

Empirische Bayes-Schätzung in nichtlinearen statistischen Modellen

R. Offinger

An Beispielen werden nichtlineare Normalverteilungsmodelle vorgestellt, die bei Finite-Element-Modellen der Strukturmechanik auftreten. Hierbei versucht man durch Anpassung von Parametern eines Finite-Element-Modells an Experimente ein besseres mathematisches Modell für das Schwingungsverhalten zu erhalten. Für einige der Parameterkomponenten hat man dabei noch Zusatzinformationen. Beispielsweise hängen die Eigenfrequenzen einer Platte von den Dickschwankungen der Platte ab und es ist naheliegend, dabei eine räumliche Korrelation anzunehmen. Mathematisch beschreiben wird dies durch eine a-priori-Verteilung für die Dickenparameter, wobei allerdings hier die a-priori-Verteilung noch unbekannte Parameter aufweist, man also die räumliche Verteilung nicht vollständig kennt. Das Problem ist also die Parameterschätzung in nichtlinearen Modellen vom empirischen Bayesschen Typ, d.h. mit unvollständiger a-priori-Information. Es werden ein approximativer Algorithmus zur Lösung des Problems, der auf der Linearisierung des Modells beruht, und einige Ergebnisse hierzu vorgestellt.

1 Einführende Beispiele

1.1 Nichtlineare Normalverteilungsmodelle

Die Überlegungen im folgenden sind auch auf allgemeine nichtlineare Modelle von der Form

$$Y = \eta(\vartheta) + \epsilon \quad \epsilon \sim \mathcal{N}(0, \Sigma) \quad (1)$$

mit a-priori-Annahme $\vartheta \sim \mathcal{N}(\mu(\gamma), B(\rho))$ anwendbar. Dabei bezeichnet \mathcal{N} die multivariate Normalverteilung, d.h. der Fehlervektor $\epsilon \in R^N$ wird als normalverteilt mit Erwartungswert 0 und Kovarianzmatrix $\Sigma \in R^{N \times N}$ angenommen. Die Funktion $\eta : R^m \rightarrow R^N$ ist i.a. nichtlinear und der Mittelwert bzw. die Streuung der a-priori-Verteilung hängen i.a. von unbekanntem Parametern γ bzw. ρ ab.

Ein einfaches akademisches Testbeispiel ist etwa $\eta(\vartheta) = \frac{1}{2}\vartheta^2(1, \dots, 1)^t$, $\epsilon \sim \mathcal{N}(0, \sigma_0 I_N)$ (d.h. N Beobachtungen $Y_i = \frac{1}{2}\vartheta^2 + \epsilon_i$ mit unabhängigen und identisch $\mathcal{N}(0, \sigma_0^2)$ -verteilten Meßfehlern ϵ_i) mit bekannter Streuung σ_0^2 des Meßfehlers und a-priori-Annahme $\vartheta \sim \mathcal{N}(\vartheta_0, \rho^2)$ mit bekanntem a-priori-Mittelwert ϑ_0 und unbekannter a-priori-Streuung ρ^2 .

1.2 Eigenwerte und Eigenvektoren eines strukturellen Eigenwertproblem als Beobachtungen

Bei der Anpassung von Finite-Element-Modellen in der Strukturmechanik versucht man mit der Beobachtung von Eigenfrequenzen und Eigenformen eines mechanischen System ein besseres mathematisches Modell durch Parameteranpassungen zu erreichen. Der Beobachtungsvektor

$$Y = \underbrace{\begin{pmatrix} \lambda_I(\vartheta) \\ x_{LI}(\vartheta) \end{pmatrix}}_{\eta(\vartheta)} + \epsilon$$

(mit $I, L \subseteq \{1, \dots, n\}$) besteht hier aus (fehlerbehafteten) Meßwerten für die Eigenwerte $\lambda_i(\vartheta)$, $i \in I$, und für gewisse Komponenten x_{ij} , $i \in I$, $j \in L$, der zugehörigen Eigenvektoren $x_i(\vartheta) = (x_{i1}(\vartheta), \dots, x_{in}(\vartheta))^t$ des strukturellen Eigenwertproblems

$$K(\vartheta)x_i = \lambda_i M(\vartheta)x_i$$

des ungedämpften Systems

$$M(\vartheta)\dot{x} + K(\vartheta)x = 0$$

mit Strukturparameter $\vartheta \in R^m$, positiv definiten Massenmatrix $M(\vartheta) \in R^{n \times n}$ und positiv definiten Steifigkeitsmatrix $K(\vartheta) \in R^{n \times n}$.

Ein einfaches Beispiel hierzu ist das in Bild 1 dargestellte Massen-Feder-System mit n Massen m_1, \dots, m_n und $n + 1$ Steifigkeiten k_1, \dots, k_{n+1} :

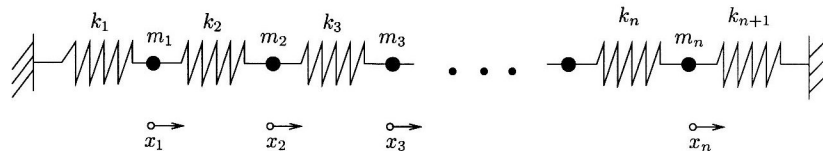


Bild 1: n-Massen-Feder-Schwinger in Reihenschaltung

Die Massen werden hier als bekannt vorausgesetzt, so daß der Strukturparameter $\vartheta = (k_1, \dots, k_{n+1})^t$ nur noch die unbekanntenen Steifigkeiten enthält. Die Massenmatrix ergibt sich zu $M(\vartheta) = \text{diag}(m_1, \dots, m_n)$ und die Steifigkeitsmatrix zu

$$K(\vartheta) = \begin{bmatrix} k_1 + k_2 & -k_2 & & & & \\ -k_2 & k_2 + k_3 & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & -k_{n-1} & k_{n-1} + k_n & -k_n \\ & & & & & -k_n & k_n + k_{n+1} \end{bmatrix}$$

Plausibel sind zwei verschiedene a-priori-Verteilungen: Zum einen

$$\vartheta \sim \mathcal{N}(\vartheta_0, \tau^2 I_{n+1})$$

d.h. die Parameterkomponenten sind unabhängig und identisch verteilt, mit bekanntem Mittelwert ϑ_0 und unbekannter Varianz τ^2 . Zum anderen mit räumlicher Korrelation zwischen den Komponenten, die mit zunehmender Entfernung exponentiell abklingt, d.h. $\text{Cov}(\vartheta_i, \vartheta_j) = \tau^2 \rho^{|i-j|}$. Genauer:

$$\vartheta \sim \mathcal{N}(\vartheta_0, B(\tau^2, \rho))$$

mit unbekanntem τ^2 und ρ und $B(\tau^2, \rho)_{ij} = \tau^2 \rho^{|i-j|}$.

1.3 Eine Platte mit nichtkonstanter Dicke

Der letzte Fall zeigt die Richtung auf, die man für praxisrelevante Modelle zu beschreiten hat. So beinhaltet ein Modell, das am Institut für Mechanik der Otto-von-Guericke-Universität untersucht wird (vgl. Zehn u.a. (2000), S.697ff.), als Parameterkomponenten die Dicken einer Platte an verschiedenen Stellen. Hierzu wurden mehrere Platten ausgemessen, um einen Eindruck zu gewinnen, welche Modelle für die räumliche Korrelation der Dickenparameter in Frage kommen. Die Dicke einer dieser Platten kann man aus Bild 2 ersehen.

Um die Verteilung der Dicken ϑ_i in den Punkten $v_i = (x_i, y_i)$, $i = 1, \dots, n$, geeignet zu beschreiben, wurde ein isotropes Modell für die Korrelationen gewählt:

$$\text{Cov}(\vartheta_i, \vartheta_j) = C(\|v_i - v_j\|)$$

Anschließend wurde versucht, verschiedene parametrische Modellen (vgl. Cressie 1993, S.61 f.) an das Variogramm $2\gamma(h) = 2C(0) - 2C(h)$ anzupassen (Beachte: $2\gamma(\|v_i - v_j\|) = \text{Var}(\vartheta_i - \vartheta_j)$, d.h. das Variogramm kann man aus den quadrierten Dickendifferenzen schätzen; die jeweiligen Mittelwerte dieser quadrierten Differenzen in Abhängigkeit vom Abstand sind in Bild 3 als Punktwolke dargestellt).

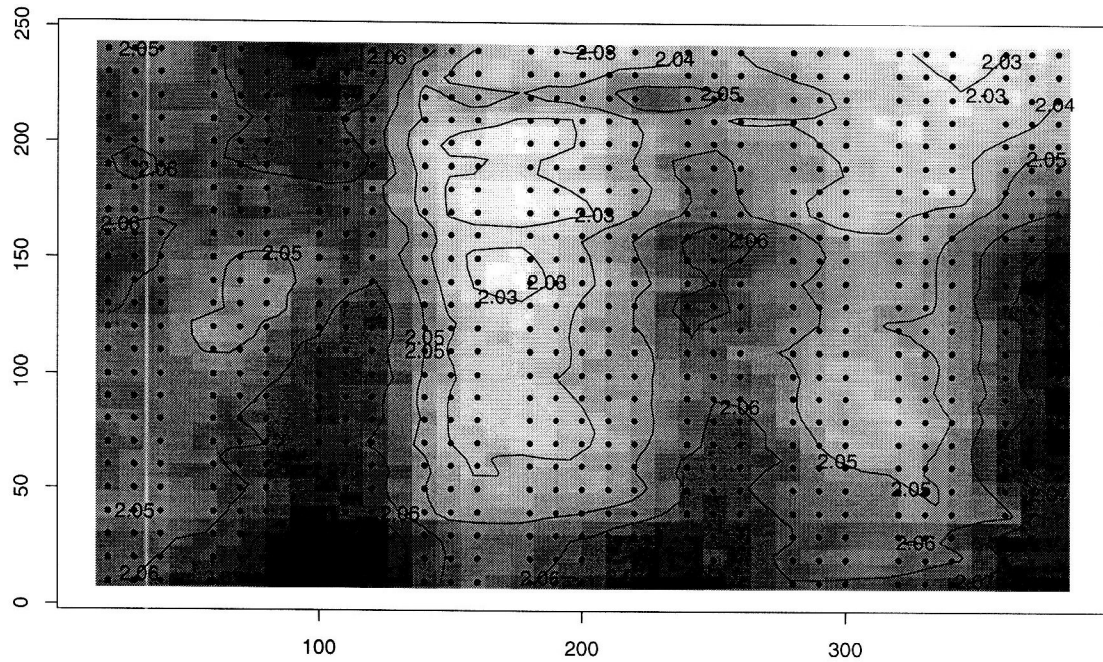


Bild 2: Eine Platte mit nichtkonstanter Dicke

Ein geeignetes parametrisches Modell war hierbei das Exponentialmodell

$$2\gamma_{a,c}(h) = c(1 - \exp(-h/a))$$

Mittels einer Methode aus der räumlichen Statistik (vgl. Cressie 1993, Formel 2.6.12), die die angenehme Eigenschaft hat, besonders für kleine Distanzen eine sehr gute Näherung zu liefern, wurden dann die Parameter a und c bestimmt, und man erhielt die Kurve in Bild 3 für das geschätzte Variogramm.

Für ein praxisrelevantes Modell wird man i.a. die Dicke der Platte an den verschiedenen Stellen nicht durch Versuche ermitteln, sondern zur Anpassung eines Finite-Element-Modells versuchen, durch einen repräsentativen Ausschnitt ein geeignetes Modell für die Kovarianzfunktion zu bestimmen. Man wählt dann also z.B. ein Exponentialmodell, wobei man für das weitere Vorgehen aber a , c und $C(0)$ als unbekannt ansetzt, also eine a-priori-Verteilung von der Form

$$\vartheta \sim \mathcal{N}((\vartheta_0, \dots, \vartheta_0), B(C(0), a, c))$$

falls man zudem keine systematischen Abweichungen von einer bekannten mittleren Dicke ϑ_0 erwartet, also $E(\vartheta_i) = E(\vartheta_j) = \vartheta_0$ voraussetzt.

2 Bayes-Methode bei vollständiger a-priori-Information

Da die hier vorgestellte Methode zur empirischen Bayes-Schätzung auf der Bayes-Schätzung bei vollständiger a-priori-Information beruht, wird hierzu im folgenden ein Verfahren vorgestellt. Man betrachtet also die Situation eines nichtlinearen Modells der Form (1) mit

$$\vartheta \sim \mathcal{N}(\mu, B)$$

und bekannten μ und B . Ein übliches Schätzverfahren beruht dann darauf, die a-posteriori-Dichte zu maximieren, was zu einem Kleinste-Quadrate-Problem mit Strafterm führt:

$$\min_{\vartheta} \|y - \eta(\vartheta)\|_{\Sigma}^2 + \|\vartheta - \mu\|_B^2$$

Zur Lösung dieses Problems erwiesen sich die Verfahren von Collins u.a. (1974) und von Friswell (1989) als ungeeignet, vergleiche Zehn u.a. (1999) und Offinger (1999).

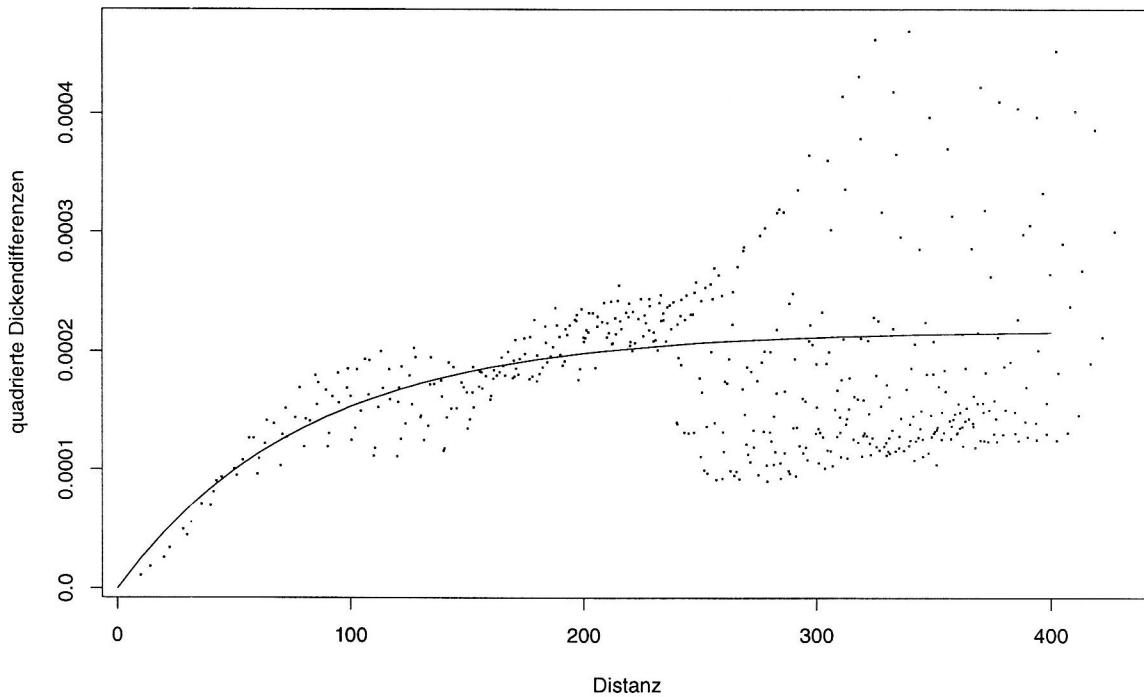


Bild 3: Variogrammschätzung

Konvergenz gegen den gesuchten Wert liefert hingegen das iterative Verfahren $\vartheta_0 = \mu$ und

$$\vartheta_{i+1} = \vartheta_i + (B^{-1} + J_i^t \Sigma^{-1} J_i)^{-1} (J_i^t \Sigma^{-1} (y - \eta(\vartheta_i)) + B^{-1}(\mu - \vartheta_i)) \quad (2)$$

mit $J_i = \frac{\partial \eta}{\partial \vartheta}(\vartheta_i)$ (zur Berechnung von Ableitungen der Eigenwerte und der Eigenvektoren im Beispiel 1b) vergleiche z.B. Offinger (1998)). Diese Formeln kann man wie folgt motivieren: Im $(i + 1)$ ten Iterationsschritt betrachtet man das in ϑ_i linearisierte approximative Modell

$$Y = \eta(\vartheta_i) + J_i \cdot (\vartheta - \vartheta_i) + \epsilon$$

und bestimmt dann in diesem die Bayes-Schätzung ϑ_{i+1} zur a-priori-Verteilung $\vartheta \sim \mathcal{N}(\mu, B)$.

Für einen Zusammenhang der Formel (2) mit den Formeln von Pázman (1992, 1993) und einer alternativen Darstellung vergl. Offinger (1999). Numerische Untersuchungen an Testproblemen vom Typ 1.1 und 1.2 zeigten, daß dieses iterative Verfahren ungefähr die gleiche Konvergenzrate wie das Verfahren von Pázman aufweist, nämlich etwas besseres Verhalten als lineare Konvergenz, i.a. aber weniger Iterationen benötigt. Zudem läßt sich mathematisch beweisen, daß zumindest ein lokales Maximum der a-posteriori-Dichte gefunden wird, falls Konvergenz eintritt.

3 Die empirische Bayes-Methode

Man betrachtet jetzt den Fall, daß in der a-priori-Verteilung unbekannte Parameter γ und ρ auftreten:

$$\vartheta \sim \mathcal{N}(\mu(\gamma), B(\rho))$$

Das übliche Vorgehen besteht nun darin, Schätzwerte $\hat{\gamma}$ und $\hat{\rho}$ für diese Parameter aus der Randdichte

$$m_{(\gamma, \rho)}(y) = \int f_{\vartheta}(y) \pi_{(\gamma, \rho)}(\vartheta) d\vartheta$$

von Y mit dem Maximum-Likelihood-Verfahren zu bestimmen, wobei f_{ϑ} die Dichte der $\mathcal{N}(\eta(\vartheta), \Sigma)$ -Verteilung und $\pi_{(\gamma, \rho)}(\vartheta)$ die Dichte der $\mathcal{N}(\mu(\gamma), B(\rho))$ -Verteilung, der a-priori-Verteilung, bezeichnet. Anschließend ergibt sich die Schätzung für ϑ gemäß dem Verfahren bei vollständiger a-priori-Information.

Ein Problem besteht nun darin, daß im nichtlinearen Fall die Integration in der Formel für die Randdichte nicht explizit lösbar ist und insbesondere nicht die Dichte einer Normalverteilung ist. Ein mögliches Vorgehen im nichtlinearen Fall wäre nun, das Integral für die Randdichte jeweils numerisch auszuwerten, z.B. mit einer Monte-Carlo-Methode. Für praxisrelevante Beispiele ist dieses Verfahren allerdings sehr aufwendig, da hierbei für viele verschiedene ϑ -Werte $\eta(\vartheta)$ bestimmt werden muß, also viele strukturelle Eigenwert-Probleme gelöst werden müssen. Weniger aufwendig ist folgendes Verfahren, das auf schrittweiser Linearisierung basiert:

Im linearen Fall ist nämlich $m_{(\gamma,\rho)}(y)$ die Dichte einer Normalverteilung: Falls

$$\eta(\vartheta) = J \cdot \vartheta + y_0 \quad \epsilon \sim \mathcal{N}(0, \Sigma) \quad \text{und} \quad \vartheta \sim \mathcal{N}(\mu(\gamma), B(\rho))$$

dann ist die Randdichte von Y die Normalverteilung $\mathcal{N}(J\mu(\gamma)+y_0, \Sigma+JB(\rho)J^t)$ und man erhält folgende Formel für die Randdichte (mit einer Konstanten c):

$$-2 \ln m_{(\gamma,\rho)}(y) = c + \ln \det(\Sigma + JB(\rho)J^t) + (y - y_0 - J\mu(\gamma))^t (\Sigma + JB(\rho)J^t)^{-1} (y - y_0 - J\mu(\gamma))$$

Zur Minimierung solcher Funktionen wurden in der Varianzkomponentenanalyse Verfahren entwickelt (vgl. z.B. Rao und Kleffe (1988) oder Searle u.a. (1992)). Hat man hiermit Schätzwerte $\hat{\gamma}$ und $\hat{\rho}$ bestimmt, erhält man im linearen Modell als Schätzung für ϑ :

$$\hat{\vartheta} = (B(\hat{\rho})^{-1} + J^t \Sigma^{-1} J)^{-1} (B(\hat{\rho})^{-1} \mu(\hat{\gamma}) + J^t \Sigma^{-1} (y - y_0))$$

4 Ein Algorithmus zur approximativen Lösung des empirischen Bayes-Verfahrens

Initialisierung: $i := 0$, $\hat{\vartheta}_0 = \mu(\gamma_0)$ für ein geeignetes γ_0 .

1. Linearisiere das Modell in ϑ_i , d.h. bestimme die Jacobi-Matrix $J_i = \frac{\partial \eta}{\partial \vartheta}(\vartheta_i)$ der Funktion η in ϑ_i :

$$Y \approx \eta(\hat{\vartheta}_i) + J_i \cdot (\vartheta - \hat{\vartheta}_i) + \epsilon$$

2. Berechne die (neue) empirische Bayes-Schätzung $(\hat{\gamma}, \hat{\rho})$ in diesem linearen Modell mit einem Verfahren aus der Varianzkomponentenanalyse, d.h. löse

$$\min_{(\gamma,\rho)} \ln \det(\Sigma + J_i B(\rho) J_i^t) +$$

$$(y - \eta(\hat{\vartheta}_i) - J_i(\mu(\gamma) - \hat{\vartheta}_i))^t (\Sigma + J_i B(\rho) J_i^t)^{-1} (y - \eta(\hat{\vartheta}_i) - J_i(\mu(\gamma) - \hat{\vartheta}_i))$$

3. Führe einen Updateschritt des Verfahrens zur Bayes-Schätzung bei vollständiger a-priori-Information durch:

$$\hat{\vartheta}_{i+1} = \hat{\vartheta}_i + (B^{-1}(\hat{\rho}) + J_i^t \Sigma^{-1} J_i)^{-1} (J_i^t \Sigma^{-1} (y - \eta(\hat{\vartheta}_i)) + B^{-1}(\hat{\rho})(\mu(\hat{\gamma}) - \hat{\vartheta}_i))$$

4. Falls Konvergenzkriterium nicht erfüllt, setze $i := i + 1$ und gehe zurück zu (1).

Dieser Algorithmus wurde untersucht an unserem (akademischen) Testbeispiel aus 1.1 bei verschiedenen Werte von σ_0^2/N für $\vartheta_0 = 1$ und bei Beobachtung $\bar{y} = \sum_{i=1}^N y_i = 1$, d.h. im Modell $Y_i = \frac{1}{2}\vartheta^2 + \epsilon_i$ mit unabhängigen und identischen Fehlern $\epsilon_i \sim \mathcal{N}(0, \sigma_0^2)$ und a-priori-Annahme $\vartheta \sim \mathcal{N}(1, \rho^2)$. Die Ergebnisse sind in Bild 4 dargestellt. Ferner ist hier zum Vergleich noch der Bayes-Schätzer für $\rho^2 = 1$, d.h. bei vollständiger a-priori-Information angegeben.

Man erkennt, daß unser Algorithmus eine sehr gute Näherung ergibt, obwohl er mit deutlich weniger Rechenaufwand als die numerische Integration verbunden ist.

Danksagung:

Der Deutschen Forschungsgemeinschaft danke ich für die finanzielle Unterstützung meiner Arbeit im Graduiertenkolleg Modellierung, Berechnung und Identifikation mechanischer Systeme (GRK 203).

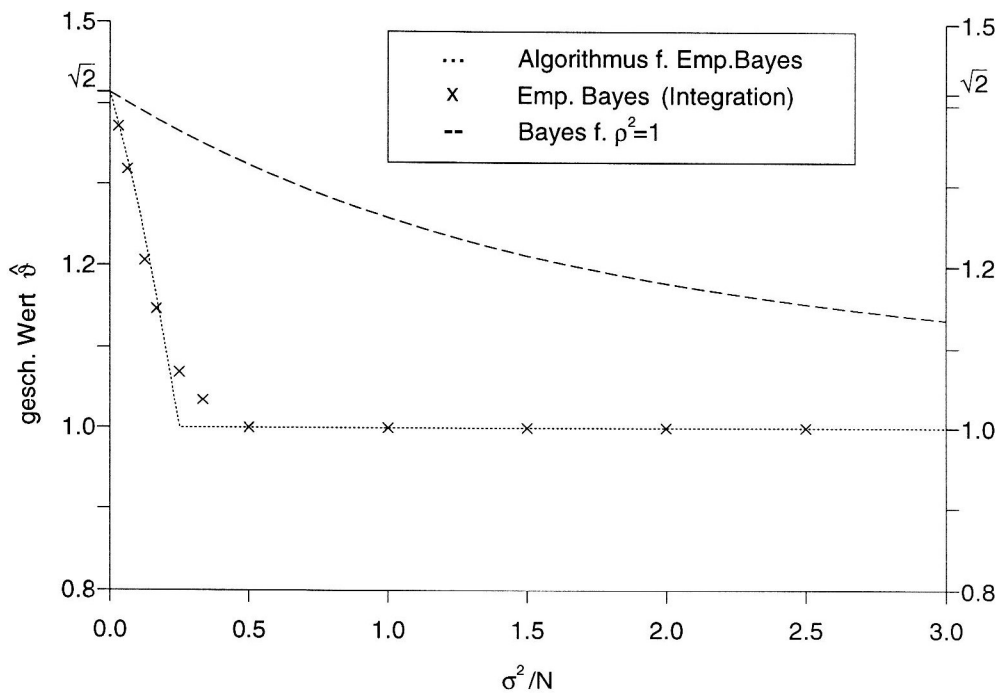


Bild 4: Ergebnisse des Algorithmus im akademischen Testbeispiel aus 1.1

Literatur

- Collins, J. D.; Hart, G. C.; Hasselman, T. K.; Kennedy, B.: Statistical Identification of Structures. AIAA Journal, 12, (1974), 185–190.
- Cressie, N. A. C.: Statistics for spatial data. John Wiley & Sons, New York u.a., (1993).
- Friswell, M. I.: The Adjustment of Structural Parameters using a Minimum Variance Estimator. Mechanical Systems and Signal Processing, 3(2), (1989), 143–155.
- Offinger, R.: Bayes-Verfahren zur Parameterschätzung in strukturmechanischen Modellen, Preprint MBI-98-1, Otto-von-Guericke-Universität Magdeburg (1998), 107–113.
- Offinger, R.: Bayes-Verfahren zur Parameterschätzung in nichtlinearen Regressionsmodellen, Preprint MBI-99-1, Otto-von-Guericke-Universität Magdeburg (1999), 13–20.
- Pázman, A.: Geometry of the Nonlinear Regression with Prior. Acta Math. Univ. Comenianae, LXI(2), (1992), 263–276.
- Pázman, A.: Nonlinear Statistical Models. Kluwer Academic Publishers, Dordrecht u.a., (1993).
- Rao, C. R.; Kleffe, J.: Estimation of variance components and applications. Elsevier Science Publishers, Amsterdam u.a., (1988).
- Searle, S. R.; Casella, G.; McCulloch, C. E.: Variance Components. John Wiley & Sons, New York u.a., (1992).
- Zehn, M. W.; Martin, O.; Offinger, R.: Influence of Parameter Estimation Procedures on the Updating Process of Large Finite Element Models, (1999), 240–250 in: Friswell, M.I. and Mottershead, J.E. and Lees, A.W. (Hrsg.): Identification in Engineering Systems, Proceedings of the Second International Conference held in Swansea, (March 1999).
- Zehn, M. W.; Saitov, A.; Schmidt G.: Berücksichtigung statistischer Modellunsicherheiten in den Dicken von CFK-Strukturen bei der indirekten Parameteridentifikation. VDI Berichte, 1550, (2000), 691–705.

Anschrift: Dipl.-Math. Robert Offinger, Institut für Mathematische Stochastik, Otto-von-Guericke-Universität Magdeburg, D-39016 Magdeburg. *e-mail:* roff@saturn.math.uni-magdeburg.de