Solving Model Discrimination Problems by OED: Applications in Enzyme Stability

Dipl. Phys. Christian Hoffmann

Simulation and Optimization Group
Interdisciplinary Center for Scientific Computing (IWR)
University of Heidelberg

christian.hoffmann[at]iwr[dot]uni-heidelberg[dot]de

October 9, 2007
Introduction:
Enzyme stability
Enzyme deactivation

**Wanted:** description of biochemical reaction $S \xrightarrow{r_E} P$

**Required:** model of deactivation reactions of enzyme $E$
**Wanted:** description of biochemical reaction \( S \xrightarrow{r_E} P \)

**Required:** model of deactivation reactions of enzyme \( E \)

- Many possibilities [Sad91]:

\[
\begin{align*}
E_1 & \underset{k_1}{\overset{k_{-1}}{\leftrightarrow}} E_2 & \underset{k_2}{\overset{k_{-2}}{\leftrightarrow}} \cdots & \underset{k_n}{\overset{k_{-n}}{\leftrightarrow}} E_n \\
E_{X1} & \underset{k_1^d}{\downarrow} & E_{X2} & \underset{k_2^d}{\downarrow} & \cdots & \underset{k_n^d}{\downarrow} & E_{Xn}
\end{align*}
\]
**Wanted:** description of biochemical reaction $S \xrightarrow{r_E} P$

**Required:** model of deactivation reactions of enzyme $E$

- But poor initial knowledge:

\[
\begin{align*}
E_1 & \xleftarrow{?} \quad ? \quad \xrightarrow{?} \quad \ldots \quad \xleftarrow{?} \quad ? \\
\downarrow & \quad \downarrow \quad \downarrow \quad \downarrow \\
? & \quad ? \quad \ldots \quad ?
\end{align*}
\]
Enzyme deactivation

**Wanted:** description of biochemical reaction $S \overset{r_E}{\rightarrow} P$

**Required:** model of deactivation reactions of enzyme $E$

- But poor initial knowledge:

  $E_1 \xrightarrow{?} ? \xrightarrow{?} ? \xrightarrow{?} \ldots \xrightarrow{?} ?$

  $\downarrow ? \quad \downarrow ? \quad \downarrow ?$

- Limited experimental methods:
  even hard to detect unknown proteins ("?"") qualitatively!
Wanted: description of biochemical reaction $S \xrightarrow{r_E} P$

Required: model of deactivation reactions of enzyme $E$

- But poor initial knowledge:
  $$
  E_1 \xrightarrow{?} ? \xrightarrow{?} \ldots \xrightarrow{?} ?
  $$

- Limited experimental methods:
  even hard to detect unknown proteins (“?”) qualitatively!

⇒ Formulate several rival model propositions.
  (examples from M.Boy, BASF AG [BDV99])
Proposed rival deactivation models

Model A

\[ E_N \xrightarrow{k} K \xleftarrow{K} E_D \]

\[ \downarrow k \quad \downarrow k \]

\[ E_X \quad E_X \]
Proposed rival deactivation models

**Model A**

\[
\begin{align*}
E_N & \xrightleftharpoons[K]{\text{}} E_D \\
& \quad k \\
E_X & \quad k
\end{align*}
\]

**Model B**

\[
\begin{align*}
E_N & \xrightleftharpoons[K]{\text{}} E_D \\
& \quad k \\
E_X & \quad k \\
E_X' & \quad k'
\end{align*}
\]
Proposed rival deactivation models

Model A

Enzyme: $E_N$ → $E_D$ via $K$

$E_N$ → $E_X$

$E_D$ → $E_X$

$E_N$ in $E_D$

Model B

Enzyme: $E_N$ → $E_D$ via $K$

$E_N$ → $E_X$

$E_D$ → $E_X$

$E_N'$ in $E_D$

Model C

$E_N$ → $E_D$ via $k$

$E_N$ → $E_X$

$E_D$ → $E_X$

$E_N$ in $E_D$
Proposed rival deactivation models

Introduction

Enzyme deactivation

Rival models

Resulting MD problem

MD

OED for MD

Numerical Methods

Results

Summary

Model A

\[
\begin{align*}
EN & \xrightarrow{K} ED \\
\downarrow k & \downarrow k \\
EX & \downarrow & \downarrow \\
\end{align*}
\]

Model B

\[
\begin{align*}
EN & \leftrightarrow \xrightarrow{K} ED \xrightarrow{K'} EN' \\
\downarrow k & \downarrow k & \downarrow k' \\
EX & \downarrow & \downarrow EX' \\
\end{align*}
\]

Model C

\[
\begin{align*}
EN & \xrightarrow{k} ED \\
\downarrow k' & \downarrow \\
EX & \\
\end{align*}
\]

and \( S + E(\cdot) \xrightarrow{r_E} P + E(\cdot) \)

Unknown parameters:
- equilibrium constants
- frequency factors
- activation energies

Only measureable: specific activity

\[ a(t) := \frac{r_E(t)}{m_{E,0}} \]
Resulting model discrimination problem

Measurements from experiment 1

Temperature measured
Model predictions for experiment 1, after fitting

**Temperature**
- Measured
- Response model A

**Specific activity [1000 IU]**

**Time [h]**

**Temperature [°C]**

Resulting model discrimination problem

Model predictions for experiment 1, after fitting

- **Temperature measured** (measured)
- **response model A** (blue)
- **response model B** (purple)

Specific activity [1000 IU] vs. Temperature [°C] vs. Time [h]
Resulting model discrimination problem

Introduction

Enzyme deactivation

Rival models

Resulting MD problem

MD

OED for MD

Numerical Methods

Results

Summary

Model predictions for experiment 1, after fitting

![Graph showing model predictions](image-url)
Models describe data with comparable precision!
Model discrimination (MD)

Introduction
- Enzyme deactivation
- Rival models
- Resulting MD problem
- MD
- OED for MD
- Numerical Methods
- Results
- Summary

One process – several models propositions
Choose which?
Model discrimination (MD)

Introduction

One process – several models propositions
Choose which?

Special challenges in systems biology:

- Little initial knowledge
- Limited experimental methods
- Complex systems, high-dimensional models
- Large number of rival models
Optimal experimental design for model discrimination
When is a model “correct”? 

Introduction
OED for MD
“Correct” models
Sequential discrimination procedure
Design criterion
Practical approach

Numerical Methods
Results
Summary
When is a model “correct”?

Experiment described by design quantities $q$. Provides measurement data:

$$\eta_q = \bar{\eta}_q + \epsilon_q, \epsilon_q \sim \mathcal{P}_q$$
When is a model “correct”? 

Experiment described by design quantities $q$. Provides measurement data:

$$\eta_q = \bar{\eta}_q + \epsilon_q, \epsilon_q \sim \mathcal{P}_q$$

Parameterized nonlinear regression models $i = 0, \ldots, N$:

$$\bar{\eta}_q \approx h^i[q, p^i], \quad \mathcal{P}_q \approx \mathcal{N}(0, \text{diag}(s))$$
When is a model “correct”? 

Experiment described by *design quantities* \( q \). Provides measurement data:

\[
\eta_q = \bar{\eta}_q + \epsilon_q, \epsilon_q \sim \bar{P}_q
\]

Parameterized nonlinear regression models \( i = 0, \ldots, N \):

\[
\bar{\eta}_q \approx h^i [q, \bar{p}^i], \quad \bar{P}_q \approx \mathcal{N}(0, \text{diag}(s))
\]

Model \( i \) describes the process “correctly”, if

1. the model responses are precise

\[
\exists \bar{p}^i : \| \bar{\eta}_q - h^i [q, \bar{p}^i] \| \leq \delta, \forall q \in Q
\]
When is a model “correct”? 

Experiment described by *design quantities* $q$. Provides measurement data:

$$
\eta_q = \bar{\eta}_q + \epsilon_q, \epsilon_q \sim \bar{P}_q
$$

Parameterized nonlinear regression models $i = 0, \ldots, N$:

$$
\bar{\eta}_q \approx h_i \left[ q, p^i \right], \bar{P}_q \approx \mathcal{N}(0, \text{diag}(s))
$$

Model $i$ describes the process “correctly”, if

1. the model responses are precise
   $$
   \exists \bar{p}^i : \| \bar{\eta}_q - h_i \left[ q, \bar{p}^i \right] \| \leq \delta, \ \forall q \in Q
   $$

2. *and* the error model is adequate ($\delta = 0$ for convenience)

   (DC)  \hspace{1cm} \mathbb{E} \| r \|_2^2 := \mathbb{E} \| \eta_q - h_i \left[ q, \bar{p}^i \right] \|_2^2 = s, \ \forall q \in Q
When is a model “correct”? 

Experiment described by design quantities $q$. Provides measurement data:

$$\eta_q = \bar{\eta}_q + \epsilon_q, \epsilon_q \sim \bar{P}_q$$

Parameterized nonlinear regression models $i = 0, \ldots, N$:

$$\bar{\eta}_q \approx h^i[q, \bar{p}^i], \quad \bar{P}_q \approx \mathcal{N}(0, \text{diag}(s))$$

Model $i$ describes the process “correctly”, if

1. the model responses are precise

$$\exists \bar{p}^i : \|\bar{\eta}_q - h^i[q, \bar{p}^i]\| \leq \delta, \quad \forall q \in Q$$

2. and the error model is adequate ($\delta = 0$ for convenience)

$$(\text{DC}) \quad \mathbb{E} \|r\|_2^2 := \mathbb{E} \|\eta_q - h^i[q, \bar{p}^i]\|_2^2 = s, \quad \forall q \in Q$$

Discrimination criterion (DC) can be used to falsify models:

Take samples, perform adequate statistical tests (F-tests and relatives)
Sequential discrimination procedure

- Sequential discrimination procedure
- OED for MD
- "Correct" models
- Parameters $p^{i,k}$
- Design criterion
- Practical approach
- Numerical Methods
- Results
- Summary

$\mathcal{N}^k$ rival models
Parameters $p^{i,k}$

$k \leftarrow k + 1$

Estimate parameters using all measurement data
$\Rightarrow$ New parameters: $\tilde{p}^{i,k}$

Perform experiment(s)
$\Rightarrow$ New measurement data
Sequential discrimination procedure

1. **$N^k$ rival models**
   - **Parameters** $\vec{p}_{i,k}$
   - $k \leftarrow k + 1$

2. **Estimate parameters**
   - using *all* measurement data
   - $\Rightarrow$ New parameters: $\vec{p}_{i,k}$

3. **Apply discrimination criteria**
   - using *all* measurement data
   - $\Rightarrow$ $N^k$ models remain

4. **Perform experiment(s)**
   - $\Rightarrow$ New measurement data

   - If $N^k = 0$
     - Formulate new models
   - If $N^k = 1$
     - Improve model
Sequential discrimination procedure

\[ N^k \text{ rival models} \]
\[ \text{Parameters } p_{i,k} \]

\[ k \leftarrow k + 1 \]

Estimate parameters using all measurement data
\[ \Rightarrow \text{New parameters: } \tilde{p}_{i,k} \]

Apply discrimination criteria using all measurement data
\[ \Rightarrow N^k \text{ models remain} \]

Design experiment(s) for model discrimination

Perform experiment(s)
\[ \Rightarrow \text{New measurement data} \]

Formulate new models

Improve model

\[ N^k = 0 \]
\[ N^k = 1 \]
Perform the experiment which will most strain the incorrect model to explain the data.” [HR65]

For a single model $i$:
perform the experiment described by a design solving

$$\max_q \| \Sigma^{-1} (\bar{\eta}_q - h^i[q, p^i]) \|_2^2, \; q \in Q$$

T-Optimal experimental design

“Optimal design with respect to a certain statistical Test”
Design criterion

„Perform the experiment which will most strain the incorrect model to explain the data.” [HR65]

For a single model \( i \):
perform the experiment described by a design solving

\[
\max_q \left\| \Sigma^{-1} (\bar{\eta}_q - h^i [q, p^i]) \right\|_2^2, \; q \in \mathcal{Q}
\]

T-Optimal experimental design

“Optimal design with respect to a certain statistical Test”

- Identifies incorrect model with maximum likelihood! [Fed71]
Perform the experiment which will most strain the incorrect model to explain the data.” [HR65]

For a single model $i$:
perform the experiment described by a design solving

$$\max_q \left\| \Sigma^{-1}(\bar{\eta}_q - h^i[q, p^i]) \right\|_2^2, \ q \in Q$$

T-Optimal experimental design
“Optimal design with respect to a certain statistical Test”

- Identifies incorrect model with maximum likelihood! [Fed71]
- **But**: specific for model $i$
Design criterion

,,Perform the experiment which will most strain the incorrect model to explain the data.” [HR65]

For a single model $i$:
perform the experiment described by a design solving

$$\max_q \left\{ \| \Sigma^{-1} (\bar{\eta}_q - h^i [q, \phi^i]) \|_2^2, \quad q \in \mathcal{Q} \right\}$$

T-Optimal experimental design

“Optimal design with respect to a certain statistical Test”

- Identifies incorrect model with maximum likelihood! [Fed71]
- **But**: specific for model $i$
- **But**: cannot be realized at all: $\bar{\eta}_q$ unknown!
Design criterion

,,Perform the experiment which will most strain the incorrect model to explain the data.” [HR65]

For a single model $i$:
perform the experiment described by a design solving

$$\max_q \| \Sigma^{-1}(\bar{\eta}_q - h^i[q, p^i]) \|^2, \quad q \in Q$$

T-Optimal experimental design
“Optimal design with respect to a certain statistical Test”

- Identifies incorrect model with maximum likelihood! [Fed71]
- **But**: specific for model $i$
- **But**: cannot be realized at all: $\bar{\eta}_q$ unknown!

⇒ T-optimal design: theoretical “upper limit”!
Use what you have: several model responses

For two models $A$, $B$:
Perform the experiment described by the design $q$ solving

\[
\max_q \| \Sigma^{-1} (h^A[q,p^A] - h^B[q,p^B]) \| , q \in Q
\]

"Perform the experiments where model predictions differ most."
Use what you have: several model responses

For two models $A, B$:
Perform the experiment described by the design $q$ solving

$$\max_q \| \Sigma^{-1} \left( h^A[q, p^A] - h^B[q, p^B] \right) \|, q \in Q$$

"Perform the experiments where model predictions differ most."

- In sequential discrimination procedure:
  Converges to $T$-optimal design almost surely. [Fed75, AD92]
Use what you have: several model responses

For two models $A$, $B$:
Perform the experiment described by the design $q$ solving

$$
\max_q \| \Sigma^{-1} (h^A[q,p^A] - h^B[q,p^B]) \|, q \in Q
$$

"Perform the experiments where model predictions differ most."

- In sequential discrimination procedure:
  Converges to T-optimal design almost surely. [Fed75, AD92]

- Generalization to $N > 2$ models:
  - Use (1) with the two best-fitting models ("ranking" [AF75])
Practical solution approach

Use what you have: several model responses

For two models $A$, $B$:
Perform the experiment described by the design $q$ solving

$$
\max_q \| \Sigma^{-1} (h^A[q, p^A] - h^B[q, p^B]) \|, \quad q \in Q
$$

"Perform the experiments where model predictions differ most."

- In sequential discrimination procedure:
  Converges to T-optimal design almost surely. [Fed75, AD92]

- Generalization to $N > 2$ models:
  - Use (1) with the two best-fitting models (“ranking” [AF75])
  - Apply (1) pairwise to all models, weight by model reliability

$$
\max_q \sum_{i, j=1}^{N} w_{ij} \| h^i - h^j \|
$$
OED for model discrimination: a complex optimal control problem [Hof05]
To be realizable, a design $q$ for MD must respect
To be realizable, a design $q$ for MD must respect

- the *model dynamics*,
  formulated as constraints for *state variables* $x^i$ (here: DAEs)

\[
\bar{\eta}_q \approx h^i[q, x^i, p^i] \\
0 = f^i(\dot{x}^i(t), x^i(t), p^i, q)
\]
Design constraints

To be realizable, a design \( q \) for MD must respect

- the *model dynamics*,
  formulated as constraints for *state variables* \( x^i \) (here: DAEs)

\[
\tilde{\eta}_q \approx h^i[q, x^i, p^i]
\]
\[
0 = f^i(\dot{x}^i(t), x^i(t), p^i, q)
\]

- *experimental constraints*:
  initial conditions, security regulations, measurement limits, …
  formulated in terms of the relevant model

\[
0 \leq b^i(x^i(t_0), \ldots, x^i(t_B), p^i, q)
\]
\[
0 \leq c^i(x^i, p^i, q)
\]
Optimal control problem of MD

Resulting OED problem for $i = 1, \ldots, N$ models:
Resulting OED problem for $i = 1, \ldots, N$ models:

$$\max_{x^1, \ldots, x^N, q} \Phi (h^1, \ldots, h^N), \text{ with } h^j = [q, x^j, p^j]$$
Resulting OED problem for \( i = 1, \ldots, N \) models:

\[
\max_{x^1, \ldots, x^N, q} \Phi (h^1, \ldots, h^N), \quad \text{with} \quad h^j = [q, x^j, p^j]
\]

s.t. constraints from model dynamics:

\[
0 = F := \begin{cases} 
  f^1 (\dot{x}^1, x^1, p^1, q) \\
  \vdots \\
  f^N (\dot{x}^N, x^N, p^N, q)
\end{cases}
\]
Optimal control problem of MD

Resulting OED problem for $i = 1, \ldots, N$ models:

$$\max_{x^1, \ldots, x^N, q} \Phi (h^1, \ldots, h^N), \text{ with } h^j = [q, x^j, p^j]$$

s.t. constraints from model dynamics:

$$0 = F := \begin{cases} f^1 (\dot{x}^1, x^1, p^1, q) \\ \vdots \\ f^N (\dot{x}^N, x^N, p^N, q) \end{cases}$$

s.t. experimental constraints:

$$0 \leq B := \begin{cases} b^1 (x^1(t_0), \ldots, x^1(t_B), p^1, q) \\ \vdots \\ b^N (x^N(t_0), \ldots, x^N(t_B), p^N, q) \end{cases}$$

$$0 \leq C := \begin{cases} c^1 (x^1, p^1, q) \\ \vdots \\ c^N (x^N, p^N, q) \end{cases}$$
Resulting OED problem for $i = 1, \ldots, N$ models:

$$\max_{x^1, \ldots, x^N, q} \Phi (h^1, \ldots, h^N), \quad \text{with} \ h^j = [q, x^j, p^j]$$

s.t. constraints from
model dynamics:

$$0 = F := \begin{cases} f^1 (\dot{x}^1, x^1, p^1, q) \\ \vdots \\ f^N (\dot{x}^N, x^N, p^N, q) \end{cases}$$

s.t. experimental
constraints:

$$0 \leq B := \begin{cases} b^1 (x^1(t_0), \ldots, x^1(t_B), p^1, q) \\ \vdots \\ b^N (x^N(t_0), \ldots, x^N(t_B), p^N, q) \end{cases}$$

$$0 \leq C := \begin{cases} c^1 (x^1, p^1, q) \\ \vdots \\ c^N (x^N, p^N, q) \end{cases}$$

High-dimensional, nonlinear, constrained optimal control problem
Specific challenges
Specific challenges

- Many models!
Specific challenges

- **Many** models!

- Stiff dynamics (e.g. severely different reaction rates)
  BDF-type solver DAESOL-II/SOLVIND for state integration
  [Alb05, AK07]
Specific challenges

- **Many** models!

- Stiff dynamics (e.g. severely different reaction rates)
  BDF-type solver DAESOL-II/SOLVIND for state integration
  [Alb05, AK07]

- Highly nonlinear dynamics (e.g. Arrhenius’ law)
  Direct boundary value problem approach,
  Multiple-Shooting state discretization [BP84]
Specific challenges

- **Many** models!

- Stiff dynamics (e.g. severely different reaction rates)
  BDF-type solver DAESOL-II/SOLVIND for state integration
  [Alb05, AK07]

- Highly nonlinear dynamics (e.g. Arrhenius’ law)
  Direct boundary value problem approach,
  Multiple-Shooting state discretization [BP84]

  ➤ high-dimensional, constrained NLP:

  \[
  \min_{v} \Phi(v), \text{ s.t. } 0 = \zeta(v), 0 \leq \xi(v)
  \]

  Solved by tailored SQP algorithm MUSCOD-II [Lei99, Die01, Sch05]
Specific challenges

- **Many** models!

- Stiff dynamics (e.g. severely different reaction rates)
  BDF-type solver DAESOL-II/SOLVIND for state integration
  \[ [\text{Alb05, AK07}] \]

- Highly nonlinear dynamics (e.g. Arrhenius’ law)
  Direct boundary value problem approach,
  Multiple-Shooting state discretization \[ [\text{BP84}] \]

- high-dimensional, constrained NLP:

\[
\min_{\mathbf{v}} \Phi(\mathbf{v}), \text{ s.t. } 0 = \zeta(\mathbf{v}), \ 0 \leq \xi(\mathbf{v})
\]

Solved by tailored SQP algorithm MUSCOD-II \[ [\text{Lei99, Die01, Sch05}] \]

Required: gradient of \( \Phi \), Jacobians of \( \zeta, \xi \)

*Derivative computation is expensive:*

involves solution of complex initial value problems
Problem contains $N$ mutually independent models:

$$\frac{df^i}{dx^j} = \frac{db^i}{dx^j} = \frac{dc^i}{dx^j} = 0, \quad \forall i \neq j$$

*Only* coupled via objective function $\Phi (h^1, \ldots, h^N)$
Structure exploitation

Problem contains $N$ mutually independent models:

$$\frac{df^i}{dx^j} = \frac{db^i}{dx^j} = \frac{dc^i}{dx^j} = 0, \forall i \neq j$$

*Only* coupled via objective function $\Phi (h^1, \ldots, h^N)$

⇒ NLP with specific *multiple setpoint structure* [Hof05], e.g.
Problem contains $N$ mutually independent models:

$$\frac{df^i}{dx^j} = \frac{db^i}{dx^j} = \frac{dc^i}{dx^j} = 0, \forall i \neq j$$

Only coupled via objective function $\Phi (h^1, \ldots, h^N)$

- NLP with specific *multiple setpoint structure* [Hof05], e.g.

$$\frac{dF}{dx} = \begin{pmatrix} J_1 & J_2 & \cdots & J_N \end{pmatrix}, \text{ with } J_i := \frac{df^i}{dx^i}$$

- Efficient derivative generation: effort $\propto N$, instead of $\propto N^2$!

Implementation: MUSCOD-II, prototype phase
Numerical results:
Enzyme deactivation
Aim: proof of concept

I need to *know* which model describes process correctly!

- Define: model X is „nature”.
- Use virtual measurement data:
  simulation trajectories from X plus noise
Test procedure

Aim: proof of concept

I need to *know* which model describes process correctly!

- Define: model X is ‘‘nature’’.
- Use virtual measurement data: simulation trajectories from X plus noise

Results:

- 12 test series, various conditions
- designed >30 experiments
- All unsuitable models correctly identified, no false positive.

Presented results: Model B is ‘‘nature’’
Model predictions for experiment 1, after fitting to experiment 1

Test results: A, B, C suitable
Results

Measurement data from experiment 2
Results

Model predictions for experiment 2, after fitting to experiments 1-2

Test results: A, B, C suitable
Measurement data from experiment 3
Results

Model predictions for experiment 3, after fitting to experiments 1-3

Test results: A unsuitable, B, C suitable
Results

Measurement data from experiment 4

Specific activity [1000 IU]
Temperature [°C]

Time [h]
Model predictions for experiment 4, after fitting to experiments 1-4

Test results: B, C suitable
Measurement data from experiment 5

- **Specific activity [1000 IU]**
- **Temperature [°C]**
- **Time [h]**

Temperature measured +
Model predictions for experiment 5, after fitting to experiments 1-5

Test results: B, C suitable
Results

Measurement data from experiment 6

Specific activity [1000 IU]

Time [h]

Temperature measured

Temperature [°C]
Model predictions for experiment 6, after fitting to experiments 1-6

Test results: C unsuitable ➡️ correct model B remains
Model discrimination problems arise if modeling with little previous knowledge.
Model discrimination problems arise if modeling with little previous knowledge.

They can be solved efficiently using targeted experimental design.
Model discrimination problems arise if modeling with little previous knowledge.

They can be solved efficiently using targeted experimental design.

The design tasks leads to high-dimensional OC problems. Their specific structure allows efficient solution even for many models.
Thank you very much for your attention!


