

Parameterschätzung aus unvollständigen Beobachtungsdaten mittels des EM-Algorithmus

Marc Luxen und Ansgar Brunn

Zusammenfassung

Dieser Artikel behandelt die Auswertung von unvollständigen Beobachtungsdaten. Zur Lösung der entsprechenden Schätzaufgabe wird der *Expectation Maximization (EM)-Algorithmus* vorgestellt. Das Problem der Schätzung unbekannter Parameter aufgrund unvollständiger Beobachtungsdaten wird in den geodätischen Kontext eingeordnet. Anhand der Linienextraktion aus digitalen Bildern wird der EM-Algorithmus exemplarisch dargestellt.

Summary

This paper gives an introduction into the problem of parameter estimation from incomplete data and presents the Expectation Maximization Algorithm (EM-Algorithm) as a method to solve such problems. The algorithm is put in relation to geodetic estimation problems. Its practicability is shown by an example of line extraction from digital images.

1 Einleitung

Der vorliegende Aufsatz stellt den *Expectation-Maximization-Algorithmus (EM-Algorithmus)*, ein in der Statistik und Wahrscheinlichkeitstheorie beheimatetes Schätzverfahren im Kontext geodätischer Anwendungen vor. Das dem EM-Algorithmus zugrunde liegende Modell ist eine Dichte $p(\mathbf{l}, \mathbf{m} \mid \mathbf{x})$, wobei \mathbf{l} bekannte Beobachtungen, \mathbf{m} fehlende Beobachtungen und \mathbf{x} unbekannte, feste Parameter sind. Bei den fehlenden Beobachtungen handelt es sich um nicht gemessene oder Messungen nicht zugängliche kontinuierliche oder diskrete Zufallsvariablen,¹ so dass der Vektor $(\mathbf{l}, \mathbf{m})^T$ der Beobachtungen insgesamt unvollständig ist.

In der Geodäsie wurden bisher vor allem Schätzprobleme mit vollständigen Beobachtungsvektoren behandelt, z. B. die Bestimmung von Geradenparametern, die Höhen- oder die Richtungsnetzausgleichung (s. Tab. 1).

Außer diesem Aufgabentyp liegen in einigen Bereichen der Geodäsie Klassifikationsprobleme mit unbekanntem Klassenparametrisierungen vor. Es handelt sich dabei um Probleme, die eine Modellzuordnung bei gleichzeitiger Parameterschätzung erfordern, z. B.

- die Bewegungsanalyse von Punktfeldern, wenn die Punktbewegungen in verschiedene Bewegungsklassen eingeteilt werden können (s. Tab. 1, Nr. 4),
- die Auswertung von GPS-Phasenmessungen (s. Tab. 1, Nr. 5),
- die Zuordnung von Bildpunkten in Stereoaufnahmen (Hornegger und Niemann 1997),
- die Bewegungserkennung und -analyse von Personen in Bildfolgen einer Überwachungskamera (Torr 1997) und
- die Extraktion von Linien aus Linienpixeln in Bildern oder Digitalen Oberflächenmodellen (s. Kap. 3).

Der EM-Algorithmus ermöglicht die Lösung solcher Aufgaben im Rahmen eines iterativen Prozesses.

Im Folgenden wird nach einer kurzen Einführung in die Problematik der Parameterschätzung aus unvollständigen Beobachtungsdaten und einer Vorstellung des EM-Algorithmus in Kapitel 2 die Lösung der Linienrekonstruktion aus Digitalbildern mittels des EM-Algorithmus

¹ Im Rahmen dieses Aufsatzes beschränken wir uns jedoch auf diskrete Zufallsvariablen als fehlende Beobachtungen.

	Aufgabe	unbekannte Parameter	tatsächliche Beobachtungen	fehlende Beobachtungen
1	Geradenbestimmung	Geradenparameter	Punktkoordinaten	-
2	Höhennetzausgleichung	Punkthöhen	Höhendifferenzen	-
3	Richtungsnetzausgleichung	Koordinaten, Teilkreisorientierungen	Richtungsmessungen	-
4	Gemeinsame Klassifikation und Parameterschätzung	Klassenparametrisierung	Merkmale	Klassenzugehörigkeit
5	Auswertung von GPS-Phasenmessungen	Strecken	Phasen	Anzahl ganzer Wellenlängen

Tab. 1: Beispiele von geodätischen Schätzproblemen mit einer Einordnung der Beobachtungen und Parameter in die Begriffe des EM-Algorithmus.

demonstriert. Dazu leiten wir in Kapitel 3 das Modell beispielhaft für die Extraktion einer geraden Linie und einer Parabel her und wenden die Schätzung auf ein Beispiel aus der Bildinterpretation an. Der Artikel schließt mit einer Zusammenfassung und einem Ausblick.

2 Der EM-Algorithmus

Im Vermessungswesen werden von den zahlreich existierenden Methoden der Parameterschätzung vor allem die Methode der kleinsten Quadrate, die Maximum-Likelihood-Methode und die beste erwartungstreue Schätzung verwendet (vgl. Koch 1997). Genau wie diese Schätzmethoden ist der Expectation Maximization Algorithmus ein auf der Maximum-Likelihood-Methode basierendes Verfahren zur Schätzung unbekannter Parameter anhand beobachteter Merkmale in einem Beobachtungsmodell.

Der EM-Algorithmus eignet sich zur *Parameterschätzung mit unvollständigen Beobachtungsdaten*, weswegen die entsprechende Problemstellung zunächst definiert und anhand eines Beispiels veranschaulicht wird.

2.1 Parameterschätzung mit unvollständigen Daten

Definition: Es sei $\mathbf{l} \in \mathbb{R}^I$ ein $I \times 1$ -Zufallsvektor von Beobachtungen, $\mathbf{m} \in \mathbb{R}^F$ ein unbekannter $F \times 1$ Zufallsvektor und

$$p(\mathbf{l}, \mathbf{m} | \mathbf{x})$$

die von den unbekanntem, festen Parametern \mathbf{x} abhängige Dichtefunktion des Vektors $(\mathbf{l}^T, \mathbf{m}^T)^T$. Dann bezeichnen wir das Problem der Schätzung der unbekanntem Parameter \mathbf{x} nach der Maximum-Likelihood (ML)-Methode, d. h. die Bestimmung der Schätzwerte

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{l}, \mathbf{m} | \mathbf{x}) = \arg \max_{\mathbf{x}} \log p(\mathbf{l}, \mathbf{m} | \mathbf{x}) \quad (1)$$

als *Schätzproblem mit unvollständigen Beobachtungsdaten*, \mathbf{m} als den Vektor *fehlender Beobachtungen* und $\mathbf{k} = (\mathbf{l}^T, \mathbf{m}^T)^T$ als den *vollständigen Beobachtungsvektor*².

Zur Veranschaulichung soll folgendes Beispiel dienen: Gegeben seien I Beobachtungen $l_i, i = 1, \dots, I$, die sich jeweils auf eines von mehreren Objekten $\mathcal{O}_1, \dots, \mathcal{O}_J$ mit den unbekanntem Parametern $\mathbf{x}_1, \dots, \mathbf{x}_J$ beziehen. Im konkreten Fall können z. B. Koordinaten l_i von Punkten

einer Gerade \mathcal{O}_1 und einer Parabel \mathcal{O}_2 (d. h. $J = 2$) gemessen sein, wobei die Parameter \mathbf{x}_1 der Gerade und \mathbf{x}_2 der Parabel gesucht sind (s. Abb. 1). Es sei a priori unklar, welche Beobachtung sich auf welches Objekt bezieht. Diese Unsicherheit wird mit \mathbf{m} modelliert.

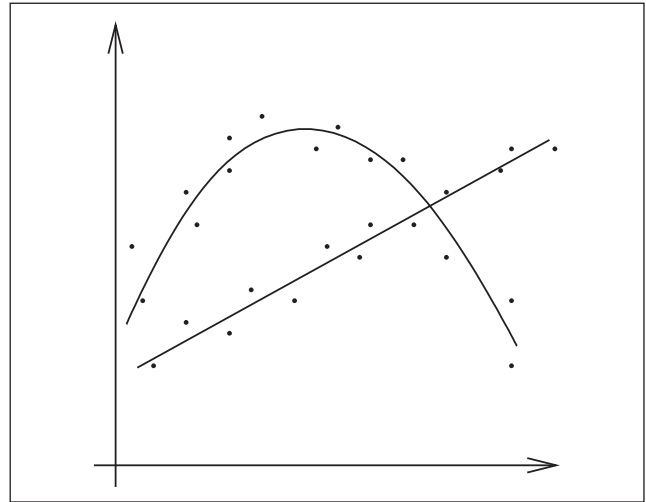


Abb. 1: Ausgleichende Gerade und ausgleichende Parabel

Im Beispiel sei die Likelihoodfunktion der i -ten Beobachtung l_i gegeben durch

$$p(l_i | \mathbf{x}) = \begin{cases} p_{1i}(l_i | \mathbf{x}_1) & \text{falls } l_i \rightarrow \mathcal{O}_1 \\ \dots\dots\dots & \\ p_{ji}(l_i | \mathbf{x}_j) & \text{falls } l_i \rightarrow \mathcal{O}_j \end{cases}, \quad (2)$$

worin $p_{ji}(l_i | \mathbf{x}_j)$ die Dichtefunktion der Beobachtung l_i angibt für den Fall, dass sie zu dem durch \mathbf{x}_j parametrisierten Objekt \mathcal{O}_j gehört. Mit dem Vektor $\mathbf{m}_i = (m_{1i}, \dots, m_{ji})^T$ stochastischer Variablen m_{1i}, \dots, m_{ji} , von denen genau eine den Wert 1 und jede andere den Wert 0 annimmt sowie dem Vektor $\mathbf{x} = (\mathbf{x}_1^T, \dots, \mathbf{x}_j^T)^T$ der unbekanntem Parameter ergibt sich der zu (2) äquivalente Ausdruck

$$p(l_i | \mathbf{m}_i, \mathbf{x}) = \prod_{j=1}^J [p_{ji}(l_i | \mathbf{x}_j)]^{m_{ji}}$$

Aufgrund des Bayes-Theorems (vgl. Koch 2000)

$$p(l_i, \mathbf{m}_i | \mathbf{x}) = p(\mathbf{m}_i | \mathbf{x}) \cdot p(l_i | \mathbf{m}_i, \mathbf{x})$$

erhält man, falls von einer Gleichverteilung $p(\mathbf{m}_i | \mathbf{x}) \propto 1$ der Zuordnungen ausgegangen wird,

$$p(l_i, \mathbf{m}_i | \mathbf{x}) \propto p(l_i | \mathbf{m}_i, \mathbf{x}).$$

Wird stochastische Unabhängigkeit der Vektoren (l_h, \mathbf{m}_h) und (l_i, \mathbf{m}_i) für $h \neq i$ vorausgesetzt, gilt mit $\mathbf{l} = (l_1, \dots, l_I)^T$ und $\mathbf{m} = (\mathbf{m}_1^T, \dots, \mathbf{m}_I^T)^T$

$$p(\mathbf{l}, \mathbf{m} | \mathbf{x}) \propto \prod_{i=1}^I \prod_{j=1}^J [p_{ji}(l_i | \mathbf{x}_j)]^{m_{ji}}$$

und wir haben es angesichts der fehlenden Information über \mathbf{m} bei der Schätzung der unbekanntem Parameter nach (1) mit einer Schätzung aus unvollständigen Daten zu tun.

² Dempster et al. 1977 verstehen das Problem der Schätzung aus unvollständigen Beobachtungsdaten allgemeiner. Nach der Auffassung der Autoren reicht die hier angegebene Definition jedoch für die meisten in der Geodäsie relevanten Anwendungen aus.

2.2 Ableitung des EM-Algorithmus

Für Ausgleichungen mit unvollständigen Beobachtungsdaten sollen nun nach (1) Schätzwerte für die unbekannt Parameter abgeleitet werden. Dies führt uns auf den EM-Algorithmus.

Offensichtlich ist (1) zur Bestimmung der unbekannt Parameter ungeeignet, da der vollständige Beobachtungsvektor $\mathbf{k} = (\mathbf{l}^T, \mathbf{m}^T)^T$ die unbekannt Elemente des $F \times 1$ Zufallsvektors \mathbf{m} enthält, so dass $p(\mathbf{l}, \mathbf{m} | \mathbf{x})$ nicht in Abhängigkeit von \mathbf{x} ausgewertet werden kann. Es liegt daher nahe, als Likelihoodfunktion statt $p(\mathbf{l}, \mathbf{m} | \mathbf{x})$ die Randdichte

$$p(\mathbf{l} | \mathbf{x}) = \int \cdots \int_{\mathbf{m} \in \mathbb{R}^F} p(\mathbf{l}, \mathbf{m} | \mathbf{x}) d\mathbf{m} \quad (3)$$

der Beobachtungen \mathbf{l} zu verwenden. Das Problem der fehlenden Beobachtungen wird durch Integration der ursprünglichen Likelihoodfunktion $p(\mathbf{l}, \mathbf{m} | \mathbf{x})$ über alle $\mathbf{m} \in \mathbb{R}^F$ gelöst, denn in $p(\mathbf{l} | \mathbf{x})$ sind die fehlenden Beobachtungen eliminiert.

Da die Auswertung des Integrals (3) rechnerisch meist aufwendig ist, geben wir die Dichte $p(\mathbf{l} | \mathbf{x})$ noch in einer zweiten Form an, die als Ausgangspunkt für die Ableitung des EM-Algorithmus dient.

Nach dem Bayes-Theorem gilt

$$p(\mathbf{l} | \mathbf{x}) = \frac{p(\mathbf{l}, \mathbf{m} | \mathbf{x})}{p(\mathbf{m} | \mathbf{l}, \mathbf{x})}.$$

Hiermit erhält die logarithmierte Likelihoodfunktion $L(\mathbf{x})$ die Form

$$L(\mathbf{x}) = \log p(\mathbf{l} | \mathbf{x}) = \log p(\mathbf{l}, \mathbf{m} | \mathbf{x}) - \log p(\mathbf{m} | \mathbf{l}, \mathbf{x})$$

und die ML-Schätzung der unbekannt Parameter folgt aus

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \max_{\mathbf{x}} L(\mathbf{x}) \\ &= \arg \max_{\mathbf{x}} [\log p(\mathbf{l}, \mathbf{m} | \mathbf{x}) - \log p(\mathbf{m} | \mathbf{l}, \mathbf{x})]. \end{aligned} \quad (4)$$

Auch in Gleichung (4) kann die rechte Seite nicht ausgewertet werden, da $(\mathbf{l}^T, \mathbf{m}^T)^T$ nicht vollständig bekannt ist. Wir lösen dieses Problem ähnlich wie in (3) durch Integration. Dazu fassen wir die logarithmierte Likelihoodfunktion $L(\mathbf{x})$ als Zufallsvariable auf und berechnen die Erwartungswerte $Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) = E_{\mathbf{m}}[\log p(\mathbf{l}, \mathbf{m} | \mathbf{x}) | \mathbf{l}, \mathbf{x}^{(\nu)}]$ und $H(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) = E_{\mathbf{m}}[\log p(\mathbf{m} | \mathbf{l}, \mathbf{x}) | \mathbf{l}, \mathbf{x}^{(\nu)}]$ der Funktionen $\log p(\mathbf{l}, \mathbf{m} | \mathbf{x})$ und $\log p(\mathbf{m} | \mathbf{l}, \mathbf{x})$ über alle $\mathbf{m} \in \mathbb{R}^F$, d. h.

$$Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) = \int \cdots \int_{\mathbf{m} \in \mathbb{R}^F} \log p(\mathbf{l}, \mathbf{m} | \mathbf{x}) \cdot p(\mathbf{m} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) d\mathbf{m}, \quad (5)$$

$$H(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) = \int \cdots \int_{\mathbf{m} \in \mathbb{R}^F} \log p(\mathbf{m} | \mathbf{l}, \mathbf{x}) \cdot p(\mathbf{m} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) d\mathbf{m}. \quad (6)$$

Der Erwartungswert $Q(\mathbf{x} | \mathbf{l}, \mathbf{x}^{(\nu)})$ wird in diesem Zusammenhang als *Kullback-Leibler-Statistik* und $-H(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)})$ als *Entropie* bezeichnet (vgl. Koch 2000, Papoulis 1991).

Zur Berechnung von Q und H setzen wir voraus, dass mit $\hat{\mathbf{x}}^{(\nu)}$ Näherungswerte für die unbekannt Parameter vorliegen, so dass mit $p(\mathbf{m} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)})$ die Verteilung des Vektors der fehlenden Beobachtungen näherungsweise angegeben werden kann. Der Erwartungswert der logarithmierten Likelihoodfunktion ergibt sich dann in Abhängigkeit von den unbekannt Parametern zu $E[L(\mathbf{x}) | \mathbf{l}, \mathbf{x}^{(\nu)}] = Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) - H(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)})$. Mit $E[L(\mathbf{x}) | \mathbf{l}, \mathbf{x}^{(\nu)}] = L(\mathbf{x})$ (vgl. Luxen 2000) folgt daraus

$$L(\mathbf{x}) = Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) - H(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}).$$

Die logarithmierte Likelihoodfunktion liegt damit in einer Form vor, in der sie nur noch von den unbekannt Parametern \mathbf{x} , den Näherungswerten $\hat{\mathbf{x}}^{(\nu)}$ und den Beobachtungen \mathbf{l} , aber nicht mehr von den fehlenden Beobachtungen \mathbf{m} abhängt. Die ML Schätzung der unbekannt Parameter ergibt sich nun aus der *Grundgleichung des EM-Algorithmus*,

$$\begin{aligned} \hat{\mathbf{x}}^{(\nu+1)} &= \arg \max_{\mathbf{x}} L(\mathbf{x}) \\ &= \arg \max_{\mathbf{x}} [Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) - H(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)})]. \end{aligned} \quad (7)$$

Der EM-Algorithmus basiert nun auf folgender Überlegung: Nach (Dempster 1977) gilt für jeden Parametersatz \mathbf{x} und jede Näherung $\hat{\mathbf{x}}^{(\nu)}$ die Beziehung $H(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) \leq H(\mathbf{x}^{(\nu)} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)})$. Somit ergibt sich, falls $Q(\mathbf{x} | \mathbf{l}, \mathbf{x}^{(\nu)}) \geq Q(\mathbf{x}^{(\nu)} | \mathbf{l}, \mathbf{x}^{(\nu)})$ gilt, auch sofort die Beziehung $L(\mathbf{x}) \geq L(\hat{\mathbf{x}}^{(\nu)})$. Dies gilt als Begründung dafür, in (7) die Entropie H zu vernachlässigen und bei der Suche nach Schätzwerten für die unbekannt Parameter nur die Kullback-Leibler-Statistik zu betrachten. Der EM-Algorithmus bestimmt also die Schätzung der unbekannt Parameter nach

$$\hat{\mathbf{x}}^{(\nu+1)} = \arg \max_{\mathbf{x}} Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}).$$

Der Schätzwert hängt von den zugrundeliegenden Näherungswerten $\mathbf{x}^{(\nu)}$ ab, so dass ein zweistufiges iteratives Verfahren erforderlich ist:

1. Im ersten Schritt, dem **Expectation Schritt** (E-Schritt) wird die Kullback-Leibler Statistik nach (5) unter Verwendung von Näherungswerten für die unbekannt Parameter berechnet. Dies erfordert in Spezialfällen lediglich die Bestimmung des Erwartungswertes $E(\mathbf{m})$ (s. Kap. 3).
2. Im zweiten Schritt, dem **Maximization Schritt** (M-Schritt) wird auf Grundlage des Ergebnisses von Schritt 1 die Kullback-Leibler Statistik durch Variation von \mathbf{x} maximiert, um zu verbesserten Schätzwerten $\hat{\mathbf{x}}^{(\nu+1)}$ zu gelangen. Sie dienen in der nächsten Iteration als neue Näherungswerte.

Diese beiden Iterationsschritte haben dem EM-Algorithmus seinen Namen gegeben. Abb. 2 gibt den Algorithmus in Form eines Struktogramms wieder.

Gegeben: Dichte $p(l, m x)$
Bestimme symbolisch die Randdichte $p(m l, x)$
Initialisierung: $v := 0, \hat{x}^{(0)}$
E-Schritt: Berechne $Q(x \hat{x}^{(v)})$ nach (5)
M-Schritt: Berechne ML-Schätzung $\hat{x}^{(v+1)} := \arg \max_x Q(x \hat{x}^{(v)})$
$v \rightarrow v + 1$
bis Abbruchkriterium erfüllt
Ausgabe der Schätzung

Abb. 2: Struktogramm des EM-Algorithmus

Ausgehend von einer gegebenen Dichte $p(l, m | x)$ des vollständigen Beobachtungsvektors wird die Randdichte $p(m | l, x)$ symbolisch bestimmt. Falls dies nicht möglich ist, erfolgt die Bestimmung numerisch in Verbindung mit dem E-Schritt. Nach Vorgabe von Näherungswerten $\hat{x}^{(0)}$ für die unbekannt Parameter wird das oben erklärte, aus dem Expectation-Schritt und dem Maximization-Schritt bestehende iterative Verfahren so lange angewendet, bis ein Abbruchkriterium, z. B. $\|\hat{x}^{(v)} - \hat{x}^{(v-1)}\| < \epsilon$ erfüllt ist.

2.3 Eigenschaften

Der EM-Algorithmus konvergiert unter relativ schwachen Bedingungen gegen ein lokales Maximum der logarithmierten Likelihoodfunktion (vgl. Wu 1982). Die Schwäche des EM-Algorithmus liegt in seiner im Allgemeinfall nur langsamen Konvergenzgeschwindigkeit.

Der Vorteil des EM-Algorithmus gegenüber der ML-Schätzung auf der Randdichte (3) besteht darin, dass die Kullback-Leibler-Statistik in vielen Anwendungen in Abhängigkeit von x und \hat{x} symbolisch angegeben und durch Variation von x maximiert werden kann. Das Integral (3) muss dagegen fast immer mit numerischen Verfahren ausgewertet und maximiert werden, was den Rechenaufwand der direkten ML-Schätzung gegenüber dem EM-Algorithmus deutlich erhöht.

3 Beispiel: Linienextraktion aus Linienpixeln

Gegeben sei ein digitales Bild, in dem $J = 2$ linienhafte Objekte \mathcal{O}_1 und \mathcal{O}_2 abgebildet sind (s. Abb. 3, links). Zur Überführung des Bildes in ein Vektorformat seien die Bildkoordinaten einzelner Punkte der beiden Objekte mittels eines Verfahrens zur Extraktion von Linienpixeln bestimmt worden (s. Abb. 3, rechts). Ohne Vorkenntnis, welches der insgesamt I Koordinatenpaare $(r, c)_i, i = 1, \dots, I$ sich auf welches Objekt bezieht, soll nun Objekt \mathcal{O}_1 durch eine Gerade und Objekt \mathcal{O}_2 durch eine Parabel approximiert werden. Der Vektor $x = (x_1^T, x_2^T)^T$ der unbekannt Parameter setzt sich somit zusammen aus den Koeffizienten $x_1 = (g_0, g_1)^T$ einer ausgleichenden Gerade und $x_2 = (h_0, h_1, h_2)^T$ einer ausgleichenden Parabel. Um die Ableitungen kurz zu halten, gehen wir davon aus, dass die Abszissen c_i der beobachteten Punkte fehlerfrei gemessen wurden und nur die Ordinaten r_i mit Messunsicherheiten behaftet sind.

3.1 Formulierung als Schätzproblem aus unvollständigen Beobachtungsdaten

Wir fassen zunächst im Vektor $l = (r_1, r_2, \dots, r_I)^T$ die gemessenen Ordinaten zusammen. Wir setzen voraus, dass die Beobachtungen r_i paarweise stochastisch voneinander unabhängig sind und je nach Objektzugehörigkeit unterschiedliche Varianzen σ_1^2 und σ_2^2 besitzen³. Als Beobachtungsgleichungen kommen folgende Modellansätze im Sinne eines Gauß-Markoff-Modells in Frage:

Modell 1 (Objekt \mathcal{O}_1 , Gerade):

$$r_i + v_{i1} = (1, c_i) \begin{pmatrix} g_0 \\ g_1 \end{pmatrix} = \mathbf{a}_i^T \mathbf{x}_1 \quad \text{mit } V(r_i) = \sigma_1^2, \quad (8)$$

Modell 2 (Objekt \mathcal{O}_2 , Parabel):

$$r_i + v_{i2} = (1, c_i, c_i^2) \begin{pmatrix} h_0 \\ h_1 \\ h_2 \end{pmatrix} = \mathbf{b}_i^T \mathbf{x}_2 \quad \text{mit } V(r_i) = \sigma_2^2. \quad (9)$$

³ U. U. sind die beiden Objekte im Bild nicht gleich gut lokalisierbar. Wir unterscheiden daher zwischen den Varianzen der zu \mathcal{O}_1 und \mathcal{O}_2 gehörenden Beobachtungen.

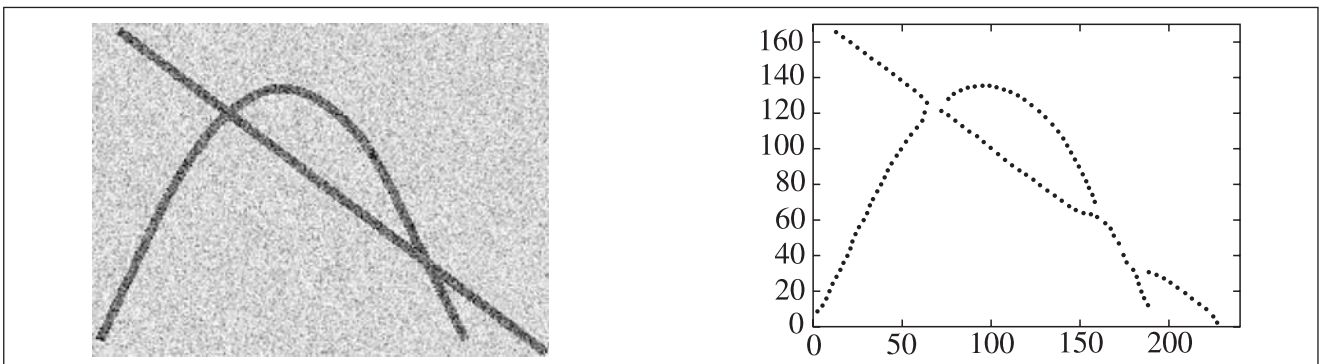


Abb. 3: Links: Verrauschtes Bild mit einer Geraden und einer Parabel. Rechts: Linienpunkte, automatisch extrahiert

In beiden Modellen verschwinden wegen stochastischer Unabhängigkeit die Kovarianzen zwischen verschiedenen Beobachtungen, d. h. $C(r_i, r_j) = 0$ für alle $i \neq j$.

Im Beobachtungsmaterial fehlen Informationen über die Zugehörigkeit der einzelnen Beobachtungen r_i zu \mathcal{O}_1 oder \mathcal{O}_2 . Daher wird formal für jede Beobachtung r_i ein *Indikatorvektor*

$$\mathbf{m}_i = \begin{pmatrix} m_{i1} \\ m_{i2} \end{pmatrix} \in \{e_1, e_2\} \quad \text{mit } e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

fehlender Beobachtungen m_{i1} und m_{i2} eingeführt, der die Zugehörigkeit jeder Beobachtung zu \mathcal{O}_1 oder \mathcal{O}_2 anzeigt. Die Indikatorvektoren \mathbf{m}_i bilden gemeinsam den $2I \times 1$ Vektor $\mathbf{m} = (m_1^T, \dots, m_I^T)^T$ der fehlenden Beobachtungen. Der vollständige Beobachtungsvektor des Schätzproblems aus unvollständigen Daten ist somit gegeben durch

$$\mathbf{k} = (\mathbf{l}^T, \mathbf{m}^T)^T = (r_1, \dots, r_I, m_1^T, \dots, m_I^T)^T.$$

3.2 Ableitung der EM-Iterationsschritte

3.2.1 Herleitung der Dichte $p(\mathbf{l}, \mathbf{m} | \mathbf{x})$

Zur Ableitung der Dichtefunktion $p(\mathbf{l}, \mathbf{m} | \mathbf{x})$ wird zunächst mit $P(\mathbf{m}_i | \mathbf{x})$ die Wahrscheinlichkeitsverteilung der diskreten Zufallsvariablen $\mathbf{m}_i \in \{e_1, e_2\}$ bezeichnet, die a priori jeder Beobachtung r_i die Wahrscheinlichkeiten $P(\mathbf{m}_i = e_1 | \mathbf{x})$ bzw. $P(\mathbf{m}_i = e_2 | \mathbf{x})$ zuordnet, dass sie zu Modell 1 bzw. zu Modell 2 gehört, sofern mit \mathbf{x} die Parameter der beiden Modelle gegeben sind. Es wird hier angenommen, dass die Zuordnungen für alle Beobachtungen r_i gleichverteilt und unabhängig von \mathbf{x} sind,⁴ d. h.

$$\begin{aligned} P(\mathbf{m}_i = e_1 | \mathbf{x}) &= P_1 & \text{und} \\ P(\mathbf{m}_i = e_2 | \mathbf{x}) &= P_2 = 1 - P_1. \end{aligned} \quad (10)$$

Die Beobachtungen r_i werden als normalverteilt vorausgesetzt. Es gilt also für Modell 1 (Gerade)

$$\begin{aligned} p(r_i | e_1, \mathbf{x}) &= p(r_i | \mathbf{m}_i = e_1, \mathbf{x}) \\ &= \frac{1}{\sqrt{2\pi\sigma_1}} \exp\left(-\frac{1}{2\sigma_1^2}[r_i - \mathbf{a}_i^T \mathbf{x}_1]^2\right) \end{aligned} \quad (11)$$

und für Modell 2 (Parabel)

$$\begin{aligned} p(r_i | e_2, \mathbf{x}) &= p(r_i | \mathbf{m}_i = e_2, \mathbf{x}) \\ &= \frac{1}{\sqrt{2\pi\sigma_2}} \exp\left(-\frac{1}{2\sigma_2^2}[r_i - \mathbf{b}_i^T \mathbf{x}_2]^2\right) \end{aligned} \quad (12)$$

Die Dichtefunktion $p(r_i, \mathbf{m}_i | \mathbf{x}) = p(r_i, \mathbf{m}_i | x_1, \sigma_1, x_2, \sigma_2)$ einer vollständigen Beobachtung $(r_i, \mathbf{m}_i^T)^T$ ergibt sich damit aufgrund des Bayes-Theorems für $j = 1, 2$ zu

$$\begin{aligned} p(r_i, \mathbf{m}_i = e_j | \mathbf{x}) &= P(\mathbf{m}_i = e_j | \mathbf{x}) \cdot p(r_i | e_j, \mathbf{x}) \\ &= P_j \cdot p(r_i | e_j, \mathbf{x}), \end{aligned} \quad (13)$$

bzw. unter Ausnutzung der Tatsache, dass in Abhängigkeit von dem für die Beobachtung r_i zutreffenden Modell entweder $\mathbf{m}_i = (1, 0)^T$ oder $\mathbf{m}_i = (0, 1)^T$ gilt:

$$p(r_i, \mathbf{m}_i | \mathbf{x}) = p(r_i, \mathbf{m}_i = e_1 | \mathbf{x})^{m_{i1}} \cdot p(r_i, \mathbf{m}_i = e_2 | \mathbf{x})^{m_{i2}}.$$

Die Vektoren $(r_i, \mathbf{m}_i^T)^T, i = 1, \dots, I$ werden als paarweise stochastisch unabhängig voneinander aufgefasst. Daher gilt $p(\mathbf{l}, \mathbf{m} | \mathbf{x}) = \prod_{i=1}^I p(r_i, \mathbf{m}_i | \mathbf{x})$ und man erhält die logarithmierte Likelihoodfunktion

$$\log p(\mathbf{l}, \mathbf{m} | \mathbf{x}) = \sum_{i=1}^I [m_{i1} \log\{P_1 \cdot p(r_i | e_1, \mathbf{x})\} + m_{i2} \log\{P_2 \cdot p(r_i | e_2, \mathbf{x})\}], \quad (14)$$

deren Erwartungswert im E-Schritt des EM-Algorithmus berechnet und im M-Schritt maximiert werden muss.

3.2.2 E-Schritt: Berechnung der Kullback-Leibler-Statistik

Sind mit $\hat{\mathbf{x}}^{(\nu)}$ vorläufige Schätzwerte für die unbekannt Parameter $\mathbf{x} = (x_1^T, \sigma_1, x_2^T, \sigma_2)^T$ bekannt, so ergibt sich die Kullback-Leibler-Statistik $Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)})$ als Erwartungswert der logarithmierten Likelihoodfunktion (14). Offensichtlich ist dieser Erwartungswert hier linear in den Indikatoren m_{i1} und m_{i2} , so dass mit $\mu_{i1}^{(\nu)} \doteq E_m[m_{i1} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}]$ und $\mu_{i2}^{(\nu)} \doteq E_m[m_{i2} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}]$ gilt

$$\begin{aligned} Q(\hat{\mathbf{x}} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) &= \sum_{i=1}^I [\mu_{i1}^{(\nu)} \log\{P_1 \cdot p(r_i | e_1, \mathbf{x})\} \\ &\quad + \mu_{i2}^{(\nu)} \log\{P_2 \cdot p(r_i | e_2, \mathbf{x})\}]. \end{aligned} \quad (15)$$

Der E-Schritt ersetzt im vorliegenden Fall also die fehlenden Beobachtungen m_{i1} und m_{i2} durch deren Erwartungswerte $\mu_{i1}^{(\nu)}$ und $\mu_{i2}^{(\nu)}$ in der ν -ten Iteration.

Da es sich bei m_{i1} und m_{i2} um diskrete Zufallsvariablen handelt, die nur die Werte 0 und 1 annehmen, gilt für $j = 1, 2$

$$\mu_{ij}^{(\nu)} = P(m_{ij} = 1 | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}),$$

d. h. die Erwartungswerte der Indikatoren geben die Wahrscheinlichkeiten für die Zugehörigkeit der Beobachtung r_i zu Objekt \mathcal{O}_1 bzw. \mathcal{O}_2 für den Fall an, dass die unbekannt Parameter durch $\hat{\mathbf{x}}^{(\nu)}$ und die Beobachtungen durch \mathbf{l} gegeben sind. Nach dem Bayes-Theorem gilt für diese Wahrscheinlichkeiten zunächst

$$P(m_{ij} = 1 | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) = \frac{p(\mathbf{l}, \mathbf{m}_i = e_j | \hat{\mathbf{x}}^{(\nu)})}{p(\mathbf{l} | \hat{\mathbf{x}}^{(\nu)})}.$$

4 Im Prinzip kann an dieser Stelle eine beliebige parametrisierbare Dichte $P(\mathbf{m}_i | \mathbf{x})$ angesetzt werden.

Der Zähler dieses Ausdrucks ergibt sich aufgrund der Bayes-Formel unter Berücksichtigung der statistischen Unabhängigkeit der Beobachtungen r_i mit (10) zunächst zu $p(\mathbf{l}, \mathbf{m}_i = e_j | \hat{\mathbf{x}}^{(\nu)}) = P_j \cdot \prod_{t=1}^I p(r_t | \mathbf{m}_i = e_j, \hat{\mathbf{x}}^{(\nu)})$ und schließlich, da nach (11) und (12) die Dichtefunktionen $p(r_t | \mathbf{m}_i = e_j, \hat{\mathbf{x}}^{(\nu)})$ für $t \neq i$ unabhängig vom Indikator \mathbf{m}_i sind, zu

$$p(\mathbf{l}, \mathbf{m}_i = e_j | \hat{\mathbf{x}}^{(\nu)}) = P_j \cdot p(r_i | \mathbf{m}_i = e_j, \hat{\mathbf{x}}^{(\nu)}) \cdot \prod_{\substack{t=1 \\ t \neq i}}^I p(r_t | \hat{\mathbf{x}}^{(\nu)}).$$

Außerdem gilt für den Nenner $p(\mathbf{l} | \hat{\mathbf{x}}^{(\nu)}) = \prod_{t=1}^I p(r_t | \hat{\mathbf{x}}^{(\nu)})$, womit man schließlich erhält

$$P(m_{ij} = 1 | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) = P_j \cdot \frac{p(r_i | \mathbf{m}_i = e_j, \hat{\mathbf{x}}^{(\nu)})}{p(r_i | \hat{\mathbf{x}}^{(\nu)})}.$$

Bei der Zugehörigkeit einer Beobachtung r_i zu \mathcal{O}_1 bzw. zu \mathcal{O}_2 handelt es sich um zwei sich gegenseitig ausschließende Ereignisse. Daher gilt $p(r_i | \hat{\mathbf{x}}^{(\nu)}) = p(r_i, \mathbf{m}_i = e_1 | \hat{\mathbf{x}}^{(\nu)}) + p(r_i, \mathbf{m}_i = e_2 | \hat{\mathbf{x}}^{(\nu)})$ und es folgt mit $\mu_{ij}^{(\nu)} = P(m_{ij} = 1 | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)})$ und (13) die Beziehung

$$\mu_{ij}^{(\nu)} = \frac{P_j \cdot p(r_i | \mathbf{m}_i = e_j, \hat{\mathbf{x}}^{(\nu)})}{P_1 \cdot p(r_i | \mathbf{m}_i = e_1, \hat{\mathbf{x}}^{(\nu)}) + P_2 \cdot p(r_i | \mathbf{m}_i = e_2, \hat{\mathbf{x}}^{(\nu)})}. \quad (16)$$

Im E-Schritt des hier zur Anwendung kommenden EM-Algorithmus sind die Erwartungswerte $\mu_{i1}^{(\nu)}$ und $\mu_{i2}^{(\nu)}$ für $i \in \{1, \dots, I\}$ nach (16) zu berechnen, wobei die hierfür benötigten Wahrscheinlichkeiten P_j und Dichten $p(r_i | \mathbf{m}_i = e_j, \hat{\mathbf{x}}^{(\nu)})$ mit (10), (11) und (12) gegeben sind. Im ν -ten Iterationsschritt ergibt sich dann nach (15) die Kullback-Leibler-Statistik in Abhängigkeit vom Parametervektor \mathbf{x} .

3.2.3 M-Schritt: Maximierung der Kullback-Leibler-Statistik

Im M-Schritt ist die Kullback-Leibler-Statistik (15) durch Variation der unbekannt Parameter $\mathbf{x} = (\mathbf{x}_1^\top, \sigma_1, \mathbf{x}_2^\top, \sigma_2)^\top$ zu maximieren. Durch Einsetzen der Dichten (11) und (12) in (15) erhält man nach Umstellung

$$Q(\mathbf{x} | \mathbf{l}, \hat{\mathbf{x}}^{(\nu)}) = \sum_{i=1}^I \left[\mu_{i1}^{(\nu)} \log \left(\frac{P_1}{\sqrt{2\pi}\sigma_1} \right) + \mu_{i2}^{(\nu)} \log \left(\frac{P_2}{\sqrt{2\pi}\sigma_2} \right) \right] - \sum_{i=1}^I \left[\frac{1}{2\sigma_1^2} \mu_{i1}^{(\nu)} (r_i - \mathbf{a}_i^\top \mathbf{x}_1)^2 + \frac{1}{2\sigma_2^2} \mu_{i2}^{(\nu)} (r_i - \mathbf{b}_i^\top \mathbf{x}_2)^2 \right].$$

Offensichtlich reicht es zur Bestimmung neuer Schätzwerte aus, ausschließlich die zweite Summe zu betrachten, da

nur diese von den unbekannt Parameter \mathbf{x}_1 und \mathbf{x}_2 abhängt. Diese Summe kann wieder in zwei Teilsummen zerlegt werden, die getrennt voneinander zu maximieren sind, d. h.

$$-\frac{1}{2\sigma_1^2} \sum_{i=1}^I \mu_{i1}^{(\nu)} [r_i - \mathbf{a}_i^\top \mathbf{x}_1]^2 \rightarrow \max, \\ -\frac{1}{2\sigma_2^2} \sum_{i=1}^I \mu_{i2}^{(\nu)} [r_i - \mathbf{b}_i^\top \mathbf{x}_2]^2 \rightarrow \max.$$

Fasst man nun die Erwartungswerte $\mu_{i1}^{(\nu)}$ bzw. $\mu_{i2}^{(\nu)}$ der fehlenden Beobachtungen in zwei Gewichtsmatrizen $\mathbf{P}_1^{(\nu)} = \text{diag}(\mu_{11}^{(\nu)}, \mu_{21}^{(\nu)}, \dots, \mu_{I1}^{(\nu)})$ bzw. $\mathbf{P}_2^{(\nu)} = \text{diag}(\mu_{12}^{(\nu)}, \mu_{22}^{(\nu)}, \dots, \mu_{I2}^{(\nu)})$ zusammen und entsprechend die Koeffizientenvektoren \mathbf{a}_i^\top bzw. \mathbf{b}_i^\top der Beobachtungsgleichungen des Modells 1, Gl. (8) bzw. des Modells 2, Gl. (9) in den Matrizen $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_I)^\top$ und $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_I)^\top$, so ergeben sich die Schätzwerte $\hat{\mathbf{x}}_1^{(\nu+1)}$ und $\hat{\mathbf{x}}_2^{(\nu+1)}$ nach

$$\hat{\mathbf{x}}_1^{(\nu+1)} = \arg \max_{\mathbf{x}_1} \left\{ -\frac{1}{2\sigma_1^2} (\mathbf{l} - \mathbf{A}\mathbf{x}_1)^\top \mathbf{P}_1^{(\nu)} (\mathbf{l} - \mathbf{A}\mathbf{x}_1) \right\} \quad (17)$$

und

$$\hat{\mathbf{x}}_2^{(\nu+1)} = \arg \max_{\mathbf{x}_2} \left\{ -\frac{1}{2\sigma_2^2} (\mathbf{l} - \mathbf{B}\mathbf{x}_2)^\top \mathbf{P}_2^{(\nu)} (\mathbf{l} - \mathbf{B}\mathbf{x}_2) \right\}. \quad (18)$$

Offensichtlich ergeben sich die neuen Schätzwerte $\hat{\mathbf{x}}_1^{(\nu+1)}$ und $\hat{\mathbf{x}}_2^{(\nu+1)}$ aufgrund zweier separater Parameterschätzungen in den Modellen 1 und 2 nach der Methode der kleinsten Quadrate, wenn die Beobachtungen r_i bei der Ausgleichung im Modell 1 mit den Gewichten $\mu_{i1}^{(\nu)}$ und bei der Ausgleichung im Modell 2 mit den Gewichten $\mu_{i2}^{(\nu)}$ versehen werden. Nach (Koch 1997) ergeben sich die Schätzungen mit

$$\hat{\mathbf{x}}_1^{(\nu+1)} = (\mathbf{A}^\top \mathbf{P}_1^{(\nu)} \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{P}_1^{(\nu)} \mathbf{l} \quad \text{und}$$

$$\hat{\mathbf{x}}_2^{(\nu+1)} = (\mathbf{B}^\top \mathbf{P}_2^{(\nu)} \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{P}_2^{(\nu)} \mathbf{l}. \quad (19)$$

Erwartungstreue Schätzwerte für die unbekannt Standardabweichungen σ_1 und σ_2 der Beobachtungen ergeben sich mit den geschätzten Verbesserungen $\hat{\vartheta}_1^{(\nu+1)} = \mathbf{A}\hat{\mathbf{x}}_1^{(\nu+1)} - \mathbf{l}$ und $\hat{\vartheta}_2^{(\nu+1)} = \mathbf{B}\hat{\mathbf{x}}_2^{(\nu+1)} - \mathbf{l}$ aus

$$(\hat{\sigma}_1^{(\nu+1)})^2 = \frac{(\hat{\vartheta}_1^{(\nu+1)})^\top \mathbf{P}_1^{(\nu)} \hat{\vartheta}_1^{(\nu+1)}}{I - 2} \quad \text{und} \\ (\hat{\sigma}_2^{(\nu+1)})^2 = \frac{(\hat{\vartheta}_2^{(\nu+1)})^\top \mathbf{P}_2^{(\nu)} \hat{\vartheta}_2^{(\nu+1)}}{I - 3}. \quad (20)$$

Mit (19) und (20) sind alle im M-Schritt des EM-Algorithmus zu berechnenden Größen bestimmt.

3.2.4 Kovarianzmatrizen der Schätzwerte und Klassifikation der Beobachtungen

Kovarianzmatrizen der Schätzungen: Nach Abschluss der EM-Iterationen werden Schätzungen $\hat{\Sigma}_{\hat{x}_1\hat{x}_1}$ und $\hat{\Sigma}_{\hat{x}_2\hat{x}_2}$ für die Kovarianzmatrizen der in der letzten Iteration geschätzten Parameter \hat{x}_1 und \hat{x}_2 angegeben. Die Schätzungen ergeben sich zu

$$\hat{\Sigma}_{\hat{x}_1\hat{x}_1} = \hat{\sigma}_1^2 \cdot (A^T P_1 A)^{-1} \quad \text{und}$$

$$\hat{\Sigma}_{\hat{x}_2\hat{x}_2} = \hat{\sigma}_2^2 \cdot (B^T P_2 B)^{-1}, \tag{21}$$

wobei $\hat{\sigma}_1$ und $\hat{\sigma}_2$ die in der letzten Iteration erhaltenen Schätzwerte für die Standardabweichungen der Beobachtungen und P_1 und P_2 die in der letzten Iteration erhaltenen Gewichtsmatrizen sind.

Klassifikation der Beobachtungen: Die angestrebte Zuordnung der Beobachtungen zu einem der beiden Objekte \mathcal{O}_1 bzw. \mathcal{O}_2 ist sehr einfach möglich. Nach (16) sind μ_{i1} und μ_{i2} die Wahrscheinlichkeiten, dass die Beobachtung r_i zu Objekt \mathcal{O}_1 bzw. zu Objekt \mathcal{O}_2 gehört.

Ist also r_i eine Beobachtung und ist o_i die Bezeichnung des Objektes, auf das sich r_i bezieht, erhält man die Klassifikation \hat{o}_i für $i = 1, \dots, I$ mit

$$\hat{o}_i = \begin{cases} \mathcal{O}_1 & , \text{ falls } \mu_{i1} > \mu_{i2} \\ \mathcal{O}_2 & , \text{ falls } \mu_{i1} \leq \mu_{i2} \end{cases}. \tag{22}$$

3.2.5 Einordnung des Ansatzes

In Abb. 4 ist der zur Lösung der gestellten Aufgabe geeignete Algorithmus mit allen erforderlichen Rechenschritten dargestellt. Offensichtlich handelt es sich hier um eine Parameterschätzung mit *iterativer Regewichtung der Beobachtungen*: Im E-Schritt jeder Iteration werden die

Gewichte $\mu_{ik}^{(\nu)}$ der Beobachtungen in den Modellen 1 und 2 auf Grundlage der vorherigen Schätzwerte $x^{(\nu)}$ neu berechnet und die Beobachtungen damit für die im M-Schritt folgende Parameterschätzung entsprechend ihrer Objektzugehörigkeit stärker oder weniger stark gewichtet.

Somit ist der sich für diese Aufgabenstellung ergebende EM-Algorithmus eng verwandt mit einer robusten Parameterschätzung (vgl. Koch 2000), bei der Ausreißer in den Beobachtungsdaten durch iterative Regewichtung der Beobachtungen sukzessive herabgewichtet werden, bis ihr Einfluss auf das Schätzergebnis vernachlässigbar gering ist. Im Unterschied zur robusten Schätzung mit Regewichtung besteht hier die Möglichkeit, zusätzlich zu einer Ausreißerklasse weitere parametrisierbare Alternativmodelle einzuführen.

3.3 Ergebnisse

Der abgeleitete Schätzalgorithmus wurde auf das in Abb. 3 (links) dargestellte synthetische, künstlich verauschte digitale Bild angewendet. Das Bild zeigt ein geradliniges und ein parabelartiges Objekt (z.B. einen Flusslauf und eine Straße). Abb. 3 (rechts) zeigt die Linienpixel, die mittels des am Institut für Photogrammetrie der Universität Bonn entwickelten Online-Merkmalsextraktionsverfahrens FEX⁵ (vgl. Fuchs 1998) aus dem Bild extrahiert wurden. Mit dem Extraktionsverfahren für Linienpixel wurden somit Punktbeobachtungen simuliert.

Mit dem im vorgehenden Abschnitt beschriebenen Verfahren wurden die Parameter der Geraden und der Parabel geschätzt. Die Näherungswerte wurden mit einer

5 WWW-Adresse: <http://www.ipb.uni-bonn.de/webfex.html>

EM-Algorithmus zur Lösung der Aufgabe:

1. **Erste Näherungswerte** $\hat{x}^{(0)}$ für die unbekanntnen Parameter x werden in einer initialen Schätzung nach (19) erhalten. Dazu werden beide Gewichtsmatrizen $P_1^{(0)} = P_2^{(0)} = I$ als Einheitsmatrizen vorgegeben.
2. **EM-Iterationen:**
 - i.) **E-Schritt:** Berechnung der Erwartungswerte $\mu_{ij}^{(\nu)}$ für $j = 1, 2$ und $i = 1, \dots, I$ mittels Gl. (16)
 - ii.) **M-Schritt:** Aufstellen der Matrizen $P_j^{(\nu)} = \text{diag}(\mu_{1j}^{(\nu)}, \mu_{2j}^{(\nu)}, \dots, \mu_{Ij}^{(\nu)})$ für $j = 1, 2$ und Berechnung der Schätzungen $\hat{x}_1^{(\nu+1)}, \hat{x}_2^{(\nu+1)}, \hat{\sigma}_1^{(\nu+1)}$ und $\hat{\sigma}_2^{(\nu+1)}$ nach den Gleichungen (19) und (20)
 - iii.) **Abbruchkriterium:** Falls $(\hat{x}_1^{(\nu+1)} - \hat{x}_1^{(\nu)})^2 + (\hat{x}_2^{(\nu+1)} - \hat{x}_2^{(\nu)})^2 < \epsilon$ gilt, wird der Algorithmus bei 3. fortgesetzt, sonst ist eine weitere EM-Iteration erforderlich.
3. **Berechnung der Kovarianzmatrizen und Klassifikation:**
Die Kovarianzmatrizen der Schätzungen \hat{x}_1 und \hat{x}_2 ergeben sich mit (20), die geschätzten Bezeichnungen \hat{o}_i der Objekte, zu denen die Beobachtungen r_i gehören, mit (22).

Abb. 4: Zusammenfassung der Rechenschritte

gleich gewichteten initialen Schätzung abgeleitet. Die Ergebnisse der iterativen EM-Schätzung sind sowohl getrennt für die Gerade und die Parabel als auch für beide zusammen in der Abb. 5 dargestellt. Die Einteilung der Beobachtungen in die beiden unterschiedlichen Beobachtungsgruppen wurde aus den Wahrscheinlichkeiten (s. Gl. 16) abgeleitet. Im Bereich der Schnittpunkte der Modelle ist die Klassifikation und damit auch die Trennung der Modelle unsicher.

4 Zusammenfassung

Der Aufsatz gibt eine Einführung in die Schätzung von Parametern bei unvollständigen Beobachtungsdaten. Er definiert das Schätzproblem, ordnet es in geodätische Fragestellungen ein und stellt den EM-Algorithmus als Methode zur Lösung dieses Typs von Schätzproblemen vor. Die Eigenschaften des EM-Algorithmus werden diskutiert und ein Beispiel aus der digitalen Bildverarbeitung gezeigt.

Durch das Beispiel werden zwei wichtige Randbedingungen der EM-Schätzung deutlich: Zum einen setzt die Verwendung des EM-Schätzers die Kenntnis über die Art und Anzahl der zur Beschreibung der Beobachtungen notwendigen Modelle voraus⁶. Zum anderen werden auch im Falle linearer Modelle Näherungswerte für die unbekannten Parameter benötigt.

Eine positive Eigenschaft der EM-Schätzung besteht darin, dass die im E-Schritt zu berechnende Kullback-Leibler-Statistik oftmals (insbesondere im Zusammenhang mit Exponentialverteilungen) unmittelbar symbolisch angegeben und maximiert werden kann. Die Schätzung ist dann im Vergleich zu anderen Schätzmethoden – z. B. der ML-Schätzung auf der Randdichte (3) – mit geringerem Rechenaufwand verbunden. Nach Auffassung der Autoren kann daher der EM-Algorithmus, der häufig im Bereich der Computer Vision zum Einsatz kommt, auch in der Geodäsie vorteilbringend eingesetzt werden.

6 Die Modellauswahl lässt sich in diesem Zusammenhang durch eine externe diskrete Optimierung (z. B. über die Beschreibungslänge, s. Rissanen 1987) erreichen.

Literatur

Dempster, A., Laird, N., Rubin, D.: Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39 (1): 1–38, 1977.

Fuchs, C.: *Extraktion polymorpher Bildstrukturen und ihre topologische und geometrische Gruppierung*. Deutsche geodätische Kommission, Bayer. Akademie der Wissenschaften, Reihe C, Heft 502, München, 1998.

Hornegger, J., Niemann, H.: Optimization Problems in Statistical Object Recognition. In: Pelillo, M., Hancock, E. R. (Hrsg.), *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Lecture Notes in Computer Science: 311–326. Springer-Verlag: Berlin, Heidelberg, New York, 1997.

Koch, K. R.: *Parameterschätzung und Hypothesentests in linearen Modellen*. Dümmler-Verlag, Bonn, 3. Aufl., 1997.

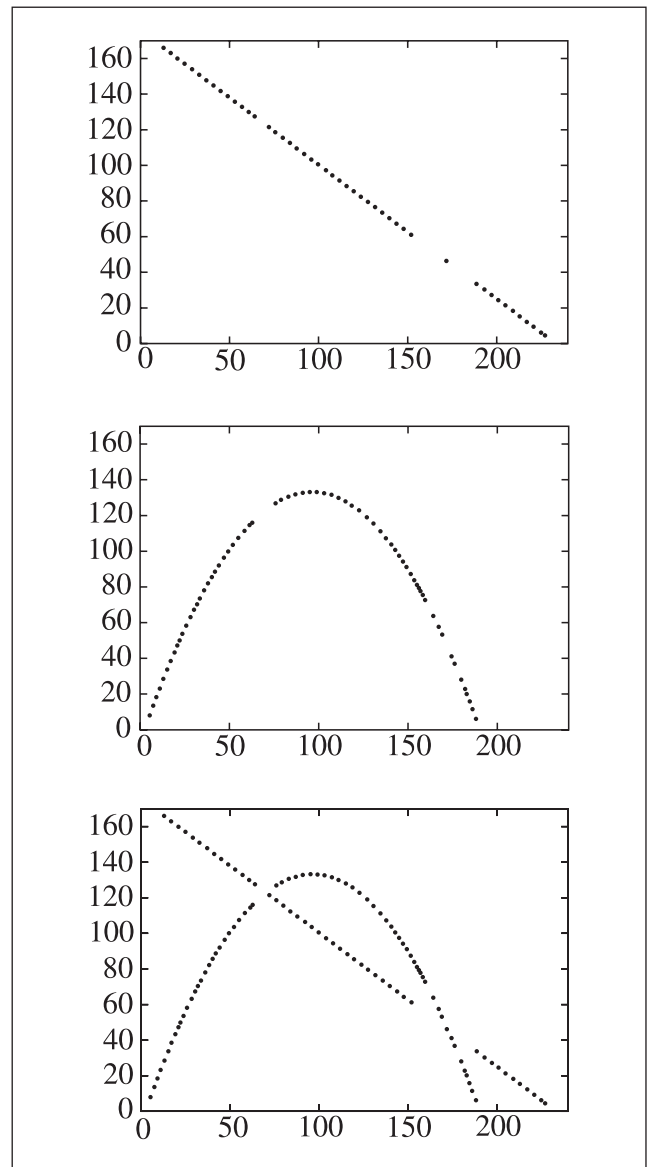


Abb. 5: Geschätzte Linienpixelpositionen der Geraden (oben), der Parabel (mitte) und beider Funktionen (unten).

Koch, K. R.: *Einführung in die Bayes-Statistik*. Springer-Verlag: Berlin, Heidelberg, New York, 2000.

Luxen, M.: *Der EM-Algorithmus bei Schätz- und Klassifikationsproblemen*. Diplomarbeit. Institut für Photogrammetrie, Universität Bonn, 2000.

Papoulis, A.: *Probability, Random Variables and Stochastic Processes*. McGraw Hill, New York, 3. Aufl., 1991.

Rissanen, I.: Minimum Description Length Principle. In: Kotz, S., Johnson, N. L. (Hrsg.), *Encyclopedia of Statistical Sciences*, Volume V: 523–527, John Wiley and Sons, New York, 1985.

Torr, P. H. S.: An Assessment of Information Criteria for Motion Model Selection. In: *Proceedings CVPR (Computer Vision and Pattern Recognition)*: 47–52, 1997.

Wu, C. F. J.: On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11 (1): 95–103, 1982.

Anschrift der Autoren

Dipl.-Ing. Marc Luxen und Dr.-Ing. Ansgar Brunn
 Institut für Photogrammetrie
 Rheinische Friedrich-Wilhelms-Universität Bonn
 Nussallee 15, D-53115 Bonn
 luxen@ipb.uni-bonn.de; brunn@ipb.uni-bonn.de
 www.ipb.uni-bonn.de