Numerical Methods for Parameter Estimation and Optimal Experiment Design in Chemical Reaction Systems†

Thomas Lohmann,* Hans Georg Bock, and Johannes P. Schlöder
Institut für Mathematik, Universität Augsburg, Universitätsstrasse 2, D-8900 Augsburg, FRG

The estimation of parameters in models of chemical kinetics, e.g., of coal pyrolysis, needs efficient and reliable numerical techniques. This paper presents a method based on a multiple shooting algorithm and a generalized Gauss–Newton method, which is designed for highly nonlinear problems and which is suitable for stiff, as well as unstable, systems that frequently occur in chemistry and chemical engineering. In order to reduce the costs of experiments, optimal experiment design methods are helpful. A new algorithm is developed that minimizes the maximum length of the joint confidence intervals of the parameters with respect to frequency factors of the measurements in order to improve the statistical reliability of the parameters. The derivatives of the covariance matrix of the parameters with respect to the frequency factors are determined semianalytically and inexpensively. Thus, the optimal frequency factors can be calculated by a special version of the sequential quadratic programming method. Results for the estimation of parameters in a kinetic model of coal pyrolysis are reported.

Parameter Identification in Chemical Reaction Systems

1.1. Parameter Estimation Problem. The modeling of many dynamical processes, e.g., descriptions of chemical kinetics of coal pyrolysis, leads to systems of nonlinear ordinary differential equations (ODE) with unknown parameters p:

$$\dot{y}(t) = f(t,y,p,\mathrm{sign}(Q_i(t,y,p))) \tag{1}$$

Such systems may be stiff and/or unstable. They may include discontinuities in the right-hand side f of the equation and/or jumps of the states $y(t)$ at points where so-called switching functions $Q(t,y,p)$ change sign. In addition, there are frequently equality or inequality constraints on the states $y(t)$ and the parameters p

$$\begin{aligned} r_1(y(t_1),...,y(t_k),p) &= 0 \\ r_2(y(t_1),...,y(t_k),p) &\geq 0 \end{aligned} \tag{2}$$

to be satisfied, which consist of initial, boundary, or interior-point conditions, e.g., conditions that characterize periodicity or positivity restrictions of the range of parameters p.

The unknown parameters p and state variables $y(t)$ have to be estimated from given observations

$$\eta_i = g_i(\bar{y}(t_1),...,\bar{y}(t_m),\bar{p}) + \epsilon_i \tag{3}$$

of functions g_i of the true parameter values $\bar{p}$ and the true states $\bar{y}(t)$ at some data points $t_1, ..., t_m \in [t_a,t_e]$. If the measurement errors ϵ_i are additive, independent, and normally distributed with zero mean and standard deviation σ_i, the minimization of the function

$$l(y,p) := \sum_{i=1}^{n_1} \frac{1}{\sigma_i^2}(\eta_i - g_i(y(t_1),...,y(t_m),p))^2 = \min_{(y,p)} \tag{4}$$

subject to the ODE system and the constraints listed above yields a maximum likelihood estimate (y^*,p^*) of the states and the parameters that reproduce the observed behavior of the real process with maximum probability.

Note that for correlated measurement errors but known covariance matrix the differences between the observations η_i and the calculated values $g_i(y(t_1),...,y(t_m),p)$ should be weighted by the inverse of the covariance matrix of measurement errors in order to obtain a maximum likelihood estimate.

1.2. Boundary Value Problem Approach. To solve ODE parameter estimation problems, one approach that is often applied consists of a repeated integration of the ODE system and an iterative improvement of the parameters.

More general and more efficient for such identification problems is the boundary value problem approach, especially multiple shooting or collocation algorithms (cf. Bock (1983, 1987)). The multiple shooting method described here is based on a selection of nodes $t_a \leq \tau_0 < ... < \tau_{m'} \leq t_e$ and on a choice of initial estimates s_k of the states $y(\tau_k)$ and the parameters p. Then the initial value problems

$$\begin{aligned} \dot{y}(t) &= f(t,y,p,\mathrm{sign}(Q_i(t,y,p))) \\ y(\tau_k) &= s_k \qquad k = 0, ..., m' - 1 \end{aligned} \tag{5}$$

are solved on the subintervals $[\tau_k,\tau_{k+1}]$. The additional free variables $s_0, ..., s_{m'}$ are implicitly determined by additional matching conditions

$$h_k(s_k,s_{k+1},p) := y(\tau_{k+1};\tau_k,s_k,p) - s_{k+1} = 0 \tag{6}$$

which ensure that a numerical solution procedure eventually delivers a continuous trajectory. Finally, by formal insertion of the discretized trajectory into the function $l(y,p)$ one obtains a large, finite dimensional, constrained least-squares problem in the variables $\chi := (s_0,...,s_{m'};p)$ which is eventually solved by a generalized Gauss–Newton method.

Practically important advantages of this approach can be summarized as follows. Adequate choice of nodes guarantees the existence of a (discontinuous) initial trajectory. The initial estimates s_k of the states $y(\tau_k)$ allow one to make use of knowledge about the solution of the estimation problem, especially the information given by observations. Finally and most important, the influence of bad initial parameter estimates $p^{(0)}$ is drastically reduced.

1.3. Generalized Gauss–Newton Method. In any case, the boundary value problem approach leads to a finite dimensional, highly nonlinear least-squares problem with equality and inequality constraints:

$$\begin{aligned} \|F_1(x)\|_2^2 &= \min_x \\ F_2(x) &= 0 \\ F_3(x) &\geq 0 \end{aligned} \tag{7}$$

† Prepared for presentation at the 1990 Summer National Meeting of the American Institute of Chemical Engineers, August 19–22, 1990, session Detailed Modeling of Hydrocarbon Pyrolysis.

0888-5885/92/2631-0054$03.00/0

This is iteratively solved by a generalized Gauss–Newton procedure

$$x^{(k+1)} = x^{(k)} + \lambda_k \Delta x^{(k)} \tag{8}$$

in which the increment $\Delta x^{(k)}$ is determined as the solution of the least-squares problem linearized about $x^{(k)}$:

$$\begin{aligned} &\|F_1(x^{(k)}) + \mathbf{J}_1(x^{(k)})\Delta x^{(k)}\|_2^2 = \min_{\Delta x^{(k)}} \\ &F_2(x^{(k)}) + \mathbf{J}_2(x^{(k)})\Delta x^{(k)} = 0 \\ &F_3(x^{(k)}) + \mathbf{J}_3(x^{(k)})\Delta x^{(k)} \geq 0 \end{aligned} \tag{9}$$

The required derivatives $\mathbf{J}_i := \partial F_i/\partial x$ are calculated by internal numerical differentiation, which means that the exact directional derivatives of the adaptive discretization scheme are internally generated by the integrators (cf. Bock (1987)). A generalized inverse $\mathbf{J}^+(x)$ can be defined that solves the linearized problem. This generalized Gauss–Newton method converges linearly but almost like a second-order method, depending on the compatibility of the model with the data. Global convergence can be ensured by suitable choice of the damping factors λ_k (cf. Bock (1983, 1987)).

1.4. Statistical Analysis of the Parameter Estimates. In order to obtain a confidence region of the parameters, one may consider (for simplicity only the equality constrained case is treated here):

$$G(\alpha,x^*) := \{x|F_2(x) = 0, \|F_1(x)\|_2^2 - \|F_1(x^*)\|_2^2 \leq \gamma^2(\alpha)\} \tag{10}$$

x^* denotes the solution of the constrained nonlinear least-squares problem, α is the level of error probability, and $\gamma(\alpha)$ is defined as the quantile

$$\gamma^2(\alpha) := \chi_{1-\alpha}^2(m - n_2) \tag{11}$$

of the χ^2 distribution with $m - n_2$ degrees of freedom. This confidence region can be approximated by linearization

$$\begin{aligned} G_L(\alpha,x^*) = \{x|&F_2(x^*) + \mathbf{J}_2(x^*)(x - x^*) = 0, \\ &\|F_1(x^*) + \mathbf{J}_1(x^*)(x - x^*)\|_2^2 - \|F_1(x^*)\|_2^2 \leq \gamma^2(\alpha)\} \end{aligned} \tag{12}$$

which is enclosed by a box (cf. Bock (1987)):

$$\begin{aligned} G_L(\alpha,x^*) \subseteq [x^*_1 - \theta_1, x^*_1 + \theta_1] \times \ldots \times \\ [x^*_m - \theta_m, x^*_m + \theta_m] \end{aligned} \tag{13}$$

$$\theta_i := \sigma_i(x^*)\,\gamma(\alpha) \tag{14}$$

$\sigma_i(x^*)$ denotes the root of the ith diagonal element of the covariance matrix

$$\mathbf{C}(x^*) := \mathbf{J}^+(x^*)\begin{pmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{n_2} \end{pmatrix}\left(\mathbf{J}^+(x^*)\right)^T \tag{15}$$

This enclosing confidence box is easily calculated by using the generalized inverse $\mathbf{J}^+(x^*)$ at the solution point x^* of the generalized Gauss–Newton method. Thus, the generalized Gauss–Newton method not only yields parameter estimates but also makes an a posteriori analysis especially easy. This statistical analysis allows one to judge the quality of the calculated parameter estimates and helps one to decide whether the solution is satisfactory or whether further experiments should be performed.

Optimal Experiment Design for Parameter Identification in ODE

2.1. Design Problem. In order to improve the statistical reliability of the model parameter estimates, one may consider an optimal design of the experimental conditions.

One important question is the optimal choice of frequencies w_j with which the functions $g_j(\bar{y}(t_1),\ldots,\bar{y}(t_m),\bar{p})$ are measured. If the measurement errors are independent and normally distributed, these frequencies w_j enter the maximum likelihood function

$$\tilde{l}(y,p,w) = \sum_{j=1}^{n_1} \frac{w_j}{\sigma_j^2}(\eta_j - g_i(y(t_1),\ldots,y(t_m),p))^2 \tag{16}$$

as design parameters. Note that in this special case the ODE system and hence the solution of the boundary value problem are independent of the design parameters. The Jacobian matrix of the corresponding discretized least-squares problem at some linearization point $\tilde{x}$ has the special form

$$\tilde{\mathbf{J}}(w) = \begin{pmatrix} \mathbf{W}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_2} \end{pmatrix}\begin{pmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \end{pmatrix} \tag{17}$$

where $\mathbf{W}^{1/2} := \text{diag}\,(w_1^{1/2},\ldots,w_{n_1}^{1/2})$ is a diagonal matrix. Denoting by $\tilde{\mathbf{J}}^+(w)$ the generalized inverse of $\tilde{\mathbf{J}}(w)$, one obtains the covariance matrix of the model parameter estimates as

$$\tilde{\mathbf{C}}(w) = \tilde{\mathbf{J}}^+(w)\begin{pmatrix} \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{n_2} \end{pmatrix}\left(\tilde{\mathbf{J}}^+(w)\right)^T \tag{18}$$

Since the use of joint confidence intervals as derived in section 1.4 seems to be one of the most common ways to characterize statistical reliability of parameter estimates, the authors suggest the minimization of the maximum scaled length of such confidence intervals (the edge lengths of the enclosing box)

$$\max_{i\in\{1,\ldots,m\}} [s_i\tilde{\sigma}_i(w)] = \min_w \tag{19}$$

as an optimal design criterion. Here s_i is a scaling factor (see below) and $\tilde{\sigma}_i(w)$ denotes the root of the ith diagonal element of the covariance matrix $\tilde{\mathbf{C}}(w)$. This approach is also computationally very convenient, since it is only based on the diagonal elements of $\tilde{\mathbf{C}}(w)$, which are easily computed. Thus, it is especially suitable for large-scale problems. Of course other design criteria, like generalizations of the A-, D-, or E-optimality criterion (minimize the trace, the determinant, or the maximum eigenvalue of $\tilde{\mathbf{C}}(w)$) or the minimization of some correlation coefficients, are also possible.

Finally, it should be noted that there are always constraints on the optimal design to be taken into account; typically

$$\begin{aligned} &\sum_{j=1}^{n_1} w_j = c \quad \text{or} \quad \sum_{j=1}^{n_1} d_j w_j = c \quad \text{and} \\ &0 \leq w_j \quad \text{or} \quad 0 \leq a_j \leq w_j \leq b_j \end{aligned} \tag{20}$$

where the weights d_j account for observation costs. Furthermore, there are frequently constraints on the covariance matrix $\tilde{\mathbf{C}}(w)$, e.g.,

$$0 \leq a_i' \leq \tilde{\sigma}_i(w) \leq b_i' \tag{21}$$

The choice of the design criterion, of the scaling factors s_i, and of the constraints depends on the special application and on the requirements of the user.

2.2. Numerical Approach. The design problem ends up in a nonlinear Chebyshev problem:

$$\max_{i\in\{1,\ldots,m\}} \{s_i\tilde{\sigma}_i(w)\} = \min_w \tag{22}$$

$$\sum_{j=1}^{n_1} d_j w_j = c, \quad 0 \leq a_j \leq w_j \leq b_j$$

Table I. Initial Parameter Estimates

	H_2	CO_2		CO		CH_4		
$A^{(0)}/\text{min}^{-1}$	1200.0	33000.0	14000.0	3300.0	150000.0	1.0×10^7	1.65×10^6	1.8×10^6
$E^{(0)}/[\text{kcal/mol}]$	22.3	19.5	23.0	18.0	30.2	31.0	31.0	35.3
$y_\infty^{(0)}/[\text{cm}^3/\text{g}]$	124.8	30.50	26.40	13.77	32.13	21.71	23.79	21.71

Table II. Transformed Initial Parameter Estimates

	H_2	CO_2		CO		CH_4		
$\alpha^{(0)}$	-23.44	-16.44	-21.90	-16.77	-28.99	-25.85	-27.65	-33.22
$\beta^{(0)}$	-3.687	0.8731	-1.479	-0.4031	-1.535	2.069	0.2671	-1.428

Table III. Calculated Optimal Parameter Estimates

	H_2	CO_2		CO		CH_4		
α^*	-22.34	-15.10	-15.53	-17.97	-19.91	-24.45	-19.60	-37.50
β^*	-3.745	0.5501	-2.793	0.5728	-3.180	1.500	-2.299	-2.450
y_∞^*	112.2	22.64	30.50	8.934	37.10	18.96	40.73	6.064

Table IV. Optimal Parameter Estimates after Back-Transformation

	H_2	CO_2		CO		CH_4		
A^*/min^{-1}	647.2	10720.0	93.57	23120.0	242.0	2.289×10^6	959.6	2.903×10^6
$E^*/[\text{kcal/mol}]$	21.00	17.62	14.35	20.39	18.40	28.81	19.22	38.92

It can be shown that the covariance matrix $\tilde{\mathbf{C}}(w)$ is differentiable with respect to the frequency factors w_j, if the derivatives $\tilde{\mathbf{J}}(w)$ and $\mathbf{J}_2$ have full rank for all w_j that fulfill the constraints, and that the derivatives $\partial\tilde{\mathbf{C}}(w)/\partial w_j$ can be determined semianalytically and inexpensively (cf. Lohmann (1991)). This allows the application of optimization procedures of higher order, e.g., a special version of the sequential quadratic programming (SQP) algorithm (cf. Powell (1978), Lohmann (1991)).

Numerical Example: A Decomposition Model of Coal Pyrolysis

3.1. Model. Consider the model equations

$$\dot{y}_{ij}(t) = dy_{ij}(t)/dt = k_{ij}(t)(y_{ij\infty} - y_{ij}(t)) \tag{23}$$

where the indices i denote the major coal pyrolysis gases hydrogen H_2, carbon dioxide CO_2, carbon monoxide CO, and methane CH_4 and where index j corresponds to the parallel reactions occurring for each gas. Campbell (1976) assumes one reaction for H_2, two for CO_2 and for CO, and three for CH_4. $y_{ij}(t)$ denotes the volume of gas i in reaction j per gram of coal at time t with total volume $y_{ij\infty}$ at the end of the reaction. The kinetic parameter $k_{ij}(t)$ is of Arrhenius form

$$k_{ij}(t) = A_{ij} \exp(-E_{ij}/RT(t)) \tag{24}$$

with the frequency factor A_{ij} and the activation energy E_{ij} of reaction j of gas i. R is the gas constant and $T(t)$ the temperature. The heating rate is assumed to be constant:

$$dT/dt = C \tag{25}$$

Thus, the model equation can be transformed into

$$dy_{ij}(T)/dT = \tilde{k}_{ij}(T)(y_{ij\infty} - y_{ij}(T)) \tag{26}$$

Since the parameters A_{ij}/C and E_{ij}/R are of very different magnitude, for numerical reasons the parameters A_{ij} and E_{ij} are replaced by

$$\alpha_{ij} = \ln\left(\frac{A_{ij}}{C}\right) - \frac{E_{ij}}{RT_a}, \quad \beta_{ij} = \ln\left(\frac{A_{ij}}{C}\right) - \frac{E_{ij}}{RT_e} \tag{27}$$

where $T_a := T(t_a)$ and $T_e := T(t_e)$ define the temperature

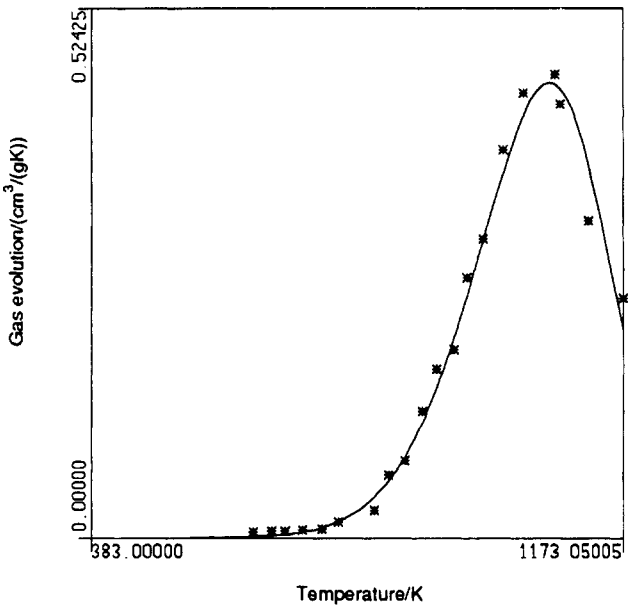

Figure 1. Optimal trajectory computed for hydrogen, H_2.

at the beginning and at the end of the reaction, respectively. With

$$\tau(T) = \frac{T_a}{T}\left(\frac{T - T_e}{T_a - T_e}\right) \tag{28}$$

$\tilde{k}_{ij}(T)$ is defined as

$$\tilde{k}_{ij}(T) = \exp(\alpha_{ij}\tau(T) + \beta_{ij}(1 - \tau(T))) \tag{29}$$

3.2. Numerical Results. The described parameter identification technique is used to solve the following problem.

Taken to estimate the parameters A_{ij}, E_{ij}, and $y_{ij\infty}$ are 20 observations of hydrogen, 22 of carbon dioxide, 23 of carbon monoxide, and 21 of methane. There are three parameters for hydrogen, six for carbon dioxide and carbon monoxide, and nine for methane, since we assume one, two, and three parallel reactions, respectively. The initial temperature is 383 K, and the heating rate is 3.33 K/min (cf. Campbell (1976)).

Table V. 95% Confidence Intervals of the Parameters

	H_2	CH_4		
α	[-24.18, -20.50]	[-37.89, -11.00]	[-28.54, -10.67]	[-116.5, 41.48]
β	[-4.076, -3.414]	[-3.122, 6.123]	[-4.477, -0.1216]	[-4.976, 0.07583]
y_∞	[97.35, 127.0]	[-2.218, 40.14]	[4.498, 76.96]	[-21.70, 33.83]
	CO_2		CO	
α	[-17.37, -12.82]	[-19.96, -11.11]	[-21.46, -14.48]	[-23.45, -16.38]
β	[-0.9628, 2.063]	[-3.513, -2.073]	[-1.361, 2.506]	[-3.560, -2.800]
y_∞	[13.15, 32.12]	[20.94, 40.06]	[4.776, 13.09]	[31.06, 43.13]

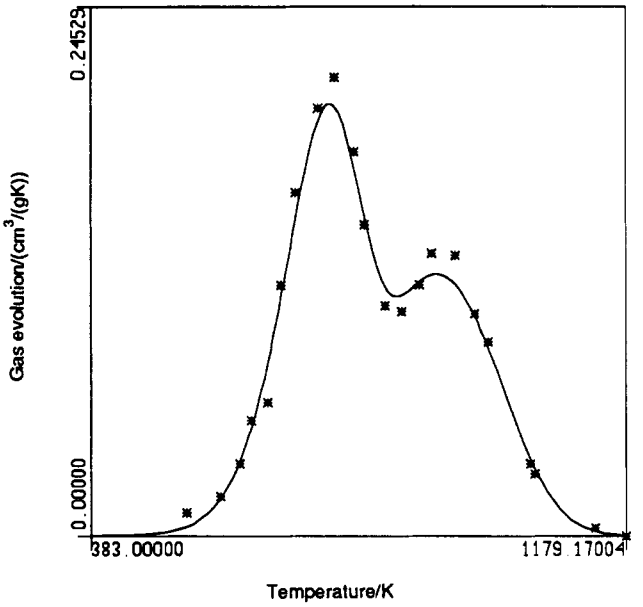

Figure 2. Optimal trajectory computed for carbon dioxide, CO_2.

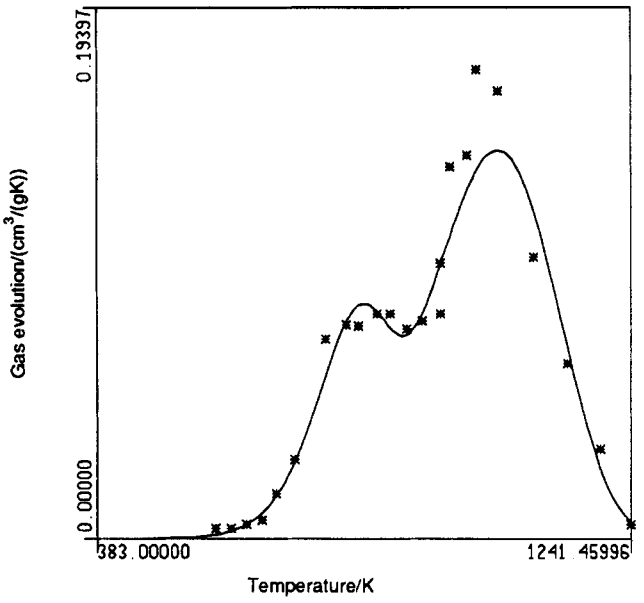

Figure 3. Optimal trajectory computed for carbon monoxide, CO.

Taking the parameter estimates $A_{ij}^{(0)}$, $E_{ij}^{(0)}$, and $y_{ij\infty}^{(0)}$ given by Campbell (1976) (cf. Table I), one obtains the initial parameters $\alpha_{ij}^{(0)}$ and $\beta_{ij}^{(0)}$ given in Table II. The optimal parameter estimates α_{ij}*, β_{ij}*, and $y_{ij\infty}$* are computed by using the program package PARFIT, which is an implementation of the described parameter identification technique (cf. Table III), and then they are transformed

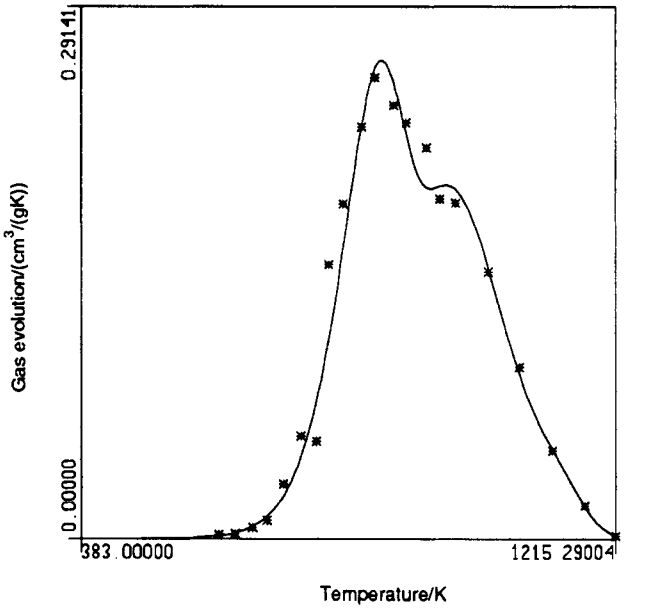

Figure 4. Optimal trajectories computed for methane, CH_4.

back to the optimal frequency factors A_{ij}* and the optimal activation energies E_{ij}* (cf. Table IV). The computed optimal trajectories are shown in the Figures 1–4. Furthermore, 95% confidence intervals of the parameters α_{ij}, β_{ij}, and $y_{ij\infty}$ are given that are automatically computed by PARFIT (cf. Table V).

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft (DFG).

Literature Cited

Bock, H. G. Recent Advances in Parameter Identification Techniques for O.D.E. In *Progress in Scientific Computing 2*; Deuflhard, P., Hairer, E., Eds.; Birkhäuser: Boston, 1983.

Bock, H. G. Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen. In *Bonner Mathematische Schriften*; Brieskorn, E., et al., Eds.; Universität Bonn: Bonn, FRG, 1987; Vol. 183.

Campbell, J. H. Pyrolysis of Subbituminous Coal As It Relates to In Situ Gasification, Parts 1 and 2. Report UCRL-52035; Lawrence Livermore Laboratory: Livermore, CA, 1976.

Lohmann, T. Numerical Computation of Optimal Experiment Designs for Parameter Identification in O.D.E. Universität Augsburg, FRG, 1991. Manuscript in preparation.

Powell, M. J. D. A Fast Algorithm for Nonlinearly Constrained Optimization Calculations. In *Numerical Analysis*; Watson, G. A., Ed.; Lecture Notes in Mathematics, Springer: Berlin, FRG, Heidelberg, FRG, New York, 1978; Vol. 630.

Received for review August 1, 1990
Revised manuscript received January 1, 1991
Accepted February 27, 1991