

Datenvalidierung und Modellauswahl

Diplomarbeit
im Fachbereich Mathematik
der Universität Bremen

In Zusammenarbeit mit dem
ABB Forschungszentrum
Baden (CH)

Elmar Plischke 17. April 2000

Inhaltsverzeichnis

Einleitung	3
Kapitel 1. Datenvalidierung	5
1. Zielsetzung	5
2. Problemstellung	6
Kapitel 2. Lineare Nebenbedingungen	9
1. Das quadratische Programm	9
2. Lineare Validierung	11
3. Eigenschaften der Validierung	15
3.1. Verteilung der validierten Werte	16
3.2. Weitere Eigenschaften der Validierung	21
4. Güte-Test der Messung	24
5. Zusammenfassung	29
Kapitel 3. Nichtlineare Nebenbedingungen	31
1. Methode der Lagrangeschen Multiplikatoren	31
1.1. Beispiel	35
2. Nichtlineare Validierung	37
2.1. Linearisierung der Nebenbedingung	37
2.2. Fixpunktverfahren	41
2.3. Lagrange-Newton Verfahren	47
2.4. Beispiele	48
3. Güte-Test im nichtlinearen Fall	50
4. Datenvalidierung als Schätzer	51
5. Zusammenfassung	53
Kapitel 4. Einfluß unterschiedlicher Modellgleichungen	55
1. Lineare Transformationen der Nebenbedingungen	55
2. Nichtlineare Transformationen der Nebenbedingungen	57
3. Transformationsinvarianz der Validierung	60
4. Zusammenfassung	63
Kapitel 5. Sensorauswahl und Modellerweiterung	65
1. Der F-Test	66
2. Modellsuchverfahren	67
3. Modellerweiterung	69
4. Linearkombinationen von Nebenbedingungen	72
5. Zusammenfassung	75

Kapitel 6. Datenvalidierung mehrerer Messungen	77
1. Ad-hoc Methoden	78
2. Datenvalidierung und Kalman-Filter	80
2.1. Lineare Dynamik	82
2.2. Kalman-Filter unter Nebenbedingungen	84
2.3. Datenglättung mittels Kalman-Filter	85
3. Nichtlineare Dynamik	86
Kapitel 7. Zusammenfassung und Ausblicke	91
Anhang A. Projektionen und Pseudo-Inverse	93
Anhang B. Verteilungen und Testtheorie	101
Literaturverzeichnis	105

Einleitung

In dieser Arbeit werden Verfahren zur Datenvalidierung untersucht. Mit diesen Verfahren erhält man Schätzungen von gemessenen Prozeßdaten, wobei man zeitliche und räumliche Zusammenhänge zwischen diesen Daten ausnutzt, um den Meßfehler einzuschränken. Die Datenvalidierung ordnet also einem Meßwert einen Schätzwert zu, der im Rahmen der Meßgenauigkeit der Messung nahe ist und außerdem die Bedingungen erfüllt, die die Zusammenhänge zwischen den Meßstellen modellieren.

Wir geben im Folgenden einen Überblick über den Inhalt der einzelnen Kapitel dieser Diplomarbeit.

In Kapitel 1 werden die Zielsetzung und Problemstellung der Datenvalidierung untersucht. Eine mathematische Formulierung des Validierungsproblems führt auf eine Minimierung einer gewichteten Minimalen-Quadrate-Zielfunktion. Dafür sind Projektionen, die in Anhang A näher untersucht werden, ein wertvolles Hilfsmittel, denn wie sich herausstellt, handelt es sich bei der Validierung um eine Projektion auf eine Menge, deren Elemente die zeitlichen und räumlichen Zusammenhänge erfüllen.

Im Kapitel 2 untersuchen wir den Fall von linearen Nebenbedingungen. Im weiteren Verlauf dieses Kapitels werden die statistischen Eigenschaften des Verfahrens untersucht, die auf einen Güte-Test für die Messung führen. Dieser erlaubt es, Ausreißer in den Daten aufzuspüren, und kann damit als Indikator für brauchbare Messungen eingesetzt werden. Der nichtlineare Fall, bei dem die Modelle keine affinen Unterräume bilden, wird in Kapitel 3 untersucht. Dabei gibt uns die Methode der Lagrangesche Multiplikatoren Bedingungen an die Existenz der Validierung. Im Anschluß untersuchen wir verschiedene Iterationsverfahren zur Lösung der Validierung bei nichtlinearen Abhängigkeiten und übertragen Ergebnisse des linearen Falls auf die nichtlineare Validierung.

Dann gehen wir im Kapitel 4 der Frage nach, welches Gleichungssystem der Nebenbedingung man anwenden soll, ohne dabei das zu Grunde liegende Modell zu verändern. Es wird untersucht, wie sich Änderungen hier auf Validierungs-Verfahren auswirken. Dabei stellt sich heraus, daß einer der vorgestellten Algorithmen unter bestimmten Voraussetzungen unabhängig von der konkreten Wahl der Nebenbedingungsgleichungen ist.

In Kapitel 5 wird ein Verfahren zur Sensorauswahl vorgestellt. Denn mit einer Datenvalidierung verbindet sich die Absicht, den zu Grunde liegenden Prozeß durch a-priori Wissen besser zu verstehen. Dort untersuchen wir, wie man sein Vorwissen möglichst optimal einsetzen kann.

Im Kapitel 6 werfen wir einen Blick auf die dynamische Datenvalidierung, bei der auch die temporale Zusammenhänge in Form von Differentialgleichungen zur Validierung benutzt werden.

In Kapitel 7 wird zusammengefaßt, wie man die hier vorgestellten Ergebnisse zur Datenvalidierung in der Praxis benutzen kann. Außerdem werden dort weitere Fragen in Zusammenhang mit der Datenvalidierung aufgeworfen, die in dieser Arbeit keinen Platz gefunden haben.

In den Anhängen erläutern wir die Zusammenhänge zwischen Pseudo-Inversen und Projektionen, außerdem erinnern wir an einige statistische Definitionen und Sätze.

KAPITEL 1

Datenvalidierung

In dieser Arbeit untersuchen wir ein Verfahren, durch das meßtechnisch gewonnene Werte auf ihre Übereinstimmung mit den zu Grunde gelegten Prozeßgleichungen verglichen und entsprechend korrigiert werden.

1. Zielsetzung

Eine Anlage stellt dem Bediener Messungen zur Verfügung, die zur Überwachung und Steuerung der dort ablaufenden Prozesse eingesetzt werden. Allerdings sind diese Messungen mit Meßungenauigkeiten behaftet, die das Meßergebnis verfälschen. Ziel ist es, diese Ungenauigkeiten aus den erfaßten Daten herauszurechnen. Dieses Verfahren nennt sich Datenvalidierung (Datenaufbereitung, engl.: data reconciliation). Läßt sich jede Meßstelle einem Bauteil der Anlage zuordnen, für dessen Verhalten Gleichungen bekannt sind, die z.B. auf physikalischen oder chemischen Gesetzen beruhen, wie z.B. Massen- und Energieerhaltung, thermodynamische Beziehungen zwischen Druck, Temperatur, Dichte und Volumen, Bauteilekennlinien oder Modellen von Reaktionsmechanismen, so lassen sich die „offensichtlichen Unstimmigkeiten“ zwischen Modell und Messung mit dem in dieser Arbeit beschriebenen Verfahren herausrechnen und eventuell auch nicht direkt meßbare Werte bestimmen.

Ein solches Vorgehen gewinnt immer mehr Bedeutung, denn Bedingungen an die Prozeßführung, wie z.B. Produktionsqualität, Emissionsbestimmungen oder Lebendaueroptimierung, erfordern eine immer höhere Genauigkeit dieser Messungen. Aber Sensoren mit höherer Genauigkeit sind teuer und die Messung von kritischen Größen ist nicht immer direkt möglich. Deswegen greift man zu Verfahren, die die Genauigkeit des gegebenen Instrumentensatzes durch den Einsatz geeigneter mathematischer Modelle erhöhen.

Wir beschränken uns im ersten Teil dieser Arbeit auf solche Gleichungen, die keine Differentialterme enthalten, erst in Kapitel 6 betrachten wir die dynamische Datenvalidierung.

Ziel des Verfahrens zur Datenvalidierung ist also eine Schätzung des wahren Zustands auf Grundlage der Meßwerte, wobei der Begriff „Zustand“ im allgemeinen Sinne verwendet wird. Aus den Ergebnissen lassen sich dann Rückschlüsse auf Modell- oder Sensorfehler ziehen. Mit der Annahme, daß jeder Sensor nur eine reelle Messung liefert, haben wir also n Sensoren und damit eine n -dimensionale Messung y . Fassen

wir die Nebenbedingung f als m -dimensionales Gleichungssystem auf, so gibt es mehr Sensordaten als Gleichungen der Nebenbedingung. Die Einschränkung auf einen festen Zeitpunkt umgeht Probleme, die bei einem zeitabhängigen Modell entstehen. Ohne zeitliche Abhängigkeit könnte man genauso gut einen zeitlichen Mittelwert mehrerer Messungen als „wahren Wert“ benutzen.

Außerdem muß man noch beurteilen können, inwieweit es gerechtfertigt ist, die Messung im Rahmen der Meßgenauigkeit durch die Validierung zu ersetzen. Mit so einer Beurteilung läßt sich dann die Güte der Messung bestimmen.

Ein weiteres Ziel diese Arbeit ist die Untersuchung geeigneter Voraussetzungen für die Datenvalidierung. Dazu untersuchen wir, welche Grundlagen gegeben sein müssen, damit die Datenvalidierung möglichst sinnvolle Ergebnisse liefert. Insbesondere geht es darum, ein geeignetes Modell und dessen mathematische Darstellung zu finden.

2. Problemstellung

Wir wollen in diesem Abschnitt eine mathematische Formulierung für das Problem der Datenvalidierung finden.

Wir nehmen an, daß ein Sensor i (Meßwertgeber) nur einen reellwertigen Meßwert $y_i \in \mathbb{R}$ liefert. Dies vernachlässigt die Tatsache, daß Sensoren nur einen beschränkten Meßbereich haben oder auch nur Werte aus einem diskrete Wertebereich liefern können. Zusätzlich werden die *Varianzen* σ_i^2 der Störungen jedes einzelnen Sensors als bekannt vorausgesetzt. Diese Varianzen lassen sich aus den technischen Daten des Sensors oder experimentell gewinnen.

Die *Messung* $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ soll durch die Validierung im Rahmen der Meßgenauigkeit so abgeändert werden, daß dieser veränderte Wert x^* zu einem gegebenen Modell paßt. Dieses Modell \mathcal{M} wird implizit gegeben durch die *lokale Gleichung*

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^m, f(x) = 0 \iff x \in \mathcal{M},$$

wobei f zweifach stetig differenzierbar und $n > m$ ist. Diese Gleichung enthält also die funktionalen Abhängigkeiten zwischen den einzelnen Sensormeßstellen. Das Modell \mathcal{M} besteht dann aus der Nullstellenmenge von f . Wir fordern, daß das Modell \mathcal{M} eine Untermannigfaltigkeit des \mathbb{R}^n bildet, damit die Voraussetzungen für die Existenz von Lösungen in Kapitel 3 automatisch erfüllt sind.

Die Darstellung von \mathcal{M} durch $f^{-1}(0)$ gilt nur lokal. Da wir uns nur für lokale Aussagen interessieren, nehmen wir an, daß f \mathcal{M} vollständig beschreibt. Daher benutzen wir die Begriffe *Modell* und *Nebenbedingung* als Synonyme, sofern Verwechslung ausgeschlossen ist.

Wir fassen die nötigen Bezeichnungen in folgender Definition zusammen. Die dabei verwendete Norm wird in Anhang A definiert.

DEFINITION 1.1. Wir führen folgende Bezeichnungen ein:

- Die Kostenfunktion $J : \mathbb{R}^n \rightarrow \mathbb{R}$, die jedem r nichtnegative Kosten zuordnet, sei gegeben durch

$$J(r) = r^\top D^{-1} r = \|r\|_{D^{-1}}^2,$$

wobei D die gegebene Kovarianzmatrix der Sensoren i ($i = 1, \dots, n$) ist, die die einzelnen reellen Meßwerte gewichtet, so daß sich D als Diagonalmatrix $D = \text{diag}(\sigma_i^2)$ der einzelnen Varianzen σ_i^2 schreiben läßt. D ist dann eine positiv definite Matrix, was wir mit $D \succ 0$ bezeichnen.

- Dann ist eine *Datenvalidierung* x^* von y unter dem Modell \mathcal{M} gegeben durch folgende Forderungen:
 1. Die Validierung erfüllt das Modell $x^* \in \mathcal{M} \iff f(x^*) = 0$.
 2. Die Kosten $J(x^* - y)$ von der Messung zum validierten Wert sind minimal.
- Das Residuum r wird gegeben als Differenz zwischen der Messung y und dem validierten Wert x^* :

$$r := x^* - y$$

Das Residuum läßt sich als Schätzung des Fehlers interpretieren, der der Messung anhaftet.

Die Kostenfunktion ist in solche Richtungen $v \in \mathbb{R}^n$ groß, in die Norm $\|Dv\|$ klein ist, d.h. Sensoren mit hoher Genauigkeit und deswegen kleiner Varianz fallen bei diesen Kosten stärker ins Gewicht.

Wir treffen die folgenden Annahmen:

- (A1) Die Sensorstörungen sind normalverteilt mit bekannter Varianz σ^2 und Erwartungswert 0.
- (A2) Das Modell f beschreibt die Wirklichkeit im Rahmen der Meßgenauigkeit hinreichend gut, der Modellfehler ist im Vergleich zum Meßfehler vernachlässigbar klein.

Die erste Bedingung deckt nicht den Fall ab, daß die Sensoren driften, denn in diesem Fall hätten wir von 0 verschiedene Erwartungswerte. Die zweite Bedingung enthält z.B. die Forderung, daß der Prozeß im Gleichgewichtszustand operiert, falls die Gleichungen der Nebenbedingung nur im Gleichgewicht gültig sind.

Ist $f(y) = 0$, so ist eine Datenvalidierung nicht notwendig, denn dann ist die Nebenbedingung erfüllt, somit liegt der Meßwert im Modell.

Die Forderungen in der Definition 1.1 lassen sich zu der folgenden Bedingung zusammenfassen.

DEFINITION 1.2. Mit den Voraussetzungen und Bezeichnungen von Definition 1.1 gilt: Die *Datenvalidierung* x^* der Messung y unter dem Modell f wird gegeben als Lösung einer Minimierungsaufgabe der Form

$$(1) \quad x^* = \arg \min_{f(x)=0} \|x - y\|_{D^{-1}}^2.$$

Dabei bezeichnet $\arg \min$ ein minimierendes Argument, das wie folgt definiert wird:

$$z = \arg \min\{A(x) \mid x \in B\} \iff z \in B \text{ und } A(z) \leq A(x) \forall x \in B.$$

Dieses minimierende Argument ist nicht eindeutig bestimmt. \square

Ziel der Datenvalidierung ist es also, einen Wert x^* im Modell zu finden, der der Messung y im Sinne *gewichteter minimaler Quadrate* am nächsten kommt. Dabei müssen Nebenbedingungen f eingehalten werden, die die Eigenschaften der zu Grunde liegenden Anlage modellieren. Es handelt sich hierbei um eine *Ausgleichsrechnung* unter Nebenbedingungen, allerdings wird hier die Messung zu einem festen Zeitpunkt an das Modell angepaßt, und nicht, wie üblich, ein Modell an eine ganze Meßreihe zu unterschiedlichen Zeiten. Die Ausgleichsrechnung wird besonders im statistischen Bereich auch als *lineare Regression* bezeichnet. In [Bro97], S. 364, werden weitere übliche Bezeichnungen für die Ausgleichsrechnung aufgeführt.

KAPITEL 2

Lineare Nebenbedingungen

Wir untersuchen zunächst affin lineare Nebenbedingungen für unser Problem aus Definition 1.2. Diese Untersuchung führt auf einen geschlossenen Ausdruck, dessen Eigenschaften wir in den folgenden Abschnitten noch weiteren untersuchen wollen.

Lineare Nebenbedingungen ergeben sich z.B. bei der Modellierung von Zu- und Abflüssen in einem Leitungssystem, denn die Summe von Zu- und Abflüssen versehen mit den entsprechenden Vorzeichen in einer Verzweigung sollte bei perfekter Messung 0 sein.

Die Nebenbedingung sei gegeben durch die Gleichung

$$f(x) = Ax - b,$$

(2) bzw. das Modell durch die Menge

$$\mathcal{M} = \{x \in \mathbb{R}^n \mid Ax - b = 0\},$$

wobei $A \in \mathbb{R}^{m \times n}$, $m < n$ eine reguläre Matrix ist, d.h. $\text{rang } A = m$, und $b \in \mathbb{R}^m$. Wir verwenden die Begriffe Modell und Nebenbedingung als Synonyme, sofern keine Verwechslungsgefahr besteht.

Die affine Gleichung f sei außerdem lösbar, also $b \in \text{Im}A$. Damit ist gewährleistet, daß das Modell \mathcal{M} nicht leer ist. Die Kostenfunktion ist nach Definition 1.1 gegeben durch

$$J(r) = r^\top D^{-1} r = \|r\|_{D^{-1}}^2,$$

wobei das Residuum $r := x^* - y$, die Schätzung des Meßfehlers, von der Messung y und der Validierung x^* abhängt.

Wir wollen zeigen, daß die Validierung, also die Lösung von (1), in diesem Fall ein *quadratisches Programm (QP)* ist. Wir schauen uns im nächsten Abschnitt die allgemeine Lösung eines QPs an und spezialisieren die Lösung für unsere Situation im Abschnitt 2.2.

1. Das quadratische Programm

Als quadratisches Programm (QP) wird eine Minimierungsaufgabe einer quadratischen Zielfunktion unter linearen Nebenbedingungen bezeichnet:

(3) Minimiere $x^\top P x + 2c^\top x$ unter der Nebenbedingung $Ax = b$,

wobei $P \succ 0$ und $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, A regulär, gegeben sind. Die Bezeichnung „Programm“ stammt dabei aus der Wirtschaftsmathematik.

Meist werden noch Ungleichungen mit in die Problemstellung aufgenommen, aber diesen Fall untersuchen wir hier nicht. Wir werden in Gleichung (5) sehen werden, ist eine Validierung ein quadratisches Programm.

Wir wollen die allgemeine Lösung des quadratischen Programms (3) herleiten (Siehe auch [Ber95, Kapitel 2.1], [Fle87, Kapitel 10]).

PROPOSITION 2.1 (QP). *Das QP (3) wird gelöst durch*

$$(4) \quad x^* = -P^{-1} (c - A^\top (AP^{-1}A^\top)^{-1} (AP^{-1}c + b)).$$

BEWEIS. Zunächst setzen wir $P = I$ und $b = 0$. Dann läßt sich das Problem (3) durch eine konstante quadratische Ergänzung der Zielfunktion in das äquivalente Problem

$$\text{Minimiere } \|x + c\|^2 \text{ unter } Ax = 0$$

überführen. Nach dem Projektionslemma (A.9) wird die Lösung x^* gegeben durch die Projektion von $-c$ auf den Kern¹ von A . Dabei erfüllt

$$x^* = -(I - A^\top (AA^\top)^{-1} A)c$$

die Bedingungen des Projektionslemmas A.9.

Für den Fall $P \neq I$ und $b \neq 0$ transformieren wir x mithilfe einer Cholesky-Zerlegung² von $P =: P^{\top/2} P^{1/2}$ und einem Urbild \bar{x} von b unter A nach $y = P^{1/2}(x - \bar{x})$, $B = AP^{-1/2}$ und $d = P^{1/2}(P^{-1}c + \bar{x})$. Die Existenz von \bar{x} ist unmittelbare Folge aus der Lösbarkeit von $Ax = b$. Dann betrachten wir eine neue Zielfunktion der Gestalt

$$\begin{aligned} y^\top y + 2d^\top y &= (x - \bar{x})^\top P(x - \bar{x}) + 2(P^{-1}c + \bar{x})^\top P(x - \bar{x}) \\ &= x^\top Px - 2\bar{x}^\top Px + \bar{x}^\top P\bar{x} + 2c^\top x + 2\bar{x}^\top Px - 2c^\top \bar{x} - 2\bar{x}^\top P\bar{x} \\ &= x^\top Px + 2c^\top x - (2c + \bar{x})^\top P\bar{x}. \end{aligned}$$

Sie unterscheidet sich also nur um Konstanten von dem ursprünglichen Problem, die auf die Minimierung keinen Einfluß ausüben. Die Nebenbedingung $By = 0$ ist äquivalent zur Bedingung $Ax = b$, denn

$$By = A(x - \bar{x}) = Ax - b.$$

Die Lösung für das Problem ist also nach dem ersten Teil gegeben durch

$$y^* = -(I - B^\top (BB^\top)^{-1} B)d.$$

¹ $\ker A = \{x \in \mathbb{R}^n \mid Ax = 0\}$

²Das Cholesky-Verfahren wird in [Sch93, Abschnitt 1.3.1] beschrieben.

Nach Rücktransformation dieser Lösung erhält man dann die Darstellung

$$\begin{aligned} x^* &= -P^{-1/2} \left(I - P^{-\top/2} A^\top (AP^{-1/2} \quad P^{-\top/2} A^\top)^{-1} AP^{-1/2} \right) \\ &\quad P^{1/2} (P^{-1}c + \bar{x}) + \bar{x} \\ &= -P^{-1}c - \bar{x} + \bar{x} + P^{-1}A^\top (AP^{-1}A^\top A(P^{-1}c + \bar{x})) \\ &= -P^{-1} (c - A^\top (AP^{-1}A^\top)^{-1} (AP^{-1}c + b)), \end{aligned}$$

womit (4) gezeigt ist. \square

Wir besitzen nun eine Lösung für das quadratische Programm. Ein alternativer Beweis dieser Lösung wird in Beispiel 3.6 mit Methoden der Lagrangeschen Multiplikatoren vorgestellt.

Im Folgenden wollen wir dieses Resultat auf das Validierungsproblem 1.2 anwenden.

2. Lineare Validierung

Wir werden nun mit dem Ergebnis aus Proposition 2.1 eine Formel für die Validierung mit affin linearen Nebenbedingungen herleiten.

Dazu müssen wir zeigen, daß die Minimierungsaufgabe (1) im linearen Fall ein QP ist. Unter Vernachlässigung konstanter Terme erhalten wir nämlich

$$(5) \quad \arg \min_{Ax=b} \|x - y\|_{D^{-1}}^2 = \arg \min_{Ax=b} \{x^\top D^{-1}x - 2y^\top D^{-1}x\}.$$

Also ist die Validierung ein quadratisches Programm. Proposition 2.1 liefert dann folgendes Resultat.

LEMMA 2.2 (Validierung). *Die Lösung des Validierungsproblems (1) wird gegeben durch*

$$(6) \quad x^* = y - DA^\top (ADA^\top)^{-1} (Ay - b) = y - DA^\top (ADA^\top)^{-1} f(y).$$

Diese Gleichung (6) bezeichnen wir im Folgenden als *Validierung oder Validierungsschritt*. Wir besitzen nun eine Formel, um (1) im linearen Fall zu lösen.

BEWEIS. Gleichung (6) ist eine Folgerung aus Proposition 2.1, indem man $P = D^{-1}$, $c = -D^{-1}y$ in die QP-Lösung (4) einsetzt. \square

Gilt schon $y = x^*$, so bewirkt der Validierungsschritt keine Veränderung mehr, denn es gilt

$$\begin{aligned} f(y) &= f(x^*) = Ax^* - b \\ &= A(y - DA^\top (ADA^\top)^{-1} (Ay - b)) - b \\ &= Ay - (Ay - b) - b = 0. \end{aligned}$$

Wir demonstrieren die Wirkungsweise des Validierungsschrittes an folgenden Beispielen. Das erste Beispiel wird uns als Leitbeispiel noch weiter durch diese Arbeit begleiten.

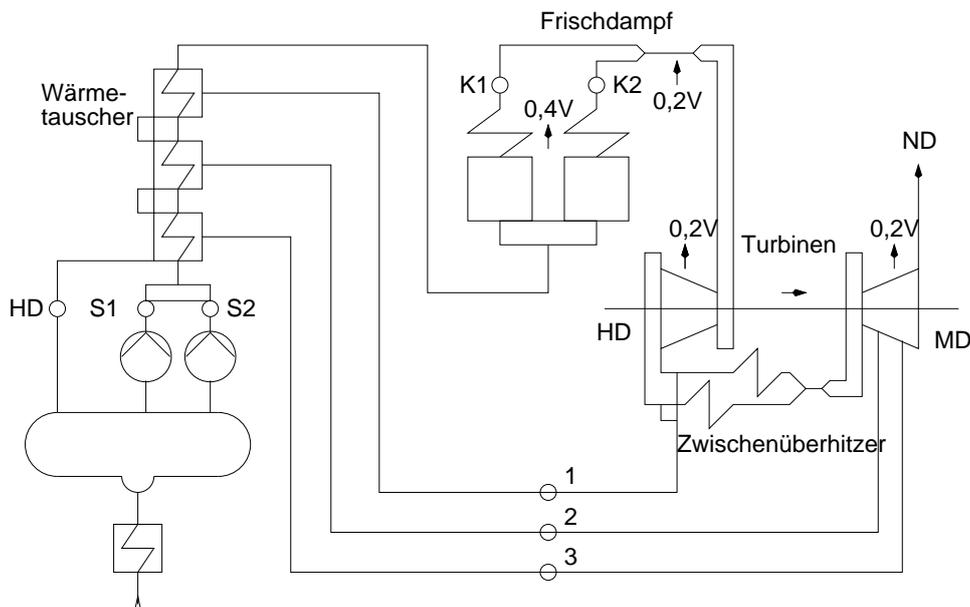


ABBILDUNG 1. Wärmeschaltbild zu Beispiel 2.3.

BEISPIEL 2.3. Frischdampf-Massenfluß einer Dampfturbinenanlage (aus [VDI98])

Durch Betrachtung der Zu- und Abflüsse im Wasser-/Dampf-Kreislauf einer Dampfturbinenanlage erhält man für den Frischdampf- und Hochdruckdampf-Massenfluß (m_{FD} bzw. m_{HD}) in diesem Beispiel die Gleichungen

$$m_{FD} = m_{K1} + m_{K2} - 0,2 \cdot m_V,$$

$$m_{FD} = m_{S1} + m_{S2} - 0,6 \cdot m_V,$$

$$m_{FD} = m_{HK} + m_1 + m_2 + m_3 + 0,4 \cdot m_V, \quad \text{außerdem}$$

$$m_{HD} = m_1 + m_2 + m_3.$$

Abbildung 1 zeigt einen Ausschnitt aus dem Wärmeschaltbild einer Dampfturbinenanlage mit den zugehörigen Meßpunkten. Leider gibt der Originalartikel nicht die Lage des Sensors für die Messung von m_{HK} an. Dabei handelt es sich wohl um den Rücklauf in den Kessel, der sich in Abbildung 1 unten links befindet.

Die Modellgleichungen für den Frischdampfmassenfluß dieses Beispiels werden am Eingang der HD-Turbine aufgestellt.

Die Verluste m_V werden nach einer Faustregel den Gleichungen zugefügt. Die Meßwerte y_i werden mit dem jeweiligen 95%-Konfidenzintervallen $y_i \pm v_i$ angegeben:

$$\begin{array}{ll} m_{K1} = 46,241 \pm 0,800 & m_{K2} = 45,668 \pm 0,790 \\ m_{S1} = 44,575 \pm 0,535 & m_{S2} = 44,319 \pm 0,532 \\ m_V = 0,525 \pm 0,105 & m_{HK} = 69,978 \pm 0,854 \\ m_1 = 10,364 \pm 0,168 & m_2 = 3,744 \pm 0,058 \\ m_3 = 4,391 \pm 0,058 & m_{HD} = 18,498 \pm 0,205. \end{array}$$

Das Modell wird gegeben durch Auflösung nach m_{FD} , denn für diesen Wert liegt keine direkte Messung vor.

$$(7) \quad f(y) = \begin{pmatrix} 1 & 1 & -1 & -1 & 0,4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 \end{pmatrix} y = Ay$$

mit $b = 0$ in der Modelldarstellung aus (2) und

$$y = (m_{K1}, m_{K2}, m_{S1}, m_{S2}, m_V, m_{HK}, m_1, m_2, m_3, m_{HD})^\top.$$

Für diese Messung gilt: $\|f(y)\| = 3.08 \neq 0$, die Meßwerte passen also nicht zum Modell. Nach der Umrechnung der Konfidenzgrenzen in Kovarianzen durch $\sigma_i^2 = \left(\frac{v_i}{1,96}\right)^2$ und dem Durchführen des Validierungsschrittes (6) mit $D = \text{diag}(\sigma_i^2)$ erfüllen die validierten Werte x^* die Modellgleichungen: Für die validierten Werte

$x^* = (44.95, 44.42, 44.83, 44.57, 0.05624, 70.82, 10.38, 3.746, 4.393, 18.52)^\top$ ergibt sich

$$\|f(x)\| = 8.87 \cdot 10^{-15} < \max \{|x_i| \mid i = 1, \dots, 10\} \text{ eps},$$

wobei $\text{eps} = 2.22 \cdot 10^{-16}$ die Maschinengenauigkeit³ des verwendeten Rechners ist, so daß das Ergebnis als „praktisch Null“ gedeutet werden kann. Dies Validierung ist allerdings ohne Angabe von Konfidenzgrenzen oder einer Kovarianz nicht besonders aussagekräftig⁴, denn es fehlt eine Aussage, wie sehr die einzelnen Komponenten der Messung durch die Validierung verändert werden und wie groß diese Änderungen ausfallen. Sinnvoller ist also die Angabe eines Bereiches, in dem diese Validierung gut zur Messung paßt.

BEISPIEL 2.4 (Lineare Redundanz). Wir betrachten den Fall, daß mehrere Sensoren an einer Meßstelle eingesetzt werden, um zusätzliche Redundanz einzuführen. Diese Redundanz wird durch ein erweitertes Modell in die Nebenbedingungen übernommen. Man erhält das erweiterte

³Die Maschinengenauigkeit ε ist der kleinste auf dem Rechner darstellbare positive Wert, für den in der Rechnerarithmetik gilt: $1.0 + \varepsilon > 1.0$. Also ist $1.0 + \varepsilon/2$ schon nicht mehr von der 1.0 zu unterscheiden.

⁴Da wir eine stetige Verteilung der Störungen angenommen haben, besitzen einzelne Punkte überhaupt keine Wahrscheinlichkeitsmasse!

Modell, indem man eine Gleichung der Form $x_i = x_j$ zu einem vorhandenen Modell hinzufügt. Wir validieren jetzt nur diese erweiterte Gleichung für $i = 1, j = 2$. Zu erwarten ist, daß die validierten Werte x^*_1 und x^*_2 gleich sind. Auf die Notation für die Verteilung von Zufallsvariablen wird in Anhang B eingegangen.

Für die Modellgleichung $f(x) = x_1 - x_2$ ergeben sich die folgende Aussagen, wenn die i -te Messung eine Fehlerverteilung $\mathcal{N}(\mu_i, \sigma_i^2)$ ($i = 1, 2$) besitzt.

Wir wollen also die Gleichung

$$f(x) = x_1 - x_2 = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Ax$$

nach Gleichung (6) validieren. Die Kovarianzmatrix hat dabei die Gestalt

$$D = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

Dann gilt:

$$ADA^\top = \sigma_1^2 + \sigma_2^2, \quad \text{damit ist dann}$$

$$DA^\top(ADA^\top)^{-1}f(x) = \frac{1}{\sigma_1^2 + \sigma_2^2} \begin{pmatrix} \sigma_1^2 \\ -\sigma_2^2 \end{pmatrix} (x_1 - x_2).$$

Also ist die Validierung gegeben durch

$$x^* = \begin{pmatrix} x_1 - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(x_1 - x_2) \\ x_2 + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}(x_1 - x_2) \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2}$$

Die Validierung dieses erweiterten Modells liefert gewichtete Mittelwerte, denn für Zufallsvektoren X und Y mit einer Normalverteilungsannahme $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}(\mu, \sigma^2)$ und $Y \stackrel{\mathcal{L}}{\sim} \mathcal{N}(\nu, \tau^2)$ ergibt sich der gewichtete Mittelwert nach Proposition B.5 zu

$$\frac{1}{\sigma^2 + \tau^2}(\tau^2 X + \sigma^2 Y) \stackrel{\mathcal{L}}{\sim} \mathcal{N}\left(\frac{\tau^2 \mu + \sigma^2 \nu}{\tau^2 + \sigma^2}, \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}\right),$$

$$\text{insbesondere ist } \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2} < \min\{\sigma^2, \tau^2\}.$$

Die gesamte Verteilung besitzt damit eine kleinere Varianz als die Varianzen der einzelnen Messungen. Es läßt sich also die Genauigkeit durch Einsatz mehrerer Sensoren an der gleichen Stelle erhöhen. Dabei ist aber zu beachten, daß diese bei einer Mittelwertbildung entsprechend ihrer Varianz gewichtet werden. Die Validierung liefert dieses Ergebnis automatisch, während man bei einer Vorverarbeitung diese Gewichtung leicht vergessen kann. Die soeben hergeleitete Formel für ein erweitertes Modell mit redundanten Sensoren erinnert entfernt an die Formel

für Parallelschaltung von Widerständen, aber wir haben hier ja auch nur Modellgleichungen parallel geschaltet.

3. Eigenschaften der Validierung

In diesem Abschnitt wollen wir unter anderem die statistischen Eigenschaften des Validierungsschrittes untersuchen. Wir haben in Beispiel 2.3 bemerkt, daß uns für praktische Aussagen eine „Genauigkeit nach Validierung“ fehlt. Diese liefert uns gerade die *induzierte Kovarianzmatrix* aus Definition 2.5.

Betrachtet man die Messung $y \in \mathbb{R}^n$ als Realisierung eines Zufallsvektors Y , wobei die einzelnen Sensorstörungen einer Normalverteilung bekannter Varianz genügen, so möchte man wissen, wie sich die Verteilung der nun als Zufallsvariable aufgefaßten Validierung X verhält. Eine Darstellung der Kovarianz von X erhalten wir in Gleichung (12). Bei linearen Nebenbedingungen ist die Validierung gerade eine lineare Transformation der Messung, so daß die Normalverteilung unter dem Verfahren erhalten bleibt.

Mit der nach Annahme (A1) vorausgesetzten Normalverteilung ist der verwendete Minimal-Quadrate-Ansatz eine *Maximum-Likelihood Schätzung*. Dieser Schätzer ist in unserem Fall (bekannte Varianz, unbekannter Mittelwert) konsistent und erwartungstreu. (Siehe [Kre91], Beispiel 12.1, Fall I). Für andere Verteilungen hat dieser dann „verallgemeinerte minimale Quadrate“-Ansatz einen Bias, der auf nicht einmal erwartungstreue Schätzer führt. Bei anderen Verteilungen müßte man untersuchen, ob eine andere Kostenfunktion wieder auf Maximum-Likelihood Schätzungen führt oder ob dann die systematischen Fehler, die durch Anwendung des vorgestellten Verfahrens unter Beibehaltung der Kostenfunktion $J(r) = r^T D^{-1} r$ entstehen, sich so stark auf die Ergebnisse auswirken, daß die dadurch gewonnenen Aussagen nutzlos sind. Das sprengt aber den Rahmen dieser Arbeit.

Wir kehren nun zur Betrachtung des linearen Falls unter normalverteilten Störungen zurück. Bei der Untersuchung der statistischen Eigenschaften der Validierung stellen sich folgende Fragen:

1. Streut Y um \hat{y} und X um \hat{x} , gilt dann $\hat{x} = \hat{y}$?
2. Wie läßt sich der Einfluß der Nebenbedingungen darstellen ?
3. Wenn Y die Kovarianz D besitzt, wie sieht die Kovarianz von X aus ?
4. Läßt sich die Güte der Messung beurteilen ?

Die erste Frage läßt sich leicht beantworten: Wir haben vorausgesetzt, daß der wahre Wert \hat{y} die Nebenbedingungen erfüllt, dieser ist also ein Fixpunkt des Validierungsschrittes, damit streut auch X um \hat{y} . Bei der zweiten und dritten Frage muß man sich überlegen, ob man die Verteilung der Residuen, also der geschätzten Sensorstörungen, oder die der Validierungen bestimmen möchte. Wie sich im nächsten Abschnitt zeigen wird, hat man mit der Bestimmung der Kovarianz der Residuen,

die den durch die Validierung herausgerechneten Anteil an Information angibt, auch die Frage nach der Kovarianz der validierten Werte gelöst. Frage 4 wird im Abschnitt 2.4 behandelt.

3.1. Verteilung der validierten Werte. In diesem Unterabschnitt wollen wir die Frage nach der Verteilung der validierten Werte beantworten. Die Kenntnis dieser Verteilung gibt uns Aufschluß über die Arbeitsweise des Validierungsschritts.

Ist Y der Zufallsvektor der Messungen und X den Zufallsvektor der Validierungen, so definieren wir:

DEFINITION 2.5. Die Kovarianzmatrix der Residuen $R := X - Y$ heißt die (durch das Modell f) *induzierte Kovarianzmatrix* D^* .

Nach Voraussetzung besitzt die Zufallsvariable der Residuen eine Normalverteilung mit Erwartungswert 0, sie läßt sich daher vollständig durch Angabe einer Kovarianzmatrix beschreiben. Wir wollen nun eine Formel für die induzierte Kovarianzmatrix D^* herleiten.

PROPOSITION 2.6. *Sind die Messungen Y normalverteilt mit Kovarianz $\text{Cov}(Y) = D$, so besitzt der Zufallsvektor R der Residuen die Verteilung*

$$R \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, D^*)$$

mit

$$D^* = DA^\top(ADA^\top)^{-1}AD.$$

BEWEIS. Nach den Annahmen aus (A1) und (A2) und mit (6) wissen wir $Y \stackrel{\mathcal{L}}{\sim} \mathcal{N}(\hat{y}, D)$ und $X = Y - DA^\top(ADA^\top)^{-1}(AY - b)$, $EX = \hat{y}$. Dann ist $R = DA^\top(ADA^\top)^{-1}A(Y - \hat{y})$, also gilt dann nach dem Transformationssatz (B.5) für Normalverteilungen

$$R \stackrel{\mathcal{L}}{\sim} \mathcal{N}(\hat{y} - EX, D^*) = \mathcal{N}(0, D^*),$$

wobei sich die Kovarianz von R aus folgender Rechnung ergibt:

$$\begin{aligned} D^* &:= \text{Cov}(R) = DA^\top(ADA^\top)^{-1}A \text{Cov}(Y)(DA^\top(ADA^\top)^{-1}A)^\top \\ &= DA^\top(ADA^\top)^{-1}ADA^\top(ADA^\top)^{-1}AD \\ &= DA^\top(ADA^\top)^{-1}AD. \end{aligned}$$

□

Wir wollen nun die induzierte Kovarianz mit der Kostenfunktion J verbinden. Für den optimalen Wert des Residuums \tilde{r} des linearen Modells gilt nach (6)

$$\tilde{r} = x^* - y = DA^\top(ADA^\top)^{-1}A\tilde{r}.$$

Damit ergibt sich für die Kostenfunktion J an dieser Stelle

$$\begin{aligned}
 J(\tilde{r}) &= \tilde{r}^\top D^{-1} \tilde{r} \\
 (8) \quad &= \tilde{r}^\top A^\top (ADA^\top)^{-1} ADD^{-1} DA^\top (ADA^\top)^{-1} A \tilde{r} \\
 &= \tilde{r}^\top A^\top (ADA^\top)^{-1} A \tilde{r}.
 \end{aligned}$$

Dieser Term $\bar{D} = A^\top (ADA^\top)^{-1} A$ bildet mit D^* einige interessante Eigenschaften, die wir in folgendem Lemma zusammenfassen.

LEMMA 2.7. *Die induzierte Kovarianz-Matrix*

$$(9) \quad D^* = DA^\top (ADA^\top)^{-1} AD$$

besitzt eine symmetrische Pseudo-Inverse (PI), die gegeben wird durch

$$(10) \quad \bar{D} = A^\top (ADA^\top)^{-1} A.$$

D^ und \bar{D} bilden sogar ein Paar von PI. Zusätzlich besitzt dieses Paar folgende Eigenschaften:*

1. $D^* = D\bar{D}D$ $D^{-1}D^* = \bar{D}D$ $D^*D^{-1} = D\bar{D}$
2. $D\bar{D}D^* = D^*$ $D^*\bar{D}D = D^*$
3. *Mit der D -orthogonalen Projektion auf $\text{Im}A^\top$,*

$$\mathcal{P}^D := A^\top (ADA^\top)^{-1} AD,$$

gelten die Gleichungen $D\mathcal{P}^D = D^$ und $\mathcal{P}^D D^{-1} = \bar{D}$.*

4. $D^{-1} - \bar{D}$ *ist pseudoinvers zu $D - D^*$.*

BEWEIS. Die meisten Tatsachen sind durch einfaches Nachrechnen zu beweisen, so haben wir

$$\begin{aligned}
 D^{-1}D^* &= D^{-1}DA^\top (ADA^\top)^{-1}AD = \bar{D}D, \\
 \bar{D}D^* &= \bar{D}DD^{-1}D^* = D^{-1}D^*D^{-1}D^* = D^{-1}D^*,
 \end{aligned}$$

und aufgrund der Symmetrie gilt

$$D^*D^{-1} = D\bar{D} = D^*\bar{D}.$$

Damit haben wir Punkt (1) und (2) der Eigenschaftsliste abgearbeitet. Die Eigenschaft, daß \bar{D} und D^* jeweils pseudoinvers zueinander sind, ist nun auch leicht zu sehen:

$$\begin{aligned}
 \bar{D}D^*\bar{D} &= \bar{D}D^*D^{-1} = \bar{D}DD^{-1} = \bar{D} \quad \text{und} \\
 D^*\bar{D}D^* &= D\bar{D}D^* = DD^{-1}D^* = D^*.
 \end{aligned}$$

Punkt (3) ist trivial, die Formel für die Projektion stammt aus Proposition A.5.

Punkt (4) ergibt sich aus Punkt (1) und (2) mittels

$$\begin{aligned}
& (D^{-1} - \bar{D})(D - D^*)(D^{-1} - \bar{D}) \\
&= (I - \bar{D}D + \bar{D}D^* - D^{-1}D^*)(D^{-1} - \bar{D}) \\
&= (I - \bar{D}D)(D^{-1} - \bar{D}) = D^{-1} - \bar{D} \\
&\quad (D - D^*)(D^{-1} - \bar{D})(D - D^*) \\
&= (I - D^*D^{-1} - D\bar{D} + D^*\bar{D})(D - D^*) \\
&= (I - D^*D^{-1})(D - D^*) = D - D^*.
\end{aligned}$$

□

Diese Ergebnisse lassen sich wie folgt deuten: Auf dem von den Zeilen von A aufgespannten linearen Unterraum übernehmen D^* und \bar{D} die Rolle von D und D^{-1} . Dabei entsteht D^* durch eine D -orthogonale Projektion von A^\top , während \bar{D} durch eine D^{-1} -orthogonale Projektion entsteht.

Weitere Eigenschaften von D^* und \bar{D} finden sich in Proposition 2.11. Nun untersuchen wir, wie die induzierte Kovarianzmatrix mit dem Validierungsschritt verbunden ist. Dazu ist zunächst zu bemerken, daß die Validierung im Spezialfall $f(x) = Ax$ die Form

$$x^* = y - DA^\top(ADA^\top)^{-1}Ay$$

besitzt. Dieses läßt sich nun mit der induzierten Kovarianz umschreiben in

$$x^* = y - D^*D^{-1}y = y - D\bar{D}y.$$

Im affinen Fall $f(x) = Ax - b, b \neq 0$, erhält man aus dieser Betrachtung folgende Aussage.

PROPOSITION 2.8. *Mit den Bezeichnungen aus (9) und (10) ist das optimale Residuum r^* Fixpunkt der Abbildung*

$$\pi(r) = D\bar{D}r = D^*D^{-1}r.$$

BEWEIS. Die allgemeine Lösung des Validierungsproblems lautet nach (6)

$$\begin{aligned}
x^* &= y - DA^\top(ADA^\top)^{-1}(Ay - b) \\
&= y - DA^\top(ADA^\top)^{-1}(Ay - Ax^*).
\end{aligned}$$

Es gilt $Ax^* = b$, denn x^* erfüllt die Nebenbedingung. Also gilt dann für das Residuum $r^* = x^* - y$, daß

$$r^* = DA^\top(ADA^\top)^{-1}Ar^* = D\bar{D}r^*,$$

wobei die Darstellung $D\bar{D} = D^*D^{-1}$ schon aus Lemma 2.7.1 bekannt ist. □

D^* beschreibt den Anteil der Unsicherheit der Messung, der durch die Nebenbedingungen herausgerechnet wird. Das ist daran zu erkennen, daß die Verteilung der Validierung X die Kovarianz $D - D^*$ besitzt, was im Folgenden hergeleitet wird.

Für den validierten Wert x^* erhält man mit Proposition 2.8 die Beziehung

$$(11) \quad \begin{aligned} x^* &= y + D^* D^{-1}(x^* - y), & \text{also} \\ (I - D^* D^{-1})(x^* - y) &= 0. \end{aligned}$$

Betrachtet man Gleichung (11) als Beziehung zwischen den Zufallsvektoren X und Y der Validierungen bzw. der Messungen, so ist X eine Linearkombination von normalverteilten Werten, und deswegen wieder normalverteilt. Da wir für die Sensorstörungen in Annahme (A1) eine Verteilung mit Erwartungswert 0 gefordert haben, gilt dann für den Erwartungswert $Ef(Y) = E(AY - b) = 0$, also ist $EAY = AEY = b$. Damit erhalten wir dann für die Varianz der validierten Werte folgendes Resultat, wobei wir $R := (ADA^\top)^{-1}$ setzen:

$$(12) \quad \text{Cov } X = D - D^*, \quad \text{denn Cov } Y = D \text{ und}$$

$$\begin{aligned} \text{Cov } X &= \text{Cov}(Y - DA^\top R(AY - b)) \\ &= E(Y - DA^\top R(AY - b) - EY)(Y - DA^\top R(AY - b) - EY)^\top \\ &= E(Y - EY - DA^\top R(AY - b))(Y - EY - DA^\top R(AY - b))^\top \\ &= E(Y - EY)(Y - EY)^\top - E(DA^\top R(AY - b)(Y - EY)^\top) \\ &\quad - E(Y - EY)(DA^\top R(AY - b))^\top \\ &\quad + E(DA^\top R(AY - b))(DA^\top R(AY - b))^\top \\ &= \text{Cov } Y - DA^\top RE(AY - b)(Y - EY)^\top \\ &\quad - E(Y - EY)(AY - b)^\top RAD \\ &\quad + DA^\top RE(AY - b)(AY - b)^\top RAD \\ &= \text{Cov } Y - DA^\top RA \text{Cov } Y \\ &\quad - \text{Cov } Y A^\top RAD + DA^\top RA \text{Cov } Y A^\top RAD \\ &= D - D^* - D^* + D^* = D - D^*. \end{aligned}$$

Dieses Ergebnis zeigt uns, daß die Verteilung der Validierung X gerade durch die Differenz $D - D^*$ gegeben wird. Dies ist also die Antwort auf die eingangs gestellte Frage 3 nach der Verteilung von X , denn für den Erwartungswert EX gilt $EX = EY$.

Wir haben wiederum nur lineare Transformationen durchgeführt, damit beschreiben die zwei Momente EX und $\text{Cov } X$ die resultierende Normalverteilung von X vollständig. $D - D^*$ gibt also eine Genauigkeit nach Validierung an, wobei D^* ebenjener Anteil ist, der durch die Validierung herausgerechnet wird. Die Tatsache, daß $D - D^*$ immer

positiv semidefinit bleibt, ist eine Folgerung aus der Proposition 2.11 des nächsten Abschnitts.

D^* beschreibt den Einfluß des Modells auf die Validierung. Wie man aber diese Information insbesondere für große Systeme sinnvoll mitteilen kann, ist noch eine offene Frage, die hier nicht beantwortet wird.

Eine weitere Anwendung für die induzierte Kovarianzmatrix wird im Abschnitt 2.4 vorgestellt, in dem wir untersuchen, ob Messung und Validierung noch nahe genug beieinander liegen.

Wir demonstrieren die Berechnung der induzierten Kovarianzmatrix an folgendem Beispiel.

BEISPIEL 2.9 (Frischdampf-Massenfluß). Wir kehren zurück zum VDI-Beispiel aus 2.3 und bestimmen die induzierte Kovarianz-Matrix.

$$\begin{aligned}
 D^* &= DA^\top(ADA^\top)^{-1}AD \\
 &= DA^\top \begin{pmatrix} 0.4777 & -0.1493 & 0 \\ -0.1493 & 0.35 & -0.009098 \\ 0 & -0.009098 & 0.02004 \end{pmatrix}^{-1} AD \\
 &= 1000 \begin{pmatrix} 67 & 65 & -17 & -17 & -0.037 & -33 & -0.7 & -0.083 & -0.083 & -0.86 \\ 65 & 64 & -17 & -16 & -0.036 & -32 & -0.68 & -0.081 & -0.081 & -0.84 \\ -17 & -17 & 20 & 20 & -0.61 & -32 & -0.69 & -0.082 & -0.082 & -0.85 \\ -17 & -16 & 20 & 20 & -0.6 & -32 & -0.68 & -0.081 & -0.081 & -0.84 \\ -0.037 & -0.036 & -0.61 & -0.6 & 0.024 & 1.6 & 0.034 & 0.004 & 0.004 & 0.042 \\ -33 & -32 & -32 & -32 & 1.6 & 120 & 2.5 & 0.3 & 0.3 & 3.2 \\ -0.7 & -0.68 & -0.69 & -0.68 & 0.034 & 2.5 & 2.7 & 0.33 & 0.33 & -3.9 \\ -0.083 & -0.081 & -0.082 & -0.081 & 0.004 & 0.3 & 0.33 & 0.039 & 0.039 & -0.47 \\ -0.083 & -0.081 & -0.082 & -0.081 & 0.004 & 0.3 & 0.33 & 0.039 & 0.039 & -0.47 \\ -0.86 & -0.84 & -0.85 & -0.84 & 0.042 & 3.2 & -3.9 & -0.47 & -0.47 & 6.1 \end{pmatrix}
 \end{aligned}$$

Diese Matrix beschreibt die funktionale Abhängigkeit der einzelnen Messungen voneinander. Man kann nun hieran erkennen, daß die Meßwerte für Sensoren 1 und 2 (m_{S1} und m_{S2}), wie auch für Sensor 6 (m_{HK}) sehr stark angeglichen wurden, da sehr große Werte auf der Hauptdiagonalen auftreten. Außerdem erkennt man eine Paarung der Sensoren 1 und 2, 3 und 4, 8 und 9, die jeweils einen gleichwertigen Einfluß auf die Validierung ausüben. Insbesondere die Sensoren 8 und 9 (m_2 und m_3) liefern gleichwertige Informationen, aber das ist klar, da sie gleiche Varianzen besitzen und sie in den Gleichungen der Nebenbedingung völlig gleichwertig behandelt werden. Weiterhin ist die Diagonalstruktur von D nun völlig aufgelöst.

Betrachtet man die Korrelationskoeffizienten ϱ_{ij} der validierten Daten,

$$\varrho_{ij} = \frac{D_{ij}^*}{\sqrt{D_{ii}^* D_{jj}^*}},$$

so ergeben sich funktionale Zusammenhänge (Korrelation ist 1 oder -1) zwischen den Sensoren 1 und 2, 3 und 4, und außerdem zwischen den Sensoren 7,8 und 9 (m_1, m_2, m_3).

3.2. Weitere Eigenschaften der Validierung. Im Folgenden wollen wir weitere Eigenschaften der Validierung untersuchen, die Folgerungen aus den Eigenschaften aus Lemma 2.7 und der Darstellung aus Gleichung (11) sind.

Vergleicht man nämlich den Term $I - D^*D^{-1}$ mit der Formel aus Proposition A.7, so erhält man folgendes Ergebnis.

PROPOSITION 2.10 (Projektionseigenschaft). *Ist x^* eine Validierung der Messung y nach (6), so ist x^* eine Projektion von y auf den affinen Unterraum $\{u \in \mathbb{R}^n \mid Au = b\}$.*

BEWEIS. Nehme ein Urbild \tilde{x} von b . Dann gilt $A(x - \tilde{x}) = 0$ für alle x , die die Bedingung $f(x) = 0$ erfüllen, insbesondere für $x = x^*$. Verschiebe y nach $z = y - \tilde{x}$. Projiziere nun z auf $\ker A$:

$$\begin{aligned} & (I - D^*D^{-1})z \\ (13) \quad & = (I - DA^\top(ADA^\top)^{-1}A)(y - \tilde{x}) \\ & = y - \tilde{x} - DA^\top(ADA^\top)^{-1}(Ay - b) \end{aligned}$$

Da \tilde{x} die Rolle des Ursprungs übernommen hat, wird die Validierung x^* gegeben durch die Summe von \tilde{x} und der Projektion von z auf den Kern von A :

$$x^* = \tilde{x} + y - \tilde{x} - DA^\top(ADA^\top)^{-1}(Ay - b) = y - DA^\top(ADA^\top)^{-1}(Ay - b)$$

Dieses ist aber gerade die Validierung aus Lemma (2.2), sie wird also durch eine Projektion gegeben. \square

Diese Situation wird in Abbildung 2 verdeutlicht: Die Validierung x^* entsteht durch Projektion der Messung y auf das Modell $Ax = b$.

Aussagen über die Eigenwerte der induzierten Kovarianzmatrix D^* und ihrer Pseudoinversen \bar{D} werden in den folgenden beiden Propositionen hergeleitet. Diese Aussagen erweitern den Eigenschaftenkatalog aus Lemma 2.7.

PROPOSITION 2.11. *Für die Eigenwerte $\lambda^* \neq 0$ von D^* gilt:*

$$\lambda_{\min}(D) \leq \lambda^* \leq \lambda_{\max}(D).$$

BEWEIS. Wir setzen

$$E = \{x \in \mathbb{R}^n \mid x^\top Dx = 1\}$$

und

$$F = \{x \in \mathbb{R}^n \mid x^\top D^*x = 1\}.$$

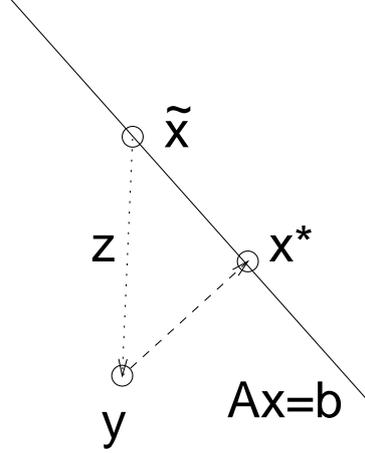


ABBILDUNG 2. Zum Beweis von Proposition 2.10.

Dann haben wir die Hauptachsenbeziehungen für den Ellipsoiden E bzw. für den Zylinder F (D^* ist nur semidefinit):

$$\begin{aligned}\lambda_{\max}^{-1}(D) &= \min_{x \in E} \|x\|^2, & \lambda_{\min}^{-1}(D) &= \max_{x \in E} \|x\|^2, \\ \lambda_{\max}^{-1}(D^*) &= \min_{x \in F} \|x\|^2, & \lambda_{\min}(D^*) &= 0,\end{aligned}$$

denn $D^{(*)}x = \lambda x \implies x^\top D^{(*)}x = \lambda \|x\|^2$. Bei der verwendeten Norm $\|\cdot\|$ handelt es sich um die euklidische Standardnorm des \mathbb{R}^n . Durch eine schiefe orthogonale Zerlegung erhalten wir die folgende Ungleichung:

$$\begin{aligned}x^\top D x &= x^\top \overbrace{DA^\top(ADA^\top)^{-1}AD}^{D^*} x + x^\top D(I - A^\top(ADA^\top)^{-1}AD)x \\ &= (D^{1/2}x)^\top \underbrace{(D^{1/2}A^\top(ADA^\top)^{-1}AD^\top/2)}_{\text{positiv semidefinit}} D^{1/2}x + \\ &\quad (D^{1/2}x)^\top \underbrace{(I - D^{1/2}A^\top(ADA^\top)^{-1}AD^\top/2)}_{\text{positiv semidefinit}} D^{1/2}x \\ &\geq x^\top DA^\top(ADA^\top)^{-1}ADx \\ &= x^\top D^* x\end{aligned}$$

Die benutzte Semidefinitheit folgt aus der Tatsache, daß die beiden oben verwendeten Matrizen symmetrisch und idempotent sind, also nur Eigenwerte 0 und 1 besitzen, wie in Proposition A.4 gezeigt wird. Zu $x \in F$ existiert daher $|\alpha| \leq 1$ mit $\alpha x \in E$, denn für $x \in F$ und $\alpha x \in E$ ergibt sich damit

$$1 = x^\top D^* x = \alpha^2 x^\top D x \implies \alpha^2 \leq 1.$$

Damit folgt

$$\begin{aligned} \forall x \in F : \quad \|x\|^2 &\geq \alpha^2 \|x\|^2 = \|\underbrace{\alpha x}_{\in E}\|^2 \geq \lambda_{\max}^{-1}(D) \\ &\implies \lambda_{\max}(D^*) \leq \lambda_{\max}(D) \end{aligned}$$

Es sei $l \in F$ Eigenvektor zum Eigenwert $\lambda \neq 0$ von D^* . Dann ist $1 = l^\top D^* l = \lambda \|l\|^2$. Die Orthogonalitätsbedingung für Eigenvektoren von symmetrischen Matrizen gibt uns die Aussage $l \perp \ker D^*$, da D^* einen nichttrivialen Kern hat, also Eigenvektoren zum Eigenwert 0 besitzt. Damit erhalten wir, daß ein $y \in \mathbb{R}^m \setminus \{0\}$ mit $l = DA^\top y$ existiert, denn

$$(\ker D^*)^\perp = \text{Im} D^* = \text{Im} DA^\top.$$

Setze $t = \lambda A^\top y$. Dann ist $t \in E \cap F$, denn

$$\begin{aligned} t^\top D t &= \lambda^2 y^\top A D A^\top y = \lambda^2 l^\top D^{-1} l = l^\top \overbrace{D^* D^{-1} D^*}^{D \bar{D} D^*} l = l^\top D^* l = 1, \\ t^\top D^* t &= \lambda^2 y^\top A D^* A^\top y = \lambda^2 l^\top \bar{D} l = l^\top \underbrace{D^* \bar{D} D^*}_{D \bar{D} D^*} l = l^\top D^* l = 1. \end{aligned}$$

Außerdem gilt: $l \perp l - t$, denn

$$l^\top (l - t) = l^\top l - l^\top t = \frac{1}{\lambda} - \lambda (y^\top A D A^\top y) = \frac{1}{\lambda} - \frac{1}{\lambda} t^\top D t = 0.$$

Also bilden die Punkte $0, t, l$ ein rechtwinkliges Dreieck mit t als Hypotenuse:

$$\lambda_{\min}^{-1}(D) \geq \|t\|^2 \geq \|l\|^2 = \lambda^{-1}.$$

Es ist $\lambda \neq 0$ ein beliebiger Eigenwert von D^* , also ist

$$\lambda_{\min}^{-1}(D) \geq \lambda_{\min}^{-1}(D^*).$$

damit haben wir die beiden Ungleichungen der Abschätzung aus der Proposition 2.11 bewiesen. \square

Die Beweis-Situation für den zweidimensionalen Fall mit

$$D = \begin{pmatrix} 0.75 & 0 \\ 0 & 0.075 \end{pmatrix}$$

und $A = (-1, 2)$ wird in Abbildung 3 veranschaulicht. Um die analoge Aussage zu Proposition 2.11 für D^{-1} und \bar{D} zu zeigen, greift man auf Proposition A.8 zurück, die die D -Projektionen mit den D^{-1} -Projektionen verbindet: Setzen wir $A = BD^{-1}$, so ergibt sich für \bar{D} die Darstellung

$$\begin{aligned} \bar{D} &= A^\top (A D A^\top)^{-1} A D = (B D^{-1})^\top ((B D^{-1}) D (B D^{-1})^\top)^{-1} (B D^{-1}) \\ &= D^{-1} B^\top (B D^{-1} B^\top)^{-1} B D. \end{aligned}$$

Aus dieser Darstellung erhalten wir mit Proposition 2.11

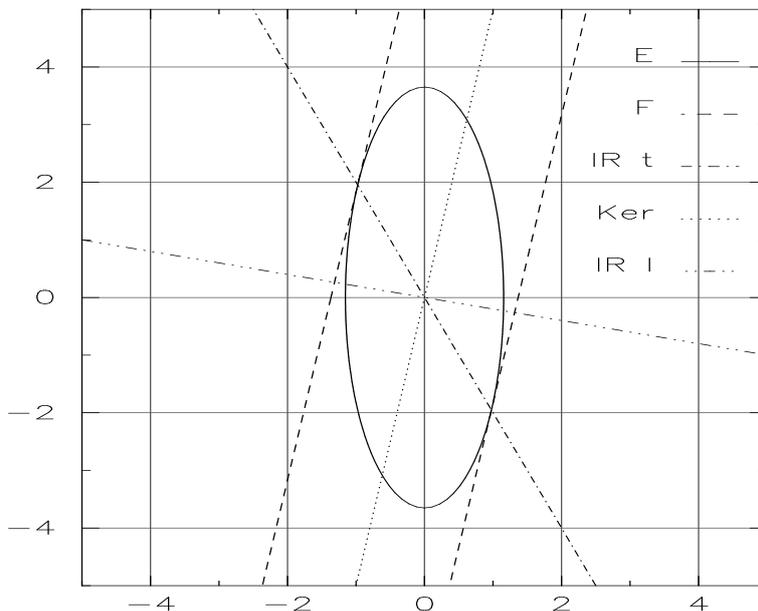


ABBILDUNG 3. Zum Beweis von Proposition 2.11.

PROPOSITION 2.12. Für die Eigenwerte $\bar{\lambda} \neq 0$ von \bar{D} gilt:

$$\lambda_{\min}(D^{-1}) \leq \bar{\lambda} \leq \lambda_{\max}(D^{-1}).$$

Die beiden Propositionen 2.11 und 2.12 liefern die Aussage, daß der Abstand des Mittelpunktes eines Ellipsoids zu einem engangliegenden Zylinder durch die Eigenwerte der Matrix der zugehörigen Bilinearform eingeschränkt wird. Im Beweis werden zwei Normen miteinander verknüpft, und zwar die euklidische Standardnorm und eine gewichtete euklidische Norm. Damit ist dann auch ersichtlich, daß für die Gewichtung $D = I$ die Aussagen trivial sind.

Für uns ist wichtig, daß der Term $D - D^*(A)$ für alle Matrizen $A \neq 0$ positiv semidefinit bleibt. Ist nämlich $z^\top(D - D^*)z < 0$, dann folgt daraus, daß $z^\top D z < z^\top D^* z$ gilt, was aber im Widerspruch zu Proposition 2.11 steht.

4. Güte-Test der Messung

In diesem Abschnitt suchen wir ein Kriterium, mit dem man beurteilen kann, ob die Messung und der validierte Wert noch hinreichend nahe beieinander liegen. Mit diesem Kriterium wird abgesichert, daß Aussagen, die auf validierten Werten basieren, möglichst sinnvoll sind. Dazu bietet sich ein Hypothesen-Test nach Definition B.9 an, der mit der induzierten Kovarianz aus Proposition 2.6 arbeitet. Dieser Test beruht auf der folgenden Idee, die in Abbildung 4 veranschaulicht wird: Wir wollen einen Konfidenzbereich um die aktuelle Messung bestimmen, in

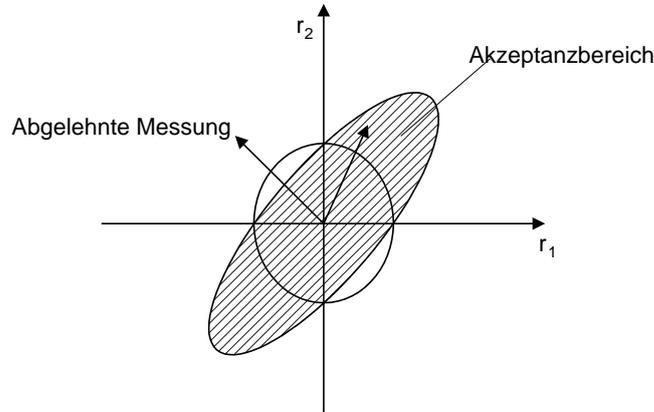


ABBILDUNG 4. Konfidenzellipsoid

dem sich der wahre Wert mit vorgegebener Wahrscheinlichkeit befindet. Dieser Bereich ist im Fall einer Normalverteilung ein Ellipsoid, der durch die induzierte Kovarianz-Matrix D^* aufgespannt wird.

Liegt nun das Residuum der Validierung in diesem Konfidenzellipsoid zum vorher festgelegten Testniveau $1 - \alpha$, mit anderen Worten, liegt die Validierung in einer hier ellipsenförmigen Umgebung der Messung, deren Größe durch α und deren Form durch D^* bestimmt wird, so ist die Messung zulässig; dies gibt uns die Nullhypothese für einen Test. Das Hauptresultat dieses Abschnittes ist, daß dieser Ellipsoid gerade durch die Kostenfunktion $J(r)$ gegeben wird.

Im Bild 4 liegt die Messung gerade im Nullpunkt, damit ist die Validierung gleich dem Residuum und wird im Bild durch einen Pfeil angedeutet.

Für die Verteilung der Norm der Residuen gilt folgender Satz. Die dabei verwendeten Bezeichnungen werden in Anhang B erläutert. Diese allgemeine Aussage wird danach auf unsere Situation angewendet.

SATZ 2.13. *Es sei $X \stackrel{\mathcal{L}}{\sim} \mathcal{N}_n(\mu, \Sigma)$ eine n -dimensionale normalverteilte Zufallsvariable mit einem Erwartungswert μ und einer Kovarianz-Matrix Σ vom Rang m . Für eine generalisierte Inverse $\bar{\Sigma}$ von Σ gilt:*

$$(14) \quad (X - \mu)^{\top} \bar{\Sigma} (X - \mu) \stackrel{\mathcal{L}}{\sim} \chi_m^2$$

BEWEIS. Wir standardisieren den Zufallsvektor X . Im eindimensionalen Fall wird eine Zufallsvariable standardisiert, indem man den Erwartungswert abzieht und durch die Wurzel aus der Varianz teilt. Die dabei entstehende Zufallsvariable $Y = \frac{X - EX}{\sqrt{\text{Cov } X}}$ hat dann Erwartungswert 0 und Varianz 1.

Im höherdimensionalen Fall suchen wir eine lineare Transformation Φ , so daß $\Phi(X)$ Erwartungswert 0 und Kovarianz I besitzt und damit

für den Fall, daß X normalverteilt ist, ein Vektor aus unabhängigen eindimensionalen Standardnormalverteilungen ist.

Aufgrund der Tatsache, daß Σ als Kovarianzmatrix positiv semidefinit ist, existiert eine reguläre Matrix $S \in \mathbb{R}^{n \times n}$ mit

$$(15) \quad \Sigma = S \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} S^\top.$$

Dieses Resultat ist eine Folgerung aus dem Trägheitssatz für symmetrische Bilinearformen über \mathbb{R} , es läßt sich aber auch direkt beweisen, denn durch durch kombinierte Zeilen-/Spalten-Operationen angewandt auf Σ erhält man eine Diagonalmatrix mit nichtnegativen Einträgen, $\Sigma = U \text{diag}(d_1, \dots, d_n) U^\top$, dann lassen sich die zu positiven Einträgen gehörende Spaltenvektoren von U so normieren, daß die entsprechenden Diagonaleinträge 1 sind. Durch Umsortierung erhält man dann die Darstellung aus (15).

Definiere C als die ersten m Spalten von S . Dann ist $\Sigma = CC^\top$.

Sei $\bar{\Sigma}$ eine Pseudo-Inverse von Σ . Dann gilt nach Proposition A.10

$$(16) \quad \begin{aligned} \Sigma \bar{\Sigma} \Sigma &= \Sigma \\ CC^\top \bar{\Sigma} CC^\top &= CC^\top. \end{aligned}$$

Da C vollen Rang besitzt, ist C injektiv ($Cx \neq Cy \implies x \neq y$), und damit links-kürzbar. C^\top ist surjektiv, also rechts-kürzbar. Dann erhalten wir aus (16) die Darstellung

$$C^\top \bar{\Sigma} C = I.$$

Es gilt $X = CZ + \mu$ ⁵ mit $Z \stackrel{\mathcal{L}}{\sim} \mathcal{N}_m(0, I)$, $\text{rang } C = m$, denn $EX = \mu$ und $\text{Cov}(X) = \Sigma = CC^\top$. Weiter erhalten wir

$$\begin{aligned} (X - \mu)^\top \bar{\Sigma} (X - \mu) &= (CZ)^\top \bar{\Sigma} (CZ) = Z^\top Z \\ &= \sum_{j=1}^m z_j^2, \quad z_j \stackrel{\mathcal{L}}{\sim} \mathcal{N}_1(0, 1) \text{ iid}^6. \end{aligned}$$

$(X - \mu)^\top \bar{\Sigma} (X - \mu)$ gehorcht nach Definition einer χ^2 -Verteilung, denn es ist eine Summe von quadrierten standardnormalverteilten Zufallsvariablen. \square

Da wir nun die Verteilung der Kosten kennen, läßt sich sehr einfach eine Teststatistik (vgl. B.9) angeben, die einen Güte-Test realisiert.

PROPOSITION 2.14. *Der Güte-Test zum Niveau $1 - \alpha$ der Validierung x^* zur Messung y wird durch folgende Teststatistik gegeben:*

$$\text{Ablehnung der Messung } y \iff f(y)^\top (ADA^\top)^{-1} f(y) > \varphi_m(\alpha),$$

⁵Die Gleichheit von Zufallsvariablen ist hier eine fast sichere Gleichheit.

⁶„iid“: independent and identically distributed, unabhängig und gleichverteilt.

wobei $\varphi_m(\cdot)$ das obere α -Quantil der χ^2 -Verteilung mit m Freiheitsgraden bezeichne und m die Anzahl der Modellgleichungen ist. Dabei hat die Nebenbedingung f die Form $f(x) = Ax - b$.

Meist wählt man $\alpha = 5\%$.

BEWEIS. Die Kostenfunktion im optimalen Residuum lautet – wie wir schon in (8) gesehen haben –

$$\begin{aligned} J(\tilde{r}) &= \tilde{r}^\top A^\top (ADA^\top)^{-1} A\tilde{r} \\ &= \tilde{r}^\top \bar{D}\tilde{r} \end{aligned}$$

Nun haben wir für $A\tilde{r}$ die Gleichung

$$A\tilde{r} = A(x^* - y) = (Ax^* - b) - (Ay - b) = f(x^*) - f(y) = -f(y),$$

da für die Validierung x^* die Nebenbedingung $f(x) = 0$ erfüllt ist. Außerdem ist die Kovarianz der Residuen gegeben durch D^* . Damit liefert das Theorem 2.13 mit $X = \tilde{r}$, $\mu = 0$ und $\Sigma = D^*$ die gewünschte Verteilungsaussage, denn die Residuen sind nach Proposition 2.6 normalverteilt.

Der Rang von D^* entspricht gerade der Anzahl der Modellgleichungen, da A als regulär vorausgesetzt ist. Damit haben wir eine Teststatistik konstruiert, die den Fehler zwischen Messung und Validierung beurteilt. \square

Die Quantile der χ^2 -Verteilung finden sich tabelliert in [Kre91], Tabelle III, hier allerdings als untere Quantile, oder in [Bro97], Kapitel 21.16.

Die folgenden Bemerkungen erweitern die Interpretation der Ergebnisse eines Tests nach Proposition 2.14.

1. Der Test 2.14 verwirft die gesamte Messung, ohne einen fehlerhaften Sensor zu identifizieren. Dabei ist das Aufspüren von fehlerhaften Sensoren gerade ein Ziel der Datenvalidierung.
2. Man kann folgende Arten von Meßfehlern unterscheiden: grobe Fehler, systematische Fehler und statistische Fehler. Mit der Validierung wird versucht, statistische Fehler auszulöschen. Der Test dient dazu, grobe Fehler aufzufinden. Systematische Fehler, die z.B. durch Wahl falscher Modellgleichungen oder durch abgedriftete Sensoren entstehen, haben wir bei unseren Betrachtungen ausgeschlossen, denn damit ist die Voraussetzung A2 verletzt. Falls aber solche Fehlerquellen vorliegen, ist die Wahrscheinlichkeit, mit der der Test eine Messung ablehnt, umso größer.
3. Eine Skalierung der Kovarianzmatrix wirkt sich linear auf den Güte-Test aus.
4. Um den Annahmehbereich des Verfahrens zu untersuchen, bietet sich eine Simulationsstudie an, d.h. man berechnet zu einem vorgegebenen Gitter aus möglichen Messungen oder aus künstlich

verrauschten Daten den Wert der Teststatistik und bestimmt damit einen ungefähren Annahmehereich. Ein Test soll möglichst unsinnige Messungen ablehnen, also einen kleinen Annahmehereich besitzen, allerdings soll auch nicht zu oft Alarm geschlagen werden, wenn der Test eine Messung ablehnt.

5. [Str95] schlägt einen Test vor, der nur die Projektionen auf die Einheitsvektoren berücksichtigt (d.h. die Diagonalelemente von D^* gegen ein Quantil der Normalverteilung getestet). Dieser Test ist i.A. nicht konservativ, d.h. er akzeptiert auch Werte, die der Test 2.14 abgelehnt hätte. In Abbildung 4 wird dieser Test durch einen Kreis angedeutet, der die Achsenabschnitte mit der Ellipse teilt, aber ein unterschiedlichen Annahmehereich besitzt. Aber solch ein Test ermöglicht es gerade, fehlerhafte Sensoren aufzuspüren, allerdings unter der zusätzlichen Voraussetzung, daß die einzelnen Störungen unabhängig voneinander sind.
6. Bei einem 95%-Testniveau ($\alpha = 0,05$) erwartet man folglich eine Fehlerquote von 5%. Eine Ablehnung der Messung kann auf schlecht gewählte Varianzen, Sensorenfehler, Modellfehler oder fehlende Voraussetzungen (stationärer Zustand) hindeuten. Andererseits läßt sich nicht von der Verträglichkeit des Modells mit den Daten auf die Gültigkeit des Modells zurückschließen.
7. Da der Erwartungswert der χ^2 -Verteilung gleich der Anzahl der Freiheitsgrade m ist, also mit der Anzahl der Nebenbedingungen wächst, kann man zu einer skalierten Verteilung übergehen. Dann gilt nach Proposition B.8 $\frac{\chi_m^2}{m} \stackrel{\mathcal{L}}{\sim} F_{m,\infty}$. Dabei ist $F_{m,\infty}$ eine Fishersche F -Verteilung mit m Zählerfreiheitsgraden und Nennerfreiheitsgrad ∞ . Diese Verteilung hat dann Erwartungswert 1 und läßt sich dann dazu einsetzen, Modelle mit unterschiedlichen Freiheitsgraden miteinander zu vergleichen. In Kapitel 5 wird die Modellauswahl eingehender untersucht.
8. Ein Vorschlag für praktisches Vorgehen besteht darin, daß man den χ^2 -Test als Güte-Bewertung benutzt. Bei Ablehnung einer Messung nimmt man an, daß nur ein Sensor gestört ist, so daß man einen Test der einzelnen Varianzen durchführen kann. Mit einer anschließenden Korrelationsanalyse kann man eine Gruppe von abhängigen Meßstellen isolieren, in der sich fehlerhafte Sensoren verbergen können. Bei Auffinden von Unregelmäßigkeiten ist möglicherweise eine Anpassung der Varianzen oder Elimination kritischer Gleichungen aus dem Modell nötig. Eine Fehlerstatistik, in der die auffälligen Sensoren tabelliert werden und die regelmäßig ausgewertet wird, gibt weiteren Aufschluß über „gefährdete Sensoren“.

Nun berechnen wir die Teststatistik für unser Leitbeispiel und wenden einige der Bemerkungen an, um die Unstimmigkeiten in der Messung aufzuspüren.

BEISPIEL 2.15. Das Frischdampf-Modell aus Beispiel 2.3 hat drei Gleichungen mit 10 Variablen. Das obere 5%-Quantil der χ^2 -Verteilung mit 3 Freiheitsgraden ist 7.815, der Test ermittelt Kosten in Höhe von 25.38. Die Messung ist also mit erheblichen Fehlern belastet. Ein Test der eindimensionalen Randverteilungen auf Normalverteilung, wie in Punkt 5 der Bemerkung vorgeschlagen, läßt die Messungen für m_{K1} und m_{K2} fragwürdig erscheinen. Erst eine Aufweichung der Varianzen dieser beiden Messungen von 0.16 auf 0.7 reduziert die Kosten auf einen annehmbaren Wert.

5. Zusammenfassung

Wir haben nun die Diskussion des linearen Falls abgeschlossen und fassen jetzt noch einmal die Hauptresultate zusammen: Für ein affin lineares Modell der Form

$$f(x) = Ax - b = 0$$

liefert uns Gleichung (6) eine affin lineare Transformation der Messung y , die uns in einem Schritt zur Validierung x^* führt. Genauere Betrachtung zeigt, daß es sich bei dieser Transformation um eine *Projektion* auf den affinen Unterraum $\{x \in \mathbb{R}^n \mid f(x) = 0\}$ handelt, was wir in Proposition 2.10 bewiesen haben. Setzen wir weiterhin eine Normalverteilung der Sensorstörungen voraus, die jeweils Erwartungswert 0 und bekannte Varianz besitzen, so läßt sich mithilfe des vorgestellten Tests beurteilen, ob wir uns mit der Validierung zu weit von der Messung entfernt haben.

KAPITEL 3

Nichtlineare Nebenbedingungen

„Keep in mind, however, that in nonlinear programming, just as there is no foolproof method, there is no foolproof advice.“
D. Bertsekas

Wir wenden uns nun nichtlinearen Modellgleichungen zu, wobei das Hauptproblem darin besteht, geeignete Verallgemeinerungen des Validierungsschrittes zu finden.

Nichtlineare Nebenbedingungen erschweren uns die Suche nach geeigneten Lösungen des Problems 1.2

$$\begin{aligned} &\text{Minimiere} && J(x) \\ &\text{unter} && f(x) = 0, \end{aligned}$$

wobei $J(x) = \|x - x^*\|_{D^{-1}}^2$ und $f \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$ gilt. Bedingungen für die Existenz und Eindeutigkeit von Lösungen lassen sich aus der Methode der Lagrangeschen Multiplikatoren ableiten, deren Hauptresultate wir im nächsten Abschnitt zusammenfassen. Allerdings geben diese Bedingungen keine Aussagen über die Lage der Lösungen. Also müssen sie so umgeformt werden, daß sich explizite Suchverfahren ergeben, die wir dann in dem folgenden Abschnitt 3.2 untersuchen.

Außerdem betrachten wir in Abschnitt 3.3, wie sich der Güte-Test aus Abschnitt 2.4 bei nichtlinearen Nebenbedingungen benutzen läßt.

Desweiteren geben in Abschnitt 3.4 ein Verfahren an, mit dessen Hilfe man Schätzwerte ermitteln kann, wenn keine direkten Messungen vorhanden sind.

1. Methode der Lagrangeschen Multiplikatoren

Als Vorbereitung auf den nichtlinearen Fall betrachten wir die Methode der Lagrangeschen Multiplikatoren, die es uns erlaubt, eine Extremwertaufgabe mit vorgegebenen Nebenbedingungen in eine Extremwertaufgabe ohne Nebenbedingungen umzuwandeln, die aber eine höhere Dimension als das Ausgangsproblem besitzt. Hier werden die Hauptergebnisse präsentiert, die den Aussagen über Extremwerte der Analysis entsprechen.

Das Minimierungsproblem unter Nebenbedingung wird gegeben durch

$$(17) \quad \begin{aligned} &\text{Minimiere} && F(r) \\ &\text{unter} && G(r) = 0, \end{aligned}$$

wobei die *Zielfunktion* $F : \mathbb{R}^n \rightarrow \mathbb{R}$ und die *Nebenbedingung* $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $m < n$ gegeben sind. Wir fordern, daß beide Funktionen mindestens \mathcal{C}^2 -glatt sind, damit die folgenden Existenz- und Eindeutigkeitsaussagen zutreffen.

DEFINITION 3.1. Die zu dem Problem (17) gehörige *Lagrangesche Funktion* $L(r, \lambda) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ wird gegeben durch

$$(18) \quad L(r, \lambda) = F(r) - 2G^\top(r)\lambda.$$

$\lambda \in \mathbb{R}^m$ heißt *Vektor der Lagrangeschen Multiplikatoren*.

BEMERKUNG. Die hier verwendete Lagrangesche Funktion ist so gewählt, daß die anfallenden Konstanten auf 1 normiert werden, wenn die Zielfunktion die Form $F(r) = J(r) = r^\top P r$ (mit $P \succ 0$) besitzt.

DEFINITION 3.2.

- Wir bezeichnen die Ableitung von G mit

$$\nabla G = \begin{pmatrix} \frac{\partial G_1}{\partial r_1} & \cdots & \frac{\partial G_1}{\partial r_n} \\ \vdots & \frac{\partial G_i}{\partial r_j} & \vdots \\ \frac{\partial G_m}{\partial r_1} & \cdots & \frac{\partial G_m}{\partial r_n} \end{pmatrix}.$$

Die Ableitung von G im Punkt r wird mit $\nabla G|_r$ bezeichnet.

- Ein Punkt $r \in \mathbb{R}^n$ heißt *regulärer Punkt* von G , wenn die Ableitung im Punkt r vollen Rang besitzt:

$$\text{rang } \nabla G|_r = m.$$

Wenn wir voraussetzen, daß die Nebenbedingung G als lokale Gleichung einer Untermannigfaltigkeit \mathcal{M} gegeben wird, was wir in Abschnitt 1.2 für die Nebenbedingung f der Validierung gefordert haben, so ist G in $r \in \mathcal{M}$ – und damit in einer ganzen Umgebung von r – regulär.

Notwendige Bedingungen für eine Lösung von (17) liefert uns

PROPOSITION 3.3. *Es sei r^* eine Lösung des Minimierungsproblems (17). Ist r^* ein regulärer Punkt von G , so gibt es einen eindeutigen Lagrangeschen Multiplikator λ^* , so daß gilt*

$$(19) \quad \nabla_r L(r^*, \lambda^*) = 0,$$

$$(20) \quad \nabla_\lambda L(r^*, \lambda^*) = 0,$$

wobei ∇_r bzw. ∇_λ für die partielle Ableitung nach den Komponenten von r bzw. λ steht.

Das Paar (x^*, λ^*) heißt *Lagrange-Paar*.

Zunächst wollen wir dieses Resultat geometrisch motivieren. Sei r^* ein Minimum von F unter der Nebenbedingung $G(r) = 0$. Verschwindet die Ableitung ∇F in r^* , so ist (19) für $\lambda = 0$ erfüllt. Dieser Fall ist aber wegen der Regularität von r^* ausgeschlossen. Setze $c = F(r^*)$.

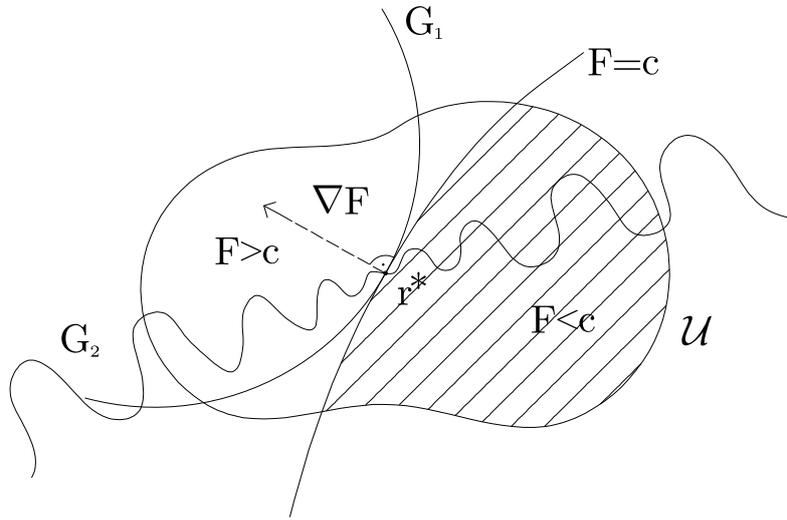


ABBILDUNG 5. F und G_1 berühren sich im kritischen Punkt r^* .

Dann trennt die Niveaulinie $F^{-1}(c)$ eine Umgebung \mathcal{U} von r^* in zwei disjunkte Teilmengen

$$\mathcal{U}^+ = \{r \in \mathcal{U} \mid F(r) > c\} \quad \text{und} \\ \mathcal{U}^- = \{r \in \mathcal{U} \mid F(r) < c\}.$$

Ist \mathcal{U} so gewählt, daß F auf \mathcal{U} in dem Punkt r^* ein Minimum relativ zu einer Kurve φ annimmt, die durch die Nebenbedingung $G(r) = 0$ gegeben wird, so müssen ∇F und ∇G linear abhängig sein, anderenfalls müßte φ das c -Niveau von F überqueren. Dann gilt $\text{spur} \varphi \cap \mathcal{U}^- \neq \emptyset$, was aber im Widerspruch zur Minimaleigenschaft von r^* steht. Also steht der Gradient von F senkrecht auf der der Kurve φ im Punkt r^* . Diese Situation wird in Abbildung 5 veranschaulicht.

BEWEIS (von Proposition 3.3). (Siehe [Ber95, Proposition 3.1.1]) Zum Beweis partitionieren wir den Vektor r mit eventueller Umordnung der Koordinaten von r in $r = (x, y) \in \mathbb{R}^{n-m} \times \mathbb{R}^m$, so daß $\nabla_y G|_{r^*}$ (der Gradient von G bezüglich y ausgewertet in r^*) invertierbar ist. Nach dem Satz über implizite Funktionen ([GF178] Satz 6.1) läßt sich y durch x mittels einer eindeutigen stetig differenzierbaren Funktion $\varphi : \mathcal{U} \rightarrow \mathbb{R}^m$ ausdrücken, wobei \mathcal{U} eine Umgebung von x ist. D.h. es gilt

$$y^* = \varphi(x^*), \quad G(x, \varphi(x)) = 0 \text{ auf } \mathcal{U}$$

und insbesondere ist

$$\nabla \varphi|_x = -(\nabla_y G|_{(x, \varphi(x))})^{-1} \nabla_x G|_{(x, \varphi(x))}.$$

Dann ist x^* ein nebenbedingungsfreies Minimum der reduzierten Kostenfunktion $F^*(x) = F(x, \varphi(x))$, auf die sich nun die allgemeine Theorie zur Lage von Extremstellen von Funktionen anwenden läßt. Die Ableitung von F^* muß also im Punkt x^* verschwinden.

$$(21) \quad \begin{aligned} 0 = \nabla F|_{x^*} &= \nabla_x F|_{r^*} + \nabla_y F|_{r^*} \nabla \varphi|_{x^*} \\ &= \nabla_x F|_{r^*} - \nabla_y F|_{r^*} (\nabla_y G|_{r^*})^{-1} \nabla_x G|_{r^*} \end{aligned}$$

Setzen wir nun

$$(22) \quad \lambda^* := \frac{1}{2} (\nabla_y F|_{r^*} (\nabla_y G|_{r^*})^{-1})^\top,$$

so erhalten wir durch Umformung von (22)

$$\nabla_y F|_{r^*} - 2\lambda^{*\top} \nabla_y G|_{r^*} = 0$$

und die Umformung von (21)

$$\nabla_x F|_{r^*} - 2\lambda^{*\top} \nabla_x G|_{r^*} = 0.$$

Diese beiden Gleichungen ergeben zusammen (19). Die Gleichung (20) ist eine andere Schreibweise für die Einhaltung der Nebenbedingung. Nun beweisen wir die Eindeutigkeit des Vektors der Lagrangeschen Multiplikatoren. Seien dazu (r^*, λ^*) und (r^*, μ^*) zwei Lagrange-Paare zur Lösung r^* mit $\lambda^* \neq \mu^*$. Dann gilt nach (19) in r^*

$$\nabla F + \lambda^{*\top} \nabla G = 0 = \nabla F + \mu^{*\top} \nabla G,$$

also ist dann $(\lambda^* - \mu^*)^\top \nabla G = 0$ eine nichttriviale Darstellung der Null aus den Zeilen von ∇G . Aufgrund der Regularität von ∇G ist dies aber nicht möglich, also ist $\lambda^* = \mu^*$. \square

Hinreichende Bedingungen für die Lösung von (17) liefert uns die folgende Proposition.

PROPOSITION 3.4. *Ist G zweifach stetig differenzierbar und erfüllt das Paar (r^*, λ^*) die Bedingungen*

$$\begin{aligned} \nabla_r L(r^*, \lambda^*) &= 0, & \nabla_\lambda L(r^*, \lambda^*) &= 0, \\ z^\top \nabla_{rr}^2 L(r^*, \lambda^*) z &> 0 & \forall z \neq 0 \text{ mit } \nabla G|_{x^*} z &= 0, \end{aligned}$$

so ist r^ ein lokales Minimum von F unter der Bedingung $G(r) = 0$. Dabei ist $\nabla_{rr}^2 L$ die zweite Ableitung (Hesse-Matrix) nach r der Lagrange-Funktion L , wobei λ als konstant angesehen wird.*

BEWEIS. Siehe [Ber95], Proposition 3.2.1.

Auch diese Proposition ist ein Analogon zu Aussagen der klassischen Analysis. Ist x nämlich eine Extremstelle von $f : \mathbb{R} \rightarrow \mathbb{R}$, so ist notwendigerweise $f'(x) = 0$. Hinreichend für die Existenz eines Minimums im Punkt x ist die Aussage $f''(x) > 0$. Diese Tatsachen lassen sich in den soeben vorgestellten Propositionen wiederfinden.

Eine weitere Aussage über Lagrange-Funktionen findet sich in Proposition 4.3.

1.1. Beispiel. Wir wollen die Methode der Lagrangeschen Multiplikatoren in einem Beispiel auf den linearen Fall anwenden und damit das quadratische Programm lösen. Dafür benötigen wir das folgende Lemma.

LEMMA 3.5 (Matrixinversionslemma). *Es sei $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ eine invertierbare Matrix. Dann gilt, falls die verwendeten Inversen existieren:*

1.

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BQCA^{-1} & -A^{-1}BQ \\ -QCA^{-1} & Q \end{pmatrix}$$

$$\text{mit } Q = (D - CA^{-1}B)^{-1},$$

2.

$$M^{-1} = \begin{pmatrix} R & -RBD^{-1} \\ -D^{-1}CR & D^{-1} + D^{-1}CRBD^{-1} \end{pmatrix}$$

$$\text{mit } R = (A - BD^{-1}C)^{-1}.$$

Diese beiden Darstellungen liefern die Sherman-Morrison-Woodbury-Formel

$$(23) \quad \begin{aligned} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} \\ = (A - BD^{-1}C)^{-1}. \end{aligned}$$

BEWEIS. 1. Durch elementare Umformungen erhält man

$$\begin{aligned} & \begin{pmatrix} A & B \\ C & D \end{pmatrix} \\ &= \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ C & I \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I & B \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \end{aligned}$$

Dieses Produkt läßt sich einfach invertieren. Mit $Q^{-1} := D - CA^{-1}B$ ergibt sich dann

$$\begin{aligned} \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} &= \begin{pmatrix} A^{-1} & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & -B \\ 0 & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} I & 0 \\ -C & I \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} A^{-1} + A^{-1}BQCA^{-1} & -A^{-1}BQ \\ -QCA^{-1} & Q \end{pmatrix}. \end{aligned}$$

2. Die zweite Darstellung erzeugt man mit

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & B \\ 0 & D \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ C & D \end{pmatrix}$$

Die Gleichung (23) ergibt sich aus der linken oberen Teilmatrix von M^{-1} in beiden Darstellungen. \square

BEISPIEL 3.6. (Lösung des quadratischen Programms mittels Lagrange-Multiplikatoren) Das quadratische Programm (QP) wird gegeben durch

$$x^* = \arg \min_{Ax=b} x^T P x + 2c^T x,$$

was wir schon in Abschnitt 2.1 untersucht haben. Man erhält mit (18) die Lagrange-Funktion

$$L(x, \lambda) = x^T P x + 2c^T x - 2(Ax - b)^T \lambda.$$

Die Bedingungen für Lösungen (x, λ) lauten dann nach Proposition 3.3:

$$\begin{aligned} \nabla_x L = 0 &: Px + c - A^T \lambda = 0 \\ \nabla_\lambda L = 0 &: Ax - b = 0 \end{aligned}$$

Dies läßt sich in Matrixschreibweise wie folgt darstellen:

$$\begin{pmatrix} P & -A^T \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} = - \begin{pmatrix} c \\ b \end{pmatrix}$$

Diese *Lagrange-Matrix* läßt sich invertieren, explizite Formeln erhält man mit dem soeben vorgestellten Matrixinversionslemma 3.5.

So finden wir dann für die Invertierung der Lagrange-Matrix folgende Formeln.

$$\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -H & T^T \\ T & U \end{pmatrix} \begin{pmatrix} c \\ b \end{pmatrix},$$

wobei die Teilmatrizen H, T und U gegeben werden durch

$$\begin{aligned} U &= (AP^{-1}A^T)^{-1}, \\ T &= UAP^{-1} \text{ und} \\ H &= P^{-1} - P^{-1}A^T U A P^{-1}. \end{aligned}$$

Also ist

$$\begin{aligned} \lambda^* &= Tc + Ub = (AP^{-1}A^T)^{-1}(AP^{-1}c + b) \\ x^* &= -Hc + T^T b = -P^{-1}(c - A^T(AP^{-1}A)^{-1}(AP^{-1}c + b)) \\ &= -P^{-1}(c - A^T \lambda^*) \end{aligned}$$

ein *Lagrange-Paar*, das das QP löst. Wir haben hier also einen alternativen Beweis für Proposition 2.1 gefunden.

In diesem Abschnitt haben wir Bedingungen eingeführt, die von Minima unter Nebenbedingungen eingehalten werden müssen. Im nächsten Abschnitt wollen wir diese Bedingungen in Verfahren umsetzen, die eine für die Datenvalidierung geeignete Umformung der Optimierungsaufgabe (17) iterativ lösen.

2. Nichtlineare Validierung

Um Nichtlinearitäten zu behandeln, gibt es zwei unterschiedliche Methoden. Zuerst einmal kann man das nichtlineare Problem in eine Folge von linearen Problemen mit bekannter Lösungsstrategie überführen. Die zweite Möglichkeit besteht darin, eine Fixpunkt-Iteration zu finden, die gegen die gesuchte Lösung des nichtlinearen Problems konvergiert. Hier wird für beide Vorgehensweisen jeweils ein Algorithmus vorgestellt. Diese unterscheiden sich auf den ersten Blick nicht allzu sehr voneinander, weisen aber doch unterschiedliche Eigenschaften auf. In der Optimierungstheorie findet man diese Verfahren unter dem Oberbegriff „Sequentielles quadratisches Programm“. Dieser Name deutet schon an, daß das nichtlineare Optimierungsproblem wieder in eine Folge quadratischer Programme zerlegt wird.

2.1. Linearisierung der Nebenbedingung. Wir untersuchen das Validierungsproblem nach 1.2, diesmal aber mit einer nichtlinearen Nebenbedingung f . Die einfachste Methode zur Lösung der Minimierungsaufgabe

$$\text{„Minimiere } J(x) \text{ unter } f(x) = 0\text{“}$$

unter einer nichtlinearen Nebenbedingung f ist eine Linearisierung der Nebenbedingung: Wir ersetzen $f(x)$ durch die Linearisierung um die Messung y :

$$\bar{f}(x, y) := f(y) + \nabla f \Big|_y (x - y).$$

Dann läßt sich das Validierungsverfahren (6) für den linearen Fall aus Lemma 2.2 benutzen.

Der validierte Wert \hat{x} erfüllt dann natürlich nur die linearisierte Nebenbedingung. Um zu höherer Genauigkeit zu gelangen, läßt sich das Verfahren wiederholen, aber nun mit einer um \hat{x} linearisierten Nebenbedingung. Man benutzt also den validierten Wert als neue „virtuelle“ Messung. Man benötigt dann noch geeignete Abbruchkriterien. Ausgehend von Gleichung (6) hat dieses Verfahren dann die folgende Form.

ALGORITHMUS 3.7 (Fluß-Verfahren). Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ die Gleichung der Nebenbedingung, $y \in \mathbb{R}^n$ eine Messung und $D \succ 0, D \in \mathbb{R}^{n \times n}$ die Kovarianzmatrix der Sensorstörungen. Dann wird das Fluß-Verfahren gegeben durch

$$\begin{aligned} x_0 &= y \\ (24) \quad A_k &= \nabla f \Big|_{x_k} \\ x_{k+1} &= x_k - DA_k^\top (A_k DA_k^\top)^{-1} f(x_k). \end{aligned}$$

BEMERKUNG. Dieser Algorithmus trägt den Namen Fluß-Verfahren, da er im Spezialfall $n = m + 1$ als ein Korrektorschritt zur numerischen Berechnung von Punkten auf einer implizit gegebenen Kurve

dient (Stichworte hierfür sind „path following“, „curve tracking“, „numerical continuation“ und „normal flow“). Der Prädiktorschritt, der von einem Punkt der Kurve aus einen weiteren Punkt in Richtung der Kurve findet, ist für unsere Problemstellung nicht von Bedeutung. Für den Fall $n = m$ ist dieses Verfahren ein multidimensionales Newton-Verfahren, die Pseudo-Inverse von A_k geht dann in eine Inverse über. Allerdings haben wir diesen Fall wegen $m < n$ ausgeschlossen.

Im Folgenden wollen wir diesen Algorithmus 3.7 motivieren, indem wir zeigen, wie dieser aus den Bedingungen an die Lagrange-Funktion des Validierungsproblems hergeleitet werden kann.

Mit der Lagrange-Funktion $L(r, \lambda) = J(r) - 2f^\top(r + y)\lambda$ gilt dann nach Proposition 3.3 für das Lagrange-Paar (x^*, λ^*)

$$(25) \quad D^{-1}r^* = \nabla f^\top \Big|_{x^*} \lambda^*,$$

$$(26) \quad f(x^*) = 0.$$

Anstatt nach Lösungen für x^* und r^* zu suchen, ersetzen wir f durch eine Linearisierung um y und suchen \tilde{r} mit

$$\tilde{r} = D\nabla f^\top \Big|_y \tilde{\lambda} \quad \text{und} \quad \bar{f}(y + \tilde{r}, y) = f(y) + \nabla f \Big|_y \tilde{r} = 0.$$

Dabei haben wir zusätzlich zur Linearisierung von f den Gradienten von f an der Stelle x^* durch den Gradienten an der Stelle y approximiert. Diese Ideen werden auch beim Newton-Raphson-Verfahren zur Nullstellensuche angewandt.

Mit $\nabla f := \nabla f \Big|_y$ folgt dann aus $\nabla f \tilde{r} = -f(y)$ und $\tilde{r} = D\nabla f^\top \tilde{\lambda}$, daß

$$\tilde{\lambda} = -(\nabla f D\nabla f^\top)^{-1} f(y).$$

Aus der Regularität von ∇f folgt, daß $\nabla f D\nabla f^\top$ positiv definit und damit invertierbar ist. Wir haben allerdings die Regularität nur in x^* gefordert, aber in einer Umgebung von x^* ist ∇f auch regulär. Es gilt dann

$$\tilde{r} = -D\nabla f^\top (\nabla f D\nabla f^\top)^{-1} f(y),$$

und damit

$$\tilde{x} = y + \tilde{r} = y - D\nabla f^\top (\nabla f D\nabla f^\top)^{-1} f(y).$$

Ist das Modell linear, so ist $\tilde{x} = x^*$, es gilt $f(\tilde{x}) = 0$.

BEMERKUNG.

- Wir haben jetzt drei Sichtweisen auf einen Schritt der Validierung kennengelernt:
 1. Die lineare Validierung ist nach Lemma 2.2 ein quadratisches Programm.
 2. Die lineare Validierung ist nach Proposition 2.10 eine Projektion auf einen affinen Unterraum.

3. Die lineare Validierung entsteht durch Anwendung eines Newton-Schrittes auf die Bedingungen aus Proposition 3.3.

- Ist x eine Nullstelle von f , so ist x auch ein Fixpunkt des Verfahrens 3.7, denn $f(x_n) = 0 \implies x_{n+1} = x_n + 0$. Das ist natürlich auch der größte Kritikpunkt an diesem Verfahren: Jeder Punkt, der nur die Nebenbedingungen einhält, ist ein Fixpunkt. Dabei wird die geforderte Minimalitätseigenschaft $J \stackrel{!}{=} \min$ nicht berücksichtigt.
- Bei dem in 3.7 vorgestellten Algorithmus handelt es sich um eine Variante eines gewichteten Newton-Verfahrens zur Nullstellensuche von f : Das Newton-Verfahren besitzt für $f : \mathbb{R} \rightarrow \mathbb{R}$ die Gestalt

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

(Siehe [Sch93, Abschnitt 5.3.2]). Da in unserem Fall eine Inverse des Gradienten f' nicht existiert, wird eine Pseudo-Inverse eingesetzt, die gerade die Kovarianzstruktur berücksichtigt. Dieses im folgenden Satz noch näher untersucht.

Unter zusätzlichen Voraussetzungen überträgt sich die günstige Konvergenzeigenschaft des Newton-Verfahrens, wie uns der folgende Satz zeigt, der uns eine Analyse des Fluß-Verfahrens liefert.

SATZ 3.8. *Es sei $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m < n$ eine differenzierbare Abbildung, deren Ableitung ∇F auf einer offenen konvexen Menge $\Omega \subset \mathbb{R}^n$ den vollen Rang $\text{rang } \nabla F = m$ besitzt, und für die folgende Eigenschaften gelten:*

1. *Es existiert ein $\gamma \geq 0$ und $p \in (0, 1]$ mit*

$$(27) \quad \|\nabla F|_x - \nabla F|_y\| \leq \gamma \|y - x\|^p$$

für alle $x, y \in \Omega$.

2. *Es gibt eine obere Schranke μ für die Norm der Pseudoinversen von ∇F :*

$$(28) \quad \|\nabla F^{(-)}\| \leq \mu \quad \text{auf } \Omega.$$

Setze außerdem

$$\Omega_\varepsilon = \{x \in \Omega \mid \text{dist}(x, \mathbb{R}^n \setminus \Omega) > \varepsilon\}$$

für ein gegebenes $\varepsilon > 0$. Dann gibt es ein $\delta > 0$, das nur von γ , p , μ und ε abhängt, so daß die Folge der (x_n) , definiert durch

$$(29) \quad x_{n+1} = x_n - \nabla F|_{x_n}^{(-)} F(x_n),$$

für einen Startwert $x_0 \in \Omega_\varepsilon$ und $\|F(x_0)\| < \delta$ wohldefiniert ist und zu einem Grenzwert x^ mit $F(x^*) = 0$ konvergiert. Außerdem gibt es eine Konstante β , für die gilt*

$$(30) \quad \|x_{n+1} - x_n\| \leq \beta \|x_n - x^*\|^{1+p} \quad \text{für } n = 0, 1, \dots$$

BEWEIS. (Siehe [WWa90, Theorem 2.1]) Ist $x \in \Omega$, so gilt für den aktuellen Schritt $s = -(\nabla F|_x)^{(-)}F(x)$ nach (28)

$$(31) \quad \|s\| \leq \mu \|F(x)\|.$$

Besitzt ∇F in x vollen Rang, so gilt für $x_+ = x + s$

$$\|F(x_+)\| = \|F(x_+) - F(x) - \nabla F|_x s\|,$$

denn $\nabla F|_x(\nabla F|_x)^{(-)} = I$.

Ist zusätzlich zu x auch $x_+ \in \Omega$, so erhalten wir die Abschätzung

$$(32) \quad \|F(x_+)\| = \|F(x_+) - F(x) - \nabla F|_x s\| \leq \frac{\gamma}{1+p} \|s\|^{1+p},$$

denn es gilt für einen Weg $x(t)$ in Ω mit $x(t) = x + ts$, $t \in [0, 1]$:

$$\begin{aligned} \|F(x_+) - F(x) - \nabla F|_x s\| &= \left\| \left(\int_0^1 \nabla F|_{x(t)} - \nabla F|_x dt \right) s \right\| \\ &\leq \int_0^1 \|\nabla F|_{x(t)} - \nabla F|_x\| dt \|s\| \\ &\stackrel{(27)}{\leq} \gamma \int_0^1 \|x(t) - x\|^p dt \|s\| \\ &= \left| \gamma \int_0^1 t^p \|s\|^p dt \right| \|s\| = \frac{\gamma}{1+p} \|s\|^{1+p}. \end{aligned}$$

Dann folgt für $s_+ = (\nabla F|_{x_+})^{(-)}F(x_+)$ aus (31) und (32)

$$\|s_+\| \leq \frac{\gamma\mu}{1+p} \|s\|^{1+p}.$$

Wähle δ nun so klein, daß gilt

$$\tau := \frac{\gamma\mu(\mu\delta)^p}{1+p} < 1 \quad \text{und} \quad \frac{\mu\delta}{1-\tau} < \varepsilon.$$

Dann gilt für $x_0 \in \Omega_\varepsilon$ und $\|F(x_0)\| < \delta$, daß die Folge der (x_n) gegeben durch (29) in Ω verbleibt, denn die Schritte $s_k = x_{k+1} - x_k = -(\nabla F|_{x_k})^{(-)}F(x_k)$ erfüllen

$$\|s_k\| < \mu\delta \quad \text{und} \quad \|s_{k+1}\| \leq \tau \|s_k\|.$$

Damit bildet (x_k) eine Cauchy-Folge mit Grenzwert $x^* \in \Omega$, denn

$$\begin{aligned} \|x_n - x_m\| &= \left\| \sum_{i=m}^{n-1} s_i \right\| \leq \|s_m\| \sum_{i=0}^{n-m-1} \tau^i \\ &\leq \|s_m\| \frac{1}{1-\tau} < \frac{\mu\delta}{1-\tau} < \varepsilon. \end{aligned}$$

Diese Betrachtung kann man auch auf den Grenzwert x^* erweitern, so daß man folgende Gleichung für alle $k \in \mathbb{N}$ erhält:

$$\begin{aligned} \|x_{k+1} - x^*\| &= \left\| \sum_{i=k+1}^{\infty} s_i \right\| \leq \frac{\|s_{k+1}\|}{1 - \tau} \\ &\leq \frac{\gamma \mu \|s_k\|^{1+p}}{(1+p)(1-\tau)} \\ &\leq \frac{\gamma \mu (\mu \|F(x_k)\|)^{1+p}}{(1+p)(1-\tau)}. \end{aligned}$$

Mit einem ähnlichen Integrationsargument wie oben erhalten wir für die Norm von $F(x_k)$ die Abschätzung

$$\|F(x_k)\| \leq \|\nabla F|_{x^*}\| \|x_k - x^*\| + \frac{\gamma}{1+p} \|x_k - x^*\|^{1+p}.$$

Die beiden letzten Aussagen ergeben zusammen die letzte Behauptung des Satzes. \square

BEMERKUNG. Das in diesem Satz untersuchte Verfahren (29) ist von der gleichen Bauart wie unser Verfahren 3.7. Also lassen sich die hier gefundenen Aussagen sofort auf unsere Situation übertragen. Insbesondere gibt uns Gleichung (30) eine Abschätzung für die Konvergenzgeschwindigkeit, diese Gleichung zeigt, daß das Verfahren (29) eine Konvergenzordnung von $p + 1$ besitzt.

BEMERKUNG. Ist F eine lineare Abbildung, so läßt sich $\gamma = 0$ in (32) wählen. Dann ergibt sich $\|s_1\| = 0$, was mit unseren bisherigen Ergebnissen übereinstimmt, denn bei einem linearen Modell ist man nach einem Schritt – hier s_0 – mit der Validierung fertig.

Das Flußverfahren 3.7 liefert bei Konvergenz einen Punkt, der die Nebenbedingungen einhält. Ob dieser die Minimalitätsbedingungen erfüllt, die für die Kosten der Validierung nach Definition 1.2 gefordert werden, ist nicht gewährleistet. Die VDI-Norm ([VDI98, Abschnitt 5.3, Gleichung (142)]) schlägt dieses Verfahren bei nichtlinearen Nebenbedingungen vor, ohne darauf hinzuweisen, daß die damit berechneten Werte nicht mit minimalen Kosten versehen sind.

Um diesen Nachteil auszugleichen, suchen wir also ein Verfahren, daß die Minimalitätsforderungen berücksichtigt.

2.2. Fixpunktverfahren. Das im Folgenden vorgestellte Verfahren entsteht, indem man die Fixpunkteigenschaft aus Proposition 2.8 auf den nichtlinearen Fall überträgt. Dieses Verfahren führt dazu, daß der validierte Wert bestimmte Minimalitätseigenschaften besitzt, denn diese Lösung wird durch genau eine Projektion der Messung berechnet. Nach dem Projektionslemma A.9 gibt dieses Verfahren also eine optimale Lösung.

Das Verfahren hat die folgende Gestalt:

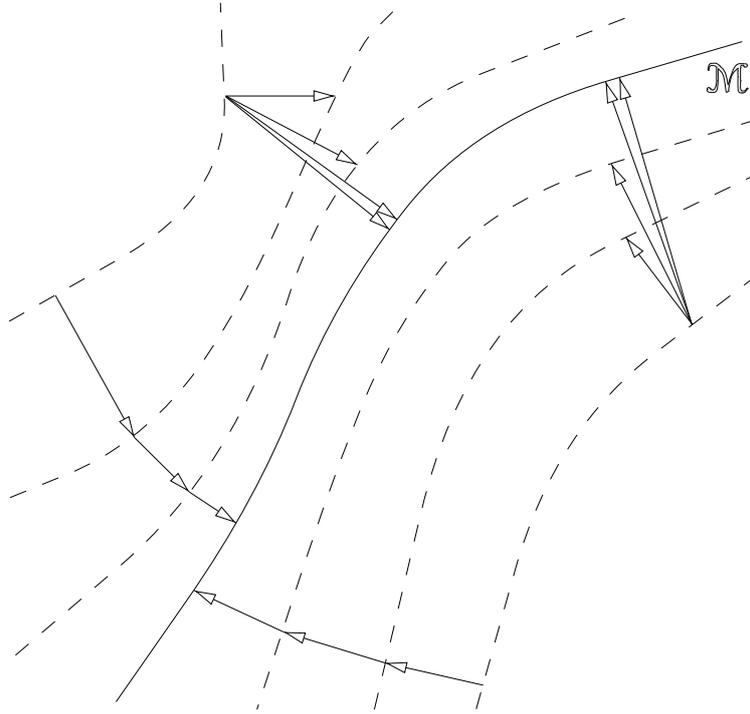


ABBILDUNG 6. Validierungsverfahren: Oben Fixpunktverfahren 3.9, unten Fluß-Verfahren 3.7

ALGORITHMUS 3.9 (Fixpunktverfahren). Wir verwenden die Bezeichnungen aus Algorithmus 3.7. Zusätzlich sei die Linearisierung von f um u gegeben durch $\bar{f}(v, u) = f(u) + \nabla f|_u (v - u)$. Dann wird das Fixpunktverfahren gegeben durch folgende Gleichungen:

$$\begin{aligned}
 (33) \quad & x_0 = y \\
 & A_k = \nabla f|_{x_k} \\
 & x_{k+1} = y - DA_k^T (A_k DA_k^T)^{-1} \bar{f}(y, x_k) \\
 & \quad = y - DA_k^T (A_k DA_k^T)^{-1} (A_k (y - x_k) + f(x_k)).
 \end{aligned}$$

Dieses Verfahren unterscheidet sich von Algorithmus 3.7 dadurch, daß die Suche nach einem Minimum immer wieder von der aktuellen Messung y ausgeht. Das Vorgehen wird in Abbildung 6 verdeutlicht.

Da das Fixpunktverfahren immer wieder auf die Messung y zurückgreift, besitzt dieses Verfahren schlechtere Konvergenzeigenschaften als 3.7. Insbesondere ist zu erwarten, daß die bestmögliche Genauigkeit im Rahmen von $\varepsilon \|y - x^*\|$ liegt, wobei ε die Maschinengenauigkeit angibt, da dann die entsprechenden Punkte in einer Umgebung von x^* im Rahmen der Rechengenauigkeit von y aus nicht mehr voneinander unterscheidbar sind.

Die Fixpunkteigenschaft findet sich in den beiden folgenden Propositionen. Dabei gibt uns die erste Proposition eine Art *Orthogonalitätsbedingung*, während Proposition 3.11 absichert, daß die Fixpunkte auch die Nebenbedingungen einhalten.

PROPOSITION 3.10. *Es sei \tilde{x} Fixpunkt des Verfahrens 3.9. Dann ist $\tilde{r} = \tilde{x} - y$ Fixpunkt der Abbildung*

$$(34) \quad \pi(r) = D\bar{D}(r)r.$$

BEWEIS. Analog zum Beweis von Proposition 2.8 im linearen Fall.

PROPOSITION 3.11. *Ist x^* Fixpunkt von 3.9, so gilt $f(x^*) = 0$*

BEWEIS. Bezeichnet man den Iterationsschritt mit $x_{k+1} = F(x_k)$, so gilt

$$AF(x) = Ay - A\bar{f}(y, x) = Ay - (f(x) + A(y - x)) = Ax - f(x).$$

Ist x Fixpunkt von F , so ist $Ax = Ax - f(x) \implies f(x) = 0$. \square

In [VSa92] wird folgende, über Proposition 3.11 hinausgehende Aussage gemacht, die falsch ist, und zwar: Ist $\|f(x_n)\|$ eine Nullfolge, so konvergiert das Verfahren 3.9.

Wir widerlegen folgende Vermutung durch die Konstruktion eines Gegenbeispiels in 3.12:

Ist $P(x) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ eine Funktion von Q -orthogonalen Projektoren, so konvergiert die *projizierte Folge*

$$(35) \quad x_{n+1} = P(x_n)x_n.$$

Da wir diese Aussage nicht haben, kann die lokale Konvergenz des Fixpunktverfahrens 3.9 nicht einfach gesichert werden.

Hätten wir nämlich die Konvergenz von Folgen der Bauart (35), so würden wir für das Verfahren 3.9 eine einfache Konvergenzaussage besitzen. Denn ist $f(x_k)$ eine Nullfolge, so können wir (33) durch

$$x_{k+1} = y - DA_k^\top (A_k DA_k^\top)^{-1} A_k (y - x_k)$$

für große Indizes $k > k_0$ ersetzen. Diese Gleichung ist aber gerade von der in (35) verwendeten Form, denn $DA_k^\top (A_k DA_k^\top)^{-1} A_k$ ist eine D -orthogonale Projektionsmatrix, die über den Gradienten A_k vom aktuellen Folgenglied x_k abhängt.

Man kann vermuten, daß eine Folge (x_n) der Bauart (35) konvergiert, denn sie ist zumindest eine Folge mit abnehmender Norm, was wir jetzt zeigen.

Die Q -orthogonale Projektion auf $\text{Im}B$, wobei B regulär und von x_n abhängig ist, wird nach Proposition A.5 gegeben durch

$$\mathcal{P} = B(B^\top QB)^{-1} B^\top Q.$$

Wie auch in Proposition 2.11 hilft uns hier ein *schiefer Satz des Pythagoras*. Denn

$$\begin{aligned}\|x\|_Q^2 &= x^\top Qx = x^\top QB(B^\top QB)^{-1}B^\top Qx \\ &\quad + x^\top Q(I - B(B^\top QB)^{-1}B^\top)Qx \\ &\geq x^\top QB(B^\top QB)^{-1}B^\top Qx \\ &= x^\top QB(B^\top QB)^{-1}B^\top QB(B^\top QB)^{-1}B^\top Qx \\ &= \|\mathcal{P}x\|_Q^2.\end{aligned}$$

Damit ist die Folge der $\|x_n\|_Q$ aus (35) monoton fallend oder konstant. Die entscheidende Lücke in einem Beweisversuch der Vermutung zu Gleichung (35) ist nun, daß man aus $\lim_n \|P(x_n)x_n\| = c$ nicht schließen kann, daß auch (x_n) eine konvergente Folge ist.

BEISPIEL 3.12 (Konstruktion eines Gegenbeispiels). Wir konstruieren eine divergente projizierte Folge der Form (35). Damit ist die Behauptung von [VSa92] widerlegt, daß Folgen solcher Bauart konvergieren. Wir wählen $n = 2$, $m = 1$ und $Q = I$, die Situation ist also der \mathbb{R}^2 versehen mit dem Standardskalarprodukt. Eine Projektion auf einen Unterraum gegeben durch den Winkel φ besitzt dann die Projektionsmatrix

$$P_\varphi = \begin{pmatrix} \cos^2(\varphi) & \sin(\varphi)\cos(\varphi) \\ \sin(\varphi)\cos(\varphi) & \sin^2(\varphi) \end{pmatrix}$$

Hat x_n die Polarkoordinaten (r_n, φ_n) mit Radius r_n und Winkel φ_n , so ergibt sich:

$$\begin{aligned}r_{n+1} &= \|x_{n+1}\| = \|P_{\varphi_{n+1}} \begin{pmatrix} \cos(\varphi_n) \\ \sin(\varphi_n) \end{pmatrix} x_n\| \\ &= (\cos^4 \varphi_{n+1} \cos^2 \varphi_n + \sin^4 \varphi_{n+1} \sin^2 \varphi_n \\ &\quad + 2 \cos \varphi_{n+1} \sin \varphi_{n+1} \cos \varphi_n \sin \varphi_n + \cos^2 \varphi_{n+1} \sin^2 \varphi_{n+1})^{1/2} r_n \\ &= \sqrt{\cos^2(\varphi_{n+1} - \varphi_n)} r_n\end{aligned}$$

Die soeben skizzierte Rechnung läßt sich einfacher geometrisch deuten. Der Punkt mit den Polarkoordinaten (r_{n+1}, φ_{n+1}) liegt auf einem Thaleskreis über der Strecke von 0 nach (r_n, φ_n) , damit ist $\cos(\varphi_{n+1} - \varphi_n) = \frac{r_{n+1}}{r_n}$. In Abbildung 7 wird diese Situation verdeutlicht: Die Folge der (x_n) wird jeweils durch eine orthogonale Projektion des vorherigen Folgengliedes gegeben, wobei diese Folge normmäßig gegen c konvergiert, aber (x_n) durch geschickte Wahl nicht konvergiert.

Wir machen daher folgenden Ansatz:

$$\begin{aligned}\varphi_{n+1} &= \varphi_n + f(r_n), \\ r_{n+1} &= \cos(\varphi_{n+1} - \varphi_n) r_n = \cos(f(r_n)) r_n.\end{aligned}$$

Wir möchten, daß $\lim r_n = c > 0$ gilt, und leiten nun Eigenschaften ab, die f für die Konvergenz dieser Folge besitzen muß.

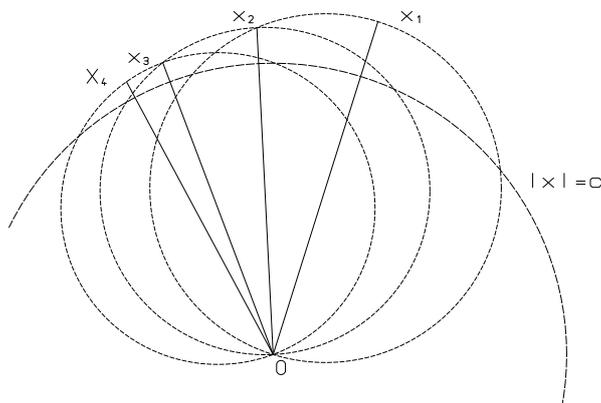


ABBILDUNG 7. Zur Konstruktion einer divergenten projizierten Folge

1. $r_0 > c, r_n > c$
2. $r_n > r_{n+1} \iff (1 - \cos(f(r_n)))r_n > 0$, um zu gewährleisten, daß (r_n) monoton fällt.
3. $f(c) = 0$, damit die Folge der (r_n) einen Fixpunkt in c besitzt.
4. Außerdem fordern wir

$$(36) \quad f(r) \leq \sqrt{2} \sqrt{1 - \frac{c}{r}}.$$

Dann gilt für $r > c$ mit einer Taylor-Entwicklung zweiter Ordnung des Kosinus

$$\begin{aligned} r_{n+1} &= \cos(f(r_n))r_n > \left(1 - \frac{f^2(r_n)}{2}\right) r_n \\ &\geq \left(1 - \frac{r_n - c}{r_n}\right) r_n = r_n - r_n + c = c. \end{aligned}$$

Für Funktionen f , die die vorhergehenden Bedingungen erfüllen, ist also (r_n) eine monoton fallende Folge, die durch den Fixpunkt c nach unten beschränkt ist, und demzufolge konvergent.

Jetzt möchten wir f so wählen, daß die Folge der (φ_n) divergiert. Dazu betrachten wir die Folge

$$s_{n+1} = \left(1 - \frac{f^2(s_n)}{2}\right) s_n,$$

die aus (r_n) durch Reihenentwicklung des Kosinus entsteht.

Nach Punkt 4 gilt die folgende Ungleichung $r_{n+1} > \left(1 - \frac{f^2(r_n)}{2}\right) r_n$. Außerdem ist c auch Fixpunkt der (s_n) -Folge. Dann ist $(r_n) > (s_n)$ für alle n , wenn für den Anfangswert $s_1 = r_1$ gilt. Dies ist eine Folgerung aus der Monotonie der beiden Folgen und aus der Tatsache, daß $r_n = s_n \implies r_{n+1} > s_{n+1}$.

Wir machen nun den Ansatz eines zeitvarianten Systems:

$$z_{n+1} = \left(1 - \frac{f^2(z_n, n)}{2}\right) z_n$$

mit $f(z_n, n) = \sqrt{2}(n(n+1)z_n)^{-1/2}$, $n = 1, 2, \dots$

Damit ist

$$z_{n+1} = \left(1 - \frac{1}{n(n+1)z_n}\right) z_n = z_n - \frac{1}{n(n+1)}.$$

Insbesondere folgt aus der Anfangsbedingung $z_1 = c + 1$, daß gilt

$$z_n = c + \frac{1}{n},$$

was wir durch folgende Induktion beweisen. Für den Induktionsanfang gilt $z_1 = c + \frac{1}{1} = c + 1$, sei also die Formel verifiziert für alle $i < n$.

Dann ist

$$z_n = z_{n-1} - \frac{1}{(n-1)n} = c + \frac{1}{n-1} - \frac{1}{(n-1)n} = c + \frac{n-1}{(n-1)n}$$

Mit dieser Anfangsbedingung $z_1 = c + 1$ gilt dann für die zeitvariante Lösung folgende Abschätzung

$$\begin{aligned} f(z_n, n) &= \sqrt{2} \sqrt{\frac{1}{n(n+1)z_n}} < \sqrt{2} \sqrt{\frac{1}{nz_n}} \\ &= \sqrt{2} \sqrt{\frac{z_n - c}{z_n}}. \end{aligned}$$

Damit ist Ungleichung (36) erfüllt. Also folgt, daß die Lösung von

$$r_{n+1} = \cos(f(r_n, n))r_n$$

mit Anfangswert $r_1 = c + 1$ gegen c konvergiert und stets $r_n \geq s_n = c + \frac{1}{n}$ gilt. Es sei \tilde{f} eine Funktion, für die $\tilde{f}(r_n) = f(r_n, n)$ für alle n gilt. Dies ist möglich, da die Folge der (r_n) monoton fällt und in c stetig durch 0 fortsetzbar ist.

Damit gilt dann für den Winkelzuwachs

$$\begin{aligned} \varphi_{n+1} - \varphi_n &= \tilde{f}(r_n) = \sqrt{\frac{2}{n(n+1)r_n}} \\ &> \sqrt{\frac{2}{c+1}} \sqrt{\frac{1}{n(n+1)}} > \sqrt{\frac{2}{c+1}} \frac{1}{n+1} \end{aligned}$$

Also bilden die (φ_n) eine divergente Folge, da $\sum \frac{1}{n}$ eine divergierende Minorante bildet.

Die projizierte Folge der $x_n = r_n \begin{pmatrix} \cos \varphi_n \\ \sin \varphi_n \end{pmatrix}$ ist damit divergent. \square

Wir suchen nun nach Bedingungen, die erfüllt sein müssen, damit das Verfahren 3.9 konvergiert. Für die Konvergenz muß notwendigerweise gelten, daß die Folge der $f(x_n)$ eine Nullfolge bildet, da Fixpunkte Nullstellen von f sind. Wie wir am Beispiel gesehen haben, ist diese Bedingung aber nicht hinreichend.

Hier liefert der Banachsche Fixpunktsatz Bedingungen für die Konvergenz, insbesondere muß es sich bei (24) aus dem Verfahren 3.7 um eine Kontraktion handeln, damit das Verfahren konvergiert (Siehe [Sch93, Satz 5.1]). Um dies nachzuprüfen, ist die Norm der Ableitung des Verfahrensschrittes

$$F(x) = y - DA^\top(x)(A(x)DA^\top(x))^{-1}(A(x)(y - x) + f(x))$$

auf $\|\nabla F\| < 1$ zu untersuchen. Ob aber die Berechnung dieser Norm ebenso aufwendig wie die Durchführung der Iteration ist, sei dahingestellt.

2.3. Lagrange-Newton Verfahren. Als dritte Möglichkeit eines Iterationsverfahrens zur Lösung des nichtlinearen Optimierungsproblems ist die *Lagrange-Newton Methode* zu nennen, die ein Newton-Verfahren zur Minimierung der Lagrange-Funktion benutzt. Um die Lagrange-Funktion

$$L_\lambda(x) = (x - y)^\top D^{-1}(x - y) - 2f(x)^\top \lambda$$

zu minimieren, sucht man eine Nullstelle der Ableitung. Diese Nullstellensuche löst man mit einem Newton-Verfahren. Dabei wird eine zweite Ableitung der Nebenbedingung benötigt, was dann auch für bessere Konvergenzeigenschaften dieses Verfahrens sorgt (Siehe [Col90] für weitere Eigenschaften). Eine mögliche Variante dieses Verfahren hat folgendes Aussehen.

ALGORITHMUS 3.13 (Lagrange-Newton). Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ die Modellgleichung, $y \in \mathbb{R}^n$ die Messung und $L_\lambda(x) = (x - y)^\top D^{-1}(x - y) - 2f(x)^\top \lambda$ die Lagrange-Funktion zum Vektor λ der Lagrange-Multiplikatoren. Dann wird das Lagrange-Newton-Verfahren gegeben durch

$$(37) \quad \begin{aligned} x_0 &= y \\ D_0 &= D \\ A_k &= \nabla f|_{x_k} \\ \lambda_{k+1} &= (A_k D_k A_k^\top)^{-1} f(x_k) \\ D_{k+1} &= (\nabla^2 L_{\lambda_k}(x_k))^{-1} \\ x_{k+1} &= x_k - D_k A_k^\top \lambda_{k+1}, \end{aligned}$$

wobei $\nabla^2 L$ die zweite Ableitung der Lagrange-Funktion in x -Richtung ist.

Die Verwendung der zweiten Ableitung läßt sich dadurch vermeiden, indem man eine Approximation zweiter Ordnung der Hesse-Matrix durch

die BFGS¹-Auffrisch-Formeln benutzt. Dieses Verfahren wird hier nicht weiter untersucht.

2.4. Beispiele. Nun untersuchen wir die vorgestellten Algorithmen an einigen Beispielen.

BEISPIEL 3.14 (Nichtlineares Beispiel). Unser Modell sei gegeben durch

$$f(a, b, c) = \begin{cases} a - \tanh(b) \\ b - c \end{cases}$$

mit der Messung (1,2,3) und der Kovarianzmatrix $D = \text{diag}(0.1, 0.2, 0.3)$. Nach einem Schritt stimmen die validierten Werte \tilde{b} und \tilde{c} überein, aber die nichtlineare Gleichung im Modell hat noch eine Abweichung von $\tanh(\tilde{a}) - \tilde{b} = 0.009$. Nach der vierten Iteration nach 3.7 werden keine weiteren Verbesserungen erzielt. Die Validierung besitzt dann den Wert $(a^*, b^*, c^*) = (0.984, 2.40, 2.40)$. Mit dem Fixpunktverfahren 3.9 erhält man auch nach 4 Schritten einen validierten Wert, der sich in der 4.ten Nachkommastelle vom vorigen Wert unterscheidet. Entfernt man die zweite Modellgleichung und die dritte Messung, so unterscheidet sich der validierte Wert dieses Modells natürlich vom vorhergehenden, daran zeigt sich, daß redundante Information auch genutzt wird.

BEISPIEL 3.15. Abbildung 8 zeigt das Iterationsverhalten von Gleichung (24) des Fluß-Verfahrens 3.7 unter einigen Gleichungen der Form $f(x, y) = y - g(x)$. Dabei wurden die Iterationen zu demselben Startwert zur Verdeutlichung miteinander verbunden. Die Kovarianz ist meist so gewählt, daß das Verfahren bevorzugt in y -Richtung nach Lösungen sucht. Hier erkennt man, daß sich das Verfahren auch bei Gleichungen wie $y = |x|^x$ noch „gesittet“ verhält, während es bei Gleichungen wie $y = \sqrt{6.5 - x^2}$ oder $y = \int_0^x e^{-t^2} dt$ zu Effekten wie Überschneidungen und Oszillation kommt. Dennoch konvergieren alle gezeigten Beispiele gegen Punkte, die die Nebenbedingungen erfüllen, die im Bild mit einer Strichpunktlinie markiert sind. Allerdings sind diese Grenzwerte keine optimalen Lösungen.

Für den Vergleich zwischen den Verfahrensgleichungen (24) und (33) betrachten wir ein Modell der Form $f(x, y) = y - x^5$ mit $D = I$ und Startwert $(x_0, y_0) = (2, 0)$, also direkt „im Knick“ von $y = x^5$. Die Ergebnisse sind in den Tabellen 1 und 2 aufgeführt. Auf den ersten Blick unterscheiden sich die Verfahren in den ersten vier Schritten nur wenig, aber während das Fluß-Verfahren 3.7 schon nach 9 Iterationen eine Lösung gefunden hat, beginnt das Fixpunkt-Verfahren 3.9 in der Nähe einer Lösung zu oszillieren und besitzt schlechte Konvergenzeigenschaften, was sich auch daran bemerkbar macht, daß $\frac{\|f(x_n, y_n)\|}{\|f(x_{n+1}, y_{n+1})\|}$

¹Die Broyden-Fletcher-Goldfarb-Shanno(BFGS) Methode ist ein sogenanntes Quasi-Newton Verfahren, das sich für viele Optimierungszwecke eignet.

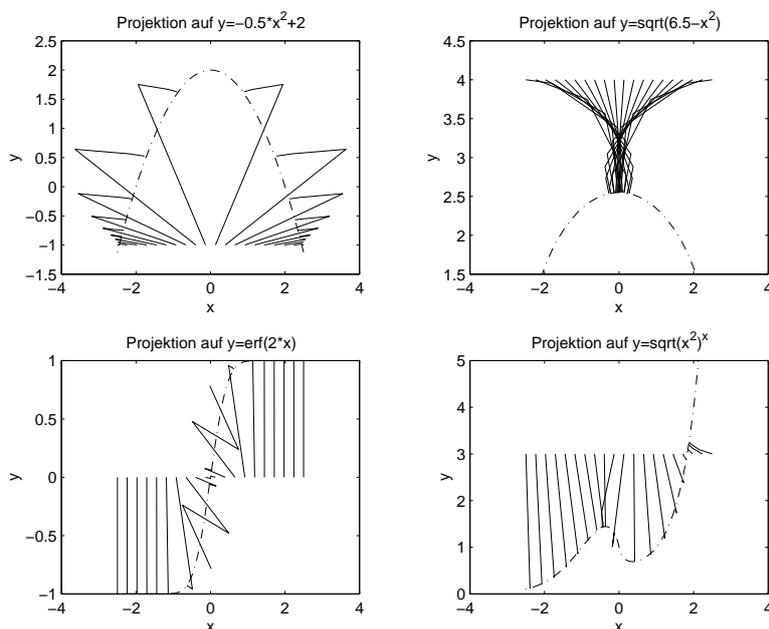


ABBILDUNG 8. Iteration des Validierungsschrittes (24) mit verschiedenen Gleichungen und Varianzen

TABELLE 1. Verfahren 3.7 unter der Nebenbedingung $f(x, y) = y - x^5$.

x	y	$\ f(x, y)\ $	\sphericalangle
2	0	32	89.3
1.6	0.005	10.5	-91
1.28	0.0147	3.43	-93.1
1.03	0.0336	1.11	-98.2
0.834	0.0683	0.335	-109
0.716	0.117	0.0703	-122
0.682	0.143	0.00402	-127
0.68	0.145	1.27e-05	-127
0.68	0.145	1.26e-10	-127
0.68	0.145	0	-127

über weite Teile der Iteration konstant bleibt. Insbesondere konvergiert das Verfahren auf dem Rechner nicht, sondern bleibt bei einer Norm von f in Höhe der Maschinengenauigkeit hängen. In den Tabellen gibt \sphericalangle den Winkel an, der aufgespannt wird durch die Verbindung des Startwertes (x_0, y_0) mit dem aktuellen Punkt der Iteration und der Geraden, die durch den Gradienten der Nebenbedingung im aktuellen Punkt gegeben wird. Dabei erkennt man, daß Algorithmus 3.9 die Orthogonalitätsbedingung aus Proposition 3.10 einhält, während dies bei

TABELLE 2. Verfahren 3.9 unter der Nebenbedingung $f(x, y) = y - x^5$.

x	y	$\ f(x, y)\ $	\sphericalangle
2	0	32	89.3
1.6	0.005	10.5	-91
1.28	0.0219	3.42	-92.5
1.03	0.0721	1.09	-95.8
0.86	0.203	0.268	-100
0.843	0.423	0.00185	-91.5
0.854	0.454	0.000816	-88.9
0.846	0.433	0.000418	-90.7
0.852	0.448	0.000194	-89.5
0.848	0.438	9.69e-05	-90.4
0.851	0.445	4.62e-05	-89.8
0.849	0.44	2.27e-05	-90.2
0.85	0.444	1.09e-05	-89.9
0.849	0.441	5.35e-06	-90.1
0.85	0.443	2.59e-06	-89.9
0.849	0.442	1.26e-06	-90

3.7 keine Rolle spielt und damit die geforderte Minimalitätsbedingung der Validierung i.A. nicht erfüllt ist. Die Verbindung zwischen Minimalitätseigenschaft und Orthogonalität liefert das Projektionslemma A.9.

Die Orthogonalitätsbedingung wird in Abbildung 5 angedeutet, der Gradient der Kostenfunktion steht an der lokalen Minimalstelle senkrecht auf der Nebenbedingung.

3. Güte-Test im nichtlinearen Fall

Auch für den Fall nichtlinearer Nebenbedingungen möchte man Aussagen über die Güte einer Messung treffen und dafür den in Proposition 2.14 vorgeschlagenen Test verwenden. Allerdings führen nichtlineare Nebenbedingungen zu einer nichtlinearen Dichte-Transformation, so daß die resultierende Verteilung der validierten Werte keine Normalverteilung ist. Insbesondere reichen dann Erwartungswert und Kovarianzmatrix nicht mehr zur Beschreibung dieser Dichte aus.

Natürlich läßt sich unabhängig vom verwendeten Verfahren der Gesamtfehler $y - x^*$ als Differenz zwischen Messung y und Validierung x^* berechnen, allerdings ist dann die zum Test verwendete Verteilungsaussage über die Kosten $(y - x^*)^\top D^{-1}(y - x^*)$ nicht mehr erfüllt.

Das Hauptproblem bei der Verwendung des Tests für eine Iteration mit dem Fluß-Verfahren nach (24) liegt darin, daß nur der aktuelle Schritt bewertet wird, und keine Aussagen für den Gesamtfehler gemacht werden können. Benutzt man also dieses Fluß-Verfahren, so kann man die

Kosten der einzelnen Schritte der Iteration aufsummieren, allerdings ist dann diese Summe $\sum \|x_{k+1} - x_k\|_{D^{-1}}^2$ nach Dreiecksungleichung nur eine obere Schranke für die Gesamtkosten, der Test nach Proposition 2.14 mit dieser Summe als Schwellwert ist damit nicht konservativ, d.h. auch „gute“ Messungen können abgelehnt werden. Außerdem sind vorhergehende Schritte der Validierung mit neueren Modellgradienten betrachtet nicht mehr optimal.

Wenn nun aber nach erfolgter Iteration $f(x) \approx 0$ gilt, und man den Fixpunkt-Algorithmus 3.9 verwendet hat, so hat man für die Fehlerverteilung wieder eine Normalverteilungsaussage, da nur die Linearisierung der Nebenbedingungen in der Iteration eingesetzt wurde. Damit ist ein Test nach Proposition 2.14 wieder sinnvoll.

4. Datenvalidierung als Schätzer

Wir wollen im Folgenden untersuchen, wie man mit der Validierung Schätzwerte erhält². So ist z.B. im Leitbeispiel 2.3 kein Meßwert für m_{FD} vorhanden, wir haben die Modellgleichungen so aufgestellt, daß dieser Wert eliminiert wird und wir mit diesem Modell Datenvalidierung betreiben können. Wollen wir aber direkt aus den erfaßten Daten einen Wert für m_{FD} ermitteln, so hilft uns das normale Validierungsverfahren nicht weiter.

Um diesen Schätzprozeß in Gang zu setzen, muß die Nebenbedingung also zusätzlich zu „echten“ Meßdaten auch noch Platzhalter für unbekannte Größen enthalten. Dieses läßt sich z.B. durch negative Einträge in der Varianzmatrix D kodieren (dort wo negative Einträge stehen, muß eine Schätzung durchgeführt werden). Für diese Platzhalter brauchen wir aber schon grobe Schätzwerte für Erwartungswert und Varianz.

Hat man z.B. für eine unbekannte Größe untere und obere Schranken, so läßt sich „uninformativ“ auf diesem Intervall $[x_{\min}, x_{\max}]$ eine Gleichverteilung annehmen. Damit erhält man die Schätzung $\hat{x} = \frac{x_{\max} + x_{\min}}{2}$ des Erwartungswertes und die Varianzschätzung $\hat{\sigma}_x^2 = \frac{(x_{\max} - x_{\min})^2}{12}$. Die Gleichverteilung heißt uninformativ, weil sie nur die Intervallgrenzen als Parameter der Verteilung besitzt.

Natürlich erfordert das von uns beschriebene Validierungsverfahren normalverteilte Werte, aber um grobe Richtwerte zu erhalten, ist dieses Verfahren wohl gerechtfertigt.

Das in Algorithmus 3.16 geschilderte Vorgehen eignet sich zu Ermittlung von Information über unzugängliche Zustände. Natürlich kann dies auch als Parameterschätzung für Modellgleichungen in einer Initialisierungsphase eingesetzt werden. Außerdem lassen sich durch die Hinzunahme weiterer freier Variablen Ungleichungen modellieren, so

²Dieses Vorgehen trägt den Namen „Co-Adaption“.

kann man z.B. $y \geq 0$ durch $y = z^2$ erzwingen, wobei z eine zusätzlich Variable im Modell ist.

Wir haben also den Meßvektor y um Schätzungen der nicht direkt meßbaren Größen und ebenso die Kovarianzmatrix um möglichst sinnvolle Einträge zu ergänzen.

Das Vorgehen zur Schätzwertgewinnung besteht darin, eine Validierung dieses ergänzten Modells durchzuführen, aber die Ergebnisse nur für die vorher unbekanntem ergänzten Werte zu übernehmen:

ALGORITHMUS 3.16 (Validierung als Schätzung). Sei I die Menge der Indizes der Platzhalter im Modell. Gesucht sind Schätzwerte an diesen Stellen.

1. Ergänze eine Messung y um grobe Schätzwerte $y_i, i \in I$ für die zu schätzenden Daten und ebenso für die zugehörigen Varianzen D_{ii} .
2. Berechne die Validierung x^* von y .
3. Setze $y_i = x_i^*$ für $i \in I$, $D_{ii} = (D - D^*)_{ii}$
4. Optional: Unterscheiden sich x_j^* sehr von y_j für $j \notin I$, dann zurück zu Schritt 2.

Durch Anwendung der Validierung erhält man einen bereinigten Schätzwert und eine verkleinerte Varianz als entsprechenden Diagonaleintrag in der Kovarianzmatrix $D - D^*$ der Validierungen. Wie sinnvoll es aber ist, diese so gewonnenen Schätzwerte wieder zur Validierung einzusetzen, sei dahingestellt, da mögliche Fehlerquellen in den tatsächlichen Messungen schon in diese Schätzwerte hineingerechnet wurden.

Für die Validierung der Messung ist dann also ein weiteres Modell nötig, in dem auf diese Schätzwerte nicht zugegriffen wird. Hierbei stellt sich die Frage, ob sich die Elimination dieser Werte aus dem Modell automatisch bewerkstelligen läßt und wie sich verschiedene Vorgehensweisen bei dieser Elimination auf die Validierung auswirken. Dieses wird hier aber nicht vertieft.

Wir wollen noch untersuchen, ob diese Schätzung durch Validierung gleiche Ergebnisse liefert wie eine explizite Vorverarbeitung, wie z.B. eine Interpolation zwischen zwei Sensorwerten, was im nächsten Beispiel genauer betrachtet wird.

BEISPIEL 3.17. Angenommen, wir können Meßwerte von zwei Enden einer Rohrleitung erhalten, deren zu Grunde liegende Meßgrößen affin linear von der Rohrlänge abhängen, sind aber an einem Zwischenwert interessiert. Ein Vorverarbeitung könnte diesen Zwischenwert als lineare Interpolation aus den anderen Werten berechnen. Unser Modell habe für die Messung $(x, y, z)^\top$ die Form

$$f(x, y, z) = y - (\vartheta x + (1 - \vartheta)z)$$

Wir nehmen an, daß wir die Werte $x = 1$ und $z = 3$ jeweils mit Varianz 0.1 erhalten haben. Wir möchten einen Schätzwert für y ermitteln,

wobei der Interpolationsparameter $\vartheta = 0.66$ bekannt sei. Wir setzen die Anfangsschätzung willkürlich auf $y = 3$ und $\sigma_y^2 = 1$.

Durch Anwendung des Validierungsschrittes (6) erhalten wir die Schätzung nach Validierung für y mit dem Wert $y_1 = 1.749$. Unter Beibehaltung der Werte für x und z und der Kovarianz D validieren wir jetzt den Vektor $(1, 1.749, 3)$ und erhalten die Validierung $(1.004, 1.684, 3.002)$. Wir sind aber nur an der zweiten Komponente interessiert. Die Validierung von $(1, 1.684, 3)$ liefert $(1.000, 1.680, 3.000)$.

Hieran erkennt man, daß man schon in diesem linearen Fall nicht mit einem Schritt auskommt, um den Wert $y = 1.680 = 0.66 \cdot 1.0 + 0.3 \cdot 3.0$ zu erreichen, der durch eine Vorverarbeitung recht einfach gewonnen wird. Zieht man bei diesem Beispiel die Kovarianz – wie eingangs vorgeschlagen – nach, so verteilt sich der Fehler recht schnell auf die erste und dritte Komponente, ohne daß die 1.680 in der zweiten Komponente erreicht werden. Dieses liegt sicherlich an der einfachen Bauweise des Beispiels, da man normalerweise erwarten kann, daß der interessierende Wert durch mehrere Gleichungen beschrieben wird.

Erhöht man die Kovarianz für y auf hohe Werte, so wird der Schätzwert für y vernachlässigt, damit wird der gewünschte Wert schneller erreicht. Allerdings erkaufte man sich dadurch numerische Instabilität, da die Kovarianzmatrix dadurch sowohl kleine als auch große Werte enthält.

Dadurch daß wir nur uns interessierende Variablen in einer Iteration übernehmen, haben wir das Fluß-Verfahren 3.7 auf einen niedrigerdimensionalen Raum eingeschränkt.

5. Zusammenfassung

Wir haben zwei Verfahren näher kennengelernt, die zur algorithmischen Bestimmung von validierten Daten unter nichtlinearen Nebenbedingungen eingesetzt werden können. Die grundlegende Arbeitsweise wurde bereits in Abbildung 6 angedeutet.

Fixpunkte des Fluß-Verfahrens 3.7 erfüllen leider nicht die Minimalitätsforderung, während die Fixpunktbedingung von 3.9 dafür sorgt, daß bedingte Extremstellen auch als Fixpunkte auftreten. Um die Vorteile beider Verfahren auszunutzen, ist also eine Mischung dieser Verfahren günstig, wobei man erst mit 3.9 iteriert und dann auf 3.7 umschaltet, um höhere Konvergenzordnung dieses Verfahrens auszunutzen, denn die Konvergenz von 3.7 ist, wie wir im Beispiel gesehen haben, nicht unbedingt gesichert. Der Umschaltzeitpunkt ist von zwei Faktoren abhängig:

- Konvergiert das Fixpunktverfahren 3.9 nicht, so kann eine Iteration nach Verfahren 3.7 noch zum Erfolg führen. Dabei kann man so vorgehen, daß, falls 3.9 nicht konvergiert, ein Schritt mit 3.7 als neue „virtuelle Messung“ berechnet wird, von der aus wieder mit 3.9 nach Fixpunkten gesucht wird.

- Hat man mit 3.9 eine gewisse Toleranzschranke für $\|f\|$ in Höhe von Maschinengenauigkeit mal Abstand zur Messung unterschritten, so sind keine Verbesserungen durch Anwendung von 3.9 zu erwarten, da Punkte in dieser Toleranz-Umgebung des aktuellen Folgengliedes x von y aus betrachtet für den Rechner ununterscheidbar sind. Um also einen Punkt zu finden, der die Nebenbedingungen noch genauer erfüllt, läßt sich wieder 3.7 einsetzen.

Dieses Vorgehen ist insofern günstig, da sich in der Anfangsphase die beiden Verfahren nur unwesentlich voneinander unterscheiden. Außerdem läßt sich der Güte-Test des linearen Falls auf diese Situation übertragen, sofern die Verbesserungen, die 3.7 liefert, nachdem mit 3.9 iteriert wurde, klein sind. Dieses Vorgehen läßt sich wie folgt zusammenfassen:

ALGORITHMUS 3.18 (Mischform). Mit der Linearisierung \bar{f} der Nebenbedingung f erhält man folgende Mischform aus 3.7 und 3.9:

$$(38) \quad \begin{aligned} x_0 &= y \\ A_k &= \nabla f \Big|_{x_k} \\ x_{k+1} &= x_m - DA_k^T (A_k DA_k^T)^{-1} \bar{f}(x_k, x_m), \end{aligned}$$

wobei $m \leq k$ einen Umschaltparameter beschreibt, mit dem sich das Verhalten dieser Mischform steuern läßt.

Dann erhalten wir für $m = k$ das Fluß-Verfahren 3.7, während für $m = 0$ das Fixpunktverfahren 3.9 entsteht.

Alle hier präsentierten Verfahren benötigen den Gradienten der Nebenbedingungen. Inwieweit sich hier Verfahren der automatischen Differentiation einsetzen lassen, die sich hier bei schwach besetztem Gradienten anbieten, wurde nicht weiter untersucht.

KAPITEL 4

Einfluß unterschiedlicher Modellgleichungen

Wir wenden uns nun der Frage zu, wie wir die Validierung durch die Wahl unterschiedlicher Modellgleichungen zum gleichen zu Grunde liegenden Modell \mathcal{M} beeinflussen können. Diese Frage drängt sich auf, da das Modell durch eine implizite Funktion gegeben ist, allerdings wurde diese Fragestellung bisher nicht weiter untersucht. So läßt sich eine explizite Funktion $y = F(x)$ auf verschiedene Weise in eine implizite Funktion $f(x, y)$ umwandeln, wie z.B.

$$\begin{aligned} f_1(x, y) &= y - F(x), \\ f_2(x, y) &= y^2 - F(x)^2 \quad (\text{elementweise bei Vektoren}) \text{ oder} \\ f_3(x, y) &= \frac{y}{F(x)} - 1, \quad \text{falls } F(x) \text{ invertierbar ist.} \end{aligned}$$

In allen hier aufgeführten Fällen gilt $(x, y) \in \mathcal{M} \iff f_i(x, y) = 0$ auf dem gemeinsamen Definitionsgebiet der drei impliziten Funktionen.

1. Lineare Transformationen der Nebenbedingungen

Zuerst zeigen wir, daß die Validierung translationsinvariant ist. Da sich die nichtlinearen Verfahren durch Iteration der Gleichung (6) ergeben, reicht es, dieses für einen Schritt der Form (6) aus dem Validierungslemma 2.2 zu untersuchen. Ist $z = y - y_0$ die verschobene Messung und $g(z) = f(z + y_0) = A(z + y_0) - b$ das verschobene lineare Modell, so ergibt sich für die Validierung dieses Wertes z

$$\begin{aligned} z^* &= z - DA^\top(ADA^\top)^{-1}g(z) \\ (39) \quad &= y - y_0 - DA^\top(ADA^\top)^{-1}(A(y - y_0 + y_0) - b) \\ &= x^* - y_0, \end{aligned}$$

also ist die Validierung des verschobenen Modells gleich der verschobenen Validierung. Damit läßt sich eine Messung auf $y = 0$ normieren. Allerdings führen dann verschiedene Messungen zu verschiedenen Modellen.

Eine weitere einfache lineare Transformation der Modellgleichung ist die Skalierung der Komponenten der impliziten Funktion f .

Auch dieses Vorgehen wirkt sich nicht auf das Validierungsergebnis aus, was folgende Proposition beweist.

PROPOSITION 4.1. *Es sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine Modellgleichung mit $n > m$ und $W \in \mathbb{R}^{m \times m}$ eine invertierbare Gewichtsmatrix. Dann ist*

die Validierung nach Gleichung (6) invariant gegenüber Skalierungen von f mit W , d.h. Wf und f besitzen für die Messung y den gleichen validierten Wert x^* .

BEWEIS. Da W eine invertierbare Matrix ist, hat Wf die gleiche Nullstellenmenge wie f . Die gewichtete Validierung x_W^* hat dann die Form

$$\begin{aligned} x_W^* &= y - D\nabla f^\top W (W\nabla f D\nabla f^\top W)^{-1} Wf(y) \\ &= y - D\nabla f^\top W W^{-1} (W\nabla f D\nabla f^\top)^{-1} Wf(y) \\ &= y - D\nabla f^\top (\nabla f D\nabla f^\top)^{-1} W^{-1} Wf(y) \\ &= y - D\nabla f^\top (\nabla f D\nabla f^\top)^{-1} f(y) \stackrel{(6)}{=} x^*. \end{aligned}$$

Dieser Schritt ist also unabhängig von der gewählten Gewichtung der Gleichungen. \square

Damit ist auch eine Iteration nach Gleichung (24) oder (33) unabhängig von einer Gewichtung, denn diese entstehen durch Iteration der Gleichung (6). Hier kann man sogar in jedem Schritt eine neue Gewichtung verwenden.

BEMERKUNG.

- Ebenso leicht läßt sich erkennen, daß sich eine Skalierung der Kovarianzmatrix D nicht auf den Validierungsschritt auswirkt. Allerdings beeinflusst solch eine Skalierung den Güte-Test 2.14.
- Eine Translation und Skalierung der Modellgleichungen und der Messung wirkt sich nicht auf das Ergebnis aus, was aus obiger Proposition 4.1 und Gleichung (39) folgt.
- Eine Skalierung der Form $f \circ S$, wobei $S \in \mathbb{R}^{n \times n}$ eine reguläre Matrix ist, ist für den Zweck der Datenvalidierung wenig sinnvoll, man kann höchstens untersuchen, wie sich das Verhalten der Datenvalidierung verändert, wenn man Messungen von redundanten Sensoren miteinander vertauscht oder bereits Mittelwerte aus diesen Sensordaten verwendet.
- Das Ergebnis von Proposition 4.1 ist nur für eine exakte Rechnung wahr. Aber diese Skalierung läßt sich dazu einsetzen, die numerischen Auslöschungen, die sich durch Subtraktion fast gleicher Werte bei einer Rechnung mit endlicher Genauigkeit ergeben, mittels einer Gewichtsmatrix auf ein vergleichbares Maß zu skalieren.

Die letzte Bemerkung bedarf noch weiterer Erklärung. Eine typische Nebenbedingungsgleichung vergleicht einen von den Eingängen eines Bauteils der Anlage abhängigen Wert z_i mit einem von den Ausgängen abhängigen Wert z_o .

Vergleichen wir den Quotienten dieser beiden Werte mit 1, so sind die numerischen Fehler unabhängig von der Größenordnung von z_i und z_o ,

allerdings ist der Gradient dieser Nebenbedingungsgleichung von keiner einfachen Bauart.

Vergleicht man dagegen die Differenz von z_i und z_o mit der Null, so besitzt der Gradient eine einfachere Struktur, allerdings wirkt sich die Größenordnung der Werte z_i und z_o auf den numerischen Fehler aus. Kennt man nun aber die Größenordnung dieser Werte, so läßt sich die Validierungsgleichung $z_i - z_o$ entsprechend skalieren, so daß für alle Gleichungen dieses Typs in den Nebenbedingungen die Fehler durch numerische Auslöschungen auf ein vergleichbares Maß gebracht werden.

2. Nichtlineare Transformationen der Nebenbedingungen

Nun wollen wir nichtlineare Umformungen der Nebenbedingungen untersuchen, so daß das Modell, also die Nullstellenmenge der Nebenbedingungen, erhalten bleibt.

Nennen wir zwei Funktionen $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ äquivalent, wenn sie die gleiche Nullstellenmenge besitzen:

$$f \sim g : \iff (f(x) = 0 \iff g(x) = 0).$$

Damit repräsentieren dann beide Funktionen das gleiche Modell $\mathcal{M} = f^{-1}(0) = g^{-1}(0)$. Wir suchen also nach einem für die Validierung geeigneten Repräsentanten aus dieser Äquivalenzklasse.

Für den Fall $m = 1$ haben wir folgende Beispiele, die das Modell \mathcal{M} bei Umformung der Gleichung $f : \mathbb{R}^n \rightarrow \mathbb{R}$ erhalten:

- Multiplikation mit einer außerhalb von \mathcal{M} nicht verschwindenden \mathcal{C}^2 -Funktion $m : \mathbb{R}^n \rightarrow \mathbb{R}$, $g(x) = m(x)f(x)$:

$$\begin{aligned} g(x) &= m(x)f(x) = 0 \\ \iff m(x) &= 0 \text{ oder } f(x) = 0 \\ \iff f(x) &= 0. \end{aligned}$$

- Zerlegt man f in $f = f_1 - f_2$, so besitzt die Funktion $g = h(f_1(\cdot)) - h(f_2(\cdot))$ die gleiche Nullstellenmenge \mathcal{M} wie f , wenn $h : \mathbb{R} \rightarrow \mathbb{R}$ eine injektive \mathcal{C}^2 -Funktion ist:

$$\begin{aligned} g(x) = 0 &\implies h(f_1(x)) = h(f_2(x)) \\ &\implies f_1(x) = f_2(x) \implies f(x) = 0 \end{aligned}$$

Im Spezialfall $f_2 \equiv 0$ ergibt sich dann $g(x) = h(f(x)) - h(0)$.

- Sind f_1 und f_2 äquivalent, so auch $|f_1| + |f_2|$ und $f_1 \cdot f_2$.

Dabei wird \mathcal{C}^2 -Glattheit gefordert, damit die erforderlichen Ableitungen für ein Validierungsverfahren existieren.

Untersuchen wir das Ganze an einem Beispiel.

BEISPIEL 4.2. Nehmen wir an, wir haben Messungen der reellwertigen Größen C, Q und U , die dem Gesetz $Q = CU$ genügen. Als Validierungsgleichungen bieten sich an

- $Q - CU$,
- $\frac{Q}{C} - U$, $(C \neq 0)$
- $\frac{Q}{CU} - 1$, $(C, U \neq 0)$
- $\ln Q - (\ln C + \ln U)$. $(Q, C, U > 0)$

Während die zweite und dritte Gleichung aus der ersten Gleichung durch Multiplikation mit nichtverschwindenden Funktionen entstehen, ist die vierte Gleichung ein Beispiel für die Transformation mit einer injektiven Funktion.

Abbildung 9 zeigt die Wirkung eines Validierungsschrittes aus Lemma 2.2 auf ein äquidistantes Gitter in der (C, U) -Ebene¹ und den Annahme-Bereich des Verfahrens für die einzelnen Gleichungen. Dabei wurde nur ein Validierungsschritt ausgeführt, es wurde also einmal Gleichung (6) auf die linearisierten Modellgleichungen angewandt. Die Änderung in Q -Richtung wurde dabei vernachlässigt. Für $Q = 16$, $C = 4$ und $U = 4$ ergeben Gleichungen 1 und 4 vernünftige Ergebnisse, denn die Annahmebereiche dieser Gleichungen (graue Schattierung in Abbildung 9) sind relativ klein und die meisten Punkte liegen nach einer Iteration im Annahmebereich, während für $Q = 16$, $C = 6.4$ und $U = 2.5$ die Gleichungen 1 und 2 zu vertrauenswürdigen Ergebnissen führen – die zu dieser Situation gehörige Abbildung wird hier nicht gezeigt.

Die Akzeptanzbereiche der einzelnen Gleichungen wurden mithilfe des oben beschriebenen Tests 2.14 berechnet. Da wir nur einen Verfahrensschritt durchgeführt haben, ist die Normalverteilungsannahme noch immer gerechtfertigt.

An diesem Beispiel läßt sich erkennen, daß man keine „einfachen“ Rezepte zur Wahl geeigneter Gleichungen ablesen kann. Wir müssen uns also um einen theoretischen Unterbau bemühen, der wiederum durch die Theorie der Lagrangeschen Multiplikatoren gegeben wird. Der Satz über implizite Funktionen liefert für die Lagrange-Funktion folgende Sensitivitätsaussage.

SATZ 4.3 (Sensitivitätstheorem). *Es seien $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $J : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ zweifach stetig differenzierbare Funktionen. Betrachte die Problem-Familie*

$$\begin{aligned} & \text{Minimiere } J(r) \\ & \text{unter der Bedingung } f(r) = u, \end{aligned}$$

parametrisiert durch $u \in \mathbb{R}^m$. Für $u = 0$ habe das Problem die reguläre Lösung x^ . Zusammen mit seinem zugehörigen Lagrange-Vektor λ^* erfülle sie die hinreichenden Bedingungen nach Proposition 3.4. Dann gibt es eine offene Umgebung U der 0, so daß für alle $v \in U$ $x(v)$ und $\lambda(v)$ ein Lagrange-Paar bilden. Dabei sind $x(\cdot)$ und $\lambda(\cdot)$ stetig*

¹ $Q = 16$, $C, U = 1, \dots, 7$ mit einer gleichförmigen Einteilung.

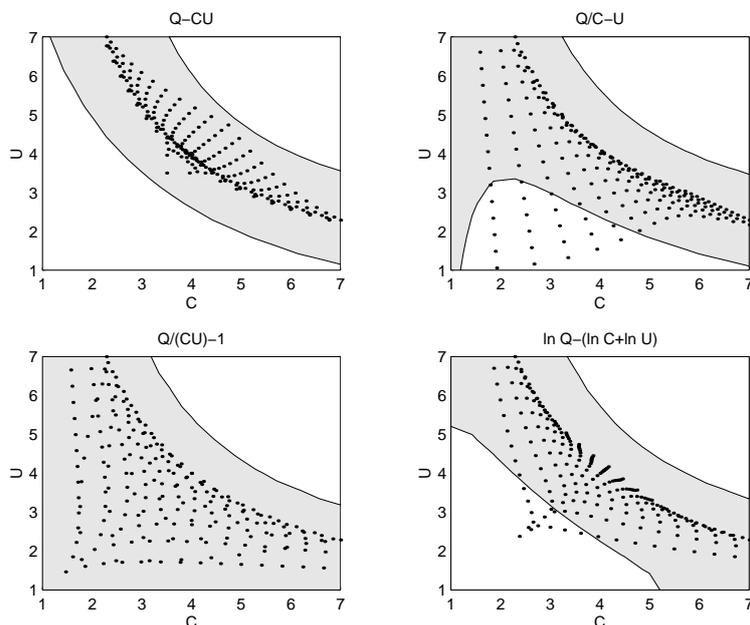


ABBILDUNG 9. Wirkung des Validierungsschrittes auf ein Gitter und Akzeptanz-Bereich des Tests ($Q = 16$, 5% Varianz)

differenzierbare Funktionen auf U . Außerdem gilt für $v \in U$:

$$\nabla p(u) = \lambda(u),$$

wobei $p(u) = J(x(u))$ die durch u parametrisierte optimale Kostenfunktion bezeichnet.

BEWEIS. Siehe [Ber95, Proposition 3.2.2].

Eine Folgerung aus diesem Satz ist also, daß uns der Vektor der Lagrangeschen Multiplikatoren eine Aussage über die Störungsanfälligkeit der einzelnen Nebenbedingungen gibt, denn nach diesem Sensitivitätstheorem ist die Ableitung der Kostenfunktion entlang der optimalen Werte für den Parameter $u = 0$ gerade der Vektor der Lagrangeschen Multiplikatoren $\lambda^* = \lambda(0)$. Wir sollten uns also für solche Gleichungen entscheiden, die für einen möglichst großen Bereich ein kleines λ besitzen, damit ist dann die Störungsanfälligkeit gering. Im Folgenden wollen wir die Verbindung zwischen dem Vektor λ der Lagrangeschen Multiplikatoren und dem Güte-Test noch etwas genauer herausarbeiten.

Zunächst suchen wir eine explizite Darstellung für λ . Diese haben wir allerdings schon für den Fall linearer Modellgleichungen erhalten. Denn wir haben im Beispiel 3.6 eine Lösung des (QP)-Problems hergeleitet

und dabei auch eine Darstellung des Vektors der Lagrangeschen Multiplikatoren λ gefunden, und zwar

$$\lambda = -(ADA^\top)^{-1}(Ay - b),$$

wobei y die Messung und $f(x) = Ax - b$ die Modellgleichung ist. Allerdings wirken sich im Unterschied zur Validierung sowohl Gewichtungen der Nebenbedingungen als auch Skalierungen der Kovarianz-Matrix auf λ aus.

Die Verbindung zum Güte-Test 2.14 ist offensichtlich, es wurde lediglich ein Faktor weggelassen, der zur Bestimmung der Teststatistik des Güte-Tests benötigt wird:

$$X = (Ay - b)^\top (ADA^\top)^{-1} (Ay - b).$$

Diese Verbindung von Güte-Test und Lagrangeschen Multiplikatoren läßt sich vom linearen Fall sofort auf die Iteration nach dem Fixpunkt-Algorithmus 3.9 übertragen, da dieser mit einer Linearisierung arbeitet. Wir erhalten also als Resultat, daß wir dann sinnvolle Modellgleichungen verwenden, wenn der Güte-Test nach Proposition 2.14 vernünftige Ergebnisse liefert. Diese zugegebenermaßen schwache Formulierung zeigt, daß man bei der Wahl passender Modellgleichungen ein gewisses Gefühl entwickeln muß.

3. Transformationsinvarianz der Validierung

Wie wir im Beispiel 4.2 gesehen habe, erhalten wir für einen Schritt der Validierung recht unterschiedliche Ergebnisse. Um so überraschender ist die folgende Aussage, die für das Fixpunktverfahren 3.9 gilt.

PROPOSITION 4.4. *Die Menge der möglichen Fixpunkte $\mathcal{F}_y(f)$ des in Algorithmus 3.9 beschriebenen Verfahrens zu gegebenem Startwert y ist unabhängig von der verwendeten Form der Nebenbedingung f zum festen Modell \mathcal{M} .*

Die möglichen Fixpunkte werden durch die Fixpunktgleichung aus Proposition 3.10 gegeben, also

$$\mathcal{F}_y(f) = \{x \in \mathcal{M} \mid (I - DA^\top (ADA^\top)^{-1} A)(x - y) = 0\}$$

mit $A = \nabla f|_x$. Die Nebenbedingungen f und g zum gleichen Modell $\mathcal{M} \subset \mathbb{R}^n$ erfüllen folgende Kriterien:

- Sie besitzen die gleiche Nullstellenmenge $f^{-1}(0) = g^{-1}(0) = \mathcal{M} \cap \mathcal{U}$ in einer Umgebung \mathcal{U} von y . Es gilt außerdem $\mathcal{M} \cap \mathcal{U} \neq \emptyset$, damit überhaupt Lösungen existieren können.
- f und g sind regulär auf \mathcal{U} . Diese Forderung ergibt sich daraus, daß das Modell \mathcal{M} eine Untermannigfaltigkeit ist.
- f und g sind zweifach stetig differenzierbar auf \mathcal{U} .

Damit sind insbesondere die Voraussetzungen für die Existenz und Eindeutigkeit von Lösungen der nichtlinearen Validierung erfüllt. Proposition 4.4 besagt dann, daß $\mathcal{F}_y f = \mathcal{F}_y g$ auf \mathcal{U} gilt. Ist insbesondere $\mathcal{F}_y(f)$ einelementig und konvergiert das Verfahren 3.9 sowohl für f als auch für g als zu Grunde liegende Nebenbedingungen, so haben beide Verfahren denselben Grenzwert.

Zum Beweis der Proposition dient folgender Satz der Differentialgeometrie.

SATZ 4.5. *Ist \mathcal{M} eine d -dimensionale Untermannigfaltigkeit des \mathbb{R}^n , so ist für jedes $p \in \mathcal{M}$ der Tangentialraum $T_p\mathcal{M}$ ein d -dimensionaler Untervektorraum. Ist $f : \mathbb{R}^n \supset \mathcal{U} \rightarrow \mathbb{R}^m$ eine lokale Gleichung für \mathcal{M} , d.h. $f^{-1}(0) = \mathcal{M} \cap \mathcal{U}$, so ist*

$$T_p\mathcal{M} = \ker \nabla f|_p.$$

BEWEIS. ([Hor98, Satz 24]) Die Richtungsvektoren aller Wege in \mathcal{M} durch p spannen den Tangentialraum $T_p\mathcal{M}$ von \mathcal{M} im Punkt p auf. Zu zeigen sind also folgende Behauptungen:

1. Ist φ ein Weg durch $p = \varphi(0)$ in \mathcal{M} , so liegt $\varphi'(0)$ im Kern von $\nabla f|_p$.
2. Zu $w \in \ker \nabla f|_p$ gibt es einen Weg in \mathcal{M} mit $\varphi(t) = p$ und $\varphi'(t) = w$.

Ad 1.) Ist φ ein Weg in \mathcal{M} durch p , so ist $f \circ \varphi = 0$ in einer Umgebung der Null, da der Weg so parametrisiert ist, daß $\varphi(0) = p$ gilt. Also ist

$$\nabla f|_{\varphi(0)} \varphi'(0) = \nabla f|_p \varphi'(0) = 0.$$

Damit ist $\varphi'(0) \in \ker \nabla f|_p \implies T_p\mathcal{M} \subset \ker \nabla f|_p$.

Ad 2.) Für die andere Inklusionsrichtung benutzen wir den Satz über implizite Funktionen: Es sei \mathcal{U} so klein gewählt, daß $\mathcal{M} \cap \mathcal{U}$ Graph einer differenzierbaren Funktion $g : \mathbb{R}^d \supset \mathcal{V} \rightarrow \mathbb{R}^m$ mit $d + m = n$ ist. Es gilt dann

$$f(x, g(x)) = 0 \quad \forall x \in \mathcal{V},$$

also

$$\nabla f_1(x, g(x)) + \nabla f_2(x, g(x)) \nabla g(x) = 0,$$

wobei ∇f_1 die partielle Ableitung in Richtung der ersten d Koordinaten und ∇f_2 die partielle Ableitung in Richtung der letzten m Koordinaten ist. Um solch ein g zu finden, muß man möglicherweise die Koordinaten vertauschen. Für $p = (x_0, g(x_0))$ hat man

$$\text{Im}(\text{id}, \nabla g|_{x_0}) \subset \ker \nabla f|_p,$$

wobei $(\text{id}, \nabla g|_{x_0}) : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^m = \mathbb{R}^n$. Nun ist $\nabla f|_p$ surjektiv, damit ist die Dimension des Kernes $n - m = d$. Dies ist auch die Dimension des Bildes von $(\text{id}, \nabla g|_{x_0})$, daraus folgt die Gleichheit des Kernes des Gradienten von f mit dem Graphen des Gradienten von g ,

$$\ker \nabla f|_p = \Gamma_{\nabla g|_{x_0}}.$$

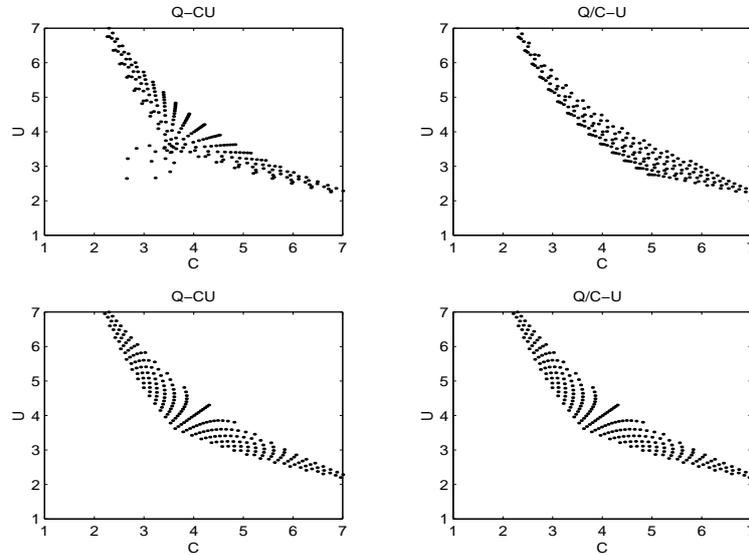


ABBILDUNG 10. Wirkung unterschiedlicher Validierungsverfahren: oben Flußverfahren, unten Fixpunktverfahren.

Ist daher $w \in \ker \nabla f_p$, so gibt es ein $v \in \mathbb{R}^d$ mit $(v, \nabla g|_{x_0} v) = w$. Die Abbildung $t \mapsto x_0 + tv$ in \mathbb{R}^d liefert den Weg

$$\varphi(t) = (x_0 + tv, g(x_0 + tv)) \in \mathcal{M} \cap \mathcal{U}$$

für hinreichend kleine t . Dann ist

$$\varphi'(t) = (v, \nabla g|_{x_0} v) = w.$$

Damit ist $T_p \mathcal{M} = \ker \nabla f_p$. \square

Der Tangentialraum ist also unabhängig von der Wahl der lokalen Gleichung und wird gegeben durch den Kern des Gradienten der lokalen Gleichung. Damit läßt sich jetzt Proposition 4.4 beweisen.

BEWEIS. (von Proposition 4.4) Um eine Basis des Tangentialraumes für die Elemente der Fixpunktmenge \mathcal{F}_y zu finden, benutzen wir eine Moore-Penrose-Pseudo-Inverse, die auf den Kern von ∇f projiziert. Diese ist unabhängig von der gewählten lokalen Gleichung des Modells, da der Kern nach vorherstehendem Satz nicht von der lokalen Gleichung abhängt. Damit ist die Validierung als spezielle Projektion unabhängig von der gewählten Gleichung. \square

Nun wollen wir Beispiel 4.2 noch einmal unter nichtlinearer Validierung untersuchen.

BEISPIEL 4.6. Wir untersuchen wie in Beispiel 4.2 wieder die Wirkung der Validierung auf ein Gitter (q, c, u) mit $q = 16$, $c, u = 1, \dots, 7$. Abbildung 10 zeigt das Verhalten der Nebenbedingungen $Q - CU$ und

$Q/C-U$ unter dem Flußverfahren 3.7 (obere Reihe) und dem Fixpunktverfahren 3.9 (untere Reihe). Die Abbruchbedingung ist so gewählt, daß entweder eine Maximalanzahl von Iterationen überschritten wird oder daß die Differenz aufeinander folgender Validierungsschätzungen x_k und x_{k+1} normmäßig klein ist. Auf die Angabe von Akzeptanzbereichen haben wir hier verzichtet, da dieses – wie wir in Abschnitt 3.3 gesehen haben – nicht ohne weiteres möglich ist.

Wie wir nach Proposition 4.4 erwartet haben, sehen die Bilder der unteren Reihe gleich aus. Die durch das Flußverfahren verformten Gitterpunkte sehen dagegen für unterschiedliche Modellgleichungen auch tatsächlich unterschiedlich aus.

Allerdings ist zu bemerken, daß das Fixpunktverfahren mehrmals die maximale Iterationszahl (hier 20) erreichte, während dies beim Flußverfahren nicht auftrat.

4. Zusammenfassung

Wir haben untersucht, wie sich verschiedene Gleichungen zum gleichen Modell \mathcal{M} unter der Validierung verhalten. Dabei haben wir festgestellt, daß die Ergebnisse des Fixpunktalgorithmus 3.9 unter schwachen Bedingungen unabhängig von der Wahl der Gleichung zu dem Modell \mathcal{M} sind. Allerdings haben wir nun das Problem, überhaupt die Konvergenz der Validierung sicherzustellen und dann die Konvergenzgeschwindigkeit für die unterschiedlichen Bauarten der Nebenbedingung zu beurteilen. Dazu muß man sich überlegen, wie gut die Approximation der Funktion durch die Linearisierung ist, es sind also Bedingungen an die zweiten Ableitungen der Nebenbedingungsgleichungen nötig. Diese Untersuchung führen wir hier nicht durch.

Sensorauswahl und Modellerweiterung

„Torturing the data until they confess.“

A. Miller

Im vorherigen Abschnitt haben wir uns die Auswirkungen angeschaut, die bei unterschiedlichen Nebenbedingungen zum gleichen Modell auftreten. Nun untersuchen wir, welches Modell besonders gut zu einem gegebenen Meßwert y paßt. Das hat nur indirekt mit der Datenvalidierung zu tun, hier wird nur nach geeigneten Voraussetzungen für die Datenvalidierung gesucht. In der Datenvalidierungsliteratur findet sich die Diskussion dieses Problems nicht. Wir wollen also verschiedene Modelle miteinander vergleichen, die auch auf unterschiedliche Sätze von Sensoren zurückgreifen können.

Angenommen, wir besitzen für den gleichen maximalen Sensorensatz verschiedene Modelle von Nebenbedingungen. Wir suchen nun nach Strategien, einen geeigneten Vertreter aus der Menge von Modellgleichungen ausfindig zu machen.

Man muß zwischen folgenden Vorgehensweisen unterscheiden (vgl. [Mil90]):

- Die Vorwärtssuche startet mit einem möglichst einfachen Modell und konstruiert durch schrittweises Hinzunehmen weiterer Variablen eine aufsteigende Modellfolge, um zu einem optimalen Modell zu gelangen.
- Die Rückwärtssuche startet demgegenüber mit einem möglichst umfangreichen (und zutreffenden) Modell und konstruiert eine absteigende Modellfolge.
- Bei der schrittweise Regression wird nach dem Hinzunehmen einer Variable geprüft, ob man eine schon im Modell enthaltene Variable wieder herausnehmen kann.
- Eine weitere Möglichkeit ist die Durchführung einer sequentiellen Ersetzung. Dabei überprüft man, ob sich Variablen durch andere Variablen ersetzen lassen. Es lassen sich auch Varianten mit verzögerter Ersetzung angeben.
- Eine Berechnung aller Teilmengen (bis zu einer gewissen Größe) von Modellgleichungen zu einem Gleichungssystem von Nebenbedingungen ist nur für einen geringen Satz von Variablen praktikierbar.

Durch wiederholte Anwendung dieser Vorgehensweisen erhält man ein Grenzmodell. Aber es gibt keinen Grund, daß die mit verschiedenen Verfahrensweisen gefundenen Modelle übereinstimmen.

Da diese Verfahren noch immer von einer möglicherweise simulierten Messung abhängen, besteht außerdem die Gefahr einer Überanpassung, d.h. für eine geringe Abweichung vom Referenzpunkt liefert das Modell schon keine sinnvollen Werte mehr.

1. Der F-Test

Ein Vorgehen zur Modellauswahl muß für alle infrage kommenden Meßwerte eine Art Siebprozeß bilden, z.B. indem man zu einzelnen Sensormessungen alle Modelle auswählt, die akzeptable Ergebnisse liefern. Aus dieser Menge oder durch eine Modellerweiterung wird wiederum für weitere erfaßte Werte ein akzeptables Modell gesucht usw. Die Auswahl der Modelle hängt dann von der Reihenfolge der Sensorwerte ab. Bei dieser Auswahl benutzen wir die Verteilungsaussage über die Kosten, die auch schon zu dem in Proposition 2.14 vorgestellten Güte-Test geführt hat. Der Quotient der χ^2 -verteilten Kosten zweier unterschiedlicher Modelle besitzt nämlich wieder eine bekannte Verteilung.

Wir betrachten für die Sensorauswahl folgende Situation.

$$\begin{aligned} \mathcal{B} &= \{y_1, \dots, y_r\} && \text{Basis-Sensor-Menge} \\ \mathcal{X} &= \{y_{r+1}, \dots, y_n\} && \text{Erweiterte Sensor-Menge} \end{aligned}$$

$$\begin{aligned} f_{\mathcal{B}} : \mathbb{R}^r &\rightarrow \mathbb{R}^s && \text{Gleichung des Grundmodells} \\ f_{\mathcal{A}} : \mathbb{R}^n &\rightarrow \mathbb{R}^m && \text{Gleichung des erweiterten Modells mit} \end{aligned}$$

$$f_{\mathcal{A}} = \begin{cases} f_{\mathcal{B}}(y_1, \dots, y_r) \\ f_{\mathcal{X}}(y_1, \dots, y_r, \dots, y_n) \end{cases}$$

Wir wollen nun das erweiterte Modell mit dem Grundmodell vergleichen. Dazu bietet sich ein Test an, der prüft, ob das Untermodell $f_{\mathcal{B}}^{-1}(0)$ unter dem Obermodell $f_{\mathcal{A}}^{-1}(0)$ gilt. Dadurch läßt sich feststellen, ob sich ein Modell vereinfachen läßt. Aber auch aus umgekehrter Sicht ist dieser Test sinnvoll, man kann damit prüfen, ob ein umfassenderes Obermodell besser geeignet ist als das Untermodell. Damit ist dieser Test für Modellauswahl mittels Vorwärts- oder Rückwärtssuche geeignet.

Dieser Test vergleicht die Abweichung der Validierung von der Messung unter dem Untermodell mit der Abweichung der Messung von der Validierung unter dem Obermodell. Das Normquadrat der Abweichungen wird gerade durch die Kosten $J_{\mathcal{A}}$ bzw. $J_{\mathcal{B}}$ gegeben.

Setzen wir dabei die echte Inklusion $f_{\mathcal{B}}^{-1}(0) \subsetneq f_{\mathcal{A}}^{-1}(0)$ der beiden Modelle voraus, so sind die Kosten unter dem Untermodell höher als die Kosten unter dem Obermodell.

Im *linearen Fall* besitzt die Teststatistik folgende Struktur (siehe [Bre73, Chapter 10]):

$$(40) \quad F := \frac{(J_{\mathcal{B}}(\tilde{r}_{\mathcal{B}}) - J_{\mathcal{A}}(\tilde{r}_{\mathcal{A}}))/(m - s)}{J_{\mathcal{A}}(\tilde{r}_{\mathcal{A}})/(n - m)} \stackrel{\mathcal{L}}{\sim} F_{ZFG, NFG},$$

diese Teststatistik gehorcht – unter der Bedingung, daß das zu ersetzende Modell zutrifft – einer F -Verteilung mit Zählerfreiheitsgrad $ZFG = m - s$ und Nennerfreiheitsgrad $NFG = n - m$, wobei $J(\tilde{r}.)$ die Kostenfunktion im optimalen Residuum des jeweiligen linearen Modells bezeichne. Die Teststatistik (40) wird nicht direkt durch den Quotienten $\frac{J_{\mathcal{B}}}{J_{\mathcal{A}}}$ der Kosten gegeben, es wird noch die Konstante $\frac{m-s}{n-m}$ abgezogen, damit wird die Verteilung von F zentralisiert.

Dieser Term (40) ist nun eine Kenngröße für die Einhaltung der Nullhypothese, dafür daß das Untermodell Gültigkeit besitzt. Denn sofern die Nullhypothese zutrifft, ist $J_{\mathcal{A}}(x_{\mathcal{A}} - y)/FG(\mathcal{A})$ eine erwartungstreue Schätzung der Varianz unter dem Obermodell, und $\Delta\text{Dev}/\Delta FG$ eine erwartungstreue Schätzung für die zusätzliche Varianz unter dem Untermodell, wobei ΔFG den Unterschied in den Freiheitsgraden der beiden Modelle und

$$\Delta\text{Dev} := \|x_{\mathcal{A}} - x_{\mathcal{B}}\|^2 = J_{\mathcal{B}}(x_{\mathcal{B}} - y) - J_{\mathcal{A}}(x_{\mathcal{A}} - y)$$

den Unterschied in der Abweichung (Deviance) der beiden Modelle angibt.

Solche Varianzschätzungen sind χ^2 -verteilt und damit ist der Quotient nach Definition B.7 dieser beiden Varianzschätzungen F -verteilt. Diese Aussagen finden sich in [Bre73, Theorem 10.34].

Dieser Test hängt sehr stark von der Normalverteilungsannahme der Störungen ab, denn daraus folgt die χ^2 -Verteilung von Zähler und Nenner von (40), und damit die F -Verteilungsannahme als Quotient zweier χ^2 -Verteilungen.

2. Modellsuchverfahren

Um bei der Auslegung einer Anlage schon ein Modellsuchverfahren zu starten, schlagen wir folgendes Verfahren vor. Man gebe für ausgesuchte Sensoren (als Index des Meßvektors) gewünschte Genauigkeiten z.B. in Form von Konfidenzintervallen vor. Wir wollen also ein Modell finden, das diese gewünschte Genauigkeit erfüllt. Dazu suchen wir ein geeignetes Modell, das aus einer Familie von um einen bekannten Betriebspunkt x^* linearisierten Modellen gewählt wird, so daß die validierten Meßwerte x^* mit 95% Wahrscheinlichkeit den wahren Wert \hat{x} in den spezifizierten Intervallen $x^*_i \pm s_i$ einfangen. Ist dieses mit der gewählten Modellmenge nicht möglich, so muß man andere Strategien hinzuziehen, z.B. Auswahl neuer Sensoren oder ein anderer Betriebspunkt, um den das Modell linearisiert wird. Ohne Beschränkung der

Allgemeinheit seien für die ersten r Sensoren Konfidenz-Intervalle vorgegeben.

Dieses soeben beschriebene Verfahren läßt sich folgendermaßen strukturieren:

ALGORITHMUS 5.1 (Modellsuche).

1. Konfidenzintervalle in Varianzen mittels $\sigma_i^2 = (\frac{s_i}{1,96})^2$ umrechnen, wobei 1,96 das zweiseitige 5%-Quantil der Normalverteilung ist, oder es werden σ_i direkt spezifiziert.
2. Starte mit einem leeren Modell F_0 mit $m_0 = 0$ Nebenbedingungen. Setze $D_0 = D$, $i = 1$, $k = 0$.
3. Gilt $(D_k)_{ii} \leq \sigma_i^2$, so reicht die Genauigkeit des aktuellen Modells aus. Wenn nicht, muß man ein neues Modell wählen wie unter 5. beschrieben.
4. Wähle die nächste Vorgabe $i \leftarrow i + 1$. Sind alle Vorgaben erfüllt ($i > r$), so ist man fertig. Sonst zurück zu Punkt 3.
5. Wähle das nächste Modell $k \leftarrow k + 1$. Gibt es keine weiteren Modelle mehr, so sind die vorgegebenen Bedingungen nicht erfüllbar und das Verfahren bricht ab.
6. Berechne die induzierte Kovarianz-Matrix D_k^* der Residuen zum Modell F_k . Setze $D_k = D - D_k^*$ (vgl. Abschnitt 2.3). Für alle bisher geprüften Meßstellen $j < i$ soll $(D_k)_{jj} \leq \sigma_j^2$ sein. Ist diese Bedingung erfüllt, so eignet sich dieses Modell für einen weiteren Durchlauf (Punkt 3). Sonst muß man wieder zurück zur Modellsuche (Punkt 5).

Der Programmablaufplan wird in einer vereinfachten Form in Abbildung 11 angegeben. Dabei wurde die Überprüfung der Diagonalelemente zu einem Schritt zusammengefaßt.

BEMERKUNG. In Schritt 5 läßt sich die Modellauswahl praktischerweise durch Hinzufügen oder Ersetzen von jeweils einer Gleichung durchführen, wobei unter Verwendung des Matrixinversionslemmas 3.5 der Rechenaufwand relativ klein gehalten werden kann. Dann ist $m_k + 1 \geq m_{k+1} \geq m_k$. Haben wir uns für diese Modellerweiterungsstrategie entschieden, so die Überprüfung der Diagonalelemente der Kovarianzmatrix der validierten Messung in Schritt 6 nicht mehr nötig. Dieses wollen wir im nächsten Abschnitt 5.3 noch genauer begründen.

Dieses soeben beschriebene Verfahren überprüft jeweils die Randverteilung der validierten Messung in Richtung der Einheitsvektoren e_i auf Normalverteilung. Ist nämlich $Z \stackrel{\mathcal{L}}{\sim} \mathcal{N}_n(0, \Sigma)$ eine multivariat-normalverteilte Zufallsvariable, so gilt für alle $t \in \mathbb{R}^n$:

$$\langle t, Z \rangle \stackrel{\mathcal{L}}{\sim} \mathcal{N}_1(0, t^T \Sigma t),$$

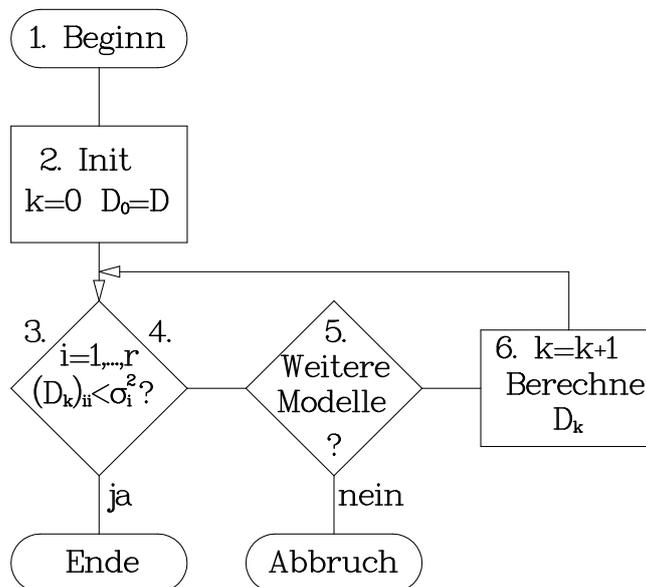


ABBILDUNG 11. Vereinfachter Ablaufplan für die Modellauswahl.

denn nach der Reduktionsmethode von Cramér-Wold erhält man für die m -dimensionale Zufallsvektoren X_n und X die folgende Aussage:

$$X_n \stackrel{\mathcal{L}}{\rightarrow} X \iff \langle t, X_n \rangle \stackrel{\mathcal{L}}{\rightarrow} \langle t, X \rangle \quad \forall t \in \mathbb{R}^m.$$

Siehe dazu auch [Bil79, Section 29].

Damit erhalten wir nach der Überprüfung der Randverteilungen auch Aussagen über die Gesamtverteilung. Für $t = e_i$ erhält man dann $\langle e_i, Z \rangle \stackrel{\mathcal{L}}{\sim} \mathcal{N}_1(0, \Sigma_{ii})$. Insbesondere läßt sich jedes $t \in \mathbb{R}^m$ durch Linearkombination von Einheitsvektoren e_i darstellen. Die Kovarianzmatrix der validierten Messung wird hier gegeben durch $\Sigma = D - D^*$ nach Gleichung (12), wobei D^* die induzierte Kovarianzmatrix des aktuellen Modells ist.

3. Modellerweiterung

Bei einer Modellsuche nach dem Verfahren 5.1 ist es nötig, beim Modellwechsel die „alten“ Konfidenzintervalle zu überprüfen.

Wenn man aber bei einer Modellsuche das Modell nur um Gleichungen erweitert, so gilt immer $D_{k+1}^* > D_k^*$, die induzierte Kovarianz der Residuen des größeren Modells ist also immer kleiner als die des kleineren Modells. Den Beweis dieser Eigenschaft liefern die nächsten Propositionen. Dabei wird zunächst ein allgemeines Resultat hergeleitet, das sich dann wieder auf die vorliegende Situation anwenden läßt.

PROPOSITION 5.2. *Sind \mathcal{X} und \mathcal{Y} lineare Unterräume des \mathbb{R}^n und gilt $\mathcal{Y} \subset \mathcal{X}$, so ist die Differenz der Q -orthogonalen Projektoren auf \mathcal{X} und*

\mathcal{Y} wiederum eine Projektion $\mathcal{P}_X^Q - \mathcal{P}_Y^Q = \mathcal{P}_{\langle (I - \mathcal{P}_Y^Q)X \rangle}^Q$, wobei X eine Basis von \mathcal{X} ist.

BEWEIS. Es sei $Y = (y_1, \dots, y_k)$ eine Basis von \mathcal{Y} , ergänze diese zu einer Basis $X = (Y, x_{k+1}, \dots, x_l)$ von \mathcal{X} und diese wiederum zu einer Basis des Gesamttraumes \mathbb{R}^n . Orthonormiere diese Basis mittels eines Gram-Schmidt-Verfahrens zum schiefen Skalarprodukt $\langle \cdot, \cdot \rangle_Q$. Mit dieser Basis gilt dann

$$\mathcal{P}_Y^Q = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \quad \text{und} \quad \mathcal{P}_X^Q = \begin{pmatrix} I_l & 0 \\ 0 & 0 \end{pmatrix}.$$

$$(I - \mathcal{P}_{\langle Y \rangle}^Q)X = \begin{pmatrix} 0 & 0 \\ 0 & I_{n-k} \end{pmatrix} \begin{pmatrix} I_l \\ 0 \end{pmatrix}$$

□

Dieses Resultat läßt sich nun auf die Modellerweiterung anwenden. Wir haben uns hierbei auf um einen gemeinsamen Bezugspunkt linearisierte Modelle beschränkt. Damit ist das erweiterte Modell ein Oberraum des alten Modells.

PROPOSITION 5.3. *Die Modellerweiterung um jeweils eine Gleichung verkleinert in jedem Schritt die Kovarianz der validierten Werte: Sei D_k^* die induzierte Kovarianzmatrix des Modells $f_k : \mathbb{R}^n \rightarrow \mathbb{R}^k$, und das erweiterte Modell gegeben durch*

$$f_{k+1} = \begin{pmatrix} f_k \\ g_{k+1} \end{pmatrix} = (g_1, \dots, g_k, g_{k+1})^\top,$$

mit $g_{k+1} : \mathbb{R}^n \rightarrow \mathbb{R}$. Dann gilt für die induzierte Kovarianzmatrix D_{k+1}^* von f_{k+1} :

$$D - D_{k+1}^* \leq D - D_k^*,$$

bzw. $D_{k+1}^* - D_k^* \succ 0$.

BEWEIS. Es sei $l = k + 1$. Aus der vorherigen Proposition erhalten wir mit $Q = D^{-1}$ eine Darstellung der Projektion \mathcal{P}_l auf das erweiterte Modell durch die Projektion auf das alte Modell \mathcal{P}_k . Dabei ist zu beachten, daß \mathcal{P}_k durch die Ableitungen der ersten k Modellgleichungen g_1, \dots, g_k beschrieben wird, also ist $g_i = \mathcal{P}_j g_i$ für $i = 1, \dots, k$, $j = i, \dots, l$. Die Ableitung der l -ten Gleichung wird als Spaltenvektor mit ∂g_l bezeichnet.

$$D_l^* - D_k^* = D\mathcal{P}_l - D\mathcal{P}_k = D\mathcal{P}_{\langle (I - \mathcal{P}_k)\partial g_l \rangle}$$

$$(41) \quad \stackrel{(55)}{=} (D - D_k^*)\partial g_l (\partial g_l^\top (D - D_k^*)\partial g_l)^{-1} \partial g_l^\top (D - D_k^*) \geq 0$$

$$\iff \partial g_l^\top (D - D_k^*)\partial g_l > 0$$

Diese Bedingung ist aber wegen der Regularitätsvoraussetzungen an die Modellgleichungen erfüllt, denn damit ist $(D - D_k^*)\partial g_l \neq 0$. □

BEMERKUNG. Aus dem Beweis erhält man eine Rekursionsformel, mit der man ohne Matrix-Inversion die D^* -Kovarianz-Matrix zum aktuellen Modell aufbauen kann. Insbesondere für die Modellsuche bietet sich dieses Verfahren an, da auf Modelle aus vorhergehenden Schritten zurückgegriffen werden kann. Es lohnt sich aber nicht, die Inverse mit diesem Verfahren „spaltenweise“ aufzubauen, weil man keine Pivot-Strategie anwenden kann (Siehe [Sch93, Abschnitt 1.1.2]). Eine Verkleinerung der Kovarianz der validierten Werte ist in diesem Fall eine Verbesserung, da mehr Unstimmigkeiten aus den Meßdaten herausgerechnet wurden.

Mit dieser Vorüberlegung können wir nun Schritt 5 und 6 aus Algorithmus 5.1 bei einer Modellerweiterung um jeweils eine Gleichung durch folgende Strategie konkretisieren.

Wir suchen diejenige modellerweiternde Gleichung mit maximalem Eigenzugewinn. Dabei ist der Eigenzugewinn der Gleichung $g : \mathbb{R}^n \rightarrow \mathbb{R}$ definiert als der von Null verschiedene Eigenwert der Matrix, die gegeben wird durch die Differenz der induzierten Kovarianz D_k^* des alten Modells von der induzierten Kovarianzmatrix $D_l^* = D_l^*(g)$ des um g erweiterten Modells.

Wir wählen also unter allen vorhandenen zusätzlichen Gleichungen g, \dots, g^r diejenige aus, für die bei festem k der Term $\lambda_{max}\{D_l^* - D_k^*\}$ am größten wird. Da $D_l^* - D_k^*$ eine degenerierte Matrix ist, die nur einen von 0 verschiedenen Eigenwert besitzt, wird dieser durch die Spur der Matrix gegeben, also $\lambda_{max} = \text{tr}(D_l^* - D_k^*)$.

Dabei reicht es nicht aus, dieses Verfahren ausgehend vom leeren Modell für jeweils eine Gleichung durchzuführen und dann die „Bestenliste“ in ein erweitertes Modell zu übernehmen, da dabei redundante Information in den Gleichungen nicht berücksichtigt wird. Wenn also zwei Gleichungen in etwa gleiche Information enthalten, so werden sie ähnlich bewertet, wird aber eine Gleichung ins Modell übernommen, so fällt der Eigenzugewinn der anderen Gleichung im nächsten Auswahl-schritt ab.

BEISPIEL 5.4. Wir betrachten wieder die Situation aus Beispiel 2.3. Dort wurde die Nebenbedingung gegeben durch $f(x) = Ax$. Wir untersuchen die einzelnen Zeilen von A nach der oben vorgeschlagenen Methode, d.h. die einzelnen Zeilen von A werden zu neuen Modellen kombiniert. Wendet man das Verfahren ausgehend vom leeren Modell ($D_0^* = 0$) auf die einzelnen Zeilen von A an, so erhält man folgende Werte:

Zeile	Zugewinn
1	0.13634
2	0.13453
3	0.0087426

Wir übernehmen Gleichung 1 in das Modell, da diese den größten Zugewinn liefert und überprüfen nun Zeilen 2 und 3. Dann erhält man:

Zeile	Zugewinn
2	0.13609
3	-6.6938e-19

Gleichung 2 besitzt also einen höheren Informationsgehalt als Gleichung 3, die nun kaum noch Zugewinn besitzt. Dabei ist der Zugewinn von Zeile 3 zwischen Auswahl-Schritt 1 und 2 rapide abgefallen.

Die Zeilen im Modell sind also schon nach der „Wichtigkeit“ angeordnet.

Wir haben jetzt Algorithmus 5.1 für den Fall einer Modellauswahl in Form einer Modellerweiterung näher untersucht und eine Vorgehensweise gefunden, die es erlaubt, aus einer Menge von Gleichungen der Form $g : \mathbb{R}^n \rightarrow \mathbb{R}$ ein Modell aufzubauen, daß die geforderte Genauigkeit nach Validierung besitzt. Natürlich lassen sich auch Gleichungen der Form $g : \mathbb{R}^d \rightarrow \mathbb{R}$ mit $d < n$ in diesen Kontext einbetten, indem man nicht modellierte Sensoren ignoriert.

4. Linearkombinationen von Nebenbedingungen

Wir wollen das oben beschriebene Vorgehen zur Modellerweiterung noch weiter untersuchen, indem wir Linearkombinationen von Modellgleichungen betrachten, um dadurch den Eigenzugewinn dieser Linearkombination zu maximieren. Dabei benötigen wir nur die Modellgradienten in einem interessierenden Referenzpunkt.

Der Term $D_l^* - D_k^*$ besitzt nach (41) die Form

$$\frac{Pvv^\top P}{v^\top Pv},$$

wobei $P = D - D_k^*$ und $v = \partial g_l$ gilt. Mit der Transformation $w = Pv$ und einer Pseudo-Inversen Q von P erhält man die äquivalente Formulierung

$$\frac{ww^\top}{w^\top Qw}.$$

In dieser Darstellung erkennt man, daß w ein Eigenvektor zum Eigenwert

$$(42) \quad \frac{w^\top w}{w^\top Qw}$$

ist. Damit läßt sich die Berechnung des Eigenzugewinns noch mehr vereinfachen: Anstatt die Spur von $D_l^* - D_k^*$ zu berechnen, müßte man nur ein Skalarprodukt ausrechnen. Da man aber im Algorithmus 5.1 an der Matrix $D - D_l^* = D - D_k^* + (D_l^* - D_k^*)$ interessiert ist, bringt diese Vereinfachung keinen Vorteil.

Wir untersuchen nun, ob man durch Linearkombination von Gleichungen eine zusätzliche Verbesserung erreichen kann. Wir nehmen also

zwei Kandidaten¹ $a \in \mathbb{R}^n$ und $b \in \mathbb{R}^n$ und suchen nach Linearkombinationen dieser Werte, die (42) maximieren.

Wir setzen also $w = \alpha a + (1 - \alpha)b$ und bestimmen $\alpha \in \mathbb{R}$ so, daß (42) maximiert wird.

Es reicht, diese Darstellung aus Konvexkombinationen zu untersuchen, was wegen der Skalierungsinvarianz der Validierung (Proposition 4.1) aus der Darstellung

$$\lambda x + \mu y = (\lambda + \mu) \left(\frac{\lambda}{\lambda + \mu} x + \left(1 - \frac{\lambda}{\lambda + \mu} \right) y \right)$$

folgt.

Mit dieser Darstellung von w hat dann (42) die folgende Gestalt.

$$\begin{aligned} h(\alpha) &= \frac{(\alpha a + (1 - \alpha)b)^\top (\alpha a + (1 - \alpha)b)}{(\alpha a + (1 - \alpha)b)^\top Q (\alpha a + (1 - \alpha)b)} \\ (43) \quad &= \frac{\alpha^2 (a^\top a - 2a^\top b + b^\top b) + 2\alpha (a^\top b - b^\top b) + b^\top b}{\alpha^2 (a^\top Q a - 2a^\top Q b + b^\top Q b) + 2\alpha (a^\top Q b - b^\top Q b) + b^\top Q b} \\ &= \frac{\alpha^2 \|a - b\|^2 + 2\alpha \langle a - b, b \rangle + \|b\|^2}{\alpha^2 \|a - b\|_Q^2 + 2\alpha \langle a - b, b \rangle_Q + \|b\|_Q^2}. \end{aligned}$$

Also ist eine Funktion der Form

$$h(\alpha) = \frac{\alpha^2 x + \alpha y + z}{\alpha^2 u + \alpha v + w}$$

mit reellen Konstanten x, y, z, u, v, w auf kritische Punkte zu untersuchen.

Dann können wir einen optimalen Wert für α bestimmen, der dann für die beste Linearkombination der Kandidaten a und b sorgt. Der Zähler von $h'(\alpha)$ hat die Gestalt

$$\begin{aligned} &(2\alpha x + y)(\alpha^2 u + \alpha v + w) - (2\alpha u + v)(\alpha^2 x + \alpha y + z) \\ &= 2\alpha^3 x u + \alpha^2 y u + 2\alpha^2 x v + \alpha y v + 2\alpha x w + y w \\ &\quad - [2\alpha^3 u x + \alpha^2 x v + 2\alpha^2 u y + \alpha v y + 2\alpha u z + v z] \\ &= \alpha^2 (xv - uy) + 2\alpha (xw - uz) + (yw - vz). \end{aligned}$$

Verswindet der Nenner von $h(\alpha)$, so erhalten wir dafür aus (43)

$$\begin{aligned} &\alpha^2 \|a - b\|_Q^2 + 2\alpha \langle a - b, b \rangle_Q + \|b\|_Q^2 = 0 \\ \implies &(\alpha \|a - b\|_Q^2 + \langle a - b, b \rangle_Q)^2 = \langle a - b, b \rangle_Q^2 - \|b\|_Q^2 \|a - b\|_Q^2 \geq 0. \end{aligned}$$

Die *Cauchy-Schwartzsche Ungleichung* liefert gerade die Abschätzung in umgekehrte Richtung, also gilt bei verschwindendem Nenner

$$\langle a - b, b \rangle_Q^2 = \|b\|_Q^2 \|a - b\|_Q^2.$$

¹Diese Kandidaten entsprechen den transformierten Gradienten der zu untersuchenden Modellgleichungen.

Verschwimmt aber der Nenner von h nicht, so haben wir also maximal zwei kritische Punkte:

$$(44) \quad \alpha_{1,2} = -\frac{xw - uz}{xv - uy} \pm \sqrt{\left(\frac{xw - uz}{xv - uy}\right)^2 - \frac{yw - vz}{xv - uy}}$$

Da h für $\alpha \rightarrow \pm\infty$ gegen die Asymptote x/u strebt, ist dort das Verhalten nicht ausschlaggebend.

Wir wollen nun das Ergebnis aus (44) wieder zurücktransformieren. Seien dazu G und G' die Gradienten von zwei Gleichungen, die für den nächsten Auswahlsschritt in Betracht gezogen werden. Dann ist

$$\begin{aligned} P &= D - D^* \\ a &= PG \\ b &= PG' \\ x &= \|a - b\|^2 = \|P(G - G')\|^2 = \|G - G'\|_{P^2}^2 \\ u &= \|a - b\|_Q^2 = \|P(G - G')\|_Q^2 = \|G - G'\|_P^2 \\ y &= 2\langle a - b, b \rangle = 2\langle G - G', G' \rangle_{P^2} \\ v &= 2\langle a - b, b \rangle_Q = 2\langle G - G', G' \rangle_P \\ z &= \|b\|^2 = \|PG'\|^2 = \|G'\|_{P^2}^2 \\ w &= \|b\|_Q^2 = \|PG'\|_Q^2 = \|G'\|_P^2 \end{aligned}$$

Insbesondere brauchen wir die Pseudoinverse Q von P nicht mehr, obwohl wir bereits eine Darstellung der Pseudoinversen Q für die Situation $P = D - D^*$ gefunden haben, nämlich $Q = D^{-1} - \bar{D}$ nach Lemma 2.7, Eigenschaft 4.

Dann erhalten wir für die Terme aus (44) die folgenden Darstellungen:

$$\begin{aligned} xw - uz &= \|G - G'\|_{P^2}^2 \|G'\|_P^2 - \|G - G'\|_P^2 \|G'\|_{P^2}^2 \\ &= (G - G')^\top P (P(G - G')G'^\top - (G - G')G'^\top P) PG', \\ yw - vz &= 2\langle G - G', G' \rangle_{P^2} \|G'\|_P^2 - 2\langle G - G', G' \rangle_P \|G'\|_{P^2}^2 \\ &= (G - G')^\top P (PG'G'^\top - G'G'^\top P) PG', \\ xv - uy &= \|G - G'\|_{P^2}^2 2\langle G - G', G' \rangle_P - \|G - G'\|_P^2 2\langle G - G', G' \rangle_{P^2} \\ &= (G - G')^\top P (P(G - G')(G - G')^\top \\ &\quad - (G - G')(G - G')^\top P) PG'. \end{aligned}$$

Mit $\Delta G = G - G'$ und $XY - YX = [X, Y]$ erhalten wir die Darstellungen :

$$\begin{aligned} xw - uz &= \Delta G^\top P [P, (\Delta G)G'^\top] PG', \\ yw - vz &= 2\Delta G^\top P [P, G'G'^\top] PG', \\ xv - uy &= 2\Delta G^\top P [P, (\Delta G)\Delta G^\top] PG'. \end{aligned}$$

Den Fall, für den der Term $(xw - uz)^2 - (xv - uy)(yw - vz)$ negative Werte annimmt, so daß komplexe Lösungen von (44) existieren, untersuchen wir hier nicht. Liegen aber reelle Lösungen vor, so maximiert eine davon $h(\alpha)$. Diese Funktion $h(\alpha)$ gibt aber ja gerade den Eigenwert der Differenz zwischen der Kovarianz des erweiterten Modells und des Ausgangsmodells an. Damit haben wir dann eine Kombination aus zwei Gleichungen gefunden, die den Eigenzugewinn maximiert.

Wir wollen kurz Verbindung zum F -Test aus (40) beleuchten: Der F -Test zur Modellerweiterung entsteht gerade durch das Verhältnis zweier χ^2 -Verteilungen, die dem Normquadrat von multivariat normalverteilten Zufallsvektoren entsprechen. Auch hier bilden wir den Quotienten aus zwei quadratischen Gleichungen, allerdings kommen noch konstante Terme hinzu, was aus der statistischen Sichtweise zu so genannten nichtzentralen Verteilungen führt.

5. Zusammenfassung

Wir haben hier ein Verfahren zur Modellauswahl untersucht, so daß das ausgewählte Modell bei der Validierung eine bestimmte Genauigkeit erfüllt. Für die dazu notwendige Modellauswahl haben wir die Erweiterung um jeweils eine Gleichung näher untersucht. Außerdem haben wir uns dafür interessiert, wie man die Erweiterungsstrategie durch Linearkombinationen von zusätzlichen Modellgleichungen noch weiter verbessern kann.

Datenvalidierung mehrerer Messungen

In diesem Kapitel wollen wir untersuchen, welche Aussagen für die Validierung möglich sind, wenn eine Meßreihe verfügbar ist, die aus Daten von aufeinander folgenden Messungen besteht. Das dafür erforderliche Vorgehen wollen wir dynamische Datenvalidierung nennen.

Wenn man sich Grundlegende Gedanken über die Validierbarkeit von Prozeßmodellen macht, so muß man verschiedene Arten von Modellen unterscheiden, wobei hier der Term Modell als Begriff für die mathematische Abstraktion von der Realität und nicht als Modell für die Validierung aufgefaßt wird:

1. Modelle, die das Eingangs-/Ausgangsverhalten eines real ablaufenden Prozesses nachbilden.
2. Modelle, die eine Vorhersage von zukünftigem Verhalten zulassen.
3. Modelle, deren interne Zustände den Zuständen der Wirklichkeit entsprechen.

Dabei vererben sich die Eigenschaften von Typ 1 auf Typ 2, und die von Typ 2 auf Typ 3.

Wir wollen aufzeigen, welche Möglichkeiten zur Datenvalidierung sich mit diesen unterschiedlichen Modellarten bieten. Modelle der ersten Art eignen sich zur Datenvalidierung, indem man für die Validierung eine Nebenbedingung der Form „gemessene Ausgänge minus den aus den gemessenen Eingangsdaten berechneten Ausgängen“ zur Datenvalidierung verwendet¹. Modelle der zweiten Art sind dann auch für dynamische Datenvalidierung geeignet, indem man zusätzlich zu den Gleichungen, die für Typ 1 entstehen, Gleichungen der Form „Vorhersage aus alter Messung minus neue Messung“ verwendet. Typ-3-Modelle bieten zusätzlich die Möglichkeit, interne Zustände mit denen der Wirklichkeit abzugleichen, sofern diese als Messung vorliegen.

Also ist ein Simulationsmodell, das zu gegebenen aktuellen Eingängen und Zuständen die aktuellen Ausgänge und die Zustände für den nächsten Zeitschritt berechnet, für die dynamische Datenvalidierung geeignet, wenn sowohl Ein- und Ausgänge als auch die Zustände durch Sensoren erfassbar sind.

¹Das Beispiel 5.9 arbeitet nach diesem Prinzip, dort treten Nebenbedingungen der Form $y = f(x)$ auf, wo sowohl für x als auch für y Messungen vorhanden sind.

In den folgenden Abschnitten wollen wir Methoden für die dynamische Datenvalidierung vorstellen.

1. Ad-hoc Methoden

In diesem Abschnitt wollen wir untersuchen, welche Aussagen wir treffen können, wenn das Modell der Nebenbedingungen den Zustand des Prozesses in einer Gleichgewichtslage beschreibt. Im Gleichgewichtsfall sind Invarianten des zu Grunde liegenden Prozesses, für die man sich bei der Datenvalidierung interessiert, wesentlich einfacher zu ermitteln und zu beschreiben, denn im Gleichgewichtsfall ist dieser Prozeß selbst zeitlich invariant.

Außerdem betrachten wir den Fall, daß die Prozeßdynamik mithilfe eines Differenzenquotienten approximiert wird.

Ohne überhaupt an Nebenbedingungen gebunden zu sein, kann man im Gleichgewichtsfall die Mittelwerte zeitlich aufeinander folgenden Messungen bilden, die den „wahren Wert“ der Gleichgewichtslage immer besser beschreiben, da die Varianz der Mittelwerte bei steigender Anzahl von Beobachtungen abnimmt.

Die Validierung bietet uns in dieser Situation, wenn wir uns nur für den Gleichgewichtsfall interessieren, den Vorteil, dank des Güte-Tests 2.14 einen Indikator für den Gleichgewichtsfall zu besitzen. Treten nämlich dynamische Effekte auf, so werden diese den Sensorfehlern zugerechnet. Damit lehnt der Test bei starken Effekten die Messung ab, so daß man bei einer Häufung von Ablehnungen für aufeinanderfolgende Messungen darauf schließen kann, daß die Voraussetzung einer Gleichgewichtslage nicht mehr gegeben ist.

Abbildung 12 zeigt, daß die Mittelwertbildung bei zeitlichen Veränderungen zu schlechteren Ergebnissen führen kann. Hier ist eine Validierung des Mittelwertes mit höheren Kosten verbunden, da die Varianz des Mittelwertes schärfer ist als die Varianzen der einzelnen Messungen, dieser Mittelwert ist also in Bezug auf den Validierungsschritt „unbeweglicher“.

Nun gilt es zu untersuchen, wie man Validierung und Mittelwertbildung möglichst optimal miteinander mischt. Es läßt sich ein ganze Reihe von Möglichkeiten aufzeigen, dies zu bewerkstelligen. Allerdings ist zu beachten, daß diese Verfahren immer mit einem Validierungsschritt enden sollten, damit man zum Modell konsistente Werte erhält. Außerdem sollte immer die Messung zum aktuellen Zeitpunkt validiert werden, damit mittels eines Güte-Tests die Zuverlässigkeit dieser Messung beurteilt werden kann. Ein weiteres Problem besteht darin, daß die induzierte Kovarianzmatrix von solchen gemischten Schritten nicht mehr positiv definit sein muß. Dieses kann man dadurch vermeiden, indem man auf die Sensorvarianzen D zurückgreift, diese ist sicherlich ungenauer als die durch Mittelwertbildung und Validierung erhaltenen Kovarianzen.

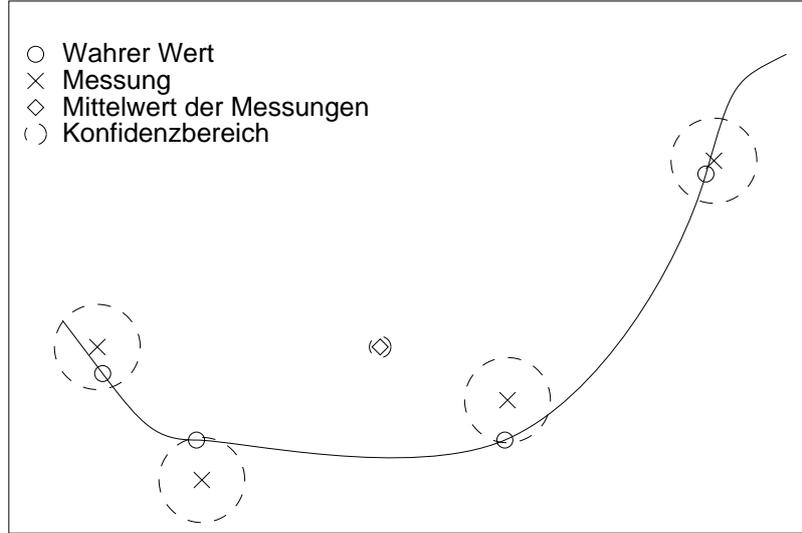


ABBILDUNG 12. Obwohl die Messungen nahe am Modell sind, ist das Mittel weit entfernt.

Eine andere Idee besteht darin, daß man für die Validierung der aktuellen Messung auf das letzte linearisierte Modell der vorhergehenden Messung zurückgreift:

$$x_t^* = y_t - DA_t^T(A_t DA_t^T)^{-1} A_t y_t \quad \text{mit } f(x_t^*) \sim 0$$

sei die zuletzt verwendete Linearisierung der Nebenbedingung f zur Messung y_t mit $A_t = \nabla f|_{x_t^*}$. Starte dann die Validierung einer neuen Messung y_{t+1} mit

$$x_{t+1,1} = y_{t+1} - DA_t^T(A_t DA_t^T)^{-1} A_t y_{t+1}.$$

Damit haben wir die „Einschwingphase“ der Validierung 3.9 übersprungen.

Ist nun aber die Prozeßdynamik in Form einer Gleichung

$$g(t, x, \dot{x}) = 0$$

gegeben, so kann man \dot{x} durch einen Differenzenquotienten approximieren. Dieses läßt sich wie folgt in den Rahmen einer Datenvalidierung einfügen. Wir führen dazu einen neuen Parameter z ein, der $g(t, x, z) = 0$ erfüllt und für den die zusätzliche Bedingung $z\Delta t - (x - x^*(t - \Delta t)) = 0$ gilt. Dabei ist Δt die Abtastrate und $x^*(t - \Delta t)$ die Validierung der vorhergehenden Messung zum Zeitpunkt $t - \Delta t$, diese Bedingung ist also eine nennerfreie Darstellung eines Differenzenquotienten.

Diese Idee hat aber den Nachteil, daß man für z keine vernünftige Verteilungsannahme besitzt. Hier läßt sich dann das in Abschnitt 3.4 aufgeführte Schätzverfahren zur Bestimmung von z benutzen.

2. Datenvalidierung und Kalman-Filter

Wir wollen untersuchen, wie man die temporale Information in den Validierungsprozeß einbinden kann. Für die dynamische Datenvalidierung wird die aktuelle Messung mit einer Vorhersage für den aktuellen Zeitpunkt aus vorigen Messungen abgeglichen. Diese Vorhersage sollte allerdings aus einer gesamten Wahrscheinlichkeitsverteilung ("bedingte Wahrscheinlichkeit") bestehen, da einzelne Punkte keine Wahrscheinlichkeitsmasse besitzen. So enthält eine Vorhersage des Erwartungswertes keine Information darüber, ob die Wahrscheinlichkeit der Vorhersage unschärfer wird, indem sich die (bedingte) Kovarianz dieser Vorhersage vergrößert.

Geht man davon aus, daß die zu Grunde liegenden Prozeßgleichungen lineare dynamische Systeme und die stochastischen Prozesse der Messungen normalverteilt sind, so gibt uns der Kalman-Filter gerade die Mischung aus Vorhersage und aktueller Messung, die auch die bedingte Kovarianz für eine Vorhersage berücksichtigt. Gaußsche Prozesse werden vollständig durch Erwartungswerte und Kovarianzen beschrieben. Da der Kalman-Filter ein spezieller linearer Beobachter² ist, läßt er sich außerdem zur Rekonstruktion von Zuständen einsetzen. Dazu muß das lineare dynamische System entdeckbar sein, um überhaupt die Existenz von stabilen Beobachtern zu gewährleisten.

Für eine Online-Behandlung der Daten bietet sich also bei einem linearen Modell ein Kalman-Filter an, der einem rekursiven Minimale-Quadrate Schätzer entspricht.

Ist man an einer nachträglichen Analyse von Meßdaten interessiert, so lassen sich auch Glättungsalgorithmen angeben, was wir im Abschnitt 6.2.3 untersuchen wollen. Dabei bedeutet Filtern, daß zur Berechnung des aktuellen Schätzwertes Daten aus vergangenen Messungen sowie die aktuelle Messung verwendet werden, Glättung verwendet auch zukünftige Werte, damit ist klar, daß eine Glättung nie im Onlinebetrieb verwendet werden kann, sondern erst zeitlich verzögert (bis sich genügend „zukünftige Werte“ angesammelt haben) zum Einsatz gelangt. Eine Vorhersage dagegen greift nur auf Werte zu, die zu zeitlich früheren Messungen gehören.

Wir betrachten zunächst eine allgemeine Situation und beschränken uns im nächsten Unterabschnitt auf lineare Modelle, bei denen sich ein Kalman-Filter einsetzen läßt.

Unser zeitvariantes Modell \mathcal{M}_t sei gegeben durch folgende Gleichungen

$$(45) \quad \begin{aligned} f(t, x) &= 0 & \text{und} \\ g(t, x) &= \dot{x} \end{aligned}$$

mit $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^m$, $f \in \mathcal{C}^2$, $m < n$, und $g : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$, $g \in \mathcal{C}^1$.

²Diese und weitere verwendete Bezeichnungen findet man in [KwS72].

Wir wollen bei der Validierung einer Messung y^t zum Zeitpunkt t die alten Validierungen x^{t-1}, x^{t-2}, \dots weiterverwenden. Mit diesen wird dann aus der dynamischen Nebenbedingung eine Vorhersage für den aktuellen Zeitpunkt gewonnen, die dann wiederum mit der aktuellen Messung verknüpft werden muß, so daß die statische Nebenbedingung eingehalten wird.

Für die Modellgleichungen aus (45) fordern wir eine Verträglichkeitsbedingung, die sich als Folgerung aus der Annahme (A2) ergibt.

DEFINITION 6.1 (Konsistenz). Die statischen und dynamischen Nebenbedingungen aus (45) heißen konsistent, falls folgende Verträglichkeitsbedingung erfüllt ist:

$$(46) \quad f(t, x) = 0 \implies \nabla_t f(t, x) + \nabla_x f(t, x)g(t, x) = 0.$$

Diese Konsistenz führt dazu, daß eine Lösung $x(t)$ der Dynamik $\dot{x} = g(t, x)$ auch Lösung der Nebenbedingung $0 = f(t, x)$ bleibt:

LEMMA 6.2. Sei (f, g) ein konsistentes Paar von Nebenbedingungen nach (45). Ist $x(t)$ eine Lösung von $\dot{x} = g(t, x)$ mit Anfangswert x_0 zur Zeit t_0 und erfüllt dieser Anfangswert die Gleichung $f(t_0, x_0) = 0$, so gilt für alle $t > 0$:

$$f(t, x(t)) = 0.$$

BEWEIS. Durch Differentiation von $f(t, x(t))$ nach t ergibt sich:

$$\begin{aligned} \nabla_t(f(t, x(t))) &= \nabla_t f(t, x(t)) + \nabla_x f(t, x(t))\dot{x}(t) \\ &= \nabla_t f(t, x(t)) + \nabla_x f(t, x(t))g(t, x(t)) \\ &= 0. \end{aligned}$$

Außerdem verschwindet $f(t, x(t))$ in $t = t_0$. Aus der Konsistenz folgt dann $f(t, x(t)) = 0$ für $t \geq t_0$. \square

Wir wollen den Begriff der Konsistenz an folgendem Beispiel untersuchen.

BEISPIEL 6.3. Betrachte das System

$$g: \quad \dot{x} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x$$

$$f(x) = \|x\|^2 - r^2.$$

Dieses System erfüllt die obengenannte Konsistenz-Forderung aus Definition 6.1, was wir nun zeigen. Da f zeitinvariant ist, haben wir also nur die Gleichung $\nabla f|_x g(t, x) = 0$ zu untersuchen:

$$2x^\top \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x = x_2 x_1 - x_1 x_2 = 0.$$

In Hinblick auf die Datenvalidierung gilt in diesem Beispiel, daß die statische Nebenbedingung f hier einen Radius vorgibt, während die dynamische Nebenbedingung g einen Winkel zwischen zwei Messungen bestimmt.

2.1. Lineare Dynamik. Wir betrachten den linearen Fall mit

$$(47) \quad \dot{x} = Ax + B\omega \quad \text{Dynamisches Modell,}$$

$$(48) \quad 0 = Fx \quad \text{Statisches Modell,}$$

wobei $x \in \mathbb{R}^n$ alle Messungen bezeichne, die mittels der Dynamik modelliert werden, ω sei Prozeßrauschen.

Dabei haben wir das Modell so vereinfacht, daß ein Referenzpunkt, der die Nebenbedingung erfüllt, schon auf Null normiert ist.

BEISPIEL 6.4. Es sei

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned}$$

ein dynamisches System mit Eingang u und Zustand x , dessen Ausgang y gemessen wird. Wir sehen aber nur ein verrauschtes Signal $y + v$, wobei das Rauschen $v \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, \text{diag}(\sigma_1^2, \dots, \sigma_n^2))$ aus normalverteilten Komponenten besteht.

Es ist verlockend, dieses System zur Validierung zu verwenden, indem man einen Beobachter benutzt, der den Zustand x des Systems schätzt. Dann wäre man in der Lage, eine Validierung der Gleichung $y - C\hat{x}$ durchzuführen, wobei \hat{x} der Zustandsschätzer des Beobachters ist.

Schätzt man aber die internen Zustände durch einen linearen Beobachter der Form

$$\dot{\hat{x}} = A\hat{x} + Bu - G(y - C\hat{x}),$$

so ist die Validierung des Modells

$$f(u, \hat{x}, \dot{\hat{x}}, y) = \begin{pmatrix} \dot{\hat{x}} - (A\hat{x} - Bu) \\ y - C\hat{x} \end{pmatrix} = \begin{pmatrix} G \\ I \end{pmatrix} (y - C\hat{x})$$

nicht möglich, denn dieses Modell ist nicht mehr regulär. Es läßt sich also nur der Beobachtungsfehler $y - C\hat{x}$ für eine Validierung verwenden.

Wir wollen nun die Verträglichkeitsbedingung auf den linearen Fall einschränken, wobei wir annehmen, daß die Matrizen A, B und C zeitunabhängig sind.

PROPOSITION 6.5. *Die Verträglichkeitsbedingung der Modellgleichungen (47) und (48) lautet wie folgt: Für alle Lösungen x der statischen Modellkomponente muß gelten:*

$$Fx = 0 \implies FAx = -FB\omega.$$

Man muß folgende Fälle unterscheiden:

- $\text{Im}A, \text{Im}B \subset \ker F$, dann ist die Konsistenzbedingung erfüllt. Diese Bedingung ist z.B. erfüllt, wenn die statische Nebenbedingung leer ist ($F = 0$).
- $\text{Im}A \not\subset \ker F$. In diesem Fall benötigt man immer eine Störung, um im statischen Modell zu verbleiben. Dieser Fall führt also zu Widersprüchen.

- $\text{Im}B \not\subset \ker F$. Hier tritt „physikalisch unmögliches Prozeßbrauchen“ auf. Dieser Fall ist also auch nicht von praktischem Interesse.

Wir können also das dynamische System auf den Kern von F einschränken und dort übliche dynamische Filter einsetzen. Diese Verfahren wollen wir noch weiter untersuchen. Dabei nehmen wir an, daß sich alle auftretenden Größen bereits im Kern von F befinden, was man leicht durch eine Projektion der einzelnen Größen erreichen kann.

PROPOSITION 6.6 (Kalman-Bucy-Filter). *Seien (x_t) und (y_t) vektorwertige stochastische Prozesse, die durch folgende Gleichungen beschrieben werden:*

$$\begin{aligned}\frac{dx_t}{dt} &= A(t)x_t + B(t)w_t, \\ y_t &= C(t)x_t + v_t,\end{aligned}$$

wobei $x_t \in \mathbb{R}^n$, $y_t \in \mathbb{R}^m$, $w_t \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, Q(t))$ und $v_t \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, R(t))$ unabhängige stochastische Prozesse sind. Die Matrizen $A(t)$, $B(t)$ und $C(t)$ sind nichtzufällige, zeitlich stetige Abbildungen. Außerdem ist der Anfangswert eine normalverteilte Zufallsvariable x_0 mit Erwartungswert \bar{x}_0 und Varianz P_0 und unabhängig von v_t und w_t . Dann wird der optimale Filter dieses Systems durch die folgenden stochastischen Differentialgleichungen für den bedingten Erwartungswert \hat{x}_t^t und für die bedingte Kovarianzmatrix P_t^t gegeben.

$$\begin{aligned}\frac{d\hat{x}_t^t}{dt} &= A(t)\hat{x}_t^t + P_t^t C^\top(t) R^{-1}(t) (y_t - C(t)\hat{x}_t^t) \\ \frac{dP_t^t}{dt} &= A(t)P_t^t + P_t^t A^\top(t) + B(t)Q(t)B^\top(t) \\ &\quad - P_t^t C^\top(t) R^{-1}(t) C(t) P_t^t\end{aligned}$$

Für die Anfangswerte gilt

$$\begin{aligned}\hat{x}_0^0 &= \bar{x}_0, \\ P_0^0 &= P_0.\end{aligned}$$

Die Matrix $K_t = P_t^t C^\top(t) R^{-1}(t)$ heißt Kalman gain matrix.

BEWEIS. Siehe [Jaz70, Proposition 7.3]

Nun müssen wir die Filtergleichungen für unseren Fall anpassen. Wenn dabei die Konsistenzbedingung 6.1 erfüllt ist, so muß nur die aktuelle Messung auf die statische Nebenbedingung (48) projiziert werden, da die Kalman-Schätzung \bar{x}_t^t aufgrund dieser Konsistenzbedingung bereits die statischen Nebenbedingungen erfüllt.

Gehen wir davon aus, daß die stochastischen Prozesse stationäre stochastische Prozesse sind, und das dynamische System einige schwache Bedingungen erfüllt (vgl. [KwS72, Section 4.4.3]), so läßt sich ein

zeitunabhängiger Kalman-Filter einsetzen, der dann mit der asymptotischen Kovarianzmatrix P_* arbeitet, die hier durch die Lösung der folgenden algebraischen Riccati-Gleichung gegeben wird:

$$0 = AP_* + P_*A^\top + BQB^\top - P_*C^\top D^{-1}CP_*^\top,$$

wobei D die Kovarianzmatrix des Meßrauschens und die Matrix Q die Kovarianz des Prozeßrauschens bezeichnet. Die Wahl dieses Prozeßrauschens ist noch relativ frei und damit ist der Kalman-Filter schwer zu justieren. Für $Q = 0$ ist allerdings die Lösung dieser algebraischen Riccati-Gleichung ein schlecht gestelltes Problem, so existiert dann die triviale Lösung $P_* = 0$.

2.2. Kalman-Filter unter Nebenbedingungen. Wir untersuchen hier, wie sich der Kalman-Filter-Ansatz retten läßt, sofern die Konsistenzbedingung nicht erfüllt ist. EOs bietet sich an, die in Abschnitt 6.1 skizzierten Verfahrenweisen auf den dynamischen Fall zu übertragen.

Man hat hier drei Ansatzpunkte, um den Validierungsschritt in den Kalman-Filter einzubauen:

- Der Meßwert wird auf die Nebenbedingung eingeschränkt.
- Der Schätzwert des Kalman-Filters wird eingeschränkt.
- Die a-priori Information, die Schätzung aus dem vorherigen Wert nach dem Einfluß der Dynamik, wird eingeschränkt.

Ist die oben definierte Konsistenz zwischen Nebenbedingung und Dynamik erfüllt, so ist die Durchführung des letzten Punktes unnötig, denn die Vorhersage der aktuellen Messung aus den alten Daten erfüllt das statische Modell.

Die anderen beiden Validierungsansätze lassen sich dann mit dem Kalman-Filter auf folgende Weise verbinden:

- Anwendung des Kalman-Filters und auf Konvergenz hoffen. Dieses entspricht dann der Mittelwertbildung aus 6.1.
- Kalman-Filter aus der aktuellen Messung berechnen und dann validieren.
- Neue Messung validieren, dann diese Validierung filtern.
- Beide Schritte mischen, d.h. mit einer Validierung der aktuellen Messung den Kalman-Filter berechnen und dann den resultierenden Wert wiederum validieren.

Eine Validierung der Schätzung garantiert uns eine Korrektur, die das Modell erfüllt. Mit einer Validierung der Werte lassen sich Meßfehler identifizieren. Beide Validierungsschritte zu mischen, kostet Zeit. Zu hoffen ist, daß sich mit Validierung die „Einschwingphase“ des Filters verkürzt.

Die eingeschränkte Kovarianz der Validierung läßt sich nicht nutzen, denn dadurch werden nur Lösungen zugelassen, die im Tangentialraum des Modells für die aktuelle Schätzung liegen. Für den Fall nichtlinearer

Nebenbedingungen ist dies wenig praktikabel, denn sobald man nur mit Erwartungswert und nicht regulärer Kovarianz rechnet, setzt man eine Normalverteilung voraus, die nur auf einem linearen Unterraum gegeben ist.

2.3. Datenglättung mittels Kalman-Filter. Bei einem Offlinebetrieb zur Datenvalidierung stellt sich die Frage, wie man die Messung zu einem Zeitpunkt T mit dem Wissen über zukünftige Messungen glätten kann. Dazu betrachten wir die statische und diskrete Version des Kalman-Filters.

PROPOSITION 6.7 (Kalman-Filter (diskret, statisch)). Seien $y_i \in \mathbb{R}^n$ und $G_i \in \mathbb{R}^{n \times k}$, $i \in \mathbb{N}$. Der Minimale-Quadrate-Schätzer

$$(49) \quad x_m = \arg \min_x \sum_{i=1}^m \|y_i - G_i x\|^2, \quad m = 1, \dots, n$$

wird durch die Iteration

$$x_m = x_{m-1} + H_m^{-1} G_m^T (y_m - G_m x_{m-1}), \quad m = 1, \dots, n$$

gegeben, wobei x_0 beliebig ist und die positiv definite Matrix H_m durch

$$H_m = H_{m-1} + G_m^T G_m$$

mit $H_0 = 0$ gegeben wird. Es gilt die allgemeine Formel für $i < m$

$$(50) \quad x_m = x_i + H_m^{-1} \sum_{j=i+1}^m G_j^T (y_j - G_j x_i).$$

BEWEIS. Für jeden Term der Summe aus (49) ergibt sich der kritische Punkt aus $(G_i^T G_i)x = G_i^T y_i$. Für x_1 mit $x_0 = 0, H_0 = 0$ gibt dies folgende Formel:

$$x_1 = (G_1^T G_1)^{-1} G_1^T y_1 = x_0 + H_1^{-1} G_1 y_1.$$

Dies bildet den Induktionsanfang für eine Induktion nach i . Sei also (50) für alle $j \leq i$ bewiesen. Dann erhält man für $m > i$:

$$\begin{aligned} x_m &= H_m^{-1} \sum_{j=1}^m G_j^T y_j \\ &= x_i + H_m^{-1} \left(\sum_{j=1}^m G_j^T y_j - H_m x_i \right) \\ &= x_i + H_m^{-1} \left(\underbrace{\sum_{j=1}^i G_j^T y_j - G_j^T G_j x_i}_{=0} + \sum_{j=i+1}^m G_j^T y_j - G_j^T G_j x_i \right). \end{aligned}$$

□

Nun wollen wir dieses Resultat so abändern, daß damit eine Datenglättung möglich ist.

Wir wollen also aus den Meßdaten zu den Zeitpunkten $1, \dots, T, T+k$ einen geglätteten Schätzwert zum Zeitpunkt T erhalten. Man erhält durch eine andere Zerlegung von $H_m = \sum_{j=1}^m G_j^\top G_j$ folgende Formel.

$$\begin{aligned}
x_m &= \left(\sum_{j=1}^m G_j^\top G_j \right)^{-1} \sum_{j=1}^m G_j^\top y_j \\
&= x_i + H_m^{-1} \left(\sum_{j=1}^m G_j^\top y_j - H_m x_i \right) \\
&= x_i + H_m^{-1} \left(\sum_{j=1}^m G_j^\top y_j - \left(H_i - \sum_{j=m+1}^i G_j^\top G_j \right) x_i \right) \quad \text{für } m < i \\
&= x_i + H_m^{-1} \left(\sum_{j=1}^i G_j^\top y_j - H_i x_i + \sum_{j=m+1}^i G_j^\top G_j x_i - G_j^\top y_j \right) \\
&= x_i + H_m^{-1} \left(\sum_{j=m+1}^i G_j^\top G_j x_i - G_j^\top y_j \right) \\
&= x_i - H_m^{-1} \left(\sum_{j=m+1}^i G_j^\top y_j - G_j^\top G_j x_i \right),
\end{aligned}$$

insbesondere gilt dann

$$x_m = x_{m+1} - H_m^{-1} G_{m+1}^\top (y_{m+1} - G_{m+1} x_{m+1}).$$

Damit läßt sich also eine Glättung der Daten zum Zeitpunkt T realisieren, wenn schon Daten der Zeitpunkte $T+1, \dots, T+k$ verfügbar sind. Die Schwierigkeit hierbei besteht aber in der Wahl von H_m .

3. Nichtlineare Dynamik

Wir kehren wieder zur Ausgangssituation (45) zurück und wollen diese weiter untersuchen. Insbesondere formulieren wir das allgemeine Validierungsproblem für den dynamischen Fall.

DEFINITION 6.8. Sei $\varphi(t; t_0, x_0) : \mathbb{R} \times (\mathbb{R} \times \mathbb{R}^n) \rightarrow \mathbb{R}^n$ Lösung der Differentialgleichung $\dot{x} = g(t, x)$ zum Anfangswert x_0 im Zeitpunkt t_0 . Dann wird die *dynamische Validierung* für den Zeitpunkt T ab dem Zeitpunkt 0 mit Abtastrate 1 zur Meßreihe (y_t) gegeben durch

$$(51) \quad \arg \min_x \left\{ \sum_{t=0}^T \|\varphi(t; T, x) - y_t\|^2 \right\}$$

unter der Nebenbedingung

$$(52) \quad f(t, \varphi(t; T, x)) = 0 \quad \text{für alle } t = 0, 1, 2, \dots, T.$$

Die Nebenbedingung läßt sich dahingehend verschärfen, daß man die Übereinstimmung auch zwischen den Abtastzeiten fordert, dies ergibt dann

$$\int_0^T \|f(t, \varphi(t; T, x))\|^2 dt = 0$$

Wir suchen also eine Schätzung, deren Lösung zu den Abtastzeiten den entsprechenden Messungen nahe kommt.

In [BBM97] wird angemerkt, daß dieses Problem zur Klasse der schlecht gestellten nichtlinearen inversen Probleme gehört, d.h. man versucht, zu einem gegebenen Effekt – hier die Meßdaten – die Ursache zu bestimmen, die aber weder eindeutig noch stetig davon abhängt. Damit können numerische Lösungen, die immer mit kleinen Diskretisierungsfehlern behaftet sind, zu nicht sinnvollen Ergebnissen gelangen.

Trotzdem setzt man auf numerische Verfahren. In [LEL92] wird das Differential-Gleichungssystem durch ein algebraischen Gleichungssystem ersetzt, das mithilfe der so genannten *orthogonalen Kollokationsmethode* bestimmt wird.

Dann läßt sich dieses nun rein algebraische Gleichungssystem wieder mit schon bekannten Methoden der sequentiellen quadratischen Programmierung lösen. Wir wollen das Verfahren der Kollokation im Folgenden erläutern.

Die Idee der Kollokation besteht darin, daß man die Lösung einer Differentialgleichung der Form

$$\dot{x} = g(t, x)$$

durch ein Polynom u n -ten Grades approximiert, das die Anfangsbedingung $u(t_0) = x_0$ erfüllt, sowie an n vorgegebenen Stellen $t_0 \leq t_1 < \dots < t_i < \dots < t_n$ die Differentialgleichung erfüllt:

$$(53) \quad u'(t_i) = g(t_i, u(t_i)).$$

Diese Approximation u soll nun eine Schätzung der Lösung zum Zeitpunkt $T \geq t_n$ ermöglichen. Sinnvollerweise normiert man das Intervall (t_0, T) auf $(0, 1)$, damit befinden sich die normierten Stützstellen $s_i = \frac{t_i - t_0}{\tau}$, $\tau = T - t_0$, im Intervall $(0, 1)$. Nun kann man mithilfe der Lagrange-Polynome zu dieser Stützstellenmenge zeigen, daß sich bei der Kollokation ein implizites Runge-Kutta-Verfahren ergibt (siehe [DeB94], Kapitel 6.3), das zu einem nichtlinearen Gleichungssystem führt. Diese impliziten Verfahren sind numerisch stabil.

Seien also L_i Lagrange-Polynome zu den Stützstellen s_i , $i = 1, \dots, n$. Ein Lagrange-Polynom L_i ist definiert durch

$$L_i(s) = \prod_{\substack{j=1 \\ j \neq i}}^n \frac{s - s_j}{s_i - s_j},$$

insbesondere ist dann $L_i(s_j) = \delta_{ij}$. Nennen wir den Wert der Ableitung von u in den Kollokationspunkten $k_i = u'(t_i) = u'(t_0 + s_i\tau)$, so erhalten wir die Darstellung

$$u'(t_0 + \vartheta\tau) = \sum_{i=1}^n k_i L_i(\vartheta),$$

denn diese Polynome $n - 1$ -ter Ordnung stimmen an n Stellen überein, sind also gleich.

Mit dieser Darstellung wollen wir die Werte von u an den Kollokationsstellen bestimmen:

$$\begin{aligned} u(t_i) &= u(t_0 + s_i\tau) = x_0 + \int_0^{s_i} u'(t_0 + s_i\tau) \\ (54) \quad &= x_0 + \tau \sum_{j=1}^n k_j \int_0^{s_i} L_j(\vartheta) d\vartheta \\ &= x_0 + \tau \sum_{j=1}^n a_{ij} k_j, \end{aligned}$$

wobei wir $a_{ij} = \int_0^{s_i} L_j(\vartheta) d\vartheta$ für $i, j = 1, \dots, n$ setzen. Setzen wir diese Darstellung in die Kollokationsbedingung (53) ein, so erhalten wir folgendes nichtlineares Gleichungssystem für die Ableitung von u in den Stützstellen:

$$k_i = g(t_0 + s_i\tau, x_0 + \tau \sum_{j=1}^n a_{ij} k_j) \quad i = 1, \dots, n.$$

Dann ist unsere approximative Lösung zum Zeitpunkt T gegeben durch

$$\begin{aligned} u(T) &= u(t_0 + \tau) = x_0 + \tau \int_0^1 u'(t_0 + \vartheta\tau) d\vartheta \\ &= x_0 + \tau \sum_{j=1}^n k_j \int_0^1 L_j(\vartheta) d\vartheta. \end{aligned}$$

Die Differentialgleichung läßt sich also in eine nichtlineare algebraische Gleichung überführen, die man den statischen Nebenbedingungen hinzufügen kann.

Wir wollen diesen Kollokationsansatz nun auf die Datenvalidierung anwenden. Dabei bietet es sich an, die Abtastzeiten als Kollokationszeiten zu verwenden und dabei mit gefensterten Daten zu arbeiten. Wir fassen also jeweils T Messungen zu einem Datenfenster zusammen, das dann durch eine dynamische Datenvalidierung aufbereitet wird, wobei die verwendeten Werte schon den jeweiligen statischen Modellen gehorchen sollten.

Für diese dann schon statisch validierten Messungen $y^i, i = 1, \dots, T$ hat man dann mit den oben eingeführten Bezeichnungen folgende Modellgleichungen für die Validierung dieser Meßwerte

$$u(t_i) - y^i = 0 \quad \text{für } i = 1, \dots, T.$$

Diese Nebenbedingung läßt sich nun mit den aus kapitel 6 bekannten Methoden behandeln, wobei wir nun alle Daten im zeithorizont als eine Messung auffassen. Es sind noch die Koeffizienten von u zu bestimmen. Ist die Abtastrate konstant, so vereinfachen sich die Lagrange-Polynome und man braucht die a_{ij} aus (54) nur einmal zu berechnen. Dieses hier vorgestellte Kollokations-Verfahren hat gegenüber den im vorherigen Abschnitt vorgestellten Verfahren den Vorteil, mehrere Messungen auf einmal zu validieren. Die dynamische Nebenbedingung wird dabei zu einer statischen Nebenbedingung für ein höherdimensionales Problem umgeformt, das alle Messungen umfaßt.

Allerdings fehlt uns bei diesem Ansatz die Information über die bedingten Varianzen, die im vorhergehenden Abschnitt eine wesentliches Merkmal der dort vorgestellten Verfahren zur dynamischen Datenvalidierung waren. Auch bei diesem Verfahren können nach erfolgter dynamischer Validierung die statischen Nebenbedingungen verletzt sein.

Zusammenfassung und Ausblicke

Wir haben in dieser Arbeit das Verfahren der Datenvalidierung untersucht, das einem Meßvektor einen Wert zuordnet, der ein gegebenes Modell von Nebenbedingungen erfüllt. Das zugehörige Vorgehen wollen wir jetzt noch einmal zusammenfassen.

Zuerst einmal benötigt man ein Modell, das für die Datenvalidierung geeignet ist. Allerdings gibt es auch bei den geeigneten Modellen Unterschiede im Verhalten unter der Datenvalidierung. Wir haben mehrere Vorgehensweisen vorgestellt, um aus geeigneten Modellen ein Modell auszuwählen, so z.B. die Erweiterungsstrategie aus Abschnitt 5.3, die sich dazu einsetzen läßt, Modelle die um einen gemeinsamen Betriebspunkt linearisiert sind, miteinander zu vergleichen.

Hat man nun ein passendes Modell gefunden, so läßt sich die Validierung auf eine Projektion auf eine Menge von zulässigen Punkten zurückführen. Für diese nichtlineare Projektion haben wir verschiedene Validierungsverfahren entwickelt. Die in dieser Arbeit näher untersuchten Verfahren lassen sich aus dem Algorithmus 3.18 gewinnen.

Wir haben außerdem in Proposition 2.14 einen Güte-Test kennengelernt, mit dem man ein Entscheidungskriterium für die Zuverlässigkeit der Messung besitzt.

Wir haben uns im vorhergehenden Kapitel 6 damit beschäftigt, wie man dynamische Datenvalidierung betreiben kann. Dazu haben wir mehrere statische validierte Messungen in einem Zeitfenster zusammengefaßt und dieses Fenster dann dynamisch validiert. Dieses Vorgehen hat den Vorteil, daß man nicht mit jeder Messung eine neue dynamische Datenvalidierung startet.

Wir wollen noch stichpunktartig aufführen, welche Punkte man noch weiter untersuchen könnte. Weitere interessante Themen sind sicherlich:

- Erweiterung der Validierung auf Ungleichungen als Nebenbedingungen.
- Graphentheorie: Das Validierungsmodell entspricht einem nicht-linearen Netzwerk, dessen Knoten aus den Modellgleichungen der einzelnen Bauteile der Anlage bestehen und dessen Kanten durch gemeinsame Zustände dieser Modelle gebildet werden.

Die Datenvalidierung soll nun dieses Netzwerk abgleichen, wenn nur partielle Informationen auf den Kanten und in den

Knoten vorhanden sind. Gibt dieser Ansatz andere Verfahren zur Datenvalidierung ?

- Nicht-Gaußsche Fehlerverteilungen. Damit die Validierung eine Maximum-Likelihood-Schätzung bleibt, muß dann eine neue Kostenfunktion benutzt werden.
- Konvergenzaussagen für die nichtlinearen Validierungsverfahren.
- Numerische Aspekte der Validierung, besonders bei schwach besetzten Gradienten der Nebenbedingung.
- Modellauswahl im dynamischen Fall.
- Dualität zwischen dynamischer Datenvalidierung und model-predictive control.

Außerdem kann man untersuchen, wie man bei anderen Arten von Modellen vorgehen kann. Hat man z.B. keine mathematische Formulierung für die Nebenbedingungen, so kann man versuchen, das Modell durch ein neurales Netz zu beschreiben.

Eine andere Art von Modellen besteht aus Fuzzy-Regeln der Art „Ändert sich der Meßwert an Punkt A nur wenig, so ändert sich auch der Meßwert an Punkt B nur wenig“.

Auch für solche Modelle möchte man eine Datenvalidierung realisieren. Dafür ist aber das hier vorgestellte Validierungsverfahren ungeeignet. Hier müssen neue Ansätze erprobt werden.

ANHANG A

Projektionen und Pseudo-Inverse

In diesem Abschnitt werden schiefe Skalarprodukte und Projektionen definiert. Wir beweisen das Projektionslemma, das ein Haupthilfsmittel in dieser Arbeit ist, da es eine Minimalitätsaussage für Projektionen liefert. Außerdem untersuchen wir den Zusammenhang zwischen Projektionen und Pseudo-Inversen. (Siehe z.B. [Ben74] oder den Übersichtsartikel [Ben76]).

DEFINITION A.1. Es sei P eine *positiv-definite symmetrische* reelle $n \times n$ Matrix (Bezeichnung: $P \succ 0$).

- Dann ist das *Skalarprodukt* bezüglich P gegeben durch

$$\langle u, v \rangle_P = u^\top P v = \langle u, P v \rangle = \langle P u, v \rangle,$$

wobei $\langle \cdot, \cdot \rangle$ das Standardskalarprodukt auf \mathbb{R}^n ist.

- Ist $\langle u, v \rangle_P = 0$, so sind u und v *P-orthogonale* oder *P-konjugierte* Vektoren.
- Die zugehörige *P-Norm* ist definiert durch

$$\|u\|_P^2 = \langle u, u \rangle_P.$$

Diese gewichtete euklidische Norm wird auch als *elliptische Norm* bezeichnet. Dieser Begriff erklärt sich aus der Tatsache, daß die Menge

$$\left\{ x \in \mathbb{R}^n \mid \|x\|_P = 1 \right\} = \left\{ \frac{y}{\sqrt{y^\top P y}} \mid \|y\| = 1, y \in \mathbb{R}^n \right\}$$

einen Ellipsoid bildet. Die Verbindung zum Standardskalarprodukt läßt sich mit der folgenden Aussage aufzeigen. Ist $P = P^{\top/2} P^{1/2}$ die *Cholesky-Zerlegung* von P , so gilt

$$\begin{aligned} \langle u, v \rangle_P &= \langle P^{1/2} u, P^{1/2} v \rangle, \\ \|u\|_P^2 &= \|P^{1/2} u\|^2. \end{aligned}$$

PROPOSITION A.2. Für die elliptische Norm $\|\cdot\|_P$ gilt die Rhombus-Identität

$$\|x + y\|_P^2 + \|x - y\|_P^2 = 2(\|x\|_P^2 + \|y\|_P^2).$$

BEWEIS. Es gilt

$$\begin{aligned} & \|x + y\|_P^2 + \|x - y\|_P^2 \\ &= x^\top Px + 2x^\top Py + y^\top Py \\ & \quad + x^\top Px - 2x^\top Py + y^\top Py \\ &= 2(\|x\|_P^2 + \|y\|_P^2). \end{aligned}$$

□

DEFINITION A.3. Es sei $B = \{b_1, \dots, b_n\}$ eine Basis des \mathbb{R}^n , eine Partitionierung der Basisvektoren sei gegeben durch $B_1 = \{b_1, \dots, b_m\}$ und $B_2 = \{b_{m+1}, \dots, b_n\}$. $P \succ 0$, $P \in \mathbb{R}^{n \times n}$ sei eine positiv definite, symmetrische Matrix.

- Die Matrix \mathcal{P} heißt *Projektion* auf $\mathcal{A} = \mathcal{L}(B_1)$ entlang $\mathcal{B} = \mathcal{L}(B_2)$, wenn gilt

$$\mathcal{P}b = \begin{cases} b, & b \in \mathcal{A} \\ 0, & b \in \mathcal{B} \end{cases}.$$

Dabei ist $\mathcal{L}(B_i)$ die lineare Hülle der Menge B_i , d.h. die Menge aller Linearkombinationen von Elementen aus B_i .

- \mathcal{P} heißt *P-orthogonale Projektion* auf den Unterraum \mathcal{A} , wenn für alle $a \in \mathcal{A}$ und für alle $x \in \mathbb{R}^n$ gilt:

$$b = \mathcal{P}x \iff b \in \mathcal{A} \text{ und } \langle b - x, a \rangle_P = 0$$

Allgemein werden Projektionen dieser Bauart als *schiefe Projektionen* bezeichnet.

Manche Autoren unterscheiden zwischen dem *Projektor*, der Projektionsabbildung, und der Projektion, dem Bild unter dem Projektor. Die Eigenschaften dieser Projektoren sind in der folgenden Proposition aufgeführt.

PROPOSITION A.4. *Es sei $Q \in \mathbb{R}^{n \times n}$ eine idempotente Matrix, d.h. $Q = QQ$. Dann gilt*

1. Q^\top und $I - Q$ sind idempotente Matrizen.
2. Q besitzt nur die Eigenwerte 0 und 1.
3. $Qx = x \iff x \in \text{Im}Q$
4. Die Vielfachheit des Eigenwertes 1 ist der Rang von Q .
5. Die Spur von Q ist der Rang von Q .
6. $Q(I - Q) = (I - Q)Q = 0$
7. $\ker Q = \text{Im}(I - Q)$
8. Q ist eine Projektion auf $\text{Im}Q$ entlang $\ker Q$.

BEWEIS. (1) und (6) sind trivial,

(2): Wegen $Q = Q^2$ muß für die Eigenwerte von Q gelten: $\lambda = \lambda^2$, also sind 0 und 1 die einzigen möglichen Eigenwerte von Q . (3): Ist $x \in \text{Im}Q$, so gibt es ein y mit $Qy = x \implies Qx = QQy = Qy = x$. Die

andere Implikation ist trivial.

(4) Die geometrische Vielfachheit des Eigenwertes 1 ist gleich seiner algebraischen Vielfachheit, denn ist $Q(Qx - x) - (Qx - x) = 0$, so ist $x = Qx$, also gibt es keine echten verallgemeinerten Eigenvektoren zum Eigenwert 1, ebenso gibt es keine echten verallgemeinerten Eigenvektoren zum Eigenwert 0, denn dann gäbe es ein y mit $Q^2y = 0$, aber $Qy \neq 0$. Also ist der Rang von Q , der Dimension des Bildes von Q , gegeben durch die Anzahl der Eigenräume zum Eigenvektor 1.

(5): Die Spur ist gerade die Summe aller Eigenwerte.

(7): Mit (1) und (3) erhalten wir

$$\begin{aligned} x \in \ker Q &\iff Qx = 0 \iff x - Qx = x \\ &\iff (I - Q)x = x \iff x \in \operatorname{Im}(I - Q). \end{aligned}$$

(8): $\operatorname{Im}Q \oplus \ker Q = \mathbb{R}^n$ und $\mathcal{P}|_{\operatorname{Im}Q} \equiv \operatorname{id}$, denn der Schnitt $\operatorname{Im}Q \cap \ker Q$ besteht nur aus der Null, was aus Eigenschaft (3) folgt. \square

PROPOSITION A.5. *Ist \mathcal{P} eine P -orthogonale Projektion auf $\mathcal{A} = \operatorname{Im}A \subset \mathbb{R}^n$ und ist $A \in \mathbb{R}^{n \times m}$ eine reguläre Matrix mit $\operatorname{rang}(A) = \min(m, n) = m$, so wird die Projektionsmatrix $\mathcal{P} \in \mathbb{R}^{n \times n}$ gegeben durch*

$$(55) \quad \mathcal{P} = ARA^\top P \quad \text{mit } R = (A^\top PA)^{-1}.$$

BEWEIS. Die Existenz von R ist dadurch gesichert, daß A regulär ist. Es sei also $b = \mathcal{P}x$. Dann ist $b \in \mathcal{A}$, denn $b = ARA^\top Px$. Außerdem gilt für $a \in \mathcal{A}$

$$\begin{aligned} (b - x)^\top Pa &= (ARA^\top Px - x)^\top PAy = x^\top (PARA^\top - I)PAy \\ &= x^\top (PARA^\top PA - PA)y = x^\top (PA - PA)y = 0 \end{aligned}$$

\square

BEMERKUNG. Besitzt A keinen vollen Rang, so ersetzt man die Inverse in (55) durch eine Pseudo-Inverse, die wir weiter unten genauer studieren.

Jetzt wollen wir eine Verbindung zwischen den Projektionen und den orthogonalen Projektionen herstellen.

PROPOSITION A.6. *Ist \mathcal{P} P -orthogonale Projektion auf $\mathcal{B} = \operatorname{Im}B$, so ist \mathcal{P} Projektion auf \mathcal{B} entlang $\mathcal{A} = \ker B^\top P$.*

BEWEIS. Gleichung (55) definiert eine idempotente Matrix. Eigenschaft (8) aus Proposition A.4 liefert die folgende Aussage: \mathcal{P} ist Projektion auf $\operatorname{Im}\mathcal{P} = \operatorname{Im}B$ entlang $\ker \mathcal{P}$. Die Rangformel¹ liefert die gesuchte Aussage $\ker \mathcal{P} = \ker(B^\top P)$. \square

PROPOSITION A.7. *Es sei $A \in \mathbb{R}^{m \times n}$, $\operatorname{Rang} A = m$, $m < n$. Dann ist*

$$\mathcal{P} = I - QA^\top(AQA^\top)^{-1}A$$

die P -orthogonale Projektion auf den Kern von A mit $Q = P^{-1}$.

¹Sei V ein endlichdimensionaler Vektorraum und $f : V \rightarrow V$ linear, dann gilt: $\dim \ker f + \dim \operatorname{Im} f = \dim V$.

BEWEIS. Offensichtlich ist $\mathcal{P}x = x$ für $x \in \ker A$. Wir müssen also nur noch die Orthogonalitätseigenschaft nachweisen:

$$\begin{aligned} (\mathcal{P} - I)^\top Px &= A^\top (AQA^\top)^{-1} AQP x \\ &= A^\top (AQA^\top)^{-1} Ax = 0 \end{aligned}$$

für $x \in \ker A$. □

In folgender Proposition wirkt sich die schiefe Projektion aus, für $P = I$ ist die Aussage trivial.

PROPOSITION A.8. *Ist \mathcal{P} P -orthogonale Projektion auf $\text{Im}A$, so ist $\mathcal{P}^\top P^{-1}$ -orthogonale Projektion auf $\text{Im}PA$.*

BEWEIS. Mit der Darstellung aus Proposition A.5 erhalten wir

$$\begin{aligned} \mathcal{P}^\top &= P A R A^\top \\ &= P A (A^\top P P^{-1} P A)^{-1} A^\top P P^{-1} \\ &= (P A) ((P A)^\top P^{-1} P A)^{-1} (P A)^\top P^{-1}. \end{aligned}$$

Ein Vergleich dieses Terms mit Formel (55) liefert die Behauptung der Proposition. □

Das folgende Lemma zeigt, daß die Projektion ein „Lotfußpunkt“, also die Lösung eines Minimale-Quadrate-Problems ist. Dieses Lemma verbindet also die Begriffe „orthogonal“ und „optimal“ miteinander.

LEMMA A.9 (Projektionlemma). *Es sei \mathcal{X} ein Vektorraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$, $x \in \mathcal{X}$, und \mathcal{Y} ein linearer Unterraum von \mathcal{X} . Dann gilt für $\hat{x} \in \mathcal{Y}$*

$$\min_{y \in \mathcal{Y}} \|x - y\|^2 = \|x - \hat{x}\|^2 \iff \langle x - \hat{x}, y \rangle = 0 \text{ für alle } y \in \mathcal{Y}$$

Wir möchten den Punkt x durch einen Punkt aus \mathcal{Y} approximieren. Dieses Lemma besagt, daß der Fehler dieser Approximation orthogonal zum Unterraum \mathcal{Y} liegt. Zerlegt man $x = \tilde{x} + \hat{x}$, wobei \tilde{x} der Fehler und \hat{x} die beste Annäherung ist, so ist \hat{x} die orthogonale Projektion (bezüglich des zu Grunde gelegten Skalarproduktes) auf den Unterraum \mathcal{Y} .

BEWEIS. (Siehe [Jaz70] Example 7.1) Setze $x = \tilde{x} + \hat{x}$. Wir zeigen zunächst $\langle x - \hat{x}, y \rangle = 0 \implies \hat{x} = \mathcal{P}_{\mathcal{Y}} x$:

Sei also $\langle \tilde{x}, y \rangle = 0$ für alle $y \in \mathcal{Y}$, $y \neq 0$:

$$\|\tilde{x} + y\|^2 = \|\tilde{x}\|^2 + 2\langle \tilde{x}, y \rangle + \|y\|^2 = \|\tilde{x}\|^2 + \|y\|^2 > \|\tilde{x}\|^2$$

Für die umgekehrte Implikation $\hat{x} = \mathcal{P}_{\mathcal{Y}} x \implies \langle \tilde{x}, y \rangle = 0$ nehmen wir an, daß es ein y mit $\langle \tilde{x}, y \rangle = \alpha \neq 0$ gäbe. Dann gilt für alle $\lambda \in \mathbb{R}$:

$$\|\tilde{x} + \lambda y\|^2 = \|\tilde{x}\|^2 + 2\lambda\alpha + \lambda^2\|y\|^2.$$

Mit $\lambda^* = -\frac{\alpha}{\|y\|^2}$ ergibt sich

$$\|\tilde{x} + \lambda^*y\|^2 = \|\tilde{x}\|^2 - \frac{\alpha^2}{\|y\|^2} < \|\tilde{x}\|^2.$$

Dies ist ein Widerspruch zur Minimalität von $\tilde{x} = x - \hat{x}$. \square

Wir wenden uns nun den Pseudo-Inversen zu, die den Begriff der Inversen auf nichtquadratische und singuläre Matrizen erweitern. Wir werden sehen, daß es sich bei Projektionen und Pseudoinversen um verwandte Konzepte handelt.

DEFINITION A.10 (Pseudo-Inverse). Es seien $A \in \mathbb{R}^{m \times n}$ und $X \in \mathbb{R}^{n \times m}$ Matrizen und $P \in \mathbb{R}^{n \times n}$, $P \succ 0$, $Q \in \mathbb{R}^{m \times m}$ und $Q \succ 0$ symmetrische positiv definite Matrizen. Dann lauten die *elliptischen Penrose-Bedingungen für X und A* :

1. $AXA = A$,
2. $XAX = X$,
3. $(PAX)^\top = PAX$,
4. $(XAQ)^\top = XAQ$.

Eine Matrix X heißt *generalisierte Inverse von A des Typs $(i j \dots l)$* , wenn sie die Bedingungen $(i), (j), \dots, (l)$ erfüllt, dabei seien $i, j, \dots, l \in \{1, 2, 3, 4\}$. Eine generalisierte Inverse vom Typ $(1 2 3 4)$ heißt eine elliptische (oder gewichtete) Moore-Penrose-Pseudo-Inverse. Allgemein spricht man von einer Pseudo-Inversen (PI) X von A , wenn mindestens die Bedingung (1) erfüllt ist.

So sind z.B. idempotente Matrizen (1 2)-selbstinvers, ist A invertierbar, so erfüllt $X = A^{-1}$ alle diese Bedingungen.

Eine Reformulierung des Projektionslemmas mit Pseudo-Inversen lautet:

PROPOSITION A.11. *Es sei $X \in \mathbb{R}^{n \times m}$ eine Pseudo-Inverse von $A \in \mathbb{R}^{m \times n}$ mit den Eigenschaften (1) und (3). Dann wird $\min_y \|x - Ay\|_P^2$ gelöst durch $y = Xx + (I - XA)z$, wobei $z \in \mathbb{R}^n$ beliebig.*

BEWEIS. Durch Differentiation von $\|x - Ay\|_P^2$ nach y erhält man die Gleichung für den kritischen Punkt y : $A^\top P A y = A^\top P x$. Dann erfüllt $y^* = Xx + (I - XA)z$ diese Gleichung:

$$\begin{aligned} A^\top P A y^* &= A^\top P A X x + A^\top P A (I - X A) z \\ &\stackrel{(3)}{=} A^\top (P A X)^\top x + A^\top P (A - X A A) z \\ &\stackrel{(1)}{=} A^\top X^\top A^\top P x + 0 \\ &\stackrel{(1)}{=} A^\top P x. \end{aligned}$$

y^* ist also kritischer Punkt von $\|x - Ay\|_P^2$. Da dieser Term ≥ 0 und quadratisch ist, handelt es sich bei y um ein Minimum. \square

Die folgende Charakterisierung dient dazu, eine Matrix unter allen Matrizen mit dieser Minimal-Quadrate-Eigenschaft eindeutig auszuzeichnen.

PROPOSITION A.12. *Es sei $X \in \mathbb{R}^{n \times m}$ eine elliptische Moore-Penrose-Pseudo-Inverse von $A \in \mathbb{R}^{m \times n}$. Dann ist X eindeutig bestimmt.*

BEWEIS. Y erfülle ebenso wie X die Bedingungen (1)-(4). Mit Bedingung (1) und (4) folgt:

$$\begin{aligned} XAQ &= (XAQ)^\top = (XAYAQ)^\top = YAQ(XA)^\top \\ &= YA(XAQ)^\top = YAXAQ = YAQ, \\ &\text{da } Q \text{ invertierbar ist, gilt } XA = YA \end{aligned}$$

Analog erhält man mit (1) und (3) und der Invertierbarkeit von P : $AX = AY$. Also ist nach Bedingung (2) $X = XAX = YAX = YAY = Y$. \square

Zur Existenz beschränken wir uns auf den Fall $P = I$. Dann lassen sich Darstellungen für die verschiedenen Inversentypen aus einer Singulärwertzerlegung gewinnen.

PROPOSITION A.13 (Singulärwertzerlegung).

Zu jeder Matrix $A \in \mathbb{R}^{n \times m}$ mit $\text{Rang } A = r \leq m < n$ existieren zwei orthogonale Matrizen $U \in \mathbb{R}^{n \times n}$ und $V \in \mathbb{R}^{m \times m}$, so daß für diese Singulärwertzerlegung (SVD) gilt

$$A = USV^\top \quad \text{mit } S = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}, \quad S \in \mathbb{R}^{m \times n}, \Sigma \in \mathbb{R}^{r \times r},$$

wobei Σ eine Diagonalmatrix mit absteigend geordneten positiven Diagonaleinträgen ist.

BEWEIS. Siehe [Sch93, Satz 7.3].

Diese Zerlegung läßt sich nun einsetzen, um damit Pseudoinversen verschiedenen Typs zu finden.

PROPOSITION A.14. *Es sei*

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^\top$$

die SVD von A . Dann gilt für die Pseudoinversen vom

- Typ (1) (Innere Inverse)

$$A^{(1)} = V \begin{pmatrix} \Sigma^{-1} & K \\ L & M \end{pmatrix} U^\top$$

- Typ (1 2)

$$A^{(12)} = V \begin{pmatrix} \Sigma^{-1} & K \\ L & L\Sigma K \end{pmatrix} U^\top$$

- Typ (1 2 3)

$$A^{(123)} = V \begin{pmatrix} \Sigma^{-1} & 0 \\ L & 0 \end{pmatrix} U^\top$$

- Typ (1 2 4)

$$A^{(124)} = V \begin{pmatrix} \Sigma^{-1} & K \\ 0 & 0 \end{pmatrix} U^\top$$

- Typ (1 2 3 4) (Moore-Penrose-Inverse)

$$A^{(1)} = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^\top,$$

wobei die verwendeten Matrizen folgende Größen besitzen :
 $K \in \mathbb{R}^{r \times (n-r)}$, $L \in \mathbb{R}^{(m-r) \times r}$ und $M \in \mathbb{R}^{(m-r) \times (n-r)}$.

BEWEIS. Siehe [Ste90].

Die Verbindung zu den Projektionen wird dadurch hergestellt, daß die Matrizen XA und AX idempotent und damit Projektoren sind.

PROPOSITION A.15. *Ist X eine (1)-Inverse von $A \in \mathbb{R}^{m \times n}$ und ist A vom Rang r , dann sind AX und XA idempotente Matrizen vom Rang r und es gilt: AX ist Projektion auf $\text{Im}A$ und XA ist Projektion auf $\text{ker}A$.*

Außerdem hat man eine explizite Formel für Pseudo-Inversen, die mit (55) zusammenhängt.

PROPOSITION A.16. *Für $A \in \mathbb{R}^{m \times n}$ mit $\text{rang}A = m$ gilt für die elliptische Moore-Penrose PI folgende Formel:*

$$(56) \quad X = (A^\top P A)^{-1} A^\top P$$

BEWEIS. Wir wissen bereits aus dem Beweis von Proposition A.11, daß gilt

$$A^\top P A y = A^\top P x.$$

Für die Aussage dieser Proposition muß man nur noch die Matrix $A^\top P A$ invertieren, was dank des vollen Ranges von A möglich ist. Wir überprüfen nun direkt die Eigenschaften. Es gilt $XA = I$, also sind die Bedingungen (1),(2) und (4) aus der Definition (A.10) trivialerweise erfüllt. $PAX = PA(A^\top P A)^{-1} A^\top P$ ist symmetrisch, also ist auch Bedingung (3) erfüllt. \square

Wir haben bereits mehrere Darstellungen von Pseudo-Inversen kennengelernt, die sich für eine algorithmische Umsetzung eignen. Im Folgenden wird ein Verfahren vorgestellt, das ohne Matrix-Inversion auskommt.

PROPOSITION A.17. *Es sei $A \in \mathbb{R}^{m \times n}$ mit den Spalten a_1, \dots, a_n und $P \in \mathbb{R}^{m \times m}$, $P \succ 0$. Setze $A_k = (a_1 \dots a_k) \in \mathbb{R}^{m \times k}$ für $k \leq n$. Um das Verfahren zu starten, setze $X_1 = 0$, falls $a_1 = 0$ (bzw. unter einer Toleranzschranke), sonst setze*

$$X_1 = (a_1^\top P a_1)^{-1} a_1^\top P.$$

Ist X_{k-1} bereits ermittelt, so suche

$$X_k = \begin{pmatrix} X_{k-1} - d_k b_k \\ b_k \end{pmatrix}$$

mit $d_k = X_{k-1} a_k$. Um b_k zu bestimmen, definieren wir $c_k = a_k - A_{k-1} d_k$. Ist $c_k \neq 0$, so setze

$$b_k = (c_k^\top P c_k)^{-1} c_k^\top P,$$

sonst

$$b_k = (1 + d_k^\top d_k)^{-1} X_{k-1}$$

Dann ist $X := X_n$ die elliptische Moore-Penrose-Inverse von A zu der positiv definiten Matrix P (Setze $Q = I$ in der Definition A.10).

BEWEIS. Dieses Verfahren entspricht einer verallgemeinerten Version der Gram-Schmidt Orthonormierung. Siehe [Kuh76] für den ungeachteten Fall $P=I$.

Numerische Experimente lassen vermuten, daß dieses einfache Verfahren numerisch nicht so stabil ist wie ein auf SVD beruhendes Verfahren. Der absolute Maximalwert der Abweichung $AXA - A$ für eine nach Proposition A.17 berechnete Pseudo-Inverse ist etwa zehnmal größer verglichen mit einer durch SVD berechneten PI. Allerdings reicht die relative Spannweite von „genauso gut“ bis 700fach schlechter. Die Werte $\max_{i,j} |AXA - A|_{ij}$ liegen aber immer noch unter 10^{-10} .

ANHANG B

Verteilungen und Testtheorie

In diesem Anhang erinnern wir an die Definition eines Hypothesentests und an einige dazu nötige Verteilungen. Außerdem erweitern wir die üblichen statistischen Begriffe auf den höherdimensionalen Fall, für den eindimensionalen Fall verweisen wir auf [Kre91].

DEFINITION B.1 (Erwartungswert). Es sei X ein n -dimensionaler Zufallsvektor, ein Spaltenvektor aus n Zufallsvariablen X_i . Dann besitzt X den Erwartungswert

$$EX = (EX_i)_{i=1,\dots,n}$$

Die Eigenschaften und Rechenregeln für Erwartungswerte übertragen sich komponentenweise auf den höherdimensionalen Fall.

DEFINITION B.2 (Kovarianz). Es seien X und Y n -dimensionale reelle Zufallsvariablen. Dann heißt

$$\text{Cov}(X, Y) = (\text{Cov}(X_i, Y_j))_{i,j=1,\dots,n} = E((X - EX)(Y - EY)^\top)$$

die *Kovarianzmatrix* von X und Y . Abkürzend schreiben wir $\text{Cov } X = \text{Cov}(X, X)$. Die Diagonaleinträge dieser Matrix sind die *Varianzen* der einzelnen Komponenten.

Für die Definition von *Dichten* und *Verteilungen* sei wieder auf [Kre91] Abschnitt 10.2 verwiesen. Wir führen folgende Schreibweisen ein: Besitzen X und Y die gleiche Verteilung, so schreiben wir $X \stackrel{\mathcal{L}}{=} Y$. Gehört X der Verteilung $\mathcal{L}(X)$, so bezeichnen wir dies mit $X \stackrel{\mathcal{L}}{\sim} \mathcal{L}(X)$. Wir wollen hier noch den Begriff der Verteilungskonvergenz definieren.

DEFINITION B.3 (Verteilungskonvergenz). Es seien P und P_n Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$ für $n \in \mathbb{N}$ mit den Verteilungsfunktionen F und F_n . Die Folge der (P_n) bzw. (F_n) heißt *verteilungskonvergent* gegen P bzw. F , falls gilt:

$$\lim_{n \rightarrow \infty} F_n(a) = F(a) \quad \text{für alle Stetigkeitsstellen } a \in \mathbb{R} \text{ von } F.$$

Hierfür führen wir die Schreibweisen

$$P_n \xrightarrow{\mathcal{L}} P \quad F_n \xrightarrow{\mathcal{L}} F$$

ein. Die Verteilungskonvergenz läßt sich auch auf Folgen (X_n) von Zufallsvariablen übertragen, indem man fordert, daß die zugehörigen Verteilungen verteilungskonvergent sind:

$$X_n \xrightarrow{\mathcal{L}} X : \iff \mathcal{L}(X_n) \xrightarrow{\mathcal{L}} \mathcal{L}(X) \quad \text{für } X_n \stackrel{\mathcal{L}}{\sim} \mathcal{L}(X_n), X \stackrel{\mathcal{L}}{\sim} \mathcal{L}(X)$$

DEFINITION B.4 (Normalverteilung). Als (eindimensionale) *Normalverteilung* mit Erwartungswert μ und Varianz σ^2 , kurz $\mathcal{N}(\mu, \sigma^2)$, bezeichnet man die Verteilung mit der Dichte

$$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x - \mu)^2}{2\sigma^2}.$$

Eine *n-dimensionale oder multivariate Normalverteilung* X läßt sich dadurch beschreiben, daß alle Linearkombinationen wieder normalverteilt sind:

$$X \stackrel{\mathcal{L}}{\sim} \mathcal{N}_n \iff \langle t, X \rangle \stackrel{\mathcal{L}}{\sim} \mathcal{N}_1 \quad \forall t \in \mathbb{R}^n.$$

Wie auch im eindimensionalen Fall ist die multivariate Normalverteilung eindeutig durch den Erwartungsvektor μ und die Kovarianzmatrix Σ festgelegt.

BEMERKUNG. Eine normalverteilte Zufallsvariable X heißt *standardnormalverteilt*, wenn $EX = 0$ und $\text{Cov } X = 1$ gilt.

Für multivariate Normalverteilungen existiert bei singulärer Kovarianzmatrix keine Dichte, damit die Definition dann zutrifft, müssen wir Konstanten als normalverteilt mit Varianz 0 ansehen.

PROPOSITION B.5 (Transformation der Normalverteilung). *Ist X ein n -dimensionaler normalverteilter Zufallsvektor mit Erwartungswert x und Kovarianz D , $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine lineare Transformation, so gilt*

$$X \stackrel{\mathcal{L}}{\sim} \mathcal{N}(x, D) \implies TX \stackrel{\mathcal{L}}{\sim} \mathcal{N}(Tx, TDT^\top);$$

BEMERKUNG. Für nicht normalverteilte ZV Y gelten diese Beziehungen für Erwartungswert und Varianz als direkte Folge aus den Definitionen:

$$\begin{aligned} E(TY) &= T(EY) \\ \text{Cov}(TY) &= T \text{Cov}(Y) T^\top \end{aligned}$$

Weitere in dieser Arbeit verwendete Verteilungen sind die F- und Chi²-Verteilung, deren Definition im Folgenden angegeben wird.

DEFINITION B.6. Sind $\xi_i \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, 1)$ ($i = 1, \dots, n$) unabhängige standardnormalverteilte Zufallsvariablen, so heißt die Verteilung von

$$\sum_{i=1}^n \xi_i^2 \stackrel{\mathcal{L}}{\sim} \chi_n^2$$

eine Chi²-Verteilung mit n Freiheitsgraden. Sie besitzt eine Dichte φ_n , die nur auf dem positiven Bereich $\mathbb{R}_{\geq 0}$ konzentriert ist, und dort gegeben ist durch

$$\varphi_n(x) = x^{\left(\frac{n}{2}\right)-1} e^{-\frac{x}{2}} \left(2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)\right)^{-1},$$

wobei Γ die Eulersche Gammafunktion ist.

Sind die Normalverteilungen nicht zentriert ($E\xi_i \neq 0$), so ergibt sich eine nichtzentrale Chi²-Verteilung. Die Nichtzentralität ist dann gegeben durch

$$\delta = \sum_1^n (E\xi_i)^2.$$

DEFINITION B.7. Sind $U \stackrel{\mathcal{L}}{\sim} \chi_n^2$ und $V \stackrel{\mathcal{L}}{\sim} \chi_m^2$ zwei unabhängige Chi²-verteilte Zufallsvariablen mit Freiheitsgrad n bzw. m , so heißt die Verteilung von

$$\frac{U/n}{V/m} \stackrel{\mathcal{L}}{\sim} F_{n,m}$$

eine F-Verteilung mit n Zählerfreiheitsgraden und m Nennerfreiheitsgraden. Die Dichte $\varphi_{n,m}$, die auf den positiven Bereich $\mathbb{R}_{\geq 0}$ konzentriert ist, wird gegeben durch

$$\varphi_{n,m} = d_{n,m} \frac{x^{(\frac{n}{2})-1}}{(nx+m)^{\frac{m+n}{2}}},$$

wobei $d_{n,m}$ eine normierende Konstante ist.

Die in den Definitionen angegebenen Formeln für die Dichten erfordern eigentlich Beweise, die hier nicht aufgeführt werden. Statt dessen zeigen wir eine asymptotische Eigenschaft der F-Verteilung.

PROPOSITION B.8. Für wachsenden Nenner-Freiheitsgrad m konvergiert die $F_{n,m}$ -Verteilung gegen eine skalierte χ^2 -Verteilung:

$$F_{n,m} \xrightarrow{\mathcal{L}} \frac{\chi_n^2}{n} \quad \text{für } m \rightarrow \infty.$$

BEWEIS. Da die Dichten normiert sind, können wir Faktoren aus den Normierungskonstanten herausziehen:

$$(57) \quad m^{\frac{m}{2}} n^{\frac{n}{2}} n^{\frac{n}{2}-1} \frac{(nx)^{\frac{n}{2}-1}}{(nx+m)^{\frac{m+n}{2}}} = \frac{(nx)^{\frac{n}{2}-1}}{\left(1 + \frac{nx}{m}\right)^{\frac{m}{2}} \left(x + \frac{m}{n}\right)^{\frac{n}{2}}}.$$

Dieser Term ist für $m \rightarrow \infty$ asymptotisch proportional zu

$$(58) \quad (nx)^{(\frac{n}{2})-1} e^{-\frac{nx}{2}},$$

denn es gilt

$$\lim_{m \rightarrow \infty} \left(1 + \frac{nx}{m}\right)^{\frac{m}{2}} = \lim_{t \rightarrow \infty} \left(1 + \frac{nx}{2t}\right)^t = \exp\left(\frac{nx}{2}\right),$$

der zweite Term im Nenner von (57) ist im Grenzwert unabhängig von x , dieser wird also zu normierenden Konstante hinzugefügt. Gleichung (58) ist proportional zu der Dichte von $\frac{\chi_n^2}{n}$, denn ist $\varphi(x)$ die Dichte einer Zufallsvariablen X , so besitzt $\frac{1}{n}X$ die Dichte $n\varphi(nx)$, was aus der allgemeinen Dichte-Transformationsformel folgt, die eine Anwendung der Substitutionsformel für Integrale ist. \square

Diese Verteilungen werden zur Konstruktion von Teststatistiken gebraucht. Die Definition eines Tests und der zugehörigen Begriffe wird in folgender Definition zusammengefaßt.

DEFINITION B.9 (Test). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $X : \Omega \rightarrow \mathbb{R}^n$ ein n -dimensionaler Zufallsvektor mit einem durch Θ parametrisierten Verteilungsmodell $\mathcal{L}(X) = \{\mathcal{L}_\theta(X) | \theta \in \Theta\}$ und einer disjunkten Zerlegung des Parameterbereiches $\Theta = H_0 + H_1$ in die *Nullhypothese* H_0 und die *Alternative* H_1 . Außerdem sei $\varphi : \mathbb{R}^n \rightarrow \{0, 1\}$ eine meßbare Abbildung, die *Entscheidungsfunktion*. Für eine Realisierung x des Zufallsvektors X wird dann der Test gegeben durch

- $\varphi(x) = 1$: Hypothese wird zugunsten der Alternative verworfen.
- $\varphi(x) = 0$: Hypothese wird beibehalten.

Weiterhin wird die Wahrscheinlichkeit, daß der Test die Nullhypothese H_0 ablehnt, als *Schärfe* des Tests bezeichnet: $\text{Pow}(\theta) = P_\theta(\varphi(X)) = \int_\Omega \varphi(X(\omega)) \mathcal{L}_\theta(dP)$. Das *Fehlerrisiko erster Art* bezeichnet die Wahrscheinlichkeit, daß der Test ablehnt, obwohl die Hypothese zutrifft = $\text{Pow}(\theta)$ für $\theta \in H_0$. Das maximale Fehlerrisiko 1. Art heißt *Niveau* des Tests : $\alpha = \sup\{\text{Pow}(\theta) | \theta \in H_0\}$

Meist wird φ als charakteristische Funktion¹ eines *Ablehnungsbereiches* gegeben, der wiederum als Schwellwertüberschreitung einer *Teststatistik* T modelliert wird:

$$\varphi(x) = \chi_{\{T(x) > t_0\}}.$$

Die Aufgabe besteht nun darin, zu einem gegebenem Testniveau eine Teststatistik mit entsprechendem Quantil zu konstruieren.

¹Die charakteristische Funktion $\chi_{\mathcal{M}}$ einer Menge \mathcal{M} ist definiert durch $\chi_{\mathcal{M}}(x) = 1$ für $x \in \mathcal{M}$ und $\chi_{\mathcal{M}}(x) = 0$ für $x \notin \mathcal{M}$.

Literaturverzeichnis

- [AMo79] B.D.O. Anderson und J.B. Moore, Optimal filtering, Prentice-Hall, Englewood Cliffs, N. J., 1979
- [Ben74] A. Ben-Israel und T.N.E. Greville, General Inverses : Theory and Application, Wiley, New York, 1974
- [Ben76] A. Ben-Israel und T.N.E. Greville, Some Topics in Generalized Inverses of Matrices, in
M.Z. Nashed (Ed.), General Inverses and Application, Academic Press, New York, 1976
- [Ber95] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, Mass., 1995
- [Bil79] P. Billingsley, Probability and Measure, Wiley, New York, 1979
- [BBM97] T. Binder, L. Blank, W. Dahmen und W. Marquardt, Towards Multiscale Dynamic Data Reconciliation, Lehrstuhl für Prozeßtechnik, RWTH Aachen, 1997
- [Bre73] L. Breiman, Statistics – With a View Toward Application, Houghton Mifflin, Boston, Mass. , 1973
- [Bro97] I. Bronstein et.al., Taschenbuch der Mathematik, 3. Auflage, Deutsch, Thun, 1997
- [Col90] T.F. Coleman, On Characterizations of Superlinear Convergence for constrained optimization, in
E.L. Allgover und K. Georg (Ed.), Computational solution of nonlinear systems of equations, Lectures in Applied Mathematics Vol. 26, American Mathematical Society, Providence, Rhode Island, 1990
- [DeB94] P. Deuffhard und F. Bornemann, Numerische Mathematik, Band II – Integration gewöhnlicher Differentialgleichungen, de Gruyter, Berlin, 1994
- [Fle87] R. Fletcher, Practical Methods of Optimization, Wiley, Chichester, 1987
- [GF178] H. Grauert und W. Fischer, Differential- und Integralrechnung II, Springer, Berlin, 1978
- [Hor98] K. Horneffer, Skript zur Differentialgeometrie, Vorlesungsskript Sommersemester 1998, Universität Bremen
- [Jaz70] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970
- [Kuh76] F. Kuhnert, Pseudoinverse Matrizen und die Methode der Regularisierung, Teubner, Leipzig, 1976
- [Kre91] U. Krengel, Einführung in die Wahrscheinlichkeitstheorie und Statistik, Vieweg, Bielefeld, 1991
- [KwS72] H. Kwakernaak und R. Sivan, Linear optimal control systems, Wiley-Interscience, New York, 1972
- [LEL92] M.J. Liebman, T.F. Edgar und L.S. Lasdon, Efficient Data Reconciliation and Estimation for Dynamic Processes using Nonlinear Programming Techniques, Comput. Chem. Eng., Vol 16(10/11), pp 963-986, 1992
- [Mil90] A.J. Miller, Subset selection in regression, Chapman and Hall, London, 1990

- [Sch93] H.R. Schwarz, Numerische Mathematik, Teubner, Stuttgart, 1993
- [Ste90] G.W. Stewart und J.-G. Sun, Matrix Pertubation Theory, Academic Press, Boston, 1990
- [Str95] S.Streit, Meßunsicherheit und Verträglichkeitsprüfungen, in Fachtagung Informationserfassung und -verarbeitung in der Energietechnik, VDI Berichte 1210, VDI Verlag, Düsseldorf, 1995
- [VDI98] VDI-Richtlinie 2048 Blatt 1 (Entwurf), Meßunsicherheiten bei Abnahmemessungen an energie- und kraftwerkstechnischen Anlagen – Grundlagen, Beuth, Berlin, 1998
- [VSa92] T. Viéville und P. Sanders, Using pseudo Kalman-Filters in the presence of constraints, INRIA Rapport de Recherche No 1669, <ftp://ftp.inria.fr/INRIA/publication/publi-ps-gz/RR/RR-1669.ps.gz>, 1992
- [WWa90] H.F. Walker und L.T. Watson: Least-change Secant Update Methods for Underdetermined Systems, SIAM J. Numer. Math., Vol 27(5), pp. 1227-1262, 1990