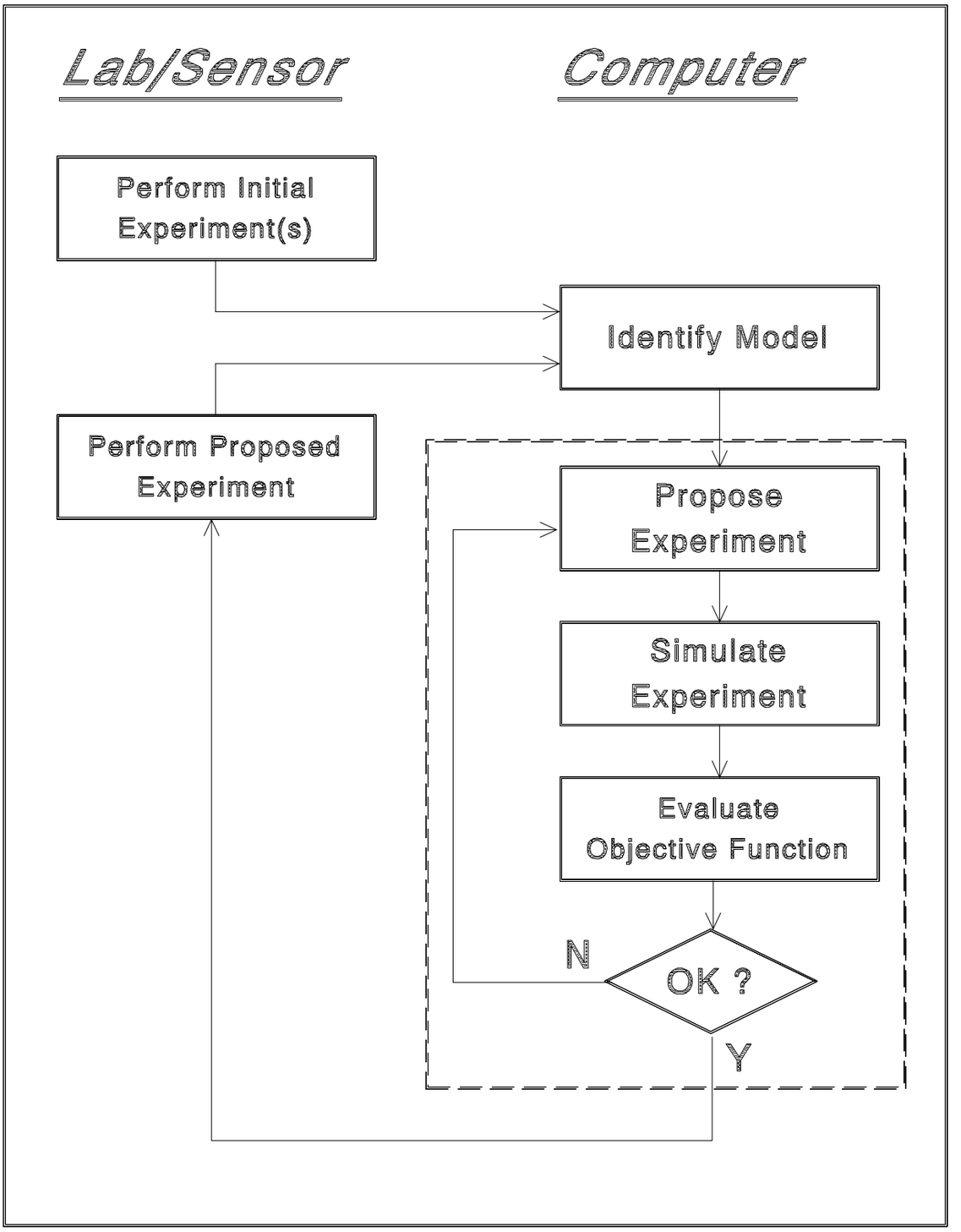
EXPERIMENTAL DESIGN

Peter A. Vanrolleghem

Laboratory of Microbial Ecology, University of Gent
Coupure Links 653, B-9000 Gent
BELGIUM
Email Peter.Vanrolleghem@rug.ac.be

The General Procedure for Experimental Design



MODELS CONSIDERED IN THIS STUDY:

express the dependence of OUR_{ex} on the biodegradation of k substrates S_i :

$$OUR_{ex} = -\sum_{i=1}^{k}(1-Y_i)\frac{dS_i}{dt} \qquad \textit{(pe 1)}$$

Different model complexities: 4 types of wastewater/sludge interaction:

Type 1 (Exponential): One pollutant, first order degradation (k=1)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}\,X}{Y_1}\,S_1 \qquad \textit{(pe 2)}$$

Type 2 (Single Monod): One pollutant, Monod type of degradation (k=1)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}\,X}{Y_1}\,\frac{S_1}{K_{m1}+S_1} \qquad \textit{(pe 3)}$$

Type 3 (Double Monod): Two pollutants, both Monod type (k=2)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}\,X}{Y_1}\,\frac{S_1}{K_{m1}+S_1}$$

$$\frac{dS_2}{dt} = -\frac{\mu_{max2}\,X}{Y_2}\,\frac{S_2}{K_{m2}+S_2} \qquad \textit{(pe 4)}$$

Type 4 (Modified IAWQ model n˚1): Three pollutants, two hydrolyse into the first substrate which is used for growth according to Monod (k=1)

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}\,X}{Y_1}\,\frac{S_1}{K_{m1}+S_1} + k_r X_r + k_s X_s$$

$$\frac{dX_r}{dt} = -\,k_r\,X_r$$

$$\frac{dX_s}{dt} = -\,k_s\,X_s \qquad \textit{(pe 5)}$$

In this model OUR_{ex} should be rewritten as:

$$OUR_{ex} = -\,(1-Y_1)\left(\frac{dS_1}{dt} - k_r X_r - k_s X_s\right) \qquad \textit{(pe 6)}$$

Theoretical background of OED/PE

Question addressed:

"*with the given experimental data, can the parameters be given unique values"*

or: "if a small deviation in the parameter set occurs,
does this have a considerable decrease of the fit as a consequence".

THE MATHEMATICS:

Consider the quadratic objective functional ("sum of squared errors"):

$$J(\theta) = \sum_{i=1}^{N} (\hat{\boldsymbol{y}}_i(\theta) - \boldsymbol{y}_i)^T \boldsymbol{Q}_i (\hat{\boldsymbol{y}}_i(\theta) - \boldsymbol{y}_i) \qquad \textit{(pe 7)}$$

in which $\boldsymbol{y}_i$ and $\hat{\boldsymbol{y}}_i(\theta)$ are vectors of N measured values and model predictions

$\boldsymbol{Q}_i$ is a square matrix with user-supplied weighting coefficients.

Parameter estimation can be formulated as:

minimization of J by optimal choice of the parameters θ.

The effect of a small deviation of the parameters $\delta\theta$ on the model fit:

$$\boldsymbol{y}(t, \theta + \delta\theta) = \boldsymbol{y}(t, \theta) + \left[\frac{\partial \boldsymbol{y}}{\partial \theta}(t)\right]_{\theta} \delta\theta = \boldsymbol{y}(t, \theta) + \boldsymbol{Y}_{\theta}(t)\, \delta\theta \qquad \textit{(pe 8)}$$

where $\boldsymbol{Y}_{\theta}(t)$: output sensitivity functions with respect to parameter variations

Expected value of J for a parameter set slightly different from the optimal one:

$$E\left[J(\theta+\delta\theta)\right] \approx \delta\theta^T \left(\sum_{i=1}^{N} (\boldsymbol{Y}_\theta(t_i))^T \boldsymbol{Q}_i \, \boldsymbol{Y}_\theta(t_i) \right) \delta\theta + \sum_{i=1}^{N} tr\left(\boldsymbol{C}_i \, \boldsymbol{Q}_i \right) \quad \textit{(pe 9)}$$

in which $\boldsymbol{C}_i$ represents the [redacted]

($\boldsymbol{Q}_i$ is typically chosen as $\boldsymbol{C}_i^{-1}$ => second term reduces to Nm,
m = dimension of measurement vector)

To optimize practical identifiability
= maximize difference between $J(\theta + \delta\theta)$ and $J(\theta)$
= maximize term between brackets = Fisher Information Matrix

$$F = \sum_{i=1}^{N} (\boldsymbol{Y}_\theta(t_i))^T \boldsymbol{Q}_i \, \boldsymbol{Y}_\theta(t_i) \quad \textit{(pe 10)}$$

expresses the information content of the experiment
= inverse of the [redacted]

$$V = F^{-1} = \left(\sum_{i=1}^{N} (\boldsymbol{Y}_\theta(t_i))^T \boldsymbol{Q}_i \, \boldsymbol{Y}_\theta(t_i) \right)^{-1} \quad \textit{(pe 11)}$$

Parameter variance

use the parameter estimation error covariance matrix V
and the residual mean square s^2:

$$s^2 = \frac{J(\theta)}{N - p} \quad \textit{(pe 12)}$$

Approximate standard errors for the parameters:

$$\sigma(\theta_i) = s\sqrt{V_{ii}} \quad \textit{(pe 13)}$$

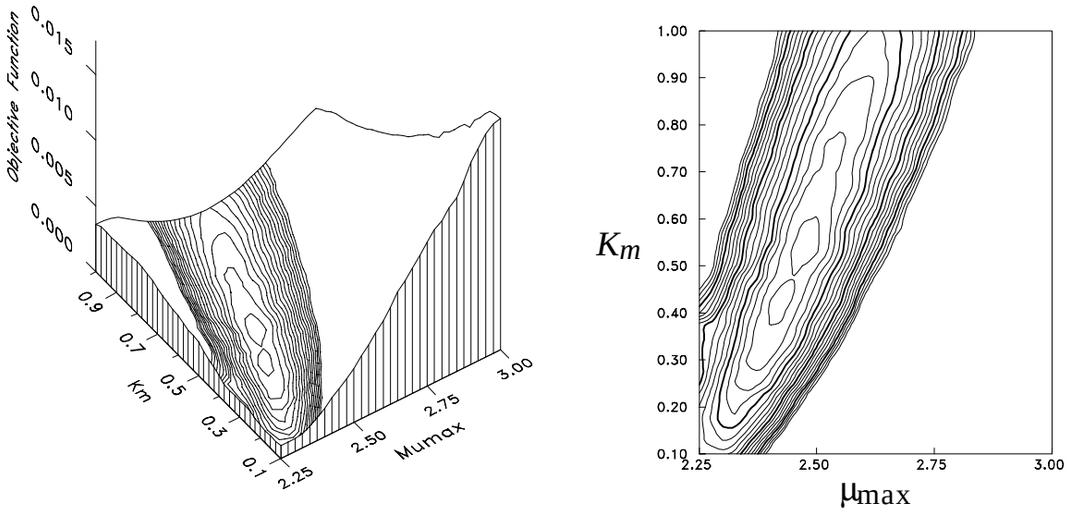
optimistic due to linear approximation of the nonlinear model

IMPORTANT:

Parameter variance can be reduced through decreased s (= increase N)
or by reduced V_{ii} (by OED/PE)

Confidence region of the parameters

J as function of μ_{max} and K_m

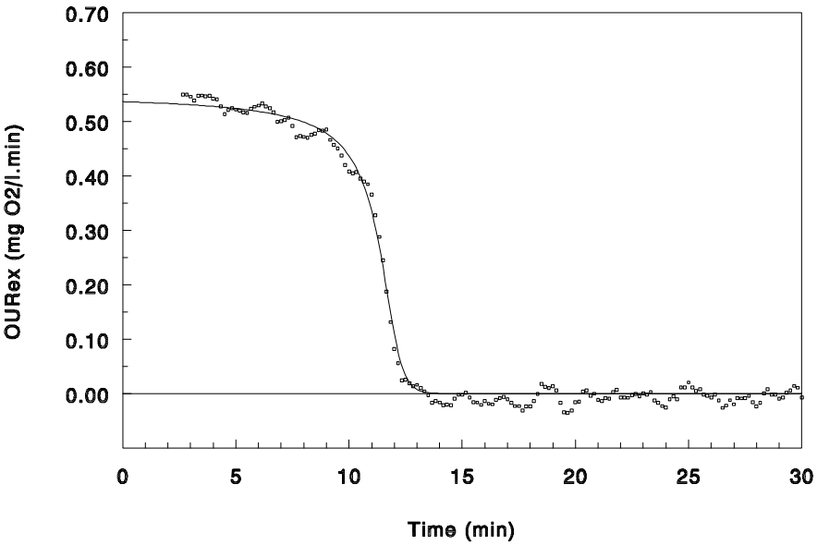


Minimum value J_{opt} at ($\hat{\mu}_{max}$, $\hat{K}_m$) within a "valley" in $J(\mu_{max}, K_m)$.

$(1-\alpha)$ Confidence region is the area delimited by
the confidence contour line = set of parameter combinations with functional:

$$J_{opt} * \left(1 + \frac{p}{N-p} F_{\alpha;p,N-p}\right) \qquad \textit{(pe 14)}$$

F is the value of the F-distribution with p and N-p d.f. and confidence level α



An Example (2 parameters α and β, measurement of S and X)

Output sensitivities deduced from the model equations:

$$S\alpha(t_i) = \left[\frac{\partial S}{\partial \alpha}\right]_{t_i} ; S\beta(t_i) = \left[\frac{\partial S}{\partial \beta}\right]_{t_i} ; X\alpha(t_i) = \left[\frac{\partial X}{\partial \alpha}\right]_{t_i} ; X\beta(t_i) = \left[\frac{\partial X}{\partial \beta}\right]_{t_i} \quad \textit{(pe 15)}$$

Measurement error covariance matrix $C(t_i)$:

$$C(t_i) = \begin{pmatrix} \sigma^2_{11}(t_i) & \sigma^2_{12}(t_i) \\ \sigma^2_{12}(t_i) & \sigma^2_{22}(t_i) \end{pmatrix} \quad \textit{(pe 16)}$$

and the corresponding weighting matrix $Q(t_i) = C^{-1}(ti)$

The Fisher Information Matrix:

$$F = \sum_{i=1}^{N} \begin{pmatrix} S\alpha(t_i) & X\alpha(t_i) \\ S\beta(t_i) & X\beta(t_i) \end{pmatrix} \begin{pmatrix} Q_{11}(t_i) & Q_{12}(t_i) \\ Q_{12}(t_i) & Q_{22}(t_i) \end{pmatrix} \begin{pmatrix} S\alpha(t_i) & S\beta(t_i) \\ X\alpha(t_i) & X\beta(t_i) \end{pmatrix} \quad \textit{(pe 17)}$$

whose elements are (t_i are omitted for ease of reading):

$$F_{11} = \sum_{i=1}^{N} Q_{11}S_\alpha^2 + 2Q_{12}S_\alpha X_\alpha + Q_{22}X_\alpha^2$$

$$F_{12} = F_{21} = \sum_{i=1}^{N} Q_{11}S_\alpha S_\beta + Q_{12}S_\alpha X_\beta + Q_{12}S_\beta X_\alpha + Q_{22}X_\alpha X_\beta$$

$$F_{22} = \sum_{i=1}^{N} Q_{11}S_\beta^2 + 2Q_{12}S_\beta X_\beta + Q_{22}X_\beta^2 \quad \textit{(pe 18)}$$

Sensitivity Functions for Monod/Modified IAWQ nr 1

OUR_{ex} measurements as only source of information

The sensitivity of OUR_{ex} with respect to μ_{max1} is (output sensitivity function):

$$\frac{\partial OUR_{ex}}{\partial \mu_{max1}} = \frac{\partial}{\partial \mu_{max1}}\left(-(1-Y_1)\,\frac{dS_1}{dt} \right) \qquad \textit{(pe 19a)}$$

$$= -(1-Y_1)\,\frac{d}{dt}\left(\frac{\partial S_1}{\partial \mu_{max1}} \right) \qquad \textit{(pe 19b)}$$

in which the state sensitivity $\frac{\partial S_1}{\partial \mu_{max1}}$ is obtained by integration of:

$$\frac{d}{dt}\left(\frac{\partial S_1}{\partial \mu_{max1}} \right) = \frac{\partial}{\partial \mu_{max1}}\left(-\frac{\mu_{max1} X}{Y_1}\,\frac{S_1}{K_{m1}+S_1} \right) \qquad \textit{(pe 20a)}$$

$$= -\frac{X}{Y_1}\left(\frac{S_1}{K_{m1}+S_1} + \frac{\mu_{max1}\,K_{m1}\,\frac{\partial S_1}{\partial \mu_{max1}}}{(K_{m1}+S_1)^2} \right) \qquad \textit{(pe 20b)}$$

where S_1 is calculated by integration of:

$$\frac{dS_1}{dt} = -\frac{\mu_{max1}\,X}{Y_1}\,\frac{S_1}{K_{m1}+S_1} \qquad \textit{(pe 21)}$$

One can proceed similarly for the sensitivity of OUR_{ex} with respect to K_{m1}. The following relations are obtained:

$$\frac{\partial OUR_{ex}}{\partial K_{m1}} = -(1-Y_1)\,\frac{d}{dt}\left(\frac{\partial S_1}{\partial K_{m1}} \right) \qquad \textit{(pe 22)}$$

$$\frac{d}{dt}\left(\frac{\partial S_1}{\partial K_{m1}} \right) = -\frac{\mu_{max1}\,X}{Y_1}\left(\frac{K_{m1}\,\frac{\partial S_1}{\partial K_{m1}} - S_1}{(K_{m1}+S_1)^2} \right) \qquad \textit{(pe 23)}$$

Output sensitivities of the IAWQ model as modified by Sollfrank & Gujer, 1991:

$$\frac{\partial OUR_{ex}}{\partial k_r} = (1 - Y_1)\frac{\partial}{\partial k_r}\left(-\frac{dS_1}{dt} + k_r X_r + k_s X_s\right) \qquad \textit{(pe 24a)}$$

$$= (1 - Y_1)\left(X_r + k_r\frac{\partial X_r}{\partial k_r} - \frac{d}{dt}\left(\frac{\partial S_1}{\partial k_r}\right)\right) \qquad \textit{(pe 24b)}$$

$$\frac{\partial OUR_{ex}}{\partial k_s} = (1 - Y_1)\left(X_s + k_s\frac{\partial X_s}{\partial k_s} - \frac{d}{dt}\left(\frac{\partial S_1}{\partial k_s}\right)\right) \qquad \textit{(pe 25)}$$

State sensitivities needed, are:

$$\frac{d}{dt}\left(\frac{\partial X_r}{\partial k_r}\right) = \frac{\partial}{\partial k_r}\left(-k_r X_r\right) = -\left(X_r + k_r\frac{\partial X_r}{\partial k_r}\right) \qquad \textit{(pe 26)}$$

$$\frac{d}{dt}\left(\frac{\partial X_s}{\partial k_s}\right) = -\left(X_s + k_s\frac{\partial X_s}{\partial k_s}\right) \qquad \textit{(pe 27)}$$

Output and state sensitivities for μ_{max1} and K_{m1} are as for the Monod model

Properties

Sensitivity functions indicate conditions where the dependence is the largest
which conditions provide most information on the parameters.

The (state and output) sensitivities are dependent on the parameter values.

a general characteristic of nonlinear models

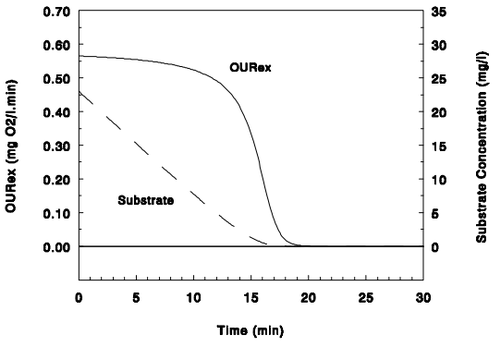
the Fisher Information Matrix is dependent on the parameter values
which has important implications for the OED/PE

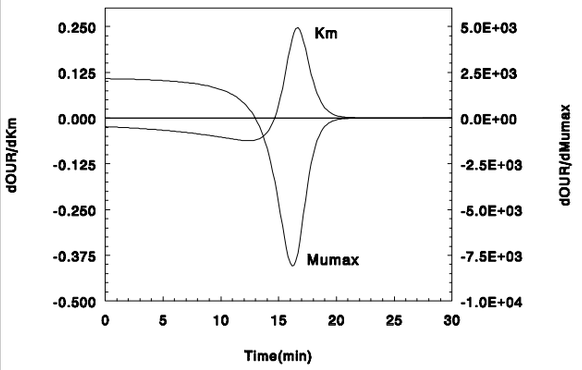
If sensitivity functions are linearly dependent
= > Fisher Information Matrix singular = > Non-informative experiment

Calculation of the rank of the Fisher Information Matrix
If no linear dependency exists, it should be full rank

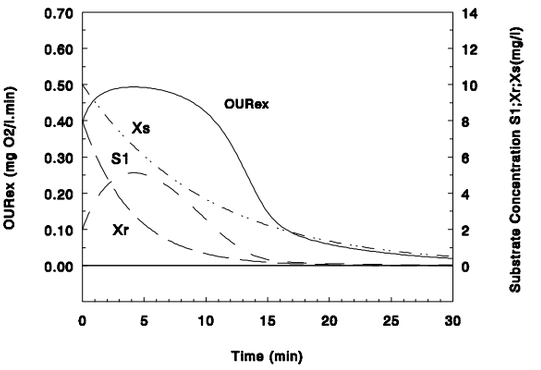
Condition number of the Fisher matrix (ratio of largest to smallest eigenvalue)
indicates whether the sensitivities are nearly linearly dependent
the higher the condition number, the lower the practical identifiability.

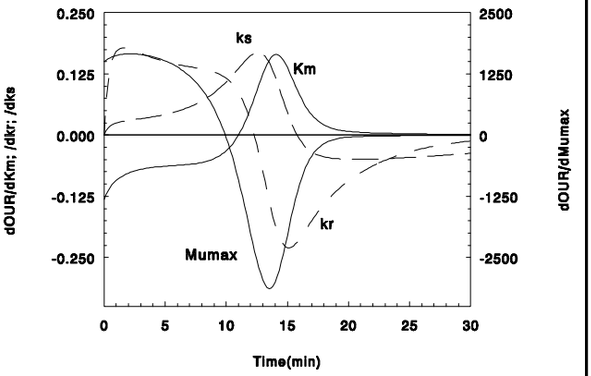
SINGLE MONOD





MODIFIED IAWQ





Practical Identifiability of the Initial conditions

For the Single Monod model:

$$\frac{\partial OUR_{ex}}{\partial S_1(0)} = -(1-Y_1)\frac{d}{dt}\left(\frac{\partial S_1}{\partial S_1(0)}\right)$$

$$= (1-Y_1)\frac{\mu_{max1} X}{Y_1}\frac{\partial}{\partial S_1(0)}\left(\frac{S_1}{K_{m1} + S_1}\right) \qquad \textit{(pe 28)}$$

Introducing the initial condition:

$$S_1(t) = S_1(0) - \frac{\int_0^t OUR_{ex}(\tau)d\tau}{1 - Y_1} \qquad \textit{(pe 29)}$$

yielding:

$$\frac{\partial OUR_{ex}}{\partial S_1(0)} = \frac{(1-Y_1)\,\mu_{max1}\,X}{Y_1}\frac{\partial}{\partial S_1(0)}\left(\frac{(1-Y_1)\,S_1(0) - \int_0^t OUR_{ex}(\tau)d\tau}{(1-Y_1)\left(K_{m1} + S_1(0)\right) - \int_0^t OUR_{ex}(\tau)d\tau}\right)$$

(pe 30)

and the final equation:

$$\frac{\partial OUR_{ex}}{\partial S_1(0)} = \frac{(1-Y_1)^2\,\mu_{max1}\,X\,K_{m1}}{Y_1}\left(\frac{(1-Y_1) - \int_0^t \frac{\partial OUR_{ex}(\tau)}{\partial S_1(0)}d\tau}{\left((1-Y_1)\left(K_{m1} + S_1(0)\right) - \int_0^t OUR_{ex}(\tau)d\tau\right)^2}\right)$$

(pe 31)

Optimal Experimental Design for Parameter Estimation

Fisher Information Matrix (≈ covariance matrix) = corner-stone of OED/PE

these matrices summarize the information content of an experiment
or the preciseness of the parameter estimates.

Different scalar measures of these matrices can be optimized:

- A-optimal design criterion: $\min \; tr(F^{-1})$
- modified A-optimal design criterion: $\max \; tr(F)$
- D-optimal design criterion: $\max \; \det(F)$
- E-optimal design criterion: $\max \; \lambda_{min}(F)$
- modified E-optimal design criterion: $\min \; \frac{\lambda_{max}(F)}{\lambda_{min}(F)}$

in which $\lambda_{min}(F)$ and $\lambda_{max}(F)$ are the smallest and largest eigenvalue of F

Interpretation

A- and D-optimal designs

minimize the arithmetic and geometric mean of the identification errors

E-criterion

minimize the largest error.
maximize the distance from the singular (non-informative) case

Modified E criterion

optimize objective functional shape (ratio of largest to smallest eigenvalue)

Modified A-criterion

No interpretation
may give rise to non-informative and unidentifiable experiments‘!!

EXAMPLES OF OED/PE FOR SINGLE MONOD MODEL

Degrees of Freedom for OED/PE

- Optimal initial substrate
- Optimal additional pulse with fixed initial substrate
- Optimal additional pulse and initial substrate
- Optimal design with multiple additional pulses

and Constraints for OED/PE

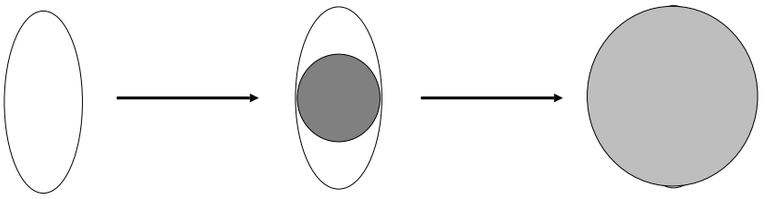
- Experiment time (hard bound)

Theoretical Example 1: initial substrate

Modified E:

substrate concentration 10 times lower than reference

shape has improved (ratio has decreased by 3.3)
at expense of parameter estimation quality: confidence regions have increased
for μ_{max1} by an order of magnitude
for K_{m1} with a factor 3



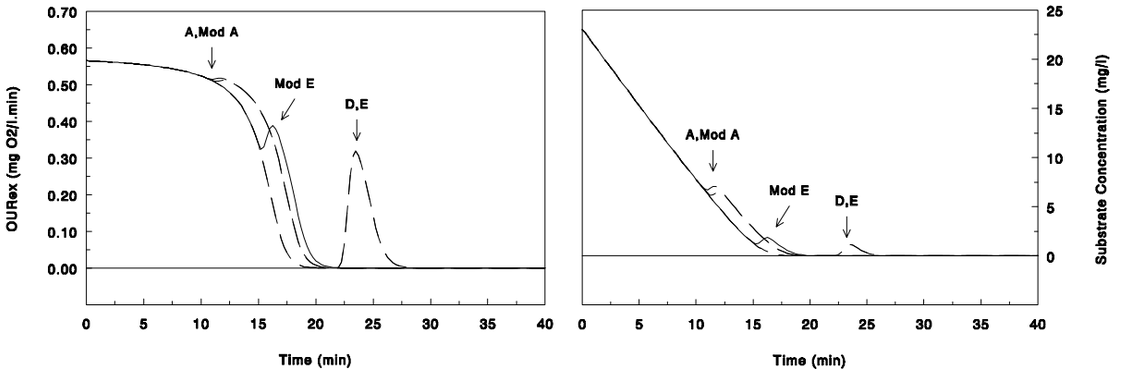
Other four criteria:

experiments with the highest possible information content
from the sensitivity functions:
experiment where exogenous oxygen uptake rate is different from zero
for the longest possible time
the substrate is almost completely oxidized by the end of the experiment

confidences in estimates of μ_{max1} and K_{m1} improve with a factor 2.4 and 1.25

Theoretical Example 2: One additional pulse

initial substrate concentration is fixed identical to the base case



Differences among the design criteria are considerable.

A- and modified A:

prolong experimental conditions with maximal substrate degradation

= > the most important improvement for μ_{max1}

D- and E-criteria:

substrate is injected only after exogenous respiration dropped completely

substrate for a longer period of time at $S \approx K_{m1}$

Modified E-criterion:

design in between both approaches

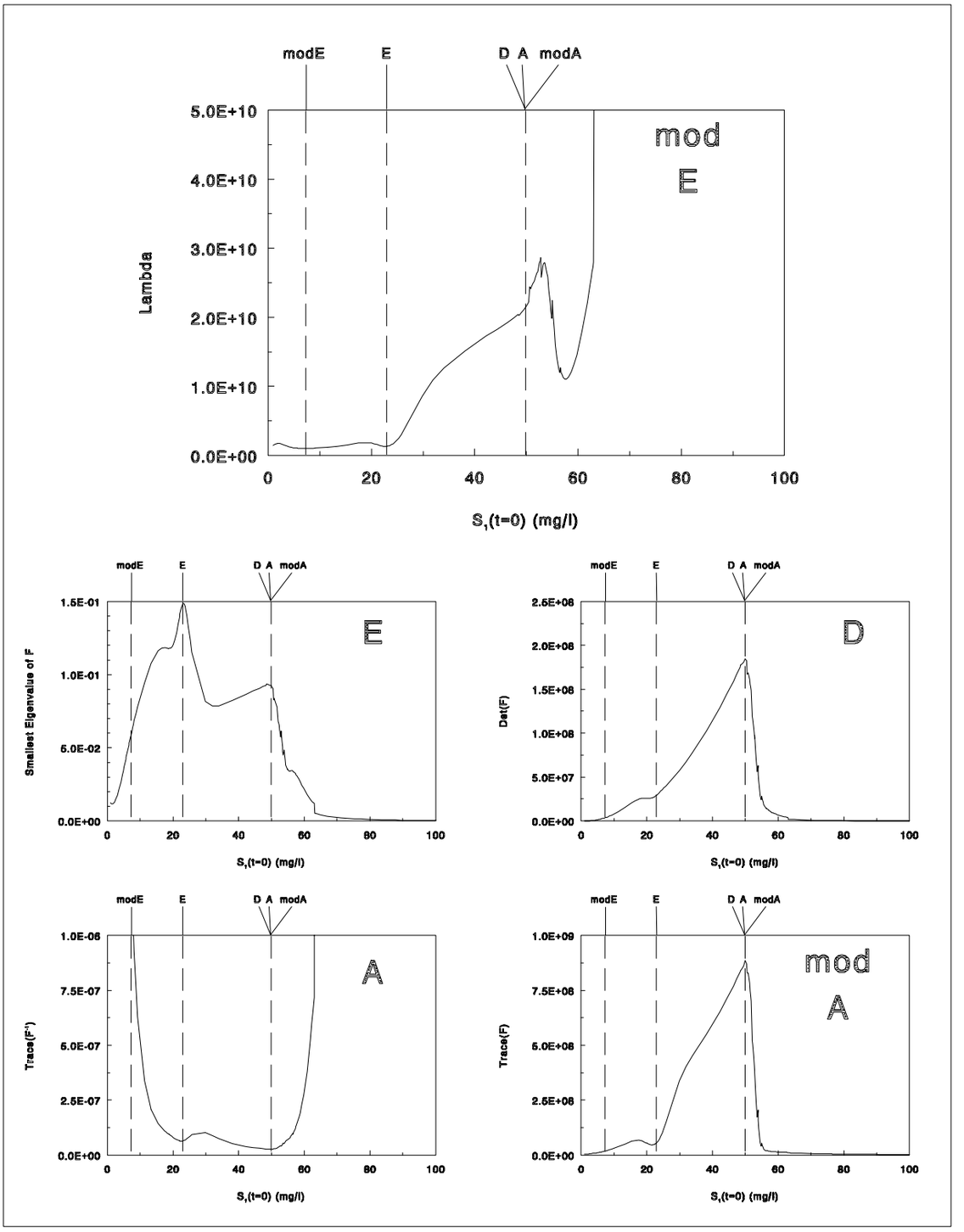
confidence interval for the μ_{max1} only decreased with 10 %,
the K_{m1} accuracy increased with more than 50 %

covariance between μ_{max1} and K_{m1} is reduced to the same extent

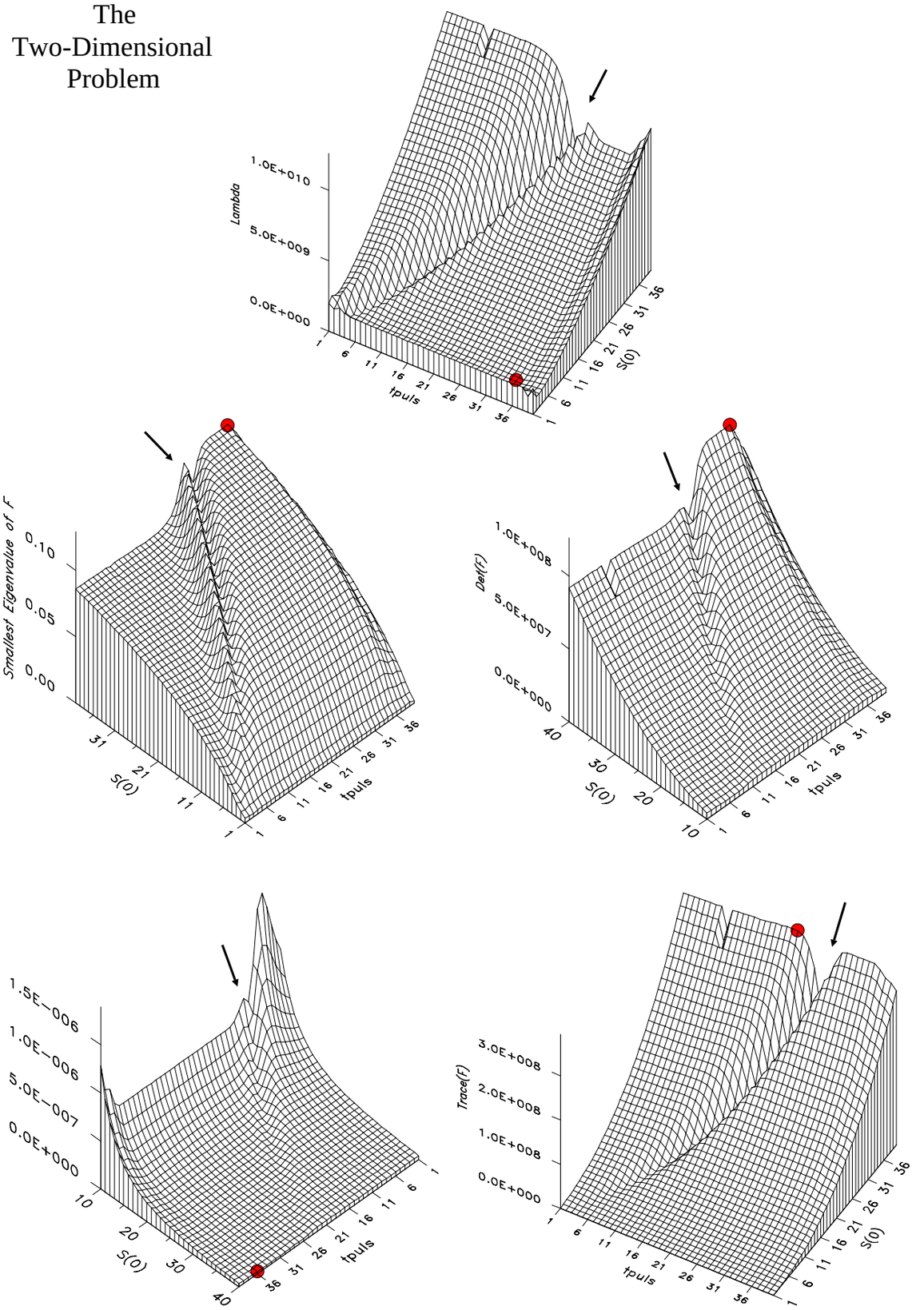
Theoretical Example 3: Additional pulse + Initial substrate

Two-dimensional optimization problem

In example below: fixed tpuls at 18.2 min; optimize $S_1(0)$



The
Two-Dimensional
Problem



Discussion

Optimal experimental designs depend significantly on criterion

Local extrema (e.g. E-optimal)
corresponding to conditions optimal for some other criteria

Modified E has rather different behaviour
probably due to the different underlying objective

For tpuls=18.2 min (= the optimum pulse time for $S_1(0)$ =23) :

$S_1(0)$ = 23 is the minimum (E-criterion)
$S_1(0)$ = 23 remains a local minimum (mod E; A)

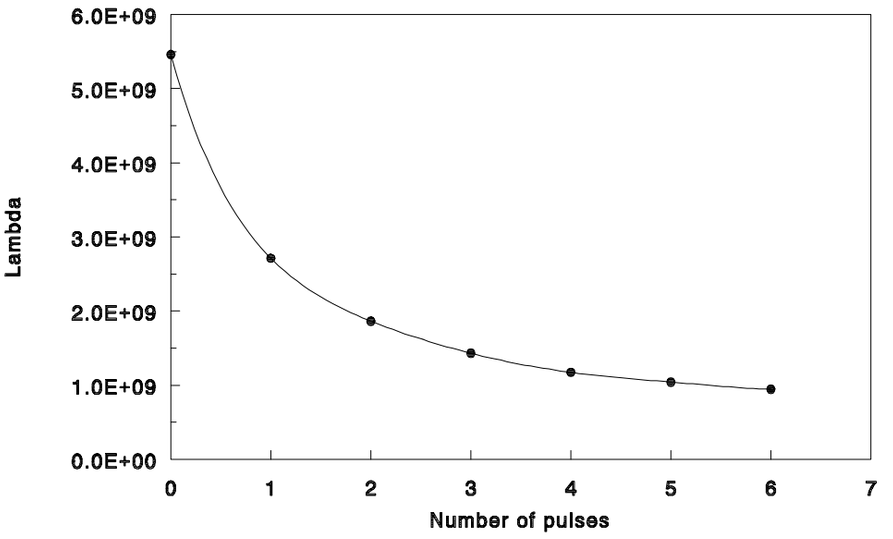
Presence of local extrema may cause convergence problems

3D-plots:

"Ripple" on the surface
corresponds to conditions in which the pulse is added
at the time the substrate initially present in the reactor is depleted

"ripple" moves further as $S_1(0)$ increases

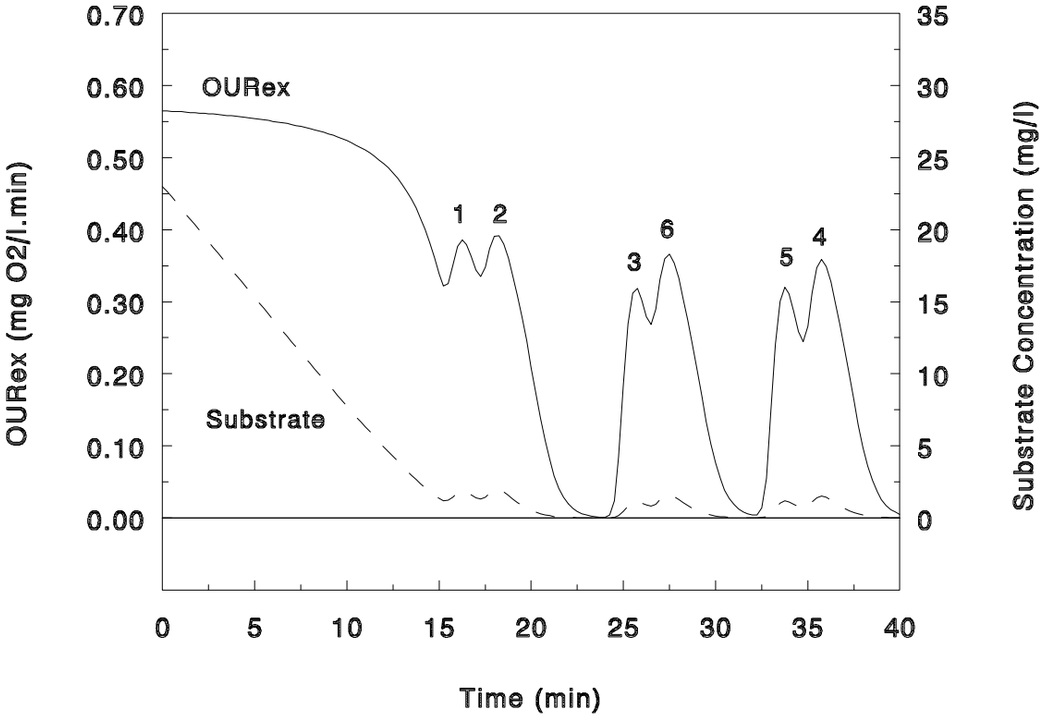
Theoretical Example 4: Multiple Pulses



Benefit of adding another degree of freedom to the design becomes marginal.

Design canoccur sequentially

= > computational burden is considerably lower
only one-dimensional optimization problems must be solved



NOMENCLATURE

$\boldsymbol{C}_i$	: Measurement error covariance matrix	(-)
F	: Fisher information matrix	(-)
J	: Objective functional	(-)
J_{opt}	: Minimal value of the objective functional	(-)
K_{mi}	: Monod half-saturation coefficient for substrate i	(mg COD/l)
k_r	: Rapid hydrolysis rate constant	(/min)
k_s	: Slow hydrolysis rate constant	(/min)
λ_{max}	: Largest eigenvalue of F (in absolute values)	(-)
λ_{min}	: Smallest eigenvalue of F (in absolute values)	(-)
m	: Dimension of measurement vector	(-)
μ_{maxi}	: Maximum specific growth rate on substrate i	(/min)
N	: Number of measurements	(-)
OED/PE	: Optimal experimental design for parameter estimation	(-)
OUR_{ex}	: Exogenous oxygen uptake rate	(mg O_2/l.min)
$\boldsymbol{Q}_i$	: Weighting matrix	(-)
s^2	: Residual mean square	(-)
S_i	: Concentration of substrate i	(mg COD/l)
$\sigma(\theta_i)$	: Standard error of parameter i	(-)
$tpuls$	: Time of pulse addition	(min)
θ_i	: Parameter i	(-)
V	: Parameter estimation covariance matrix	(-)
X	: Concentration of biomass	(mg COD/l)
X_r	: Rapidly hydrolyzable substrate concentration	(mg COD/l)
X_s	: Slowly hydrolyzable substrate concentration	(mg COD/l)
Y_i	: Yield coefficient on substrate i	(mg COD_X/mg COD_{Si})
$\boldsymbol{y}$	: Measurement vector	(-)
$\hat{\boldsymbol{y}}(\theta_i)$	: Model prediction vector	(-)
$\boldsymbol{Y}_\theta$	: Output sensitivities with respect to parameters θ	(-)

SOME USEFUL REFERENCES

Bard Y. (1974) Nonlinear Parameter Estimation. Academic Press, London. pp. 341.

Beck J.V. and Arnold K.J. (1977) Parameter Estimation in Engineering and Science. Wiley, New York. pp. 501.

Box G.E.P. and Hill W.J. (1967) Discrimination among mechanistic models. Technometrics, 9, 57-71. Cochran W.G. (1973) Experiments for nonlinear functions. J. Am. Statist. Assoc., 68, 771-781.

Boyle W.C. and Berthouex P.M. (1974) Biological wastewater treatment model building: Fits and misfits. Biotechnol. Bioeng., 16, 1139-1159.

Brent R.P. (1973) Algorithms for Minimization without Derivatives. Prentice-Hall, Englewood Cliffs, New Jersey. pp. 195.

Froment G.F. and Hosten L.H. (1981) Catalytic kinetics: Modelling. In: Catalysis - Science and Technology. Eds. Anderson J.R. and Boudart M., Springer-Verlag, Heidelberg. 97-170.

Hill P.D.H. (1978) A review of experimental design procedures for regression model discrimination. Technometrics, 20, 15-21.

Ljung L. (1987) System Identification - Theory for the User. Prentice-Hall, Englewood Cliffs, New Jersey. pp. 519.

Munack A. (1991) Optimization of sampling. In: Biotechnology, a Multi-volume Comprehensive Treatise, Vol. 4. Measuring, Modelling and Control. Ed. Schügerl K., VCH, Weinheim, 251-264.

Munack A. (1992) Some improvements in the identification of bioprocesses. In: Modeling and Control of Biotechnical Processes. Eds. Karim M.N. and Stephanopoulos G., Pergamon Press, Oxford. 89-94.

Press W.H., Flannery B.P., Teukolsky S.A. and Vetterling W.T. (1986) Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, Cambridge, UK. pp. 818.

Robinson J.A. (1985) Determining microbial parameters using nonlinear regression analysis: Advantages and limitations in microbial ecology. Adv. Microb. Ecol., 8, 61-114.

Schoukens J., Guillaume P. and Pintelon R. (1993) Design of broadband excitation signals. In: Perturbation Signals for System Identification. Ed. Godfrey K., Prentice- Hall, Englewood Cliffs, New Jersey. 1-37.

Soderstrom T. and Stoica P. (1989) Model validation and model structure determination. In: System Identification. Prentice-Hall, Englewood Cliffs, New Jersey. 422- 467.

Why Experimental Design ?

LIMITED RESOURCES:

- Time
- Expensive Equipment

INCREASED DEGREES OF FREEDOM FOR EXPERIMENTATION

QUALITY OF INFORMATION W.R.T. PURSUED GOAL

What are the Pursued Goals ?

Two basic purposes:

- Establish the form of an adequate mathematical model of the process
- Estimating precisely the values of its parameters

History

Not a new theory -> already in 1922 -> Fisher -> MPN-test for bacteriology

Most of the work on Static Models (regressions, ...)

Only in the last 15 years: Dynamic Models -> Parameter estimation

Initial Work: *Non-sequential Design*:
Parallel experimentation (e.g. agriculture, screening)
Reason: long experimentation time

Later: *Sequential Design* :
Take advantage of previously acquired insights

Three Steps

- Definition of Objective Function
  (= Mathematical Translation of the Objective)
- Enumeration of Degrees of Freedom and Constraints
- Extremization of the Objective Function

Some Degrees of Freedom to consider:

- What to measure --> *System Definition* is important (Variables,...)
- Where to measure --> Problem of *Sensor Location*
- When to measure --> *Sampling* Strategy
- Which manipulations --> *Excitation* Signal Design (sine-wave, pulse)
- What data treatment --> Noise rejection, Outlier detection/removal
  Elimination of uninteresting dynamics

Some Advice:

"While several different design variables associated with models and methods can be tried out on the computer, the experimental data can be changed only by a new experiment, which could be a costly and time-consuming procedure.

Therefore it is worthwhile to design the experiment thoughtfully so as togenerate data that are sufficiently informative"

"It is wise to let the experiment resemble the situation under which the model is to be used"

" Careful experiment design yielding data with good information is the basis of a succesful identification application"

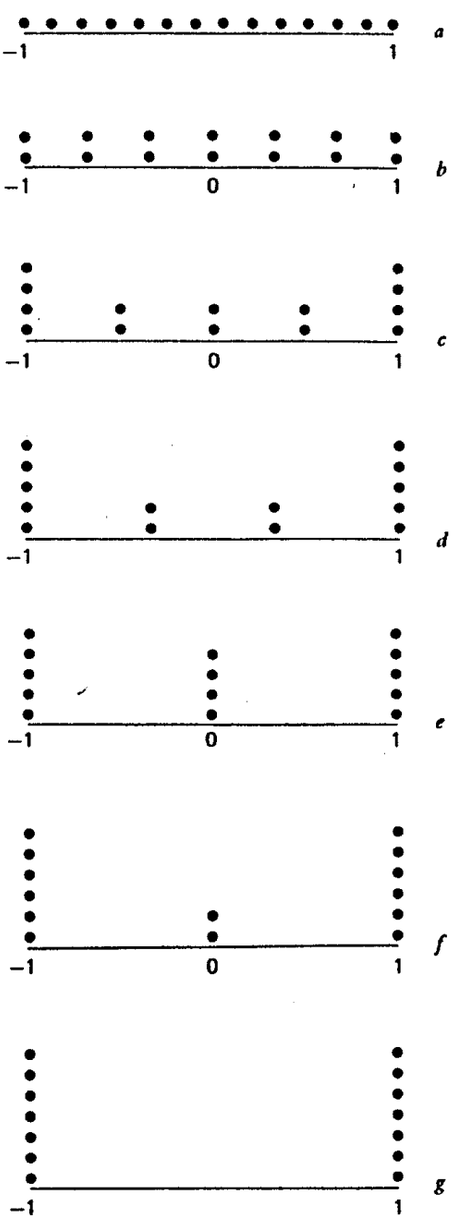
A simple example: Structure selection/Parameter estimation

REGRESSION $Y = bX + a$ *or* $Y = cX^2 + bX + a$

Degrees of Freedom: Choice of the values of the independent variables X
Constraints: Number of values = 14

Standard Deviation of b:

$$\sigma_b = \frac{\sigma}{\sum\left(X_i - \bar{X}\right)^2}$$



Structure selection based on ANOVA:

Variance	Sum of Squared Errors
explained by Y=bX+c	SSR1
explained in addition by X^2-term	SSR2-SSR1
Residual Error	SSE2
Total Error	SSTOT

Depending on the a priori knowledge on Model Structure different designs result:

Design	Degrees of Freedom		$\frac{\sigma_b}{\sigma}$	Number of sites
	SSR2-SSR1	SSE2		
a	12	0	0.43	14
b	5	7	0.4	7
c	3	9	0.33	5
d	2	10	0.31	4
e	1	11	0.32	3
f	1	11	0.29	3
g	0	12	0.27	2

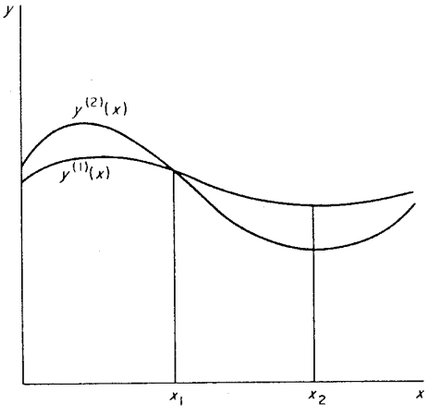
Theoretical Background of OED/SD

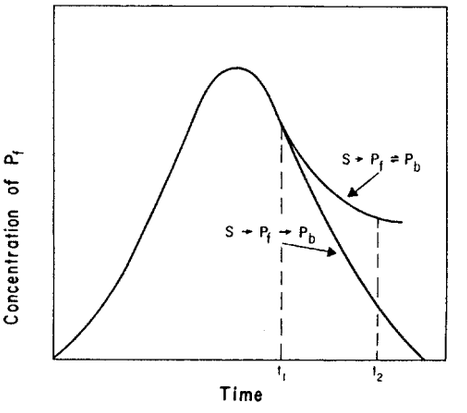
Problem Statement:

Design an Experiment to Discriminate between rival Models M_1 and M_2

Hunter-Reiner Approach (1965)

Choose Experimental Conditions χ_i such that $\left(\hat{M}_1(\chi_i) - \hat{M}_2(\chi_i)\right)^2$ *is max.*

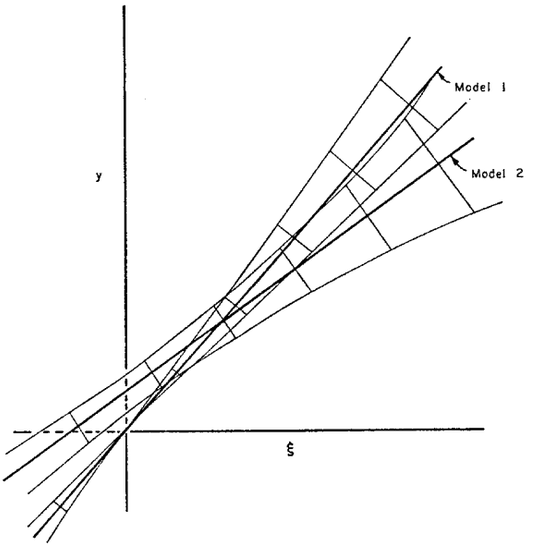




Here: take x_2 for its *max. Discriminative Power*

x_1 will tell us nothing !!

BUT: the *uncertainty on the model predictions* $\hat{M}^{(k)}(\chi_i)$ must be taken into account



Example

Discriminate between

$Y = bX + a$ and $Y = bX$

The Box-Hill Approach (1977)

with n Experiments performed, design the $(n+1)$th Experiment

Divergence between m rival Models $M^{(k)}$ at Experimental Conditions χ_i:

$$D(\chi_i) = \sum_{k=1}^{m-1} \sum_{j=k+1}^{m} \Pi_{k,n}\, \Pi_{j,n} \left\{ \frac{\sigma^2_{ik} - \sigma^2_{ij}}{\left(\sigma^2+\sigma^2_{ik}\right)\left(\sigma^2+\sigma^2_{ij}\right)} - \frac{\left(\hat{M}^{(k)}(\chi_i) - \hat{M}^{(j)}(\chi_i)\right)^2}{\frac{1}{\sigma^2+\sigma^2_{ik}} + \frac{1}{\sigma^2+\sigma^2_{ij}}} \right\}$$

where σ^2_{ik} is given by:

$$\sigma^2_{ik} = V\left(\hat{M}^{(k)}(\chi_i)\right) = \chi_i^T \left(X_k^T X_k\right)^{-1} \chi_i$$

or for nonlinear models approximated by:

$$\sigma^2_{ik} = V\left(\hat{M}^{(k)}(\chi_i)\right) = \left[\frac{\partial M^{(k)}}{\partial \theta}\right]^T_{\chi_i} \left(X_k^T X_k\right)^{-1} \left[\frac{\partial M^{(k)}}{\partial \theta}\right]_{\chi_i}$$

and $\Pi_{k,n}$ is the Probability that Model k is adequate after n Experiments $\Pi_{k,n}$ is calculated according to:

$$\Pi_{k,n} = \frac{p_k\ \Pi_{k,n-1}}{\sum_{j=1}^{m} p_j\ \Pi_{j,n-1}}$$

where p_k is the probability of the nth observation, (y_n) under Model k:

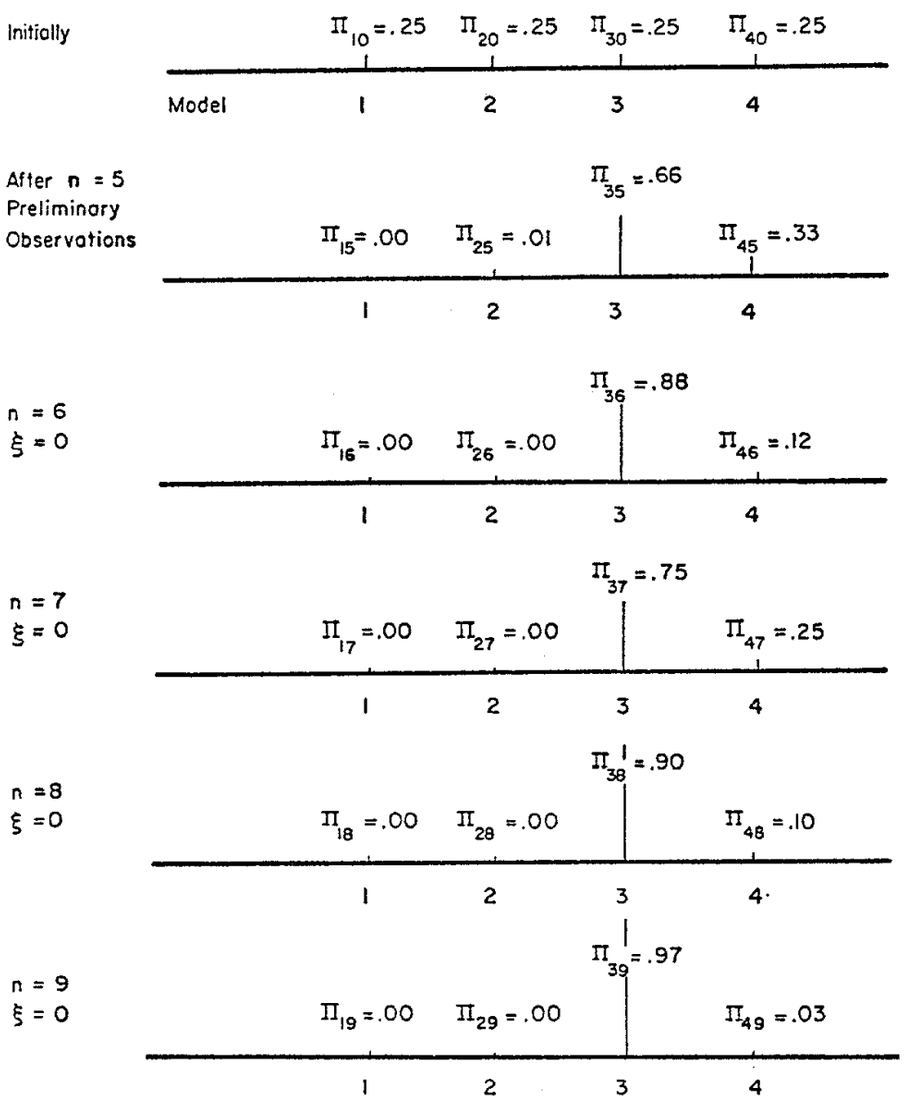
$$p_k = p_k\,(y_n) = \frac{1}{\left[2\pi\left(\sigma^2+\sigma^2_{n,k}\right)\right]^{0.5}} \exp\left[-\frac{1}{2\left(\sigma^2+\sigma^2_{n,k}\right)}\left(y_n - \hat{M}^{(k)}{}_n\right)^2\right]$$

and

$$\Pi_{k,n=0} = \frac{1}{m}$$

Example

- Model 1: $Y_1(x) = \theta_{11}\, x$
- Model 2: $Y_2(x) = \theta_{21} + \theta_{22}\, x$
- Model 3: $Y_3(x) = \theta_{31} + \theta_{32}\, x + \theta_{33}\, x^2$
- Model 4: $Y_4(x) = \theta_{41}\, x + \theta_{42}\, x^2$

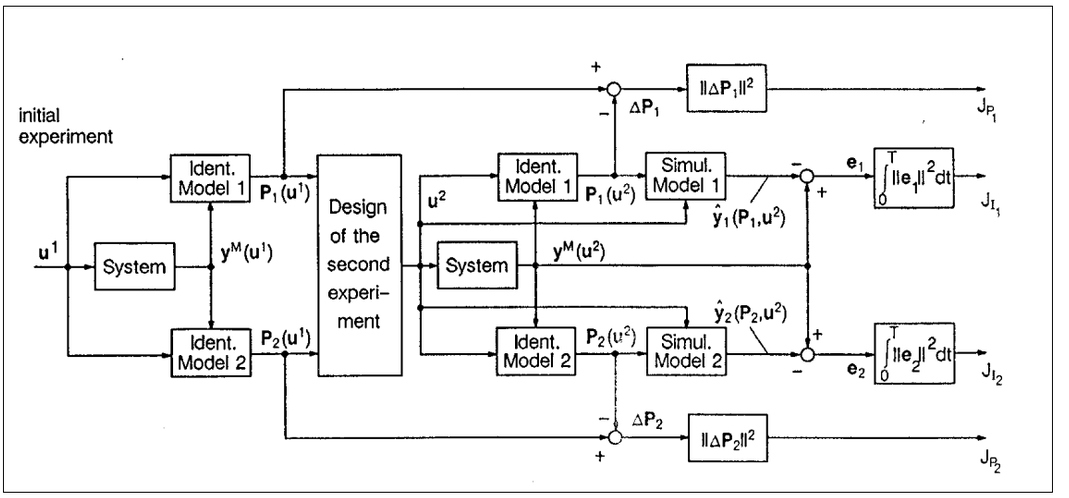


The Munack Approaches (1992)

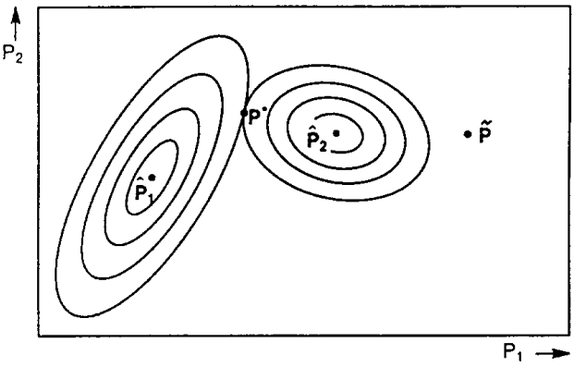
Discrimination between 2 Dynamic Models

Two approaches to discriminate between 2 models:

- Maximize the difference in responses (trajectory)
- Look for dependence of parameters on experimental conditions



Parameter distancewhich takes the estimation accuracy into account is calculated by the Minimal Mahalanobis Distance:



Distance in Trajectory should also take into account that Model complexity should be minimal:

Penalty term for model complexity: AIC= f(Fit)+ g(number of parameters)

Munack Method 1: Maximize Trajectory Difference

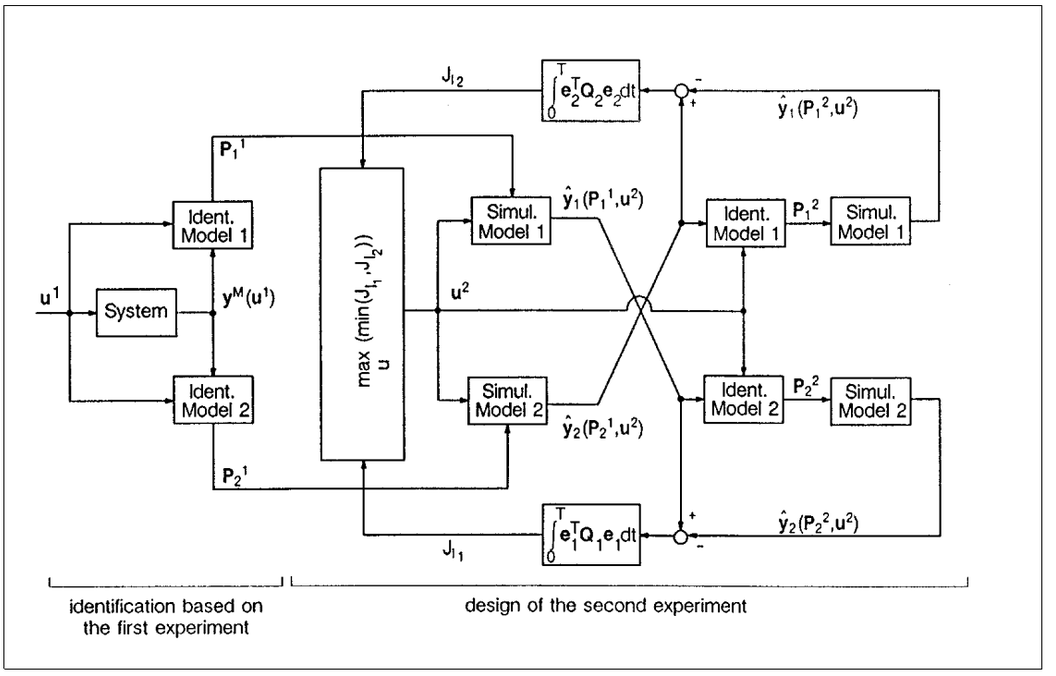
Perform hypothetical experiments through simulation with the 2 models and do the analysis as if these data are originating from the real process.

OPTIMIZATION LOOP

Propose Experiment

1) consider Model 1 correct -> Fit model 1 and 2 to Model 1 generated data
2) consider Model 2 correct -> Fit model 1 and 2 to Model 2 generated data

Calculate Difference between the (simulated) Trajectories
Maximize the smallest of the two Differences obtained



Munack Method 2: Maximize Parameter Change

Perform hypothetical experiments through simulation with the 2 models
and do the analysis as if these data are originating from the real process.

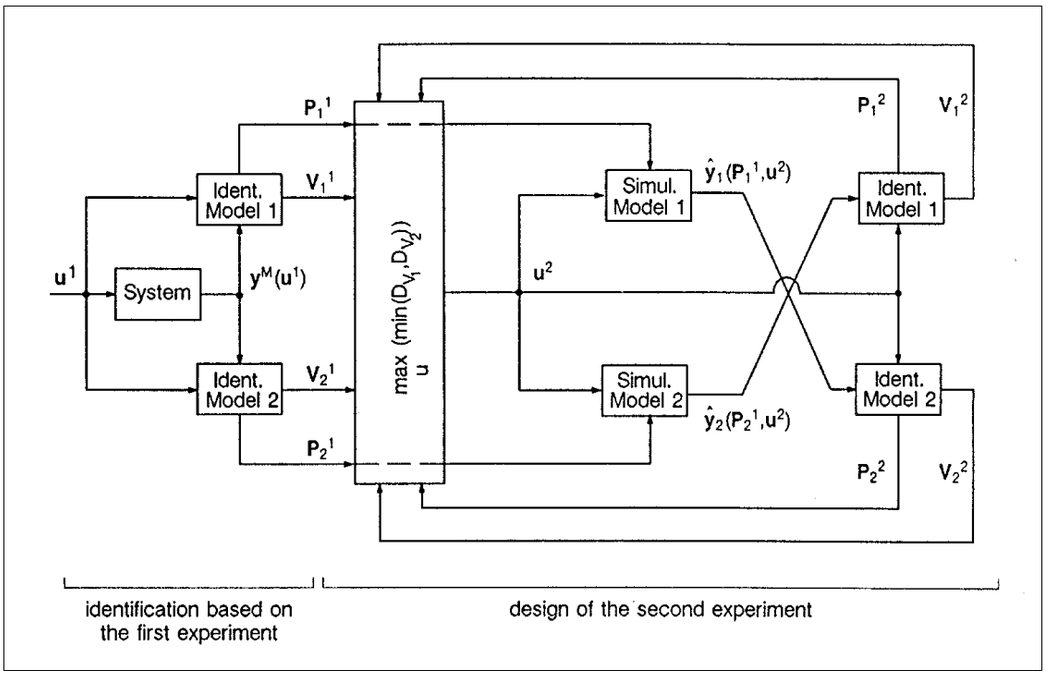
OPTIMIZATION LOOP

Propose Experiment

1) consider Model 1 correct -> Fit model 1 and 2 to Model 1 generated data
2) consider Model 2 correct -> Fit model 1 and 2 to Model 2 generated data

Compare parameter sets for the first (real) experiment
and the second (simulated) experiment

Maximize the smallest of the Mahalanobis Distances



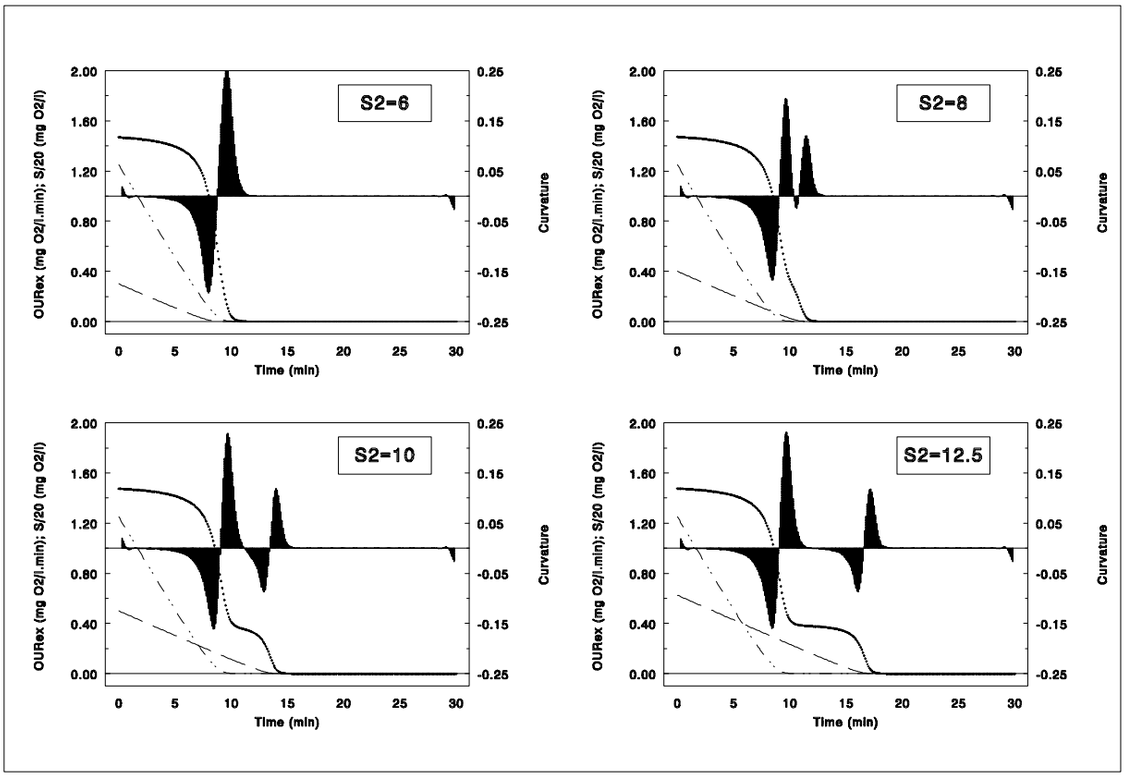
Heuristic Discrimination

Define a criterion function based on experience

Example:

Design mixture of BOD and NH_4Cl to guarantee reliable discrimination between Single Monod and Double Monod models

Objective Function: Reliability of Inflection Point Determination



Dual Problem of Model discrimination/Parameter estimation

Very few results

Standard approach:

- experiments for Model Discrimination
- Experiments for Parameter Estimation

Hill, Hunter and Wichern (1968):

Choose experimental conditions χ_j to maximize:

$$W \frac{D}{D_{\max}} + \left(1 - W\right) \sum_{j=1}^{m} \Pi_{j,n} \frac{E_j}{E_{\max}}$$

where $D_{\max}$ and $E_{\max}$ are the maximum attainable values of D and E_j over the allowed χ-region

$\Pi_{j,n}$ is the probability of model j after n experiments

the weighting W attached to Discrimination is suggested to be:

$$\left(\frac{m}{m-1} \left(1 - \Pi_{b,n}\right) \right)^{\lambda}$$

in which $\Pi_{b,n}$ is the largest of the $\Pi_{j,n}$

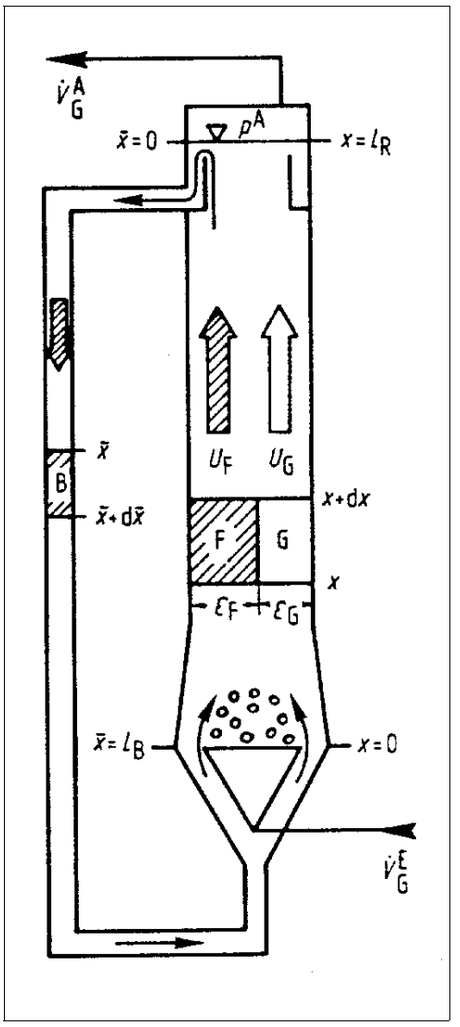
the positive power λ gives the experimenter a flexible control over the rate at which interest shifts from OED/SD to OED/PE

OED/PE for Choice of measured Variable and Sensor Location

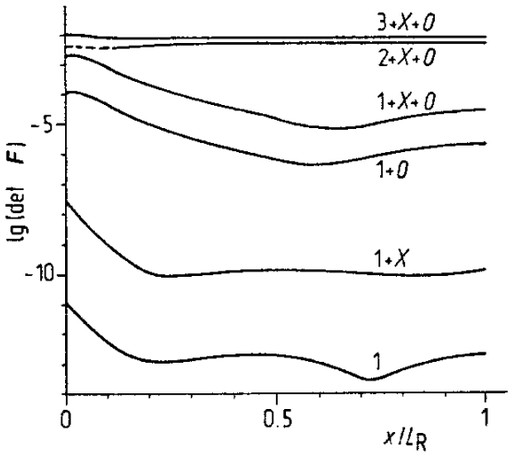
Application: Tower-loop reactor

Available Sensor Technology:

- DO probe
- Off-gas O_2 analysis
- Turbidimeter -> Biomass Concentration



Fisher Information Matrix



Sampling Frequency

Sampling inevitably leads to information losses --> Minimize the Loss

"Sample as frequently as possible"

- Identification to large data-sets may be numerically sensitive
- The high-frequency dynamics may dominate the identification process
- RESAMPLING is always possible, not the other way round !

Constant or Time-varying Sample Interval

- Equidistant sampling is more convenient (automation)
- Changing interval can be advantageous, e.g. when designed
  Example (Vialas et al., 1983): measure when sensitivity to a parameter is high, e.g. at the end of a batch experiment when $S \approx K_m$

Relation to Time Constants of the Process

Sampling Interval $< \frac{\tau_{min}}{10}$;

Experiment Length (==> N) > $10 * \tau_{max}$

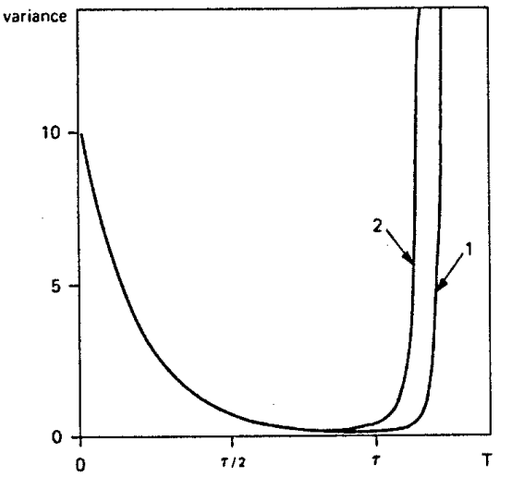
Example

First order system with time constant τ

$$\frac{dx}{dt} = \frac{1}{\tau} x(t) + u(t)$$
$$y(t) = x(t) + v(t)$$

Too high sampling interval is
far worse than too low interval !

at high sampling frequency:
numerical problems
model fits in high-frequency bands
poor return for extra work



Data Pretreatment

Before Model Identification starts

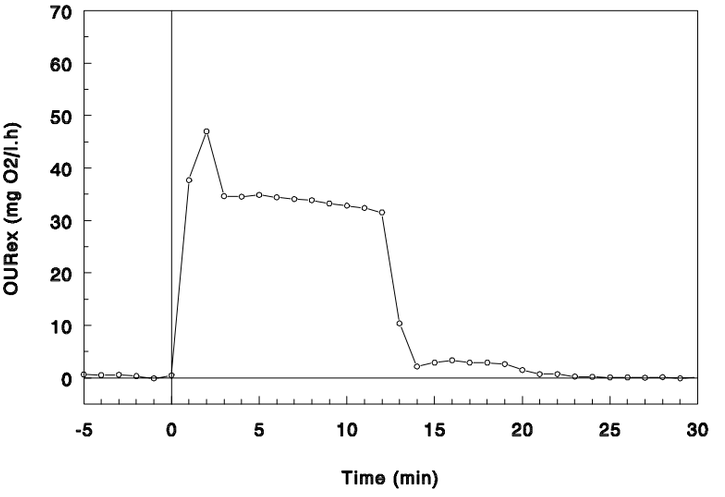
Aim: to increase the quality of the output of the identification stage

Outlier Detection and Removal

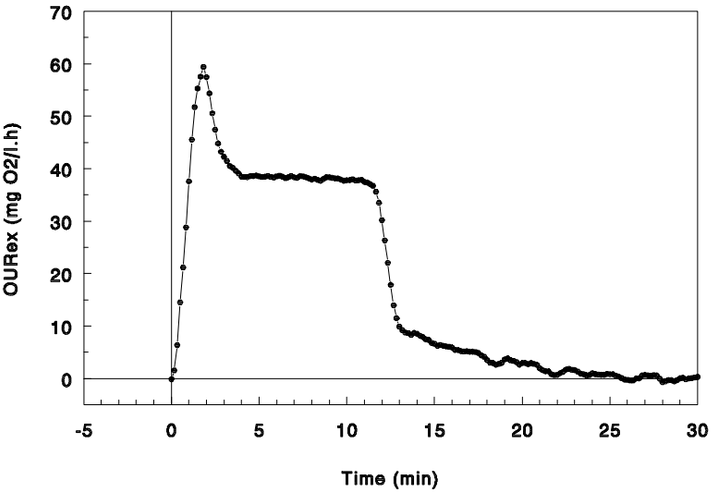
Diagnostic checking : Plot your raw data -> Inspection for deficiencies

Removing data is a kind of cheating and may be dangerous

Example: Respirometer with OUR sampling interval of 1 minute



Respirometer with OUR sampling interval of 10 seconds



Elimination of Uninteresting Dynamics

High-frequency disturbances

occur for instance when sampling rate is too high

-> Solution: Resample the data (picking every sth data point)

Slow disturbances

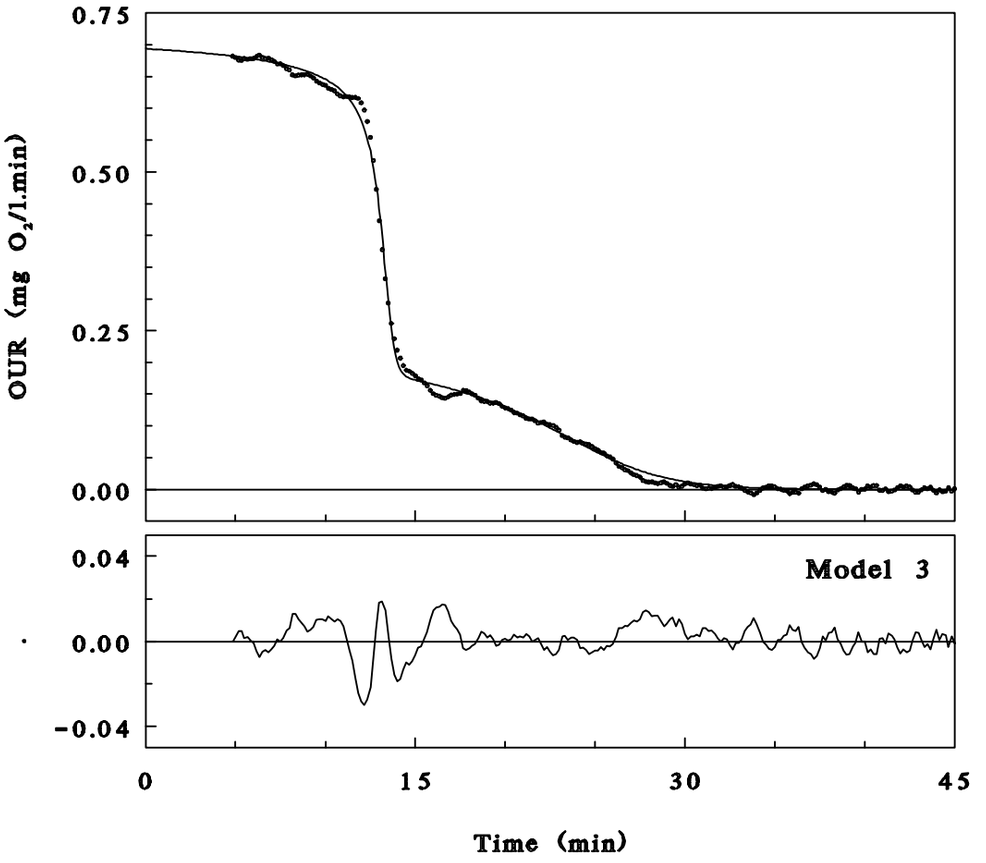
Examples: off-sets, drifts, periodic variations

Typically < external sources that we are not interested in for modelling

e.g. temperature effects (day/night -> conductivity)

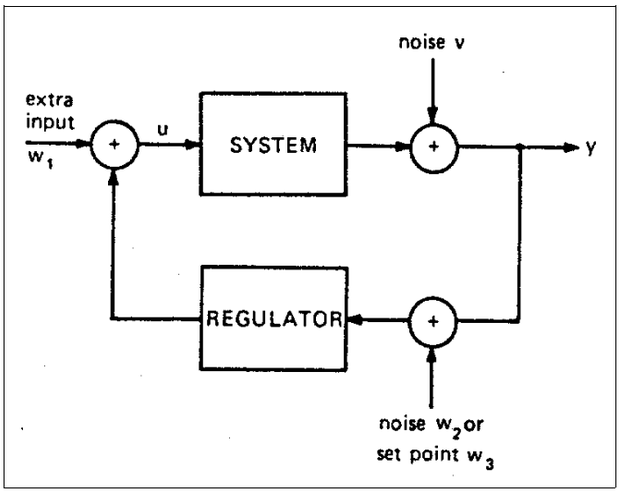
Example:

Respirometer -> OUR-data: Due to the use of derivatives the small oscillation present in the DO-data is emphasized



Experimental Design under Closed-loop Conditions

Hot topic in Control Engineering Research



Illustrative Example:

First order system:

$$y(t) + a\,y(t-1) = b\,u(t-1) + e(t)$$

controlled by a proportional regulator:

$$u(t) = f\,y(t)$$

Closed-loop system dynamics:

$$y(t) + \left(a - b\,f\right) y(t-1) = b\,u(t-1) + e(t)$$

Conclusion:any model parameters subject to:

$$\hat{a} = a + \gamma f$$

$$\hat{b} = b + \gamma$$

will yield adequate description of system dynamics => Non-identifiability !!

Solutions:

- Temporarily open the loop
- Modify the controller, e.g. make f timevarying
  = = > breaking the dependency

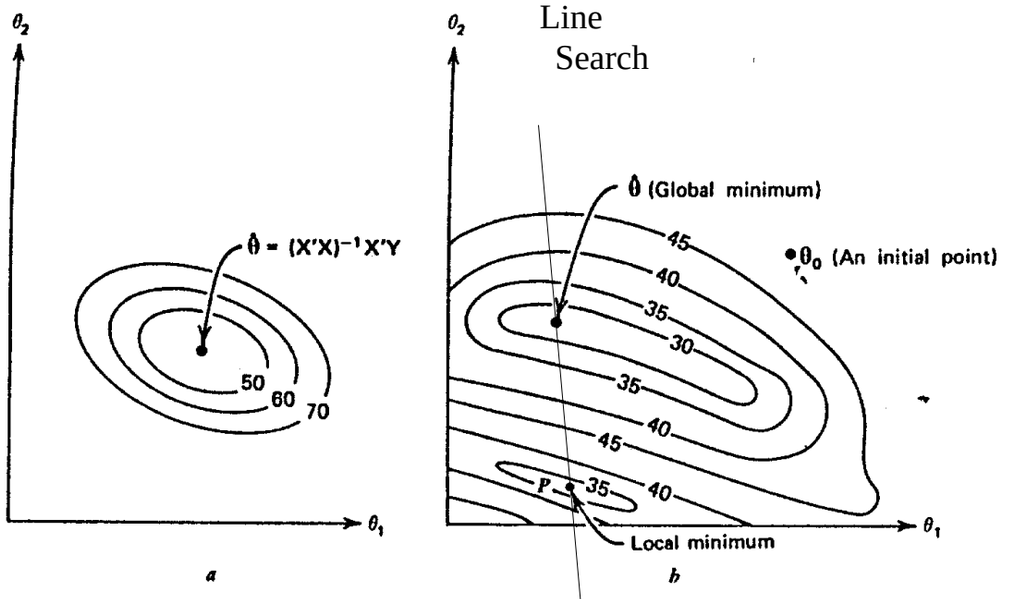
Nonlinear Parameter Estimation without Derivatives

Problem statement:

Algorithms using first and second order derivatives of the objective function to the variables to be optimized (e.g. parameter values or the design criterion)

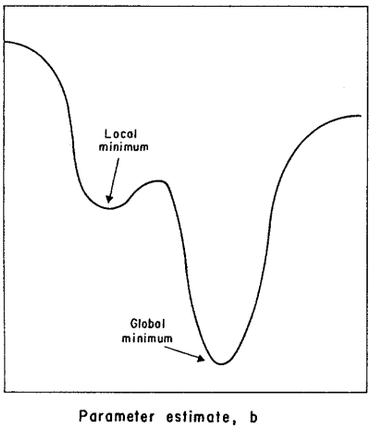
= = > involved mathematics to evaluate sensitivity functions
prone to analytical errors !
numerical approximation --> dangerous due to numerical errors

Nonlinearity in the Parameters: what trouble does it cause?



==> Sensitivity Functions
involve the Parameters

==> Objective Function Shape
depends on Parameter Values



LOCAL MINIMA

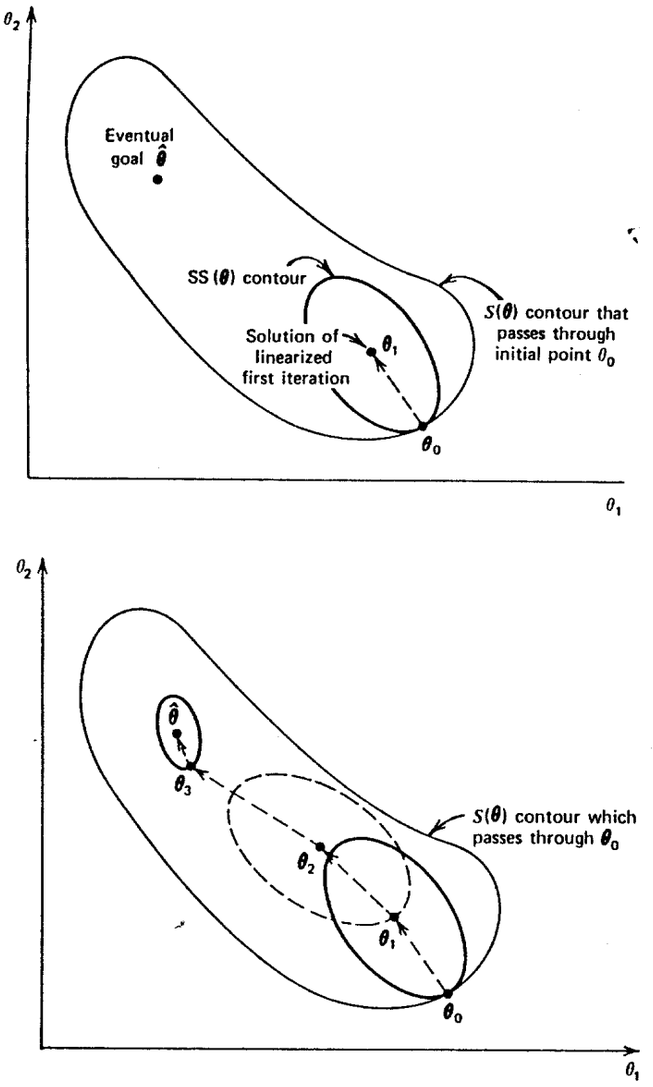
= = > No guaranteed Convergence

Solution:

- Start from widely varying starting values θ_0
- Perturn a local minimum --> Jump away from it

Linear Approximation Approach

- choose an initial value
- calculate linear approximation of the objective function (Jacobian)
- find minimum
- return to step 1



Disadvantage: you easily get stuck in a local minimum

Random Search

"Randomly jump around in Parameter Space"

Exhaustive Search --> VERY INEFFICIENT

Si mplex (Nelder & Mead, 1965)

Not very efficient, but easy to implement

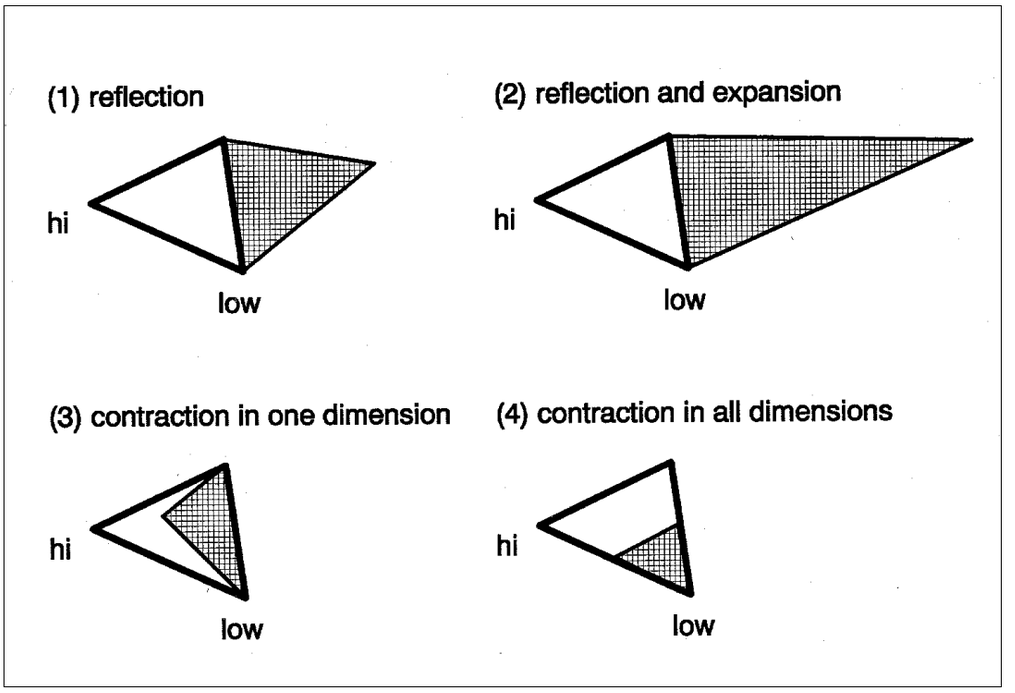
Robust against local minima

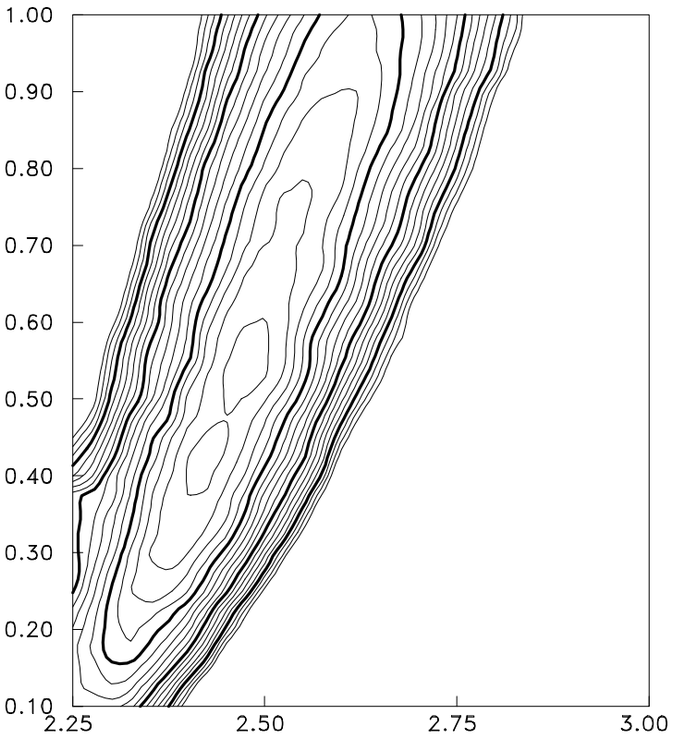
Algorithm:

Choose initial Simplex = Number of (n+1) sets of n parameters $\left(\theta_1\ \theta_2 \ldots \theta_n\right)$

Evaluate the objective function at all simplex points

Replace the worst point by a new one using four basic steps:

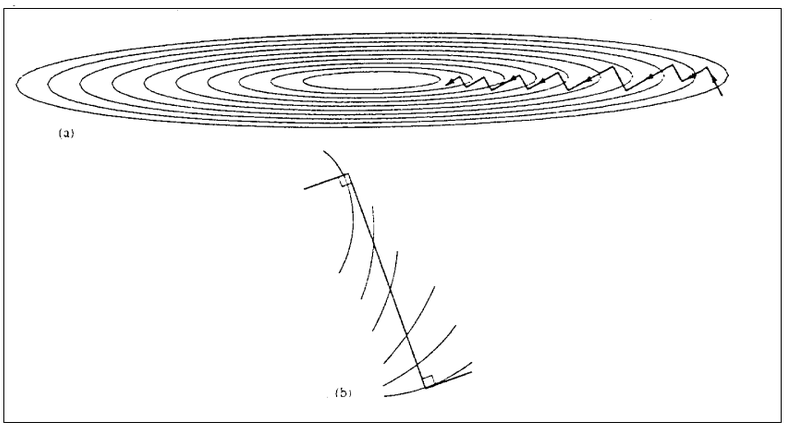



1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
2.25
2.50
2.75
3.00

Direction Set (Powell, 1964; Brent, 1973)

Steepest Decent:

Direction of parameter adjustment is based on the gradient information

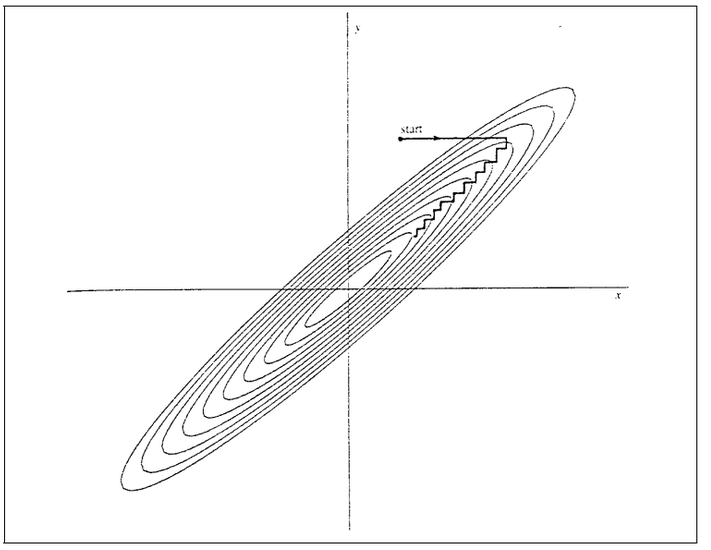


If no gradients are calculated (derivative-free algorithms)

--> propose a set of search directions to be modified as minimization proceeds

1) No modification of directions

Direction Set = Unit Vectors $\left[e_1\, e_2 \ldots e_n \right]$ with $e_i = \left(\, 0\,0\,0 \ldots 0\,1\,0 \ldots 0 \, \right)$



Line minimization along successive directions, cycling through
the whole set of directions until objective function stops decreasing

Powell/Brent's Conjugate Directions

Find the directions $\left[u_1 u_2 \ldots u_n \right]$

where the improvement in objective function is maximal

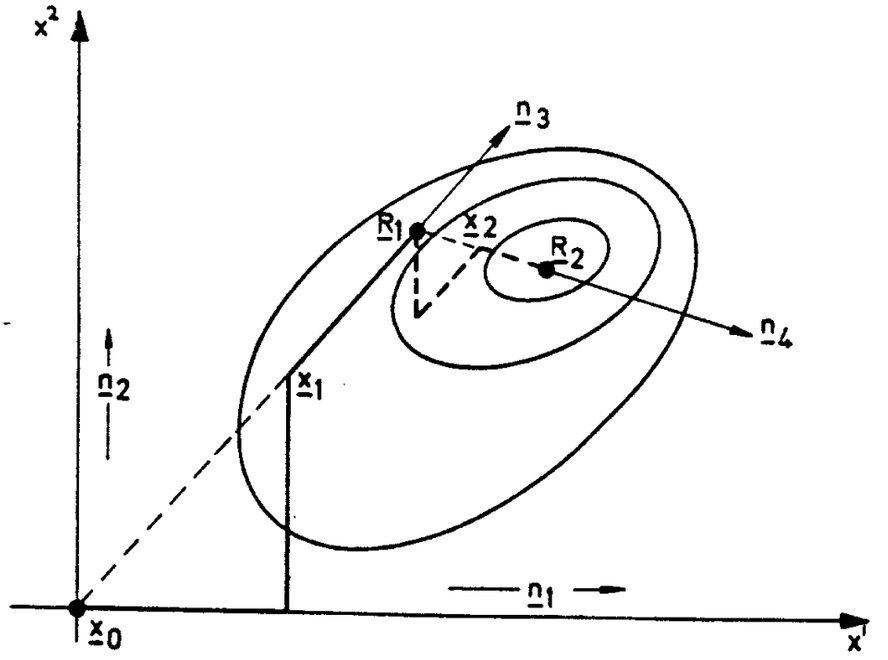
But: After a line minimization along a direction u_i

= = > This direction is not interesting anymore (gradient = 0 !!)

= = > Another direction is to be chosen

<u>General procedure:</u>

- Save starting position θ_0
- Perform n minimizations along the directions $\left[u_1 u_2 \ldots u_n \right]$ --> θ_n
- discard direction u_1 and rename u_i to u_{i-1}, i=2,n
- $u_n = \theta_n - \theta_0$
- move θ_n to minimum along u_n --> new θ_0



Dealing with Constraints

e.g. Affinity constants > 0; Experiment Length < 2 hr

Easiest way --> Penalty added to the Objective Functional
may lead to convergence problems...