^{m \times p})$ differenzierbar.

Dann ist $H : D \rightarrow \mathbb{R}^{r \times s}$ mit

$$H(Y) = G(F(Y))$$

differenzierbar und es gilt

$$\mathbf{D}H(Y) = \mathbf{D}G(F(Y)) \mathbf{D}F(Y).$$

BEWEIS:

siehe ([22], S. 96) □

Lemma 9.3 Seien für $D \subset \mathbb{R}^{n \times q}$ offen die Funktionen $F : D \rightarrow \mathbb{R}^{m \times p}$ und $G : D \rightarrow \mathbb{R}^{r \times s}$ differenzierbar.

(a) Für $m = r$, $p = s$ und $H(Y) = F(Y) + G(Y)$ gilt:

$$\mathbf{D}H(Y) = \mathbf{D}F(Y) + \mathbf{D}G(Y).$$

(b) Für $p = r$ und $H(Y) = F(Y)G(Y)$ gilt:

$$\mathbf{D}H(Y) = (G'(Y) \otimes I_m) \mathbf{D}F(Y) + (I_s \otimes F(Y)) \mathbf{D}G(Y).$$

BEWEIS:

(a) Folgt aus (13) in ([22], S. 148) und dem Identifikationssatz für Matrix-Funktionen.

(b) Nach (15) in ([22], S. 148) gilt

$$dH(Y) = [dF(Y)]G(Y) + F(Y)[dG(Y)]$$

und mit (19) in ([22], S. 148) sowie Lemma A.17

$$\text{vec } dH(Y) = (G'(Y) \otimes I_m) \text{vec } dF(Y) + (I_m \otimes F(Y)) \text{vec } dG(Y).$$

Mit dem Identifikationssatz erhält man schließlich die Behauptung. □

In der Abbildung 9.1 sind die für den weiteren Verlauf benötigten Jacobi-Matrizen angegeben. Dabei ist stets $F : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$, $A, B \in \mathbb{R}^{n \times n}$ und

$N_n = \frac{1}{2}(I_n^2 + K_{nn})$. Dabei ist K_{nn} die Vertauschungsmatrix aus Definition A.10 (vgl. [22], S. 175, S. 182ff).

F(Y)	DF(Y)
A	0
AYB	$B' \otimes A$
YY'	$2 N_n(Y \otimes I_n)$
Y^{-1}	$-(Y')^{-1} \otimes Y^{-1}$

Abbildung 9.1: Ableitungen matrixwertiger Funktionen

Lemma 9.4 Für positiv definite Matrizen Y sei $F(Y) = Y^{\frac{1}{2}}$. Dann gilt

$$\mathbf{DF}(Y) = \left[(Y^{\frac{1}{2}} \otimes I_n) + (I_n \otimes Y^{\frac{1}{2}}) \right]^{-1}.$$

BEWEIS:

Es gilt

$$\begin{aligned} \mathbf{D}(F(Y) F(Y)) &= I_{n^2} \\ \stackrel{\text{Lemma 9.3}}{\Leftrightarrow} & (Y^{\frac{1}{2}} \otimes I_n) \mathbf{DF}(Y) + (I_n \otimes Y^{\frac{1}{2}}) \mathbf{DF}(Y) = I_n^2 \\ \stackrel{\text{Lemma A.15}}{\Leftrightarrow} & \mathbf{DF}(Y) = \left[(Y^{\frac{1}{2}} \otimes I_n) + (I_n \otimes Y^{\frac{1}{2}}) \right]^{-1}. \end{aligned}$$

□

9.2 Fixpunktgleichung

Man befinde sich in der Situation des Schätzproblems aus Definition 2.1 mit $\text{rg } X = \text{rg } B = k$ und definiere $S = X'X$ und $F = S^{-1}B'BS^{-1}$. Kennt man gemäß Satz 5.2 eine nicht negativ definite Matrix $A \in \mathbb{R}^{k \times k}$ und eine Zahl $a > 0$, die die Bedingungen

$$(1) M_0 = \sqrt{a}(F + A)^{\frac{1}{2}} - S^{-1} \text{ n.n.d.},$$

$$(2) \quad \frac{1}{\sqrt{a}}(F + A)^{-\frac{1}{2}}A = SA,$$

$$(3) \quad \sqrt{a} \operatorname{tr}(F + A)^{\frac{1}{2}} = 1 + \operatorname{tr} S^{-1}$$

erfüllen, so kann mittels A der MILE ausgedrückt werden. Durch Auflösen von (3) nach \sqrt{a} und Einsetzen in (2) erhält man die Matrixgleichung

$$\frac{\operatorname{tr}(F + A)^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} (F + A)^{-\frac{1}{2}}A = SA,$$

welche äquivalent zu

$$\frac{\operatorname{tr}(F + A)^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1}(F + A)^{-\frac{1}{2}}A = A \quad (9.1)$$

ist. Dadurch dass A nicht negativ definit ist, gibt es ein $Y \in \mathbb{R}^{k \times k}$ mit $A = YY'$. Damit wird (9.1) äquivalent zu

$$\frac{\operatorname{tr}(F + YY')^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1}(F + YY')^{-\frac{1}{2}}Y = Y.$$

Durch die Definition der Funktion $F : \mathbb{R}^{k \times k} \rightarrow \mathbb{R}^{k \times k}$ mit

$$F(Y) = \frac{\operatorname{tr}(F + YY')^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1}(F + YY')^{-\frac{1}{2}}Y \quad (9.2)$$

kann die Berechnung des MILE durch einen Fixpunkt von F erfolgen, aber nur dann, wenn dieser Fixpunkt Y die Bedingung

$$\frac{1 + \operatorname{tr} S^{-1}}{\operatorname{tr}(F + YY')^{\frac{1}{2}}} (F + YY')^{\frac{1}{2}} - S^{-1} \quad \text{n.n.d} \quad (9.3)$$

erfüllt.

Zur Bestimmung des Fixpunktes ist es möglich ein iteratives Verfahren in Betracht zu ziehen. Für eine Startmatrix $Y_{(0)}$ wende man die Iterationsvorschrift

$$Y_{(n+1)} = F(Y_{(n)}) \quad n = 0, 1, \dots \quad (9.4)$$

an. Ob dieses Vorgehen von Erfolg gekrönt ist, hängt von der Konvergenz der Folge $Y_{(n)}$ ab. Nach ([23], Theorem 10.1.3) liegt lokale Konvergenz vor,

d.h. die Folge $Y_{(n)}$ konvergiert in einer Umgebung $U(Y^*)$ des Fixpunktes Y^* gegen den Fixpunkt Y^* , wenn

$$\varrho(\mathbf{DF}(Y^*)) < 1$$

gilt. Dabei ist $\varrho(M)$ der Spektralradius einer Matrix M . Da die Funktion F aus (9.2) stetig differenzierbar ist, wird die Ableitung von F berechnet.

Satz 9.5 *Sei F wie in (9.2). Dann ist*

$$\begin{aligned} \mathbf{DF}(Y) = & 2 \left(Y' \otimes \frac{1}{1 + \operatorname{tr} S^{-1}} S^{-1} \right) \left[\operatorname{vec} (F + YY')^{-\frac{1}{2}} (\operatorname{vec} I_k)' \right. \\ & \left. - \operatorname{tr} (F + YY')^{\frac{1}{2}} (F + YY')^{-\frac{1}{2}} \otimes (F + YY')^{-\frac{1}{2}} \right] \\ & \cdot \left\{ [(F + YY')^{\frac{1}{2}} \otimes I_k] + [I_k \otimes (F + YY')^{\frac{1}{2}}] \right\}^{-1} N_k(Y \otimes I_k) \\ & + \left[I_k \otimes \frac{\operatorname{tr} (F + YY')^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} (F + YY')^{-\frac{1}{2}} \right]. \end{aligned}$$

BEWEIS:

Definiere $G(Y) = (F + YY')^{\frac{1}{2}}$. Dann ist

$$F(Y) = \frac{\operatorname{tr} G(Y)}{1 + \operatorname{tr} S^{-1}} S^{-1} G^{-1}(Y) Y.$$

Mit (15), (19), (20) aus ([22], S. 148) gilt

$$d[\operatorname{tr} G(Y)] G^{-1}(Y) = [\operatorname{tr} dG(Y)] G^{-1}(Y) + \operatorname{tr} G(Y) dG^{-1}(Y)$$

und aus Lemma A.16 sowie Lemma A.17 folgt

$$\begin{aligned} & \operatorname{vec} d[\operatorname{tr} G(Y)] G^{-1}(Y) \\ & = \operatorname{vec} G^{-1}(Y) (\operatorname{vec} I_k)' \operatorname{vec} dG(Y) + \operatorname{tr} G(Y) \operatorname{vec} dG^{-1}(Y). \end{aligned}$$

Durch den Identifikationssatz erhält man

$$\begin{aligned} & \mathbf{D}[\operatorname{tr} G(Y)] G^{-1}(Y) \\ & = \operatorname{vec} G^{-1}(Y) (\operatorname{vec} I_k)' \mathbf{D}G(Y) + \operatorname{tr} G(Y) \mathbf{D}G^{-1}(Y). \end{aligned}$$

Weiter gilt mit Lemma 9.2

$$\mathbf{D}G^{-1}(Y) = [-G^{-1}(Y) \otimes G^{-1}(Y)] \mathbf{D}G(Y)$$

und mit Lemma 9.4

$$\begin{aligned} \mathbf{D}G(Y) &= \{[G(Y) \otimes I_k] + [I_k \otimes G(Y)]\}^{-1} \mathbf{D}(F + YY') \\ &= 2 \{[G(Y) \otimes I_k] + [I_k \otimes G(Y)]\}^{-1} N_k(Y \otimes I_k). \end{aligned}$$

Es gilt nun mit Lemma 9.3

$$\begin{aligned} \mathbf{D}F(Y) &= \left(Y' \otimes \frac{1}{1 + \operatorname{tr} S^{-1}} S^{-1} \right) \mathbf{D}[\operatorname{tr} G(Y)] G^{-1}(Y) \\ &\quad + \left[I_k \otimes \frac{\operatorname{tr} G(Y)}{1 + \operatorname{tr} S^{-1}} S^{-1} G^{-1} \right] \\ &= 2 \left(Y' \otimes \frac{1}{1 + \operatorname{tr} S^{-1}} S^{-1} \right) [\operatorname{vec} G^{-1}(Y) (\operatorname{vec} I_k)' \\ &\quad - \operatorname{tr} G(Y) G^{-1}(Y) \otimes G^{-1}(Y)] \\ &\quad \cdot \{[G(Y) \otimes I_k] + [I_k \otimes G(Y)]\}^{-1} N_k(Y \otimes I_k) \\ &\quad + \left[I_k \otimes \frac{\operatorname{tr} G(Y)}{1 + \operatorname{tr} S^{-1}} S^{-1} G^{-1}(Y) \right] \end{aligned}$$

und schließlich folgt die Behauptung. \square

Beim Vorliegen des Falls $j = k$ ist der gesuchte Fixpunkt $Y^* = 0$ (vgl. (5.4)). Für diese Situation kann durch den folgenden Satz die Konvergenz einer Folge $Y_{(n)}$ gegen den Fixpunkt, die gemäß (9.4) erhalten wird, in einer Umgebung von $Y^* = 0$ gezeigt werden.

Satz 9.6 *Im Schätzproblem aus Definition 2.1 mit $\operatorname{rg} B = \operatorname{rg} X = k$ liege der Fall $j = k$ vor. Dann gilt für die Funktion F aus (9.2)*

$$\varrho(\mathbf{D}F(0)) < 1.$$

BEWEIS:

Da der Fall $j = k$ vorliegt, gilt für M_0 aus Satz 5.2:

$$M_0 = \frac{1 + \operatorname{tr} S^{-1}}{\operatorname{tr} F^{\frac{1}{2}}} F^{\frac{1}{2}} - S^{-1} \quad \text{n.n.d.}$$

und $\operatorname{rg} M_0 = k$. Daraus folgt

$$\begin{aligned} & \frac{1 + \operatorname{tr} S^{-1}}{\operatorname{tr} F^{\frac{1}{2}}} F^{\frac{1}{2}} - S^{-1} \quad \text{p.d.} \\ \Leftrightarrow & S - \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} F^{-\frac{1}{2}} \quad \text{p.d.} \\ \Leftrightarrow & S \left[I_k - \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} \right] \quad \text{p.d.} \\ \Leftrightarrow & S^{\frac{1}{2}} \left[I_k - \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} \right] S^{-\frac{1}{2}} \quad \text{p.d.} \\ \Leftrightarrow & \left[I_k - S^{\frac{1}{2}} \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} S^{-\frac{1}{2}} \right] \quad \text{p.d.} \\ \Leftrightarrow & \lambda_{\max} \left(S^{\frac{1}{2}} \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} S^{-\frac{1}{2}} \right) < 1. \end{aligned}$$

Da $S^{\frac{1}{2}} \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} S^{-\frac{1}{2}}$ positiv definit ist gilt

$$\lambda_{\min} \left(S^{\frac{1}{2}} \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} S^{-\frac{1}{2}} \right) > 0.$$

Durch die Ähnlichkeit von $\frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}}$ zu $S^{\frac{1}{2}} \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} S^{-\frac{1}{2}}$ erhält man

$$\lambda_{\max} \left(\frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} \right) < 1. \quad (9.5)$$

und

$$\lambda_{\min} \left(\frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}} \right) > 0. \quad (9.6)$$

Weiter ist mit

$$\mathbf{DF}(0) = I_k \otimes \frac{\operatorname{tr} F^{\frac{1}{2}}}{1 + \operatorname{tr} S^{-1}} S^{-1} F^{-\frac{1}{2}}$$

durch Lemma A.15 und (9.5) sowie (9.6)

$$\varrho(\mathbf{DF}(0)) < 1.$$

□

Trotz der lokalen Konvergenz bleibt die Frage nach der Wahl der Startmatrix $Y_{(0)}$ offen. Außerdem ist das Verhalten der Funktion für die Fälle $j < k$ nicht geklärt. Im nächsten Abschnitt wird durch Simulation vieler Beispiele das Konvergenzverhalten der Funktion F betrachtet.

9.3 Simulation

In zwei Simulationen wird nun untersucht, wie sich die Funktion F aus (9.2) bei Betrachtung vieler Beispiele verhält. Diese Beispiele sind zufällig erzeugt, das heißt, die Anzahl der Einflussgrößen, die Einträge der Matrizen S und B sind gleichverteilte Zufallszahlen. Bei der Anzahl der Einflussgrößen fand die diskrete Gleichverteilung auf der Menge $\{2, 3, 4, 5, 6\}$ Verwendung. Wegen der Anmerkung aus Kapitel 2.1 ist die Gewichtsmatrix B diagonal und die Einträge der Diagonalen sind in der ersten Simulation auf dem Intervall $[0.2, 5]$ gleichverteilte Zufallszahlen. Die Matrix S wird durch $S = T'T$ (T quadratisch) dargestellt, wobei die Einträge von T gleichverteilte Zufallszahlen aus dem Intervall $[-5, 5]$ sind. Erfüllt eine so entstandene Matrix S die Forderung $\lambda_{\min}(S) > 0.1$, findet sie in der Simulation Verwendung, ansonsten nicht. Damit ist gesichert, dass S und B positiv definit sind. Auf diese Art werden 100000 Beispiele für ein Schätzproblem aus Definition 2.1 mit

$\text{rg } B = \text{rg } X = k$ erzeugt. Mit der Startmatrix $Y_{(0)} = I_k$ wird die Iterationsvorschrift (9.4) durchgeführt. Da die Konvergenz der aus der Iteration erhaltenen Folge $Y_{(n)}$ nicht gesichert ist, wird das Iterationsverfahren nach spätestens 100000 Iterationen abgebrochen. Falls vorher ein Y existiert, für das $\varrho(F(Y) - Y) < 10^{-10}$ gilt, so wird Y als Fixpunkt angesehen und die Ableitung $\mathbf{DF}(Y)$ für dieses Y bestimmt. Durch Einsetzen in (9.3) überprüft man, ob

$$M_0 = \frac{1 + \text{tr } S^{-1}}{\text{tr}(F + YY')^{\frac{1}{2}}} (F + YY')^{\frac{1}{2}} - S^{-1} \quad (9.7)$$

nicht negativ definit ist. Die Matrix M_0 wird als nicht negativ definit angesehen, wenn $\lambda_{\min}(M_0) > -1 \cdot 10^{-6}$ ist. Falls $\lambda_{\min}(M_0) > 10^{-6}$ wird M_0 als positiv definit angesehen und man kann dann auf das Vorliegen des Falls $j = k$ schließen. Diese Simulation wurde mit dem Computeralgebrasystem Mathematica 4.1 durchgeführt (siehe Anhang D).

Bei nur 11 Beispielen wurde die maximale Fixpunktiterationsanzahl von 100000 erreicht und es konnte kein Fixpunkt gefunden werden. Bei 99% der den 99989 erfolgreichen Beispielen wurden nicht mehr als 2035 Iterationen benötigt, bei 95% der erfolgreichen Beispiele sogar nicht mehr als 511 Iterationen. In der Abbildung 9.2 sind weitere Quantile der Verteilung der Iterationsanzahl in den 99989 erfolgreichen Beispielen dargestellt.

Alle so bestimmten Fixpunkte halten die Bedingung, dass M_0 aus (9.7) nicht negativ definit ist, und sind somit zur Berechnung des MILE geeignet.

Bei der Untersuchung der Ableitung ist einmal von Interesse, ob der Spektralradius der Ableitung im Fixpunkt kleiner eins ist. Er wird als kleiner eins angenommen, wenn $\varrho(\mathbf{DF}(Y)) - 1 < -1 \cdot 10^{-6}$ ist und als eins angenommen wenn $|\varrho(\mathbf{DF}(Y)) - 1| < 10^{-6}$ gilt. Besondere Aufmerksamkeit liegt dabei auf den Fällen $j < k$, weil darüber bisher in der Theorie noch keine

Quantil	Anzahl
0	8
0.25	30
0.5	64
0.75	141
0.95	511
0.99	2035
1	94135

Abbildung 9.2: Verteilung der Iterationsanzahl in Simulation 1

Aussagen gemacht wurden. In der Kontingenztafel aus Abbildung 9.3 ist die Anzahl der erfolgreichen Beispiele für den Fall $j = k$ bzw. $j < k$ in Bezug zu $\varrho(\mathbf{DF}(Y)) < 1$ bzw. $\varrho(\mathbf{DF}(Y)) = 1$ angegeben.

	$\varrho < 1$	$\varrho = 1$
$j < k$	0	49085
$j = k$	50904	0

Abbildung 9.3: Fall und Ableitung im Fixpunkt bei Simulation 1

Für den Fall $j = k$ verhält sich der Spektralradius der Ableitung im Fixpunkt wie erwartet. Für die Fälle $j < k$ erhält man in diesen Beispielen jedoch Spektralradien der Ableitung im Fixpunkt, die nahe eins sind. Dies könnte ein Indiz für die Schwierigkeiten des Nachweises der lokalen Konvergenz für die Fälle $j < k$ sein.

In einer zweiten Simulation konnten ähnliche Ergebnisse erzielt werden. Diesmal handelte es sich bei den Einträgen der Diagonalen der Gewichtsmatrix B um auf dem Intervall $[0.2, 1]$ gleichverteilte Zufallszahlen. Die Einträge in T mit $S = T'T$ sind gleichverteilte Zufallszahlen auf dem Intervall $[-1, 1]$.

Alle anderen Schranken und Verfahren blieben bestehen. Jedoch wurde bei 45 Beispielen die Fixpunktiteration ergebnislos abgebrochen. Hierbei werden bei 99% der 99955 erfolgreichen Beispiele nicht mehr als 7089 bzw. bei 95% nicht mehr als 1760 Iterationen benötigt. In der Abbildung 9.4 sind weitere Quantile der Verteilung der Iterationsanzahl in den 99955 erfolgreichen Beispielen dargestellt.

Quantil	Anzahl
0	32
0.25	142
0.5	223
0.75	430
0.95	1760
0.99	7089
1	97923

Abbildung 9.4: Verteilung der Iterationsanzahl in Simulation 2

Auch in dieser Simulation ist für jeden erhaltenen Fixpunkt die Matrix M_0 aus (9.7) nicht negativ definit. Wie Abbildung 9.5 zeigt, wird für den Spektralradius der Ableitung im Fixpunkt der Zusammenhang aus der ersten Simulation erneut vorgefunden.

	$\varrho < 1$	$\varrho = 1$
$j < k$	0	84375
$j = k$	15580	0

Abbildung 9.5: Fall und Ableitung im Fixpunkt bei Simulation 2

Diese beiden Simulation geben Anhaltspunkte, wie sich die Funktion F aus (9.2) bezüglich der Konvergenz zu dem gesuchten Fixpunkt verhalten

könnte und wo es Ansatzpunkte gibt diese Konvergenz nachzuweisen. Aber eine Verallgemeinerung, dass dieses Verfahren mit hoher Wahrscheinlichkeit konvergiert, ist ohne weiteres nicht möglich.

Wendet man dieses Iterationverfahren auf das Beispiel aus Kapitel 8.2 an, also mit den Matrizen

$$X'X = \begin{pmatrix} 3 & \frac{9}{11} & 0 \\ \frac{9}{11} & \frac{15}{44} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

und

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

so erhält man nach 594 Iterationen einen Fixpunkt. Dieser erhaltene Fixpunkt erfüllt die Nebenbedingung (9.3). Es liegt der Fall $j = 2$ vor und der Spektralradius der Ableitung in diesem Fixpunkt ist nahe eins (siehe Anhang D.3).

Kapitel 10

Zusammenfassung und Ausblick

Ausgehend von der Formulierung des Schätzproblems im linearen Modell mit ellipsoider Vorinformation mit Mitteln der statistischen Entscheidungstheorie kann die Minimax-Schätzung als sinnvolles und optimales Schätzverfahren motiviert und definiert werden. Zur Bestimmung des MILE werden die bisher verwendeten Zugänge Bayes'scher Ansatz, spektraler Ansatz und Darstellung von Hoffmann und Läuter kurz erläutert und in Beziehung gesetzt. Beim spektralen Ansatz liegt besondere Aufmerksamkeit auf der Vielfachheit des maximalen Eigenwerts der Matrix $B(\hat{C}_M X - I_k) \left[B(\hat{C}_M X - I_k) \right]'$. Dabei sind B und X wie in Definition 2.1. Mit $\hat{\beta}_M = \hat{C}_M y$ wird der Minimax-Schätzer bezeichnet. Für die Entscheidung, dass der maximale Eigenwert einfach ist, konnten die Bedingungen verschärft werden, so dass eine Reduzierung der Überprüfung von Ungleichungen stattfindet. Dieses nutzt man dann auch zur Strukturierung einer Verallgemeinerung eines Beispiels aus ([8], S. 9ff). Dort geht es um ein lineares Modell mit ellipsoider Vorinformation, drei Einflussgrößen und die Matrizen $X'X$ und B besitzen hier einen gemeinsamen Eigenvektor. Durch Strukturierung können in Abhängigkeit der Einträge der

Matrizen $X'X$ und B die bereits gelösten und ungelösten Fälle unterschieden werden. Eine solche Situation wird näher untersucht. Dazu wird noch die Einschränkung gemacht, dass aus der Matrix $X'X$ der Eintrag $s_{12} > 0$ ist. Als Ergebnis erhält man, dass die Berechnung des Minimax-Schätzers auf eine auf dem Intervall $I \subset (0, \frac{1}{\sqrt{2}})$ eindeutige Nullstelle einer reellen Funktion geführt werden kann. Dies führt dazu, dass man mit der Intervallschachtelung ein konvergentes numerisches Verfahren für dieses Problem hat. Außerdem kann man die maximale Anzahl an Iterationen für eine bestimmte Genauigkeit angeben. Für $s_{12} < 0$ besteht die Hoffnung, dass diese Vorgehensweise auch funktioniert, wenngleich das Intervall, das die Nullstelle enthält, von $(-\frac{1}{\sqrt{2}}, 0)$ überdeckt wird. Die ungelöste Situation aus Struktur II ist noch nicht untersucht. Es ist aber zu befürchten, dass der Rechenaufwand ähnlich immens wird. Abschließend kann für das Beispiel aus ([8], S. 9ff) mit $\mu = 1$ gezeigt werden, dass die Nullstelle nicht in Abhängigkeit der Einträge der Matrizen $X'X$ und B dargestellt werden kann.

Die Matrixgleichung mit Nebenbedingung aus der Darstellung von Hoffmann und Läuter kann zu einer Fixpunktgleichung umgeformt werden. Dabei ist die Nullmatrix immer ein Fixpunkt, aber sie erfüllt nicht immer die Nebenbedingung. Für den Fall, dass die Nullmatrix die Nebenbedingung erfüllt, liegt lokale Konvergenz der Fixpunktiteration vor. In einer Simulation von vielen Beispielen mit zwei bis sechs Einflussgrößen können bei der Wahl der Einheitsmatrix als Startmatrix sehr häufig durch die Fixpunktiteration diejenigen Fixpunkte erhalten werden, die der Nebenbedingung genügen. In den anderen Fällen fand ein Abbruch wegen zu hoher Iterationsanzahlen statt. Es ergibt sich daraus die Frage nach der Konvergenz des Iterationsverfahrens im Allgemeinen. Dabei richtet sich ein besonderer Blick auf die Fixpunkte, die die Nebenbedingung erfüllen. Desweiteren ist ein Vergleich zu den bisher

bestehenden numerischen Verfahren von Interesse.

Anhang A

Matrizenrechnung

Definition A.1 (Moore-Penrose-Inverse) Sei $X \in \mathbb{R}^{m \times n}$. Die Matrix $X^+ \in \mathbb{R}^{n \times m}$ heißt Moore-Penrose-Inverse von X , wenn

$$(1) \quad XX^+X = X,$$

$$(2) \quad X^+XX^+ = X^+,$$

$$(3) \quad XX^+ = (XX^+)',$$

$$(4) \quad X^+X = (X^+X)'$$

gilt.

Lemma A.2 Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch und $\lambda_{\max}(A)$ der maximale Eigenwert von A . Dann gilt

$$\max_{\{x \in \mathbb{R}^n \mid x'x \leq 1\}} x'Ax = \lambda_{\max}(A).$$

BEWEIS:

Da $A \in \mathbb{R}^{n \times n}$ symmetrisch ist, existiert die Spektralzerlegung von A durch

$A = P\Lambda P'$. Dabei sind P eine orthogonale Matrix und

$$\Lambda = \begin{pmatrix} \lambda_{\max}(A) & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

mit $\lambda_{\max}(A) \geq \lambda_2 \geq \cdots \geq \lambda_n$. Damit gilt, dass $\lambda_{\max}(A)I_n - A$ nicht negativ definit ist. Für alle x mit $x'x \leq 1$ folgt

$$\begin{aligned} x'Ax &= x'P\Lambda P'x \leq \lambda_{\max}(A)x'PP'x \\ &= \lambda_{\max}(A)x'x \leq \lambda_{\max}(A). \end{aligned}$$

Sei nun x_{\max} ein normierter Eigenvektor zum maximalen Eigenwert von A , also es gelte $Ax_{\max} = \lambda_{\max}(A)x_{\max}$, dann folgt

$$x'_{\max}Ax_{\max} = \lambda_{\max}(A)x'_{\max}x_{\max} = \lambda_{\max}(A)$$

und somit die Behauptung. □

Lemma A.3 Sei $B \in \mathbb{R}^{m \times k}$ mit $\text{rg } B = m$. Dann ist

$$(B'B)^+ = B'(BB')^{-2}B.$$

BEWEIS:

Es gilt

- (1) $B'BB'(BB')^{-2}BB'B = B'B$,
- (2) $B'(BB')^{-2}BB'BB'(BB')^{-2}B = B'(BB')^{-2}B$,
- (3) $B'BB'(BB')^{-2}B = B'(BB')^{-1}B$ ist symmetrisch,
- (4) $B'(BB')^{-2}BB'B = B'(BB')^{-1}B$ ist symmetrisch

und damit ist die Behauptung richtig. □

Lemma A.4 Sei $X \in \mathbb{R}^{n \times k}$ und $M \in \mathbb{R}^{k \times k}$ n.n.d.. Dann gilt:

$$(I_k + X'XM)^{-1} = I_k - X'(I_n + XMX')^{-1}XM.$$

BEWEIS:

Es ist

$$\begin{aligned} & (I_k + X'XM) \left[I_k - X'(I_n + XMX')^{-1}XM \right] \\ &= I_k + X'XM - X'(I_n + XMX')^{-1}XM \\ &\quad - X'XMX'(I_n + XMX')^{-1}XM \\ &= I_k + X' \left[I_n - (I_n + XMX')^{-1} - XMX'(I_n + XMX')^{-1} \right] XM \\ &= I_k + X' \left[I_n - (I_n + XMX')(I_n + XMX')^{-1} \right] XM = I_k. \end{aligned}$$

Wegen der Eindeutigkeit der Inversen folgt die Behauptung. \square

Lemma A.5 Sei $B \in \mathbb{R}^{k \times k}$ p.d. und $X \in \mathbb{R}^{n \times k}$ mit $\text{rg } X = k$. Weiter sei ϱ ein Eigenwert von $B(I_k + X'X)^{-1}B'$ mit zugehörigem normierten Eigenvektor u . Dann gilt:

(a) $\varrho < u'BB'u$,

(b) $u'(BB')^{-1}u < \frac{1}{\varrho}$.

BEWEIS:

zu zeigen: $BB' - B(I_k + X'X)^{-1}B'$ ist positiv definit.

$$\begin{aligned} BB' - B(I_k + X'X)^{-1}B' &= B \left[I - (I + X'X)^{-1} \right] B' \\ &\stackrel{\text{Lemma A.4}}{=} B \left[I - I + X'(I + X'X)^{-1}X \right] B' \\ &= BX'(I + X'X)^{-1}XB' \\ &= (XB')'(I + X'X)^{-1}XB \quad \text{p.d.} \end{aligned}$$

Daraus folgt sofort (a) und (b). \square

Lemma A.6 Seien $A, B \in \mathbb{R}^{k \times k}$ regulär und es gelte $AB = BA$. Dann ist

$$A^{-1} - B^{-1} = (AB)^{-1}(B - A).$$

BEWEIS:

Es gilt

$$\begin{aligned} (AB)^{-1}(B - A) &= A^{-1}B^{-1}(B - A) \\ &= A^{-1} - (AB)^{-1}A \\ &= A^{-1} - B^{-1}. \end{aligned}$$

□

Definition A.7 Sei $D \subset \mathbb{R}^{m \times n}$ offen. Eine Funktion $F : D \rightarrow \mathbb{R}$ heißt in $X \in D$ mit $X = (x_{ij})$ stetig partiell differenzierbar, falls für alle $i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$ die partiellen Ableitungen $\frac{\partial F}{\partial x_{ij}}$ existieren und stetig sind. Die Matrix der partiellen Ableitung wird dann in der Form

$$\frac{\partial F}{\partial X} = \begin{pmatrix} \frac{\partial F}{\partial x_{11}} & \cdots & \frac{\partial F}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial F}{\partial x_{m1}} & \cdots & \frac{\partial F}{\partial x_{mn}} \end{pmatrix}$$

dargestellt.

Lemma A.8 Seien A, B, C, X Matrizen derart, dass die Funktionen $F(X) = \text{tr } AX$ und $G(X) = \text{tr } XBX'C$ existieren. Dann gilt

(a)

$$\frac{\partial F}{\partial X} = A',$$

(b)

$$\frac{\partial G}{\partial X} = C'XB' + CXB.$$

BEWEIS:

siehe ([22], S.178) □

Definition A.9 (Vec-Operator) Sei $A \in \mathbb{R}^{m \times n}$ mit $A = (a_{ij})$. Dann heißt

$$\text{vec } A = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \\ a_{12} \\ \vdots \\ a_{mn} \end{pmatrix}$$

Vec-Operator.

Definition A.10 (Vertauschungsmatrix) Sei $A \in \mathbb{R}^{m \times n}$. Die Matrix $K_{mn} \in \mathbb{R}^{mn \times mn}$, für die

$$K_{mn} \text{vec } A = \text{vec } A'$$

gilt, heißt Vertauschungsmatrix.

Definition A.11 (Kronecker-Produkt) Seien $A \in \mathbb{R}^{m \times n}$ und $B \in \mathbb{R}^{p \times q}$ Matrizen mit $A = (a_{ij})$. Dann ist durch

$$A \otimes B = \begin{pmatrix} a_{11} B & \cdots & a_{1n} B \\ \vdots & & \vdots \\ a_{m1} B & \cdots & a_{mn} B \end{pmatrix}$$

das Kronecker-Produkt von A und B gegeben.

Lemma A.12 Sei $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times q}$. Dann ist

$$(A \otimes B)' = A' \otimes B'.$$

BEWEIS:

siehe ([22], S. 30) □

Lemma A.13 Seien A, B, C und D Matrizen, so dass AC und BD existieren. Dann gilt

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

BEWEIS:

siehe ([22], S. 28) □

Lemma A.14 Seien $A \in \mathbb{R}^{m \times m}$ und $B \in \mathbb{R}^{p \times p}$ regulär. Dann gilt:

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

BEWEIS:

siehe ([22], S. 28) □

Lemma A.15 Seien $A \in \mathbb{R}^{m \times m}$ mit Eigenwerten $\lambda_1, \dots, \lambda_m$ und $B \in \mathbb{R}^{p \times p}$ mit Eigenwerten μ_1, \dots, μ_p gegeben. Dann werden die $m \cdot p$ Eigenwerte von $(A \otimes B)$ aus den Produkten $\lambda_i \mu_j$ für $i = 1, \dots, m$ und $j = 1, \dots, p$ erhalten.

BEWEIS:

siehe ([22], S. 28, Theorem 1) □

Lemma A.16 Sei $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times q}$. Dann gilt

$$\text{tr } A'B = (\text{vec } A)' \text{vec } B.$$

BEWEIS:

siehe ([22], S. 30) □

Lemma A.17 Seien A, B und C Matrizen, derart dass ABC existiert. Dann folgt

$$\text{vec}(ABC) = (C' \otimes A) \text{vec } B.$$

BEWEIS:

siehe ([22], S. 30) □

Anhang B

Minimax-Theorem

Das hier beschriebene Minimax-Theorem von Sion findet im Beweis von Lemma 3.9 Anwendung. Der Beweis dieses Minimax-Theorems orientiert sich an Sion [28].

Definition B.1 *Seien X und Y konvexe Mengen. Eine Funktion $f : X \times Y \rightarrow \mathbb{R}$ heißt quasi-konkav-konvex, wenn $\{x \in X \mid f(x, y) \geq c\}$ für $y \in Y$, $c \in \mathbb{R}$, sowie $\{y \in Y \mid f(x, y) \leq d\}$ für $x \in X$, $d \in \mathbb{R}$ konvexe Mengen sind.*

Definition B.2 *Seien X und Y konvexe Mengen. Eine Funktion $f : X \times Y \rightarrow \mathbb{R}$ heißt von unten und von oben halbstetig, falls $\limsup_{x \rightarrow x_0} f(x, y) \leq f(x_0, y)$ für alle $y \in Y$ und $\liminf_{y \rightarrow y_0} f(x, y) \geq f(x, y_0)$ für alle $x \in X$ gilt.*

Definition B.3 *Sei X eine Menge. Mit $\text{conv}(X)$ werde ihre konvexe Hülle bezeichnet.*

Satz B.4 *Seien X und Y konvexe und kompakte Mengen und $f : X \times Y \rightarrow \mathbb{R}$ eine quasi-konkav-konvexe und von unten und oben halbstetige Funktion.*

Dann gilt:

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) = \inf_{y \in Y} \sup_{x \in X} f(x, y).$$

BEWEIS:

Es werde angenommen, dass ein $c \in \mathbb{R}$ existiert mit

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) < c < \inf_{y \in Y} \sup_{x \in X} f(x, y).$$

Definiere $A_x = \{y \in Y \mid f(x, y) > c\}$ und $B_y = \{x \in X \mid f(x, y) < c\}$. Dann gilt $Y \subset \bigcup_{x \in X} A_x$ und $X \subset \bigcup_{y \in Y} B_y$. Da X und Y kompakt und A_x und B_y offen sind, genügen endlich viele A_x um Y und endlich viele B_y um X zu überdecken. Deswegen kann man eine endliche Menge $X_1 \subset X$ derart auswählen, dass für jedes $y \in Y$ und somit auch für jedes $y \in \text{conv}(Y)$ ein $x \in X_1$ mit $f(x, y) > c$ existiert. Ähnlich wählt man eine endliche Menge $Y_1 \subset Y$ aus, so dass für alle $x \in X$ und somit auch für jedes $x \in \text{conv}(X_1)$ ein $y \in Y_1$ mit $f(x, y) < c$ existiert. Nun wird X_1 derart verkleinert, dass eine Menge $X_2 \subset X_1$ gefunden wird, so dass für alle $y \in \text{conv}(Y_1)$ ein $x \in X_2$ mit $f(x, y) > c$ existiert. Danach wird Y_1 durch $Y_2 \subset Y_1$ verkleinert, so dass für alle $x \in \text{conv}(X_2)$ ein $y \in Y_2$ mit $f(x, y) < c$ existiert. Diese Reduzierung der Mengen wird wechselseitig weiter durchgeführt. Nach einer endlichen Anzahl von Schritten erhält man Mengen $\tilde{X} \subset X$ und $\tilde{Y} \subset Y$. Dabei gilt für alle $y \in \text{conv}(\tilde{Y})$, dass es ein $x \in \tilde{X}$ mit $f(x, y) > c$ gibt und für alle $x \in \text{conv}(\tilde{X})$ gibt es ein $y \in \tilde{Y}$ mit $f(x, y) < c$. Nach Lemma 3.3 aus [28] existiert ein $x_0 \in \text{conv}(\tilde{X})$ mit $f(x_0, y) < c$ für alle $y \in \tilde{Y}$ und nach Lemma 3.3' aus [28] existiert ein $y_0 \in \text{conv}(\tilde{Y})$ mit $f(x, y_0) > c$ für alle $x \in \tilde{X}$. Da f quasi-konkav-konvex ist, gibt es ein $x_0 \in \text{conv}(\tilde{X})$ mit $f(x_0, y) < c$ für alle $y \in \text{conv}(\tilde{Y})$ und es gibt ein $y_0 \in \text{conv}(\tilde{Y})$ mit $f(x, y_0) > c$ für alle $x \in \text{conv}(\tilde{X})$. Daraus folgt $c < f(x_0, y_0) < c$ und man erhält einen Widerspruch zur Annahme. Somit ist die Behauptung richtig (siehe [28], Theorem 3.4). \square

Lemma B.5 *Seien X und Y konvexe Mengen. Zusätzlich sei X kompakt und $f : X \times Y \rightarrow \mathbb{R}$ eine quasi-konkav-konvexe Funktion und von unten und oben halbstetig. Dann gilt:*

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) = \inf_{y \in Y} \sup_{x \in X} f(x, y).$$

BEWEIS:

Man nehme an, dass ein $c \in \mathbb{R}$ mit

$$\sup_{x \in X} \inf_{y \in Y} f(x, y) < c < \inf_{y \in Y} \sup_{x \in X} f(x, y).$$

existiert. Dann gibt es durch die Kompaktheit von X eine endliche Menge $\tilde{Y} \subset Y$, dass für jedes $x \in X$ ein $y \in \tilde{Y}$ mit $f(x, y) < c$ existiert. Definiere $\tilde{f} = f|_{X \times \text{conv}(\tilde{Y})}$. Wegen der Endlichkeit von \tilde{Y} ist \tilde{Y} kompakt. Daraus folgt mit Theorem III.1.4.3 aus [15] die Kompaktheit von $\text{conv}(\tilde{Y})$. Dadurch erhält man

$$\sup_{x \in X} \inf_{y \in \text{conv}(\tilde{Y})} \tilde{f}(x, y) < c < \inf_{y \in \text{conv}(\tilde{Y})} \sup_{x \in X} \tilde{f}(x, y).$$

Dies ist wegen Satz B.4 ein Widerspruch (siehe [28], Corollary 3.3). \square

Anhang C

Konvexe Analysis

Definition C.1 Sei $F : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ eine konvexe Funktion. Das Subdifferential ∂F von F an der Stelle X ist die Menge

$$\begin{aligned} \partial F(X) \\ = \{ \Phi \in \mathbb{R}^{m \times n} \mid \text{tr } \Phi'(Y - X) \leq F(Y) - F(X) \text{ für alle } Y \in \mathbb{R}^{m \times n} \}. \end{aligned}$$

Lemma C.2 Seien $G : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ und $H : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ konvexe Funktionen. Dann ist das Subdifferential von $F(X) = G(X) + H(X)$ gegeben durch

$$\partial F(X) = \partial G(X) + \partial H(X).$$

BEWEIS:

siehe ([15], Theorem VI.4.1.1) □

Lemma C.3 Sei $F : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ eine konvexe und stetig partiell differenzierbare Funktion an der Stelle X_0 . Dann ist

$$\partial F(X) = \left\{ \frac{\partial F}{\partial X}(X_0) \right\}.$$

BEWEIS:

siehe ([15], Corollary VI.2.1.4) □

Lemma C.4 Seien $(F_i : i \in I)$ eine Folge konvexer Funktionen mit $F_i : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ und I eine beliebige Indexmenge. Weiter sei $F(X) := \sup\{F_i(X) \mid i \in I\} < \infty$ und mit

$$I(X) = \{i \in I \mid F_i(X) = F(X)\}$$

werde die aktive Indexmenge bezeichnet. Dann ist

$$\partial F(X) = \text{conv} \left(\bigcup_{i \in I(X)} \partial F_i(X) \right).$$

Dabei bezeichne $\text{conv}(A)$ die konvexe Hülle einer Menge A .

BEWEIS:

siehe ([15], Theorem VI.4.4.2) □

Satz C.5 Seien $F : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ eine konvexe Funktion. An der Stelle X^* wird F genau dann minimiert, wenn $0 \in \partial F(X^*)$ ist.

BEWEIS:

Es gilt für alle $X \in \mathbb{R}^{m \times n}$:

$$\begin{aligned} F(X) &\geq F(X^*) \\ \Leftrightarrow F(X) - F(X^*) &\geq 0 \\ \Leftrightarrow F(X) - F(X^*) &\geq \text{tr } 0'(X - X^*) \\ \Leftrightarrow 0 &\in \partial F(X^*). \end{aligned}$$

vgl. ([15], Theorem VI.2.2.1) □

Anhang D

Mathematica Programme

D.1 1. Simulation

```
(* Überprüfung einer positiv definiten Matrix *)
TestPd[mat_] :=
  Module[{ewert}, ewert = Min[Re[Eigenvalues[mat]]];
  If[ewert > 10(-1), True, False]]

(* Erzeugung von Zufallsmatrizen *)
ZufallPd[n_] :=
  Module[{mat}, mat = DiagonalMatrix[Table[0, {n}]];
  While[TestPd[Transpose[mat].mat] == False,
    mat = Table[Random[Real, {-5, 5}], {i, 1, n}, {j, 1, n}]];
  Transpose[mat].mat]
ZufallDiag[n_] :=
  DiagonalMatrix[Table[Random[Real, {0.2, 5}], {i, 1, n}]]

(* Vec - Operator, Kroneckerprodukt und Dimension *)
Needs["LinearAlgebra`MatrixManipulation`"]
Vec[v_] := Partition[Flatten[Transpose[v]], 1]
Kronecker[x_, y_] := BlockMatrix[Outer[Times, x, y]]
Spalte[x_] := Length[x[[1]]]
Zeile[x_] := Length[Transpose[x][[1]]]
```

```

(* Vertauschungsmatrix und N_n *)
index[n_] := Flatten[Vec[Partition[Table[i, {i, 1, n^2}], n]]]
comm[n_] := Table[IdentityMatrix[n^2][[index[n][[i]]]], {i, 1, n^2}]
Nn[n_] := (1/2)(IdentityMatrix[n^2] + comm[n])

(* Hoffmann - Läuter - Gleichung *)
Clear[f];
f[x_, is_, b_] :=
Module[{fo, const, y, ty, iy, f},
  fo = is.Transpose[b].b.is;
  y = MatrixPower[fo + x.Transpose[x], 0.5];
  ty = Tr[y];
  iy = Inverse[y];
  const = 1/(1 + Tr[is]) is;
  f = ty const.iy.x]

(* Ableitung *)
Clear[df];
df[x_, is_, b_] :=
Module[{fo, const, y, ty, iy, f, g, df},
  fo = is.Transpose[b].b.is;
  y = MatrixPower[fo + x.Transpose[x], 0.5];
  ty = Tr[y];
  iy = Inverse[y];
  const = 1/(1 + Tr[is]) is;
  g = ty const.iy;
  df = 2Kronecker[Transpose[x], const].
    (Vec[iy].Transpose[Vec[IdentityMatrix[Zeile[x]]]] -
     ty Kronecker[iy, iy]).
    Inverse[Kronecker[y, IdentityMatrix[Zeile[x]]]
    + Kronecker[IdentityMatrix[Zeile[x]], y]].
    Nn[Zeile[x]].
    Kronecker[x, IdentityMatrix[Zeile[x]]]
    + Kronecker[IdentityMatrix[Spalte[x]], g]]

(* Fixpunktiteration *)
fix[is_, b_] :=
Module[{list, i, x, y},
  x = IdentityMatrix[Length[b]];
  i = 0;
  While[i < 100000 &&
    Max[Abs[Eigenvalues[f[x, is, b] - x]]] > 10^(-10),
    y = f[x, is, b];

```

```

    x = y;
    i = i + 1];];
list = {i, x, x.Transpose[x]}]

(* Nebenbedingung der Matrixgleichung *)
mo[a_, is_, b_] :=
Module[{fo, y, ty, m},
  fo = is.Transpose[b].b.is;
  y = MatrixPower[fo + a, 0.5];
  ty = Tr[y];
  m = (1 + Tr[is])/ty y - is]

(* Simulation *)
SeedRandom[7774];
stop = 100000;
simulation =
Table[n = Random[Integer, {2, 6}];
  s = ZufallPd[n];
  is = Inverse[s];
  b = ZufallDiag[n];
  erg = fix[is, b];
  i = Part[erg, 1];
  a = Part[erg, 3];
  d = mo[a, is, b];
  em = Min[Re[Eigenvalues[d]]];
  Which[em < (-1)10^(-6), jk = -1, em > 10^(-6),
    jk = 1, True, jk = 0];
  ableitung =
    Max[Abs[Eigenvalues[df[Part[erg, 2], is, b]]]];
  If[IntegerQ[1/1000] == True, Print[1]];
  {1, i, n, jk, ableitung} ,
  {1, 1, stop}];

(* Anzahl der abgebrochenen Iterationen *)
Needs["Statistics`Master`"]
Needs["Statistics`DescriptiveStatistics`"]
iteration = Part[Transpose[simulation], 2];
Count[iteration, 100000]

(* konvergente Iterationen *)
pos = Position[iteration, 1000000];
fixpunkt = Delete[simulation, pos];
it = Part[Transpose[fixpunkt], 2];

```

```

(* Fixpunkt erfüllt die Nebenbedingung *)
Count[Part[Transpose[fixpunkt], 4], 0] +
  Count[Part[Transpose[fixpunkt], 4], 1]

(* Verteilung der Iterationsanzahl *)
q0 = Min[it];
q1 = Quantile[it, 0.25];
q2 = Quantile[it, 0.5];
q3 = Quantile[it, 0.75];
q4 = Quantile[it, 0.95];
q5 = Quantile[it, 0.99];
q6 = Max[it];
quantile =
  {"Quantil", "0", "0.25", "0.5", "0.75",
   "0.95", "0.99", "1"},
  {"Anzahl", q0, q1, q2, q3, q4, q5, q6}};
quantile // TableForm

(* Maximum der Ableitung im Fixpunkt *)
abl = Part[Transpose[fixpunkt], 5];
Max[abl] > 1.000001

(* Verhalten der Ableitung bei j = k bzw. j < k *)
ableitung = Ceiling[abl - 0.999999];
kon = Transpose[{Part[Transpose[fixpunkt], 4], ableitung}];
tafel =
  {"", "rho<1", "rho=1"},
  {"j<k", Count[kon, {0, 0}],
   Count[kon, {0, 1}]}, {"j=k", Count[kon, {1, 0}],
   Count[kon, {1, 1}]}};
tafel // TableForm

```

D.2 2. Simulation

```

(* Überprüfung einer positiv definiten Matrix *)
TestPd[mat_] :=
  Module[{ewert}, ewert = Min[Re[Eigenvalues[mat]]];
  If[ewert > 10(-1), True, False]]

```

```

(* Erzeugung von Zufallsmatrizen *)
ZufallPd[n_] :=
  Module[{mat}, mat = DiagonalMatrix[Table[0, {n}]];
  While[TestPd[Transpose[mat].mat] == False,
    mat = Table[Random[Real, {-1, 1}], {i, 1, n}, {j, 1, n}];
  Transpose[mat].mat]
ZufallDiag[n_] :=
  DiagonalMatrix[Table[Random[Real, {0.2, 1}], {i, 1, n}]]

(* Vec - Operator, Kroneckerprodukt und Dimension *)
Needs["LinearAlgebra`MatrixManipulation`"]
Vec[v_] := Partition[Flatten[Transpose[v]], 1]
Kronecker[x_, y_] := BlockMatrix[Outer[Times, x, y]]
Spalte[x_] := Length[x[[1]]]
Zeile[x_] := Length[Transpose[x][[1]]]

(* Vertauschungsmatrix und N_n *)
index[n_] := Flatten[Vec[Partition[Table[i, {i, 1, n^2}], n]]]
comm[n_] := Table[IdentityMatrix[n^2][[index[n][[i]]]], {i, 1, n^2}]
Nn[n_] := (1/2)(IdentityMatrix[n^2] + comm[n])

(* Hoffmann - Läuter - Gleichung *)
Clear[f];
f[x_, is_, b_] :=
  Module[{fo, const, y, ty, iy, f},
    fo = is.Transpose[b].b.is;
    y = MatrixPower[fo + x.Transpose[x], 0.5];
    ty = Tr[y];
    iy = Inverse[y];
    const = 1/(1 + Tr[is]) is;
    f = ty const.iy.x]

(* Ableitung *)
Clear[df];
df[x_, is_, b_] :=
  Module[{fo, const, y, ty, iy, f, g, df},
    fo = is.Transpose[b].b.is;
    y = MatrixPower[fo + x.Transpose[x], 0.5];
    ty = Tr[y];
    iy = Inverse[y];
    const = 1/(1 + Tr[is]) is;
    g = ty const.iy;
    df = 2Kronecker[Transpose[x], const].

```

```

      (Vec[iy].Transpose[Vec[IdentityMatrix[Zeile[x]]]] -
        ty Kronecker[iy, iy]).
      Inverse[Kronecker[y, IdentityMatrix[Zeile[x]]]
      + Kronecker[IdentityMatrix[Zeile[x]], y]].
      Nn[Zeile[x]].
      Kronecker[x, IdentityMatrix[Zeile[x]]]
      + Kronecker[IdentityMatrix[Spalte[x]], g]]

(* Fixpunktiteration *)
fix[is_, b_] :=
  Module[{list, i, x, y},
    x = IdentityMatrix[Length[b]];
    i = 0;
    While[i < 100000 &&
      Max[Abs[Eigenvalues[f[x, is, b] - x]]] > 10(-10),
      y = f[x, is, b];
      x = y;
      i = i + 1;];
    list = {i, x, x.Transpose[x]}]

(* Nebenbedingung der Matrixgleichung *)
mo[a_, is_, b_] :=
  Module[{fo, y, ty, m},
    fo = is.Transpose[b].b.is;
    y = MatrixPower[fo + a, 0.5];
    ty = Tr[y];
    m = (1 + Tr[is])/ty y - is]

(* Simulation *)
SeedRandom[7774];
stop = 100000;
simulation =
  Table[n = Random[Integer, {2, 6}];
    s = ZufallPd[n];
    is = Inverse[s];
    b = ZufallDiag[n];
    erg = fix[is, b];
    i = Part[erg, 1];
    a = Part[erg, 3];
    d = mo[a, is, b];
    em = Min[Re[Eigenvalues[d]]];
    Which[em < (-1)10(-6), jk = -1, em > 10(-6),
      jk = 1, True, jk = 0];

```

```
ableitung =
  Max[Abs[Eigenvalues[df[Part[erg, 2], is, b]]]];
If[IntegerQ[1/1000] == True, Print[1]];
{1, i, n, jk, ableitung} ,
{1, 1, stop}];

(* Anzahl der abgebrochenen Iterationen *)
Needs["Statistics`Master`"]
Needs["Statistics`DescriptiveStatistics`"]
iteration = Part[Transpose[simulation], 2];
Count[iteration, 100000]

(* konvergente Iterationen *)
pos = Position[iteration, 100000];
fixpunkt = Delete[simulation, pos];
it = Part[Transpose[fixpunkt], 2];

(* Fixpunkt erfüllt die Nebenbedingung *)
Count[Part[Transpose[fixpunkt], 4], 0] +
  Count[Part[Transpose[fixpunkt], 4], 1]

(* Verteilung der Iterationsanzahl *)
q0 = Min[it];
q1 = Quantile[it, 0.25];
q2 = Quantile[it, 0.5];
q3 = Quantile[it, 0.75];
q4 = Quantile[it, 0.95];
q5 = Quantile[it, 0.99];
q6 = Max[it];
quantile =
  {"Quantil", "0", "0.25", "0.5", "0.75",
   "0.95", "0.99", "1"},
  {"Anzahl", q0, q1, q2, q3, q4, q5, q6}};
quantile // TableForm

(* Maximum der Ableitung im Fixpunkt *)
abl = Part[Transpose[fixpunkt], 5];
Max[abl] > 1.000001

(* Verhalten der Ableitung bei j = k bzw. j < k *)
ableitung = Ceiling[abl - 0.999999];
kon = Transpose[{Part[Transpose[fixpunkt], 4], ableitung}];
tafel =
```

```

{"", "rho<1", "rho=1"},
{"j<k", Count[kon, {0, 0}],
Count[kon, {0, 1}]}, {"j=k", Count[kon, {1, 0}],
Count[kon, {1, 1}]}};
tafel // TableForm

```

D.3 Beispiel

```

(* Eingabe der Matrizen *)
S = {{3, 9/11, 0}, {9/11, 15/44, 0}, {0, 0, 1}};
B = DiagonalMatrix[{1, 1/2, 1}];
S // MatrixForm
B // MatrixForm

(* Vec - Operator, Kroneckerprodukt und Dimension *)
Needs["LinearAlgebra`MatrixManipulation`"]
Vec[v_] := Partition[Flatten[Transpose[v]], 1]
Kronecker[x_, y_] := BlockMatrix[Outer[Times, x, y]]
Spalte[x_] := Length[x[[1]]]
Zeile[x_] := Length[Transpose[x] [[1]]]

(* Vertauschungsmatrix und N_n *)
index[n_] := Flatten[Vec[Partition[Table[i, {i, 1, n^2}], n]]]
comm[n_] := Table[IdentityMatrix[n^2] [[index[n] [[i]]]], {i, 1, n^2}]
Nn[n_] := (1/2)(IdentityMatrix[n^2] + comm[n])

(* Hoffmann - Läuter - Gleichung *)
Clear[f];
f[x_, is_, b_] :=
Module[{fo, const, y, ty, iy, f},
  fo = is.Transpose[b].b.is;
  y = MatrixPower[fo + x.Transpose[x], 0.5];
  ty = Tr[y];
  iy = Inverse[y];
  const = 1/(1 + Tr[is]) is;
  f = ty const.iy.x]

(* Ableitung *)
Clear[df];
df[x_, is_, b_] :=

```

```

Module[{fo, const, y, ty, iy, f, g, df},
  fo = is.Transpose[b].b.is;
  y = MatrixPower[fo + x.Transpose[x], 0.5];
  ty = Tr[y];
  iy = Inverse[y];
  const = 1/(1 + Tr[is]) is;
  g = ty const.iy;
  df = 2Kronecker[Transpose[x], const].
    (Vec[iy].Transpose[Vec[IdentityMatrix[Zeile[x]]]] -
     ty Kronecker[iy, iy]).
  Inverse[Kronecker[y, IdentityMatrix[Zeile[x]]]
  + Kronecker[IdentityMatrix[Zeile[x]], y]].
  Nn[Zeile[x]].
  Kronecker[x, IdentityMatrix[Zeile[x]]]
  + Kronecker[IdentityMatrix[Spalte[x]], g]]

(* Fixpunktiteration *)
fix[is_, b_] :=
Module[{list, i, x, y},
  x = IdentityMatrix[Length[b]];
  i = 0;
  While[i < 100000 &&
    Max[Abs[Eigenvalues[f[x, is, b] - x]]] > 10(-10),
    y = f[x, is, b];
    x = y;
    i = i + 1;];
  list = {i, x, x.Transpose[x]}]

(* Nebenbedingung der Matrixgleichung *)
mo[a_, is_, b_] :=
Module[{fo, y, ty, m},
  fo = is.Transpose[b].b.is;
  y = MatrixPower[fo + a, 0.5];
  ty = Tr[y];
  m = (1 + Tr[is])/ty y - is]

(* Iterationsanzahl,
Überprüfung der Erfüllung der Nebenbedingung,
Angabe des Falls und der Ableitung im Fixpunkt *)
is = Inverse[S];
erg = fix[is, B];
i = Part[erg, 1];
a = Part[erg, 3];

```

```
d = mo[a, is, B];
e = Re[Eigenvalues[d]];
em = Min[e];
Which[em < (-1)10(-6), jk = -1, em > 10(-6),
      jk = 1, True, jk = 0];
If[jk == -1, erf = "nein", erf = "ja"];
fall = Count[Ceiling[e - 10(-6)], 1];
ableitung = Max[Abs[Eigenvalues[df[Part[erg, 2], is, B]]]];
ausgabe =
  {"Iterationsanzahl:", i},
  {"Nebenbedingun erfüllt:", erf},
  {"Fall:", "j=" <> ToString[fall]},
  {"Spektralradius der Ableitung:", ableitung}};
ausgabe // TableForm
```

Abbildungsverzeichnis

8.1	Graph von f	120
8.2	Graph von g	121
8.3	Restpolynome von g_5	122
9.1	Ableitungen matrixwertiger Funktionen	128
9.2	Verteilung der Iterationsanzahl in Simulation 1	135
9.3	Fall und Ableitung im Fixpunkt bei Simulation 1	135
9.4	Verteilung der Iterationsanzahl in Simulation 2	136
9.5	Fall und Ableitung im Fixpunkt bei Simulation 2	136

Literaturverzeichnis

- [1] S. Bosch (1999); *Algebra*
Springer, Berlin.
- [2] N. Christopeit, K. Helmes (1996); *Linear Minimax Estimation with Ellipsoidal Constraints*
Acta Applicandae Mathematica, **43**, 3-15.
- [3] H. Drygas (1991); *On an Extension of the Girko Equality in Linear Minimax Estimation*
in: A.Pázman, J.Volanfová (eds.), Proceedings of Probstat '91 Conference, Bratislava, 3-10.
- [4] H. Drygas, H. Läuter (1994); *On the Representation of the Minimax Linear Estimator in the Convex Linear Model*
in: T. Caliński, R. Kala (eds.), Proceedings of the International Conference on Linear Statistical Inference LINSTAT '93, 13-26.
- [5] H. Drygas (1995); *A New and General Proof of the Spectral Equation in Minimax Linear Estimation*
Mathematische Schriften Kassel, Preprint Nr.2.
- [6] H. Drygas (1996); *Spectral Methods in Linear Minimax Estimation*
Acta Applicandae Mathematica, **43**, 17-42.

- [7] H. Drygas, J. Pilz (1996); *On the Equivalence of Spectral Theory and Bayesian Analysis in Minimax Linear Estimation*
Acta Applicandae Mathematica, **43**, 43-57.
- [8] H. Drygas (1998); *Determination of the Minimax Linear Estimation (MILE) in the Three Parameter Case*
Mathematische Schriften Kassel, Preprint Nr.3.
- [9] H. Drygas (1999); *Linear Minimax Estimation in the Three Parameter Case*
Tatra Mt. Math. Publ., **17**, 311-318.
- [10] T. S. Ferguson (1973); *Mathematical Statistics; A Decision Theoretic Approach*
Academic Press, New York.
- [11] N. Gaffke, B. Heiligers (2000); *Numerical methods for linear minimax estimation*
Probability and Statistics, **20**, 51-62.
- [12] V. L. Girko (1990); *S-Estimators*
Vychisl. Prickl. Mat., **71**, 90-97.
- [13] V. L. Girko (1996); *Spectral Theory of Estimation*
Acta Applicandae Mathematica, **43**, 59-69.
- [14] D. A. Harville (1997); *Matrix Algebra from a Statistician's Perspective*
Springer, New York.
- [15] J. B. Hiriart-Urruty, C.Lemaréchal (1993); *Convex Analysis and Minimization Algorithms I*
Springer, Berlin.

- [16] K. Hoffmann (1977); *Admissibility of Linear Estimators with Respect to Restricted Parameter Sets*
Math. Operationsforschung u. Statistik, Ser. Statistics, **8**, No. 4, 425-438.
- [17] K. Hoffmann (1979); *Characterization of Minimax Linear Estimators in Linear Regression*
Math. Operationsforschung u. Statistik, Ser. Statistics, **10**, No. 1, 19-26.
- [18] S. Kotz, N. Johnson (1982); *Encyclopedia of Statistical Science*
Wiley, New York.
- [19] J. Kuks, W. Ol'man (1972); *Minimax Linear Estimation of Regression Coefficients*
Eesti NSV Teaduste Akademii toimetised, **21**, 66-72.
- [20] E. Kunz (1991); *Algebra*
Vieweg, Braunschweig.
- [21] H. Läuter (1975); *A Minimax Linear Estimator for Linear Parameters Under Restrictions in Form of Inequalities*
Math. Operationsforschung u. Statistik, **6**, No. 5, 689-695.
- [22] J. R. Magnus, H. Neudecker (1988); *Matrix Differential Calculus with Applications in Statistics and Econometrics*
Wiley, Chichester.
- [23] J. M. Ortega, W.C. Rheinboldt (1970); *Iterative Solution of Nonlinear Equations in Several Variables*
Academic Press, New York.
- [24] J. Pilz (1991); *Bayesian Estimation and Experimental Design in Linear Regression Models*

Wiley, Chichester.

- [25] S. J. Press (1972); *Applied Multivariate Analysis*
Holt, Rinehart & Winston, New York.
- [26] C. R. Rao (1973); *Lineare Statistische Methoden und ihre Anwendung*
Academic Verlag, Berlin.
- [27] B. Schipp (1986); *Minimax-Schätzer im Regressionsmodell bei linearer
und nichtlinearer Vorinformation*
Diplomarbeit, Fachbereich Statistik, Universität Dortmund.
- [28] M. Sion (1958); *On general minimax theorem*
Pacific J. Math., **8**, 171-176.
- [29] P. Stahlecker, J. Lauterbach (1989); *Approximate Linear Minimax Esti-
mation in Regression Analysis With Ellipsoidal Constraints*
Commun. Statist. Theory Math., **18**, 2755-2784.
- [30] G. Stroth (1998); *Algebra*
de Gruyter, Berlin.
- [31] H. Witting (1985); *Mathematische Statistik I*
Teubner, Stuttgart.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfsmittel angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden.