Parameter Estimation in Continuous-Time Dynamic Models with Uncertainty

Kim McAuley

Department of Chemical Engineering
Queen’s University
Kingston, Canada
Journal Articles


1. Chemical engineers develop fundamental dynamic models based on knowledge of chemical and physical phenomena.

2. Parameter estimation is a difficult problem
   - Two sources of uncertainty
     - Measurement noise
     - Disturbances that influence future behaviour
1. Chemical engineers develop fundamental dynamic models based on knowledge of chemical and physical phenomena.

2. Parameter estimation is a difficult problem
   - Two sources of uncertainty
     - Measurement noise
     - Disturbances that influence future behaviour

Proposed parameter estimation techniques account for both sources of uncertainty:
- Iterative Principal Differential Analysis
- Approximate Maximum Likelihood Estimation
Why Model Chemical Reactors?

Objectives of Chemical Companies: $$$
- Produce chemicals and polymers with targeted properties
- Make different product grades efficiently in a single reactor
- Devise improved reactor operating strategies
- Bring new products to market quickly
- Develop process knowledge for trouble-shooting

Models can help companies to:
- Train operators
- Design and test automatic control schemes
- Optimize grade changeover policies
- Simulate effects of process conditions and equipment design on product properties and production rates
- Plan experiments
- Test theories about what has gone wrong
- Capture, store and distribute knowledge
Fundamental Models of Chemical Processes

Where do model equations come from?

• Material balances on chemical species, and energy balances
  - Modeler converts mythology and assumptions into mathematical expressions
  - Algebraic equations, ODEs, PDEs

• Additional equations that describe:
  - Rates of chemical reactions
  - Movement of chemical species from one phase to another
Fundamental Models of Chemical Processes

**Example** - Polyethylene model for INEOS (BP Chemicals)

- 22 nonlinear ODEs
- 45 parameters

Model predicts:
- Reactant gas composition (ethylene, hexene, hydrogen)
- Polymer production rate
- Polymer properties

Using reactant feed rates and reactor temperature

Model for scale-up from laboratory to commercial reactors
- Use knowledge from model to reduce the number of steps and experiments required
The Parameter Estimation Problem in Dynamic Chemical Reactor Models

- Experimental situation
  - Measurements at irregular sampling times
  - Results from replicate experiments vary due to
    - Disturbances that enter the reactor and influence future behaviour
    - Uncertainties in initial reactor conditions and input-variable trajectories

- Model equations
  - Typically 10-100 ODEs and 15-50 parameters
  - Many simplifying assumptions
  - Unknown initial values for some state variables
Four Replicates of a Dynamic Experiment

- Any model that we fit through these data will result in correlated residuals.
- In dynamic systems, random errors at one time influence future responses.
- How should we account for this deviation from traditional least-squares assumptions during parameter estimation and model testing?
Traditional Parameter Estimation in a Differential Equation

\[
\frac{dx}{dt} = f(x, u, \theta), \quad x(0) = x_0
\]

\[
y_i = x_i + \varepsilon_i \quad (i = 1, \ldots n) \quad \varepsilon_i \sim N(0, \sigma_m^2)
\]

- Estimate the model parameters \( \theta \), given noisy observations \( y \) and known system inputs \( u \).

\[
J = \sum_{i=1}^{n} (y_i - \hat{x}_i(\theta))^2
\]

- We assume: 1) model structure is perfect
  2) \( u \) and \( x_0 \) are perfectly known
  3) measurements have random error
Traditional Parameter Estimation in a Differential Equation

\[
\frac{dx}{dt} = f(x, u, \theta), \quad x(0) = x_0
\]

\[
y_i = x_i + \epsilon_i \quad (i = 1, \ldots, n) \quad \epsilon_i \sim N(0, \sigma_m^2)
\]

• Estimate the model parameters \( \theta \), given noisy observations \( y \) and known system inputs \( u \).

\[
J = \sum_{i=1}^{n} (y_i - \hat{x}_i(\theta))^2
\]

• Requires repeated numerical solution of ODE each time the optimizer guesses new parameter values

• If initial conditions are unknown, they are estimated along with the parameters
Our First Algorithm (iPDA)

- Fit an empirical curve \( x_\sim(t) \) to the dynamic data using B-splines

\[
J_1 = \sum (y(t_i) - x_\sim(t_i, \beta))^2
\]
Dynamic Data and B-Spline Curve

tray temperature response - functional data object and raw data
Our First Algorithm (iPDA)

- Fit an empirical curve \( x_\sim(t) \) to the dynamic data using B-splines

\[
J_1 = \sum (y(t_i) - x_\sim(t_i, \beta))^2
\]

- Determine parameter values \( \theta \) to satisfy ODE as much as possible with \( \beta \) fixed

\[
J_2 = \int_{t_0}^{t_f} \left( \frac{dx_\sim}{dt} - f(x_\sim, u, \theta) \right)^2 dt
\]
Our First Algorithm (iPDA)

- Fit an empirical curve $x(t)$ to the dynamic data using B-splines
  $$J_1 = \sum (y(t_i) - x(t_i, \beta))^2$$

- Determine parameter values $\theta$ to satisfy ODE as much as possible with $\beta$ fixed
  $$J_2 = \int_{t_0}^{t_f} \left( \frac{dx}{dt} - f(x, u, \theta) \right)^2 dt$$

- Adjust spline parameters $\beta$ using a model-based penalty with $\theta$ fixed
  $$J_3 = \sum (y(t_i) - x(t_i, \beta))^2 + \lambda \int_{t_0}^{t_f} \left( \frac{dx}{dt} - f(x, u, \theta) \right)^2 dt$$

- Iterate between steps 2 and 3 until convergence
Is iPDA any good?

- No need for repeated numerical solution of ODE
  - No stability problems for bad parameter values
- No initial conditions required
- Easy to handle non-uniformly sampled data
Is iPDA any good?

- No need for repeated numerical solution of ODE
  - No stability problems for bad parameter values
- No initial conditions required
- Easy to handle non-uniformly sampled data
- During the parameter-estimation step, minimize residuals between spline curve and fundamental model using the differentiated form of the model

\[
J_2 = \int_{t_0}^{t_f} \left( \frac{dx_\perp}{dt} - f(x_\perp, u, \theta) \right)^2 dt
\]

**Model error**

- During the spline-fitting step, minimize deviations from the data

\[
J_3 = \sum (y(t_i) - x_\perp(t_i, \beta))^2 + \lambda \int_{t_0}^{t_f} \left( \frac{dx_\perp}{dt} - f(x_\perp, u, \theta) \right)^2 dt
\]

**Measurement error**
An Epiphany

- IPDA is equivalent to selecting $\theta$ and $\beta$ simultaneously to minimize:

$$J = \sum (y(t_i) - x_{\sim}(t_i, \beta))^2 + \lambda \int_{t_0}^{t_f} \left( \frac{dx_{\sim}}{dt} - f(x_{\sim}, u, \theta) \right)^2 dt$$
An Epiphany

- IPDA is equivalent to selecting $\theta$ and $\beta$ simultaneously to minimize:

$$J = \sum (y(t_i) - x_{\sim}(t_i, \beta))^2 + \lambda \int_{t_0}^{t_f} \left( \frac{dx_{\sim}}{dt} - f(x_{\sim}, u, \theta) \right)^2 dt$$

- This is the solution, but what is the underlying statistical problem?
- What is an appropriate value of $\lambda$?
- What happens in multi-response problems?
- What if some states aren’t measured?
- How can we enforce known initial conditions?
AMLE, a Proposed Parameter-Estimation Technique for Stochastic DEs

\[
\frac{dx}{dt} = f(x,u,\theta) + \eta(t), \quad x(0) = x_0
\]

\[
y_i = x_i + \varepsilon_i \quad (i = 1,...,n) \quad \varepsilon_i \sim N(0,\sigma^2_m)
\]

\[
E(\eta(t)\eta(t-\tau)) = Q \delta(\tau)
\]

- Two noise sources
  - Measurement noise
  - Stochastic process disturbances that can account for
    - Uncertainties in \(u\)
    - Unknown or unmeasured inputs
    - Structural imperfections in model
Random Process Disturbance

![Graph showing process disturbance vs time](image)

- Process disturbance
  - \( \eta(t) \) (kmol/m\(^3\)/min)
  - Time (min)
AMLE, a Proposed Parameter Estimation Technique for Stochastic DEs

\[
\frac{dx_\sim}{dt} = f(x_\sim, u, \theta) + \eta(t), \quad x_\sim(0) = x_0
\]

\[
y_i = x_{\sim_i} + \varepsilon_i \quad (i = 1, \ldots n)
\]

• Our approach:
  - Assume that the solution to the differential equations can be represented using B-splines or other basis functions:

\[
x(t) \cong x_\sim(t) = \sum_{i=1}^{b} \varphi_i(t) \beta_i
\]

\[
\frac{dx_\sim(t)}{dt}
\]

is used to convert ODEs into algebraic equations.
Approximate Maximum-Likelihood Estimation

- Assume the solution of the dynamic system can be well approximated by B-splines with unknown coefficients $\beta$
- Estimate the fundamental model parameters $\theta$ and the unknown spline coefficients $\beta$

- Select $\hat{\theta}$ and $\hat{\beta}$ to minimize

\[
J = \sum_{i=1}^{n} (y_i - x_{i_\sim})^2 + \lambda \int \left( \frac{dx_{\sim}(t)}{dt} - f(x_{\sim}(t), u(t), \theta) \right)^2 dt
\]

- Objective function arises from maximizing conditional joint density function of the states and measurements, given the parameters
What Weighting to Use?

\[ J = \sum_{i=1}^{n} (y_i - x_i^\sim)^2 + \lambda \int \left( \frac{dx^\sim(t)}{dt} - f(x^\sim(t), u(t), \theta) \right)^2 dt \]

- Heuristically
  - A large \( \lambda \) is appropriate when
    - Model is accurate and data are noisy
  - A small \( \lambda \) is appropriate when
    - Data are good and model is inaccurate
What Weighting to Use?

\[ J = \sum_{i=1}^{n} (y_i - x_i)^2 + \lambda \int \left( \frac{dx(t)}{dt} - f(x(t), u(t), \theta) \right)^2 dt \]

- Heuristically
  - A large \( \lambda \) is appropriate when
    - Model is accurate and data are noisy
  - A small \( \lambda \) is appropriate when
    - Data are good and model is inaccurate

\[ \lambda_{opt} = \frac{\sigma_m^2}{Q} \]

Very large \( \lambda \) corresponds to traditional least-squares parameter estimation, which assumes a perfect model and no disturbances.
Objective Function for a Multivariate Example with Known Variances

\[
J = \frac{1}{\sigma_{m1}^2} \sum_{j=1}^{n_1} \left( y_1(t_{m1j}) - x_{1\sim}(t_{m1j}) \right)^2 + \frac{1}{Q_1} \int_{t=0}^{t_f} \left( \frac{dx_{1\sim}}{dt} - f_1(x_{1\sim}, x_{2\sim}, u, \theta) \right)^2 dt \\
+ \frac{1}{\sigma_{m2}^2} \sum_{j=1}^{n_2} \left( y_2(t_{m2j}) - x_{2\sim}(t_{m2j}) \right)^2 + \frac{1}{Q_2} \int_{t=0}^{t_f} \left( \frac{dx_{2\sim}}{dt} - f_2(x_{1\sim}, x_{2\sim}, u, \theta) \right)^2 dt
\]

Straightforward to write \( J \) for models with many ODEs (or DAEs) and for problems with unmeasured states.
Reactor Example with Nonstationary Disturbance in Material Balance

\[ \frac{dC}{dt} = f_1(C, T, u, \theta) + w + \eta_1 \quad C(0) = 1.569 \text{ (kmol/m}^3\text{)} \]

\[ \frac{dT}{dt} = f_2(C, T, u, \theta) + \eta_2 \quad T(0) = 341.37 \text{ (K)} \]

\[ \frac{dw}{dt} = \eta_3 \quad w(0) = 0 \text{ (kmol/m}^3\text{/t)} \]

\[ y_1(t_i) = C(t_i) + \varepsilon_1(t_i) \quad (i = 1, \ldots, 64) \]

\[ y_2(t_j) = T(t_j) + \varepsilon_2(t_j) \quad (j = 1, \ldots, 213) \]

\( w \) could be a drifting flow rate or feed concentration disturbance (or a leak)
Reactor Example

- Objective function for parameter estimation is:

\[
\frac{1}{\sigma_{m1}^2} \sum_{j=1}^{64} \left( y_1(t_{m1j}) - C_\sim(t_{m1j}) \right)^2 + \frac{1}{Q_{p1}} \int_{t=0}^{64} \left( \frac{dC_\sim}{dt} - f_1(T_\sim, C_\sim, u, \theta) - w_\sim(t) \right)^2 dt \\
+ \frac{1}{\sigma_{m2}^2} \sum_{j=1}^{213} \left( y_2(t_{m2j}) - T_\sim(t_{m2j}) \right)^2 + \frac{1}{Q_{p2}} \int_{t=0}^{64} \left( \frac{dT_\sim}{dt} - f_2(T_\sim, C_\sim, u, \theta) \right)^2 dt \\
+ \frac{1}{Q_{p3}} \int_{t=0}^{64} \left( \frac{dw_\sim}{dt} \right)^2 dt
\]
Input Sequence $u(t)$ for Simulated Experiments

- **Input flow rate**
- **Input concentration**
- **Input temperature**
- **Input coolant temperature**
- **Coolant flow rate**
Parameter Estimation Results from Monte-Carlo Simulations of CSTR Example

- 4 parameters, $a$, $b$, $E/R$ and $k_{ref}$ were estimated using AMLE and traditional method nonlinear least squares (NLS)
Parameter Estimation Results from Monte-Carlo Simulations of CSTR Example

- Parameter estimates are better using AMLE
- Confidence intervals for parameter and state estimates are readily computed from inverse of FIM
State Trajectory Estimates

Concentration Trajectory

NLS

AML
State Estimates

- Temperature trajectory

NLS

AMLE
State Estimates

- Non-stationary disturbance $w$
Selecting Weighting Factors in J

- Modeler can estimate $\sigma_m^2$ from repeated measurements or from information from instrument supplier.

- Modeler will know that model is imperfect, and about the physical sources of disturbances, but **won’t know the noise intensity** $Q$.

- When $Q$ is unknown, we must estimate it.
  - The correct value of $Q$ results in spline fits that are consistent with $\sigma_m^2$.
  - Iterate between parameter estimation and $Q$ estimation until convergence.

- Estimate of $Q$ for each ODE provides information to modeler about disturbances and model mismatch.
Features of iPDA and AMLE Methods

• Good for systems with
  - Unknown or uncertain initial conditions
  - Irregular sampling
  - Unmeasured states
  - Meandering (nonstationary) disturbances

• No need for repeated numerical solution of ODEs
  - Collocation methods that account for model error
  - Optimization problems readily solved in AMPL/IPOPT
  - ODEs are satisfied (or not) using soft constraints in the objective function
Testing of AMLE

• Application to a nylon polymerization reactor model with data from my lab
  - 6 unknown parameters
  - 2 measured states and 1 unmeasured state
  - unknown initial conditions
  - known measurement variances, but unknown $Q$ values

• Seeking graduate students to estimate parameters in larger models
Acknowledgments

• Graduate Students:
  – M. S. Varziri, A. A. Poyton

• Collaborators:
  – P. J. McLellan and J. O. Ramsay
  – L.T. Biegler for advice on AMPL and IPOPT

• Funding
  – MITACS, Cybernetica, DuPont, Hatch, Matrikon, SAS, NSERC
  – BP Chemicals, MMO, Exxon, Nova Chemicals
1. Chemical engineers develop fundamental dynamic models based on knowledge of chemical and physical phenomena

2. Parameter estimation is a difficult problem
   - Two sources of uncertainty
     • Measurement noise
     • Disturbances that influence future behaviour

Proposed parameter estimation techniques account for both sources of uncertainty:
- Iterative Principal Differential Analysis
- Approximate Maximum Likelihood Estimation