

- We have been constructing and explaining the fundamental principles governing chemical reactor behavior.
- Drawing quantitative conclusions from these principles requires not only the models, however, but the values of the parameters in the models as well.
- In most practical applications, the model parameters are not known.
- Evaluating the model parameters therefore usually requires something outside the scope of the theory: experimental data.
- In this chapter we consider the issues involved in gathering data and estimating parameters from data with the expressed purpose of identifying a chemical reactor model.

## Experimental Methods

Please read section 9.1 on your own.

- We now turn to the issue of how we extract the information contained in the data.
- The two large questions we must address are: what model *structure* is appropriate to describe the reacting system of interest
- Having selected a structure, what model *parameters* best represent the data we have collected, and how certain are we about these parameter values.
- The first question has occupied us up to this point; it has been the central focus of Chapters 1–8.
- In this section, we focus finally on the second question, how to estimate model parameters from data.

## Some history on parameter estimation

- Parameter estimation has a distinguished place in the history of science and engineering.
- Accurate prediction of the motions of the planets based on astronomical measurements was one of the early motivating problems of parameter estimation.
- Solving this problem led Gauss to invent the least-squares method in the late 1700s.

Gauss's summary of this effort more than 175 years ago remains valid today:

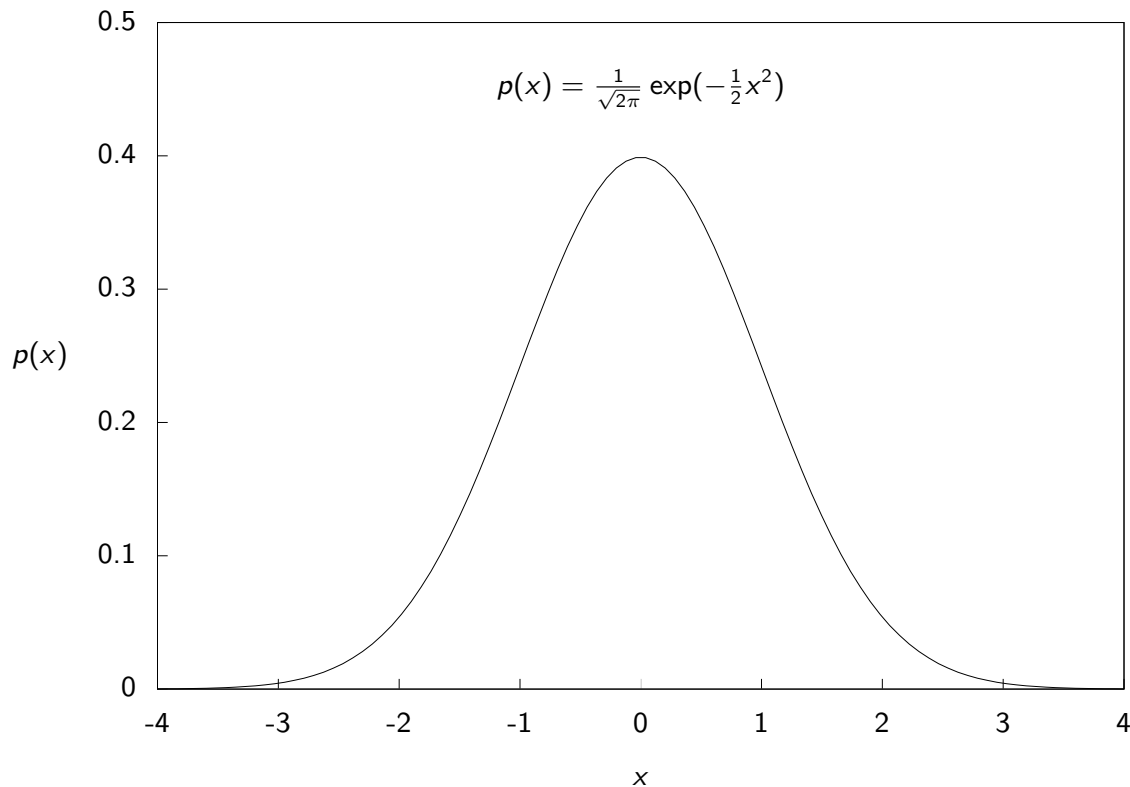
*One of the most important problems in the application of mathematics to the natural sciences is to choose the best of these many combinations, i.e., the combination that yields values of the unknowns that are least subject to errors. Theory of the Combination of Observations Least Subject to Errors. C. F. Gauss, 1821 [6].*

## Review of the Normal Distribution

- Probability and statistics provide one useful set of tools to model the uncertainty in experimental data.
- It is appropriate to start with a brief review of the normal distribution, which plays a central role in analyzing data.
- The normal or Gaussian distribution is ubiquitous in applications. It is characterized by its mean,  $m$ , and variance,  $\sigma^2$ , and is shown in the following figure

## Normal Distribution

The univariate normal with mean zero and unit variance.



7 / 116

## Normal Distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}\right) \quad (9.6)$$

We adopt the following notation to write Equation 9.6 more compactly

$$x \sim N(m, \sigma^2)$$

which is read “the random variable  $x$  is distributed as a normal with mean  $m$  and variance  $\sigma^2$ .” Equivalently, the probability density  $p(x)$  for random variable  $x$  is given by Equation 9.6.

8 / 116

- For distributions in more than one variable, we let  $\mathbf{x}$  be an  $n_p$ -vector and the generalization of the normal is

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n_p/2} |\mathbf{P}|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{m})^T \mathbf{P}^{-1} (\mathbf{x} - \mathbf{m}) \right]$$

in which the  $n_p$ -vector  $\mathbf{m}$  is the mean and the  $n_p \times n_p$ -matrix  $\mathbf{P}$  is called the covariance matrix. The notation  $|\mathbf{P}|$  denotes determinant of  $\mathbf{P}$ .

- We also write for the random variable  $\mathbf{x}$  vector

$$\mathbf{x} \sim N(\mathbf{m}, \mathbf{P})$$

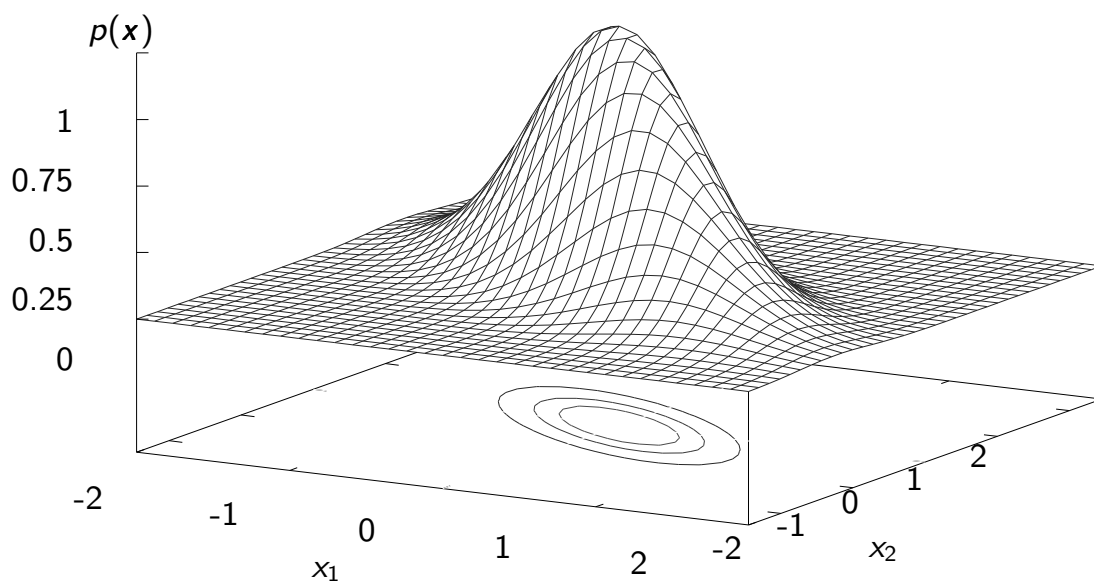
- The matrix  $\mathbf{P}$  is a real, symmetric matrix. The next figure displays a multivariate normal for

$$\mathbf{P}^{-1} = \begin{bmatrix} 3.5 & 2.5 \\ 2.5 & 4.0 \end{bmatrix}$$

9 / 116

## Multivariate normal

$$p(\mathbf{x}) = \exp \left( -1/2 (3.5x_1^2 + 2(2.5)x_1x_2 + 4.0x_2^2) \right)$$



- As shown in the figure, lines of constant probability in the multivariate normal are lines of constant

$$(\mathbf{x} - \mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x} - \mathbf{m})$$

- To understand the geometry of lines of constant probability (ellipses in two dimensions, ellipsoids or hyper-ellipsoids in three or more dimensions) we examine the eigenvalues and eigenvectors of the  $\mathbf{P}$  matrix.

## Eigenvalues and Eigenvectors

- An eigenvector of a matrix  $\mathbf{A}$  is a nonzero vector  $\mathbf{v}$  such that when multiplied by  $\mathbf{A}$ , the resulting vector points in the same direction as  $\mathbf{v}$ , and only its magnitude is rescaled.
- The rescaling factor is known as the corresponding eigenvalue  $\lambda$  of  $\mathbf{A}$ .
- Therefore the eigenvalues and eigenvectors satisfy

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}$$

- We normalize the eigenvectors so

$$\mathbf{v}^T \mathbf{v} = \sum_i v_i^2 = 1$$

# Eigenvectors of the Covariance Matrix

- The eigenvectors show us the orientation of the ellipse given by the normal distribution.
- Consider the ellipse in the two-dimensional  $\mathbf{x}$  coordinates given by the quadratic

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = b$$

- If we march along a vector  $\mathbf{x}$  pointing in the eigenvector  $\mathbf{v}$  direction, we calculate how far we can go in this direction until we hit the ellipse  $\mathbf{x}^T \mathbf{A} \mathbf{x} = b$ .
- Substituting  $\alpha \mathbf{v}$  for  $\mathbf{x}$  in this expression yields

$$(\alpha \mathbf{v}^T) \mathbf{A} (\alpha \mathbf{v}) = b$$

- Using the fact that  $\mathbf{A} \mathbf{v} = \lambda \mathbf{v}$  for the eigenvector gives

$$\alpha^2 \lambda \mathbf{v}^T \mathbf{v} = b$$

because the eigenvectors are of unit length, we solve for  $\alpha$  and obtain

$$\alpha = \sqrt{\frac{b}{\lambda}}$$

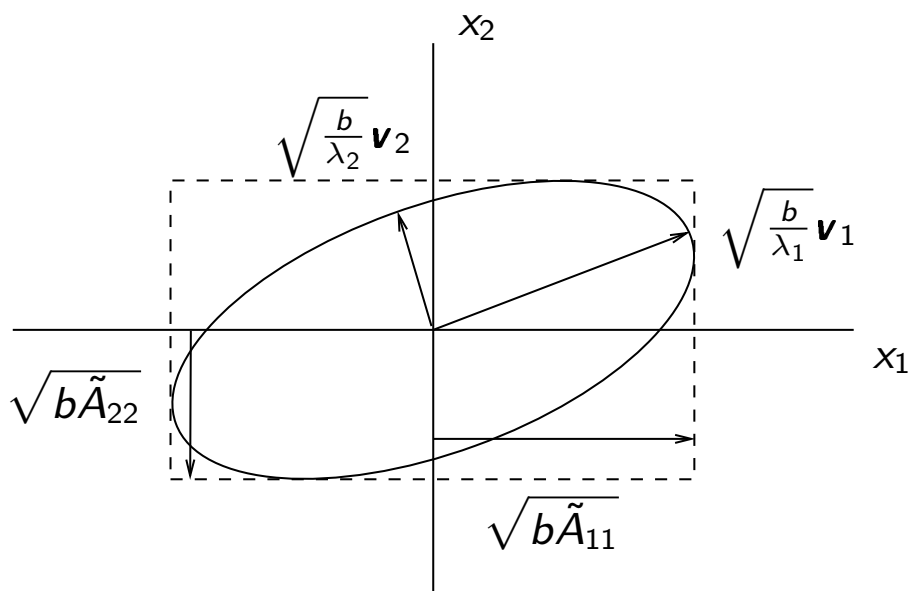
which is shown in the next figure

13 / 116

## The Geometry of $\mathbf{x}^T \mathbf{A} \mathbf{x} = b$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = b$$

$$\mathbf{A} \mathbf{v}_i = \lambda_i \mathbf{v}_i$$



14 / 116

- Each eigenvector of  $\mathbf{A}$  points along one of the axes of the ellipse. The eigenvalues show us how stretched the ellipse is in each eigenvector direction.
- If we want to put simple bounds on the ellipse, then we draw a box around it as shown in the figure
- Notice the box contains much more area than the corresponding ellipse and we have lost the correlation between the elements of  $\mathbf{x}$ . This loss of information means we can put different tangent ellipses of quite different shapes inside the same box.
- The size of the bounding box is given by

$$\text{length of } i\text{th side} = \sqrt{b\tilde{A}_{ii}}$$

in which

$$\tilde{A}_{ii} = (i, i) \text{ element of } \mathbf{A}^{-1}$$

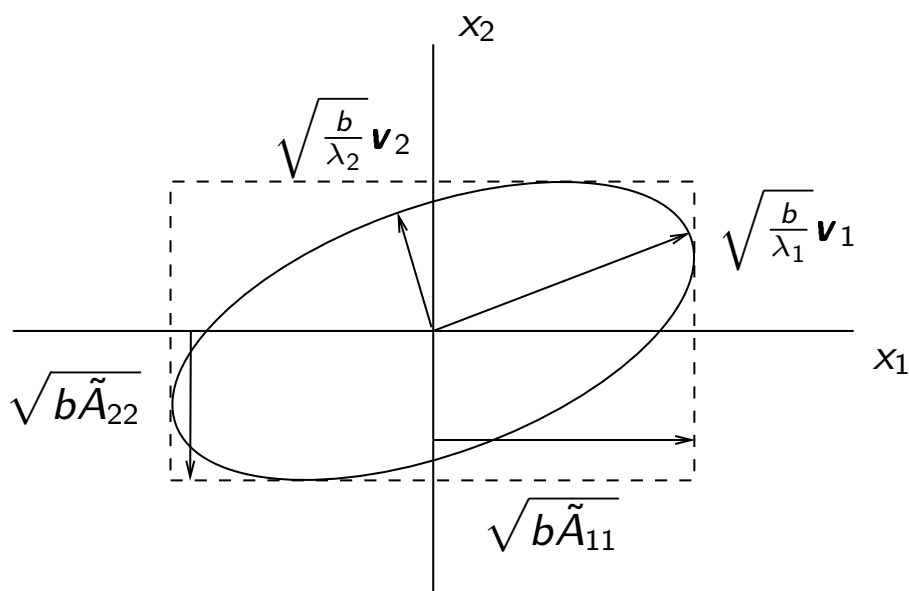
15 / 116

## Summary of the Quadratic Form $\mathbf{x}^T \mathbf{A} \mathbf{x} = b$

- The eigenvectors are aligned with the ellipse axes and the eigenvalues scale the lengths.
- The lengths of the sides of the box that is tangent to the ellipse are proportional to the square root of the diagonal elements of  $\mathbf{A}^{-1}$ .

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = b$$

$$\mathbf{A} \mathbf{v}_i = \lambda_i \mathbf{v}_i$$



16 / 116



- Consider the problem of fitting a straight line to data

$$y_i = mx_i + b$$

- $y_i$  is the measurement at  $x_i$ ,  $i = 1, \dots, n_d$  and  $n_d$  is the number of data points.
- Using matrix vector notation, we can write the equation for all the data as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}$$

17 / 116

## Least-Squares Estimation in Matrix Notation

- The parameters to be estimated are placed in the  $\boldsymbol{\theta}$  vector

$$\boldsymbol{\theta} = \begin{bmatrix} m \\ b \end{bmatrix}$$

- The  $\mathbf{y}$  vector and  $\mathbf{X}$  matrix are given by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_d} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_{n_d} & 1 \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} m \\ b \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}$$

18 / 116

- We do not expect the best fit line to pass through all the data points, so we modify the model to account for measurement error

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e} \quad (9.7)$$

in which  $\mathbf{e}$  is a random variable.

- We model the measurement error as a normal distribution with mean  $\mathbf{0}$  and variance  $\sigma^2$ .

$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (9.8)$$

- The best estimate of  $\boldsymbol{\theta}$  in a least-squares sense is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (9.9)$$

a formula that you have used often.

- However, we also can examine the distribution of parameter values given the observed measurements corrupted by the measurement errors.

## Translating Measurement Error to Estimated Parameter Error

- Imagine we create replicate datasets by drawing measurement errors  $\mathbf{e}$  from the distribution given in Equation 9.8.
- For each dataset we apply Equation 9.9 and produce a parameter estimate. The distribution of measurement errors creates a distribution of parameter estimates.
- In fact, for models linear in the parameters, we can show the parameter estimates also are normally distributed

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \mathbf{P})$$

in which the mean is the true value of the parameters and the covariance is

$$\mathbf{P} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

- We also can calculate the parameter “confidence intervals.” We merely compute the size of the ellipse containing a given probability of the multivariate normal.
- That can be shown to be the chi-squared probability function [3].
- Given the number of estimated parameters,  $n_p$ , and the confidence level,  $\alpha$ , then

$$\frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{\sigma^2} \leq \chi^2(n_p, \alpha) \quad (9.10)$$

- The  $\chi^2$  distribution is tabulated in many statistics handbooks [2] and is available in many computing environments.

## A Classic Kinetics Example

### Example 9.1: Estimating the rate constant and activation energy

- To illustrate the ideas we examine a classic problem: how to estimate the preexponential factor and activation energy of a rate constant.
- Assume a reaction rate has been measured at several different temperatures in the range  $300 \text{ K} \leq T \leq 500 \text{ K}$ .
- Estimate the preexponential factor and activation energy of the rate constant, and quantify your uncertainty in the estimated parameters.

□

- We first model the rate (rate constant) as

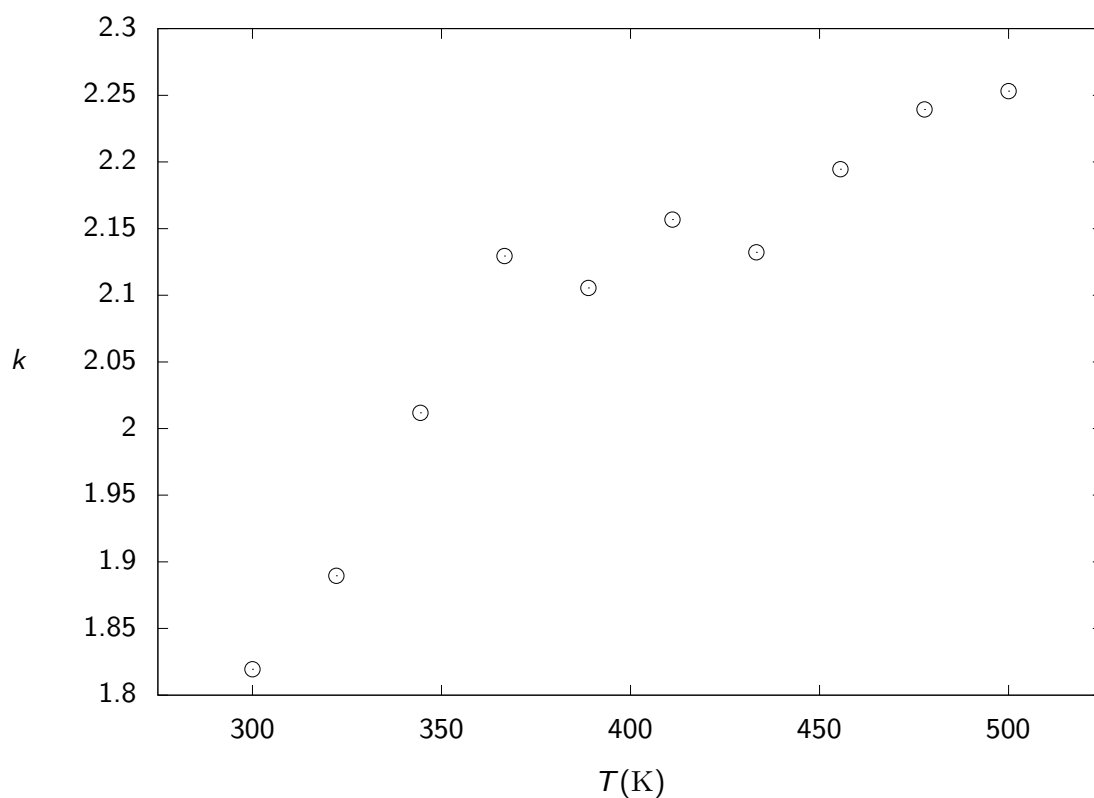
$$k = k_0 \exp(-E/T) \quad (9.11)$$

in which  $k_0$  is the preexponential factor and  $E$  is the activation energy, scaled by the gas constant.

- Here is a typical experimental measurement

23 / 116

## Experimental Data



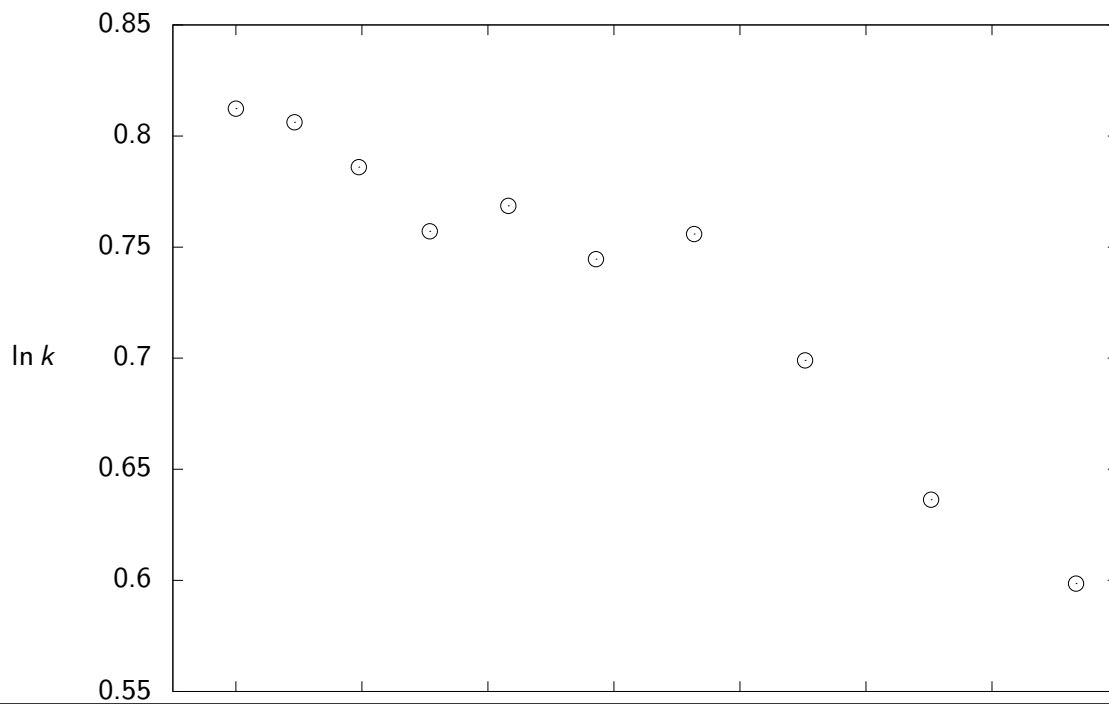
24 / 116

## Logarithmic Transformation

- Notice that the measurement of the rate constant is somewhat noisy, a likely outcome if we differentiate the concentration data to obtain the rate.
- To make the estimation problem linear, we transform the data by taking the logarithm of Equation 9.11

$$\ln k = \ln k_0 - E/T$$

- The transformed data are shown here



25 / 116

## Least Squares Parameter Estimates

- To generate these data we assume the model is correct and suppress the units of the parameter values

$$\ln k_0 = 1, \quad E = 100$$

The measurements of  $\ln k$  are corrupted with normally distributed errors having variance 0.001,

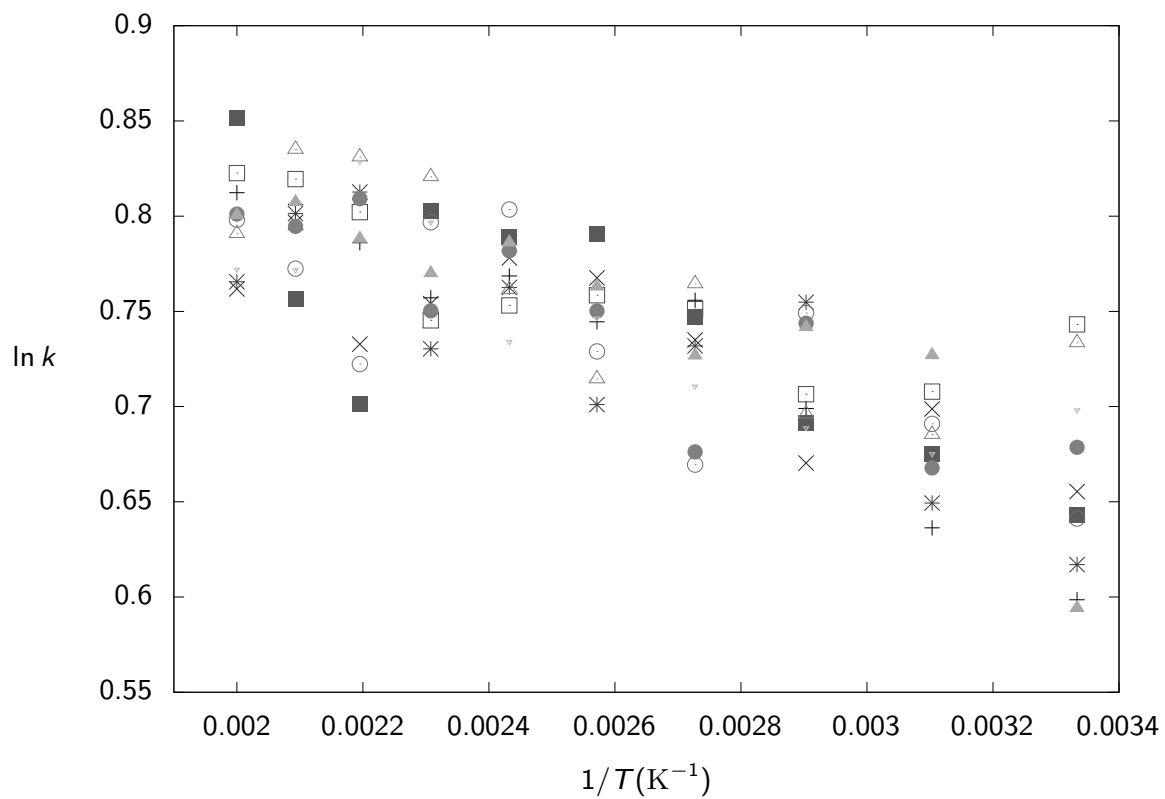
$$e \sim N(0, 0.001)$$

- We apply Equation 9.9 to estimate the parameter using the transformed model, so  $x_i = 1/T_i$  and  $y_i = \ln k_i$ .

$$\hat{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

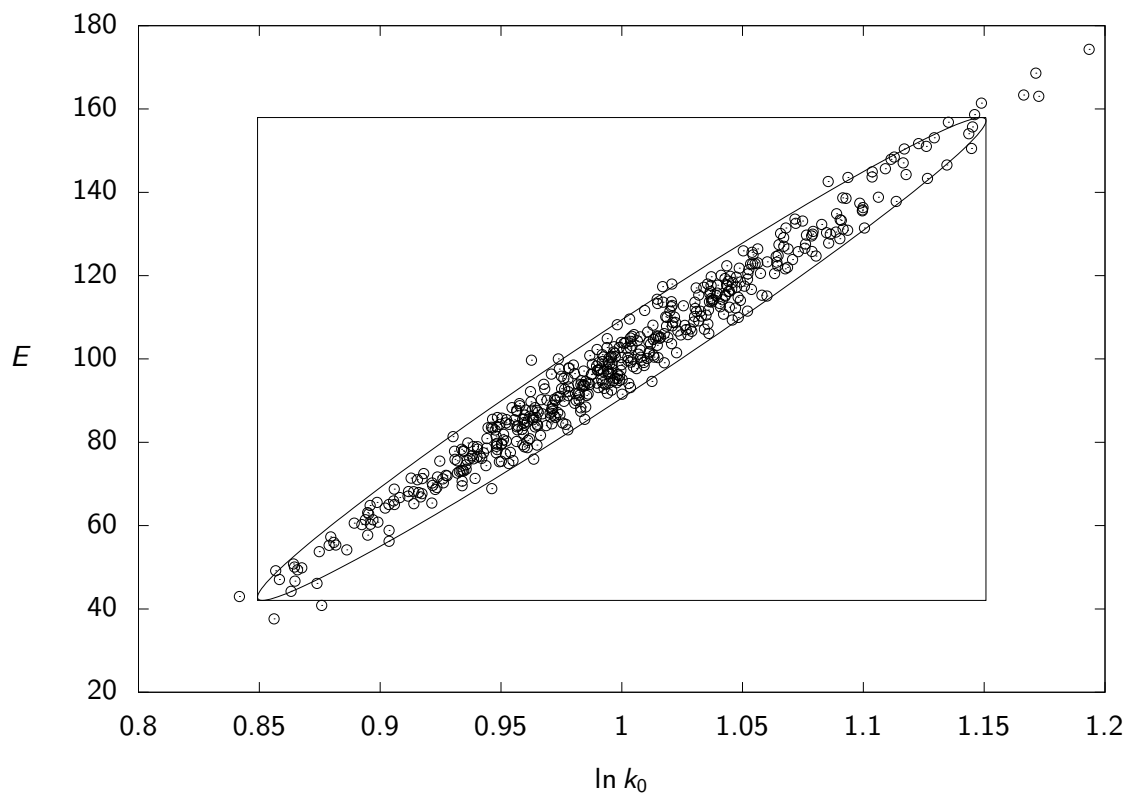
26 / 116

- To quantify the uncertainty, imagine we replicate the experiment



- Each experiment that we perform allows us to estimate the slope and intercept.
- Then we can plot the distribution of parameters.

- This figure shows the parameter estimates for 500 replicated experiments.



29 / 116

## Quantifying the Uncertainty

- Notice the points are clustering in an elliptical shape.
- We construct the 95% confidence interval from Equation 9.10.

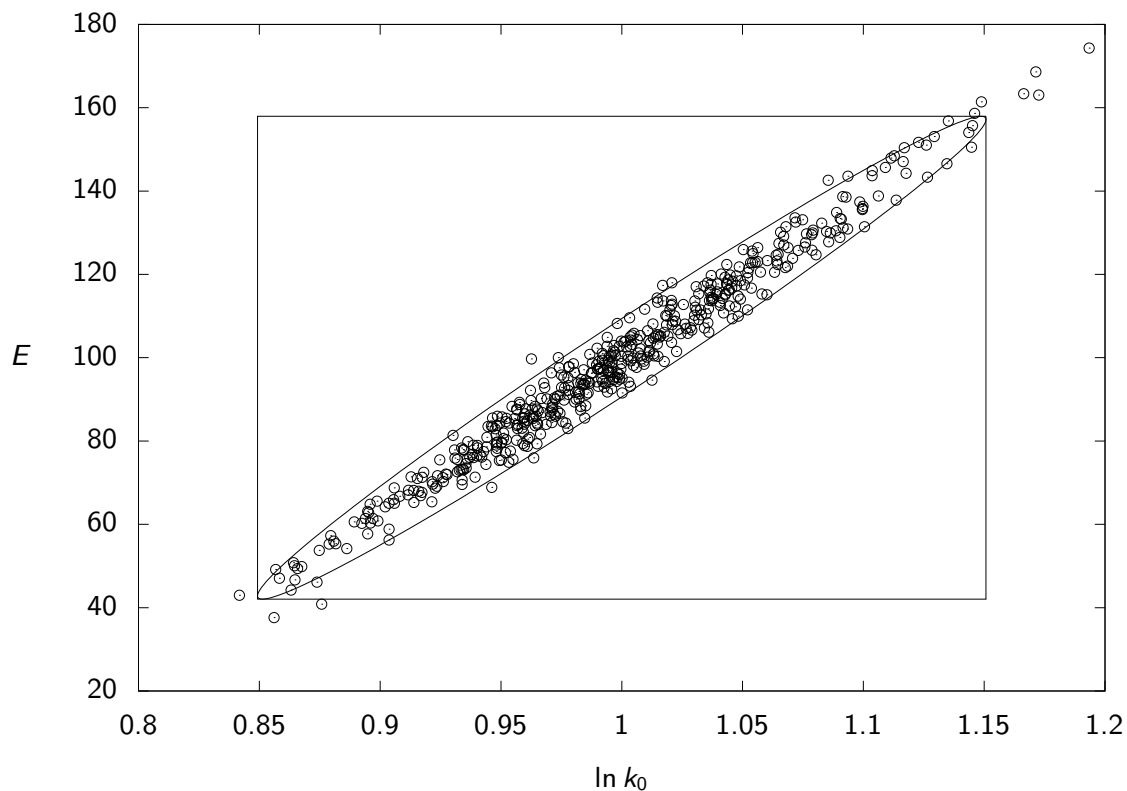
$$\frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{\sigma^2} \leq \chi^2(n_p, \alpha)$$

- In this problem  $n_p = 2$  and  $\alpha = 0.95$ , and we obtain from a statistics table  $\chi^2(2, 0.95) = 5.99$ , so we plot

$$\frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{0.001} \leq 5.99 \quad (9.12)$$

30 / 116

- The ellipse is also shown in the figure.



31 / 116

## Verifying our Confidence Intervals by Monte Carlo

- In fact, 24 out of the 500 points, or 4.8% of the estimated parameters, lie outside the 95% confidence ellipse, which indicates Equation 9.12 is fairly accurate with this many random experimental trials.
- One often sees parameters reported with plus/minus limits. For this problem, one might report

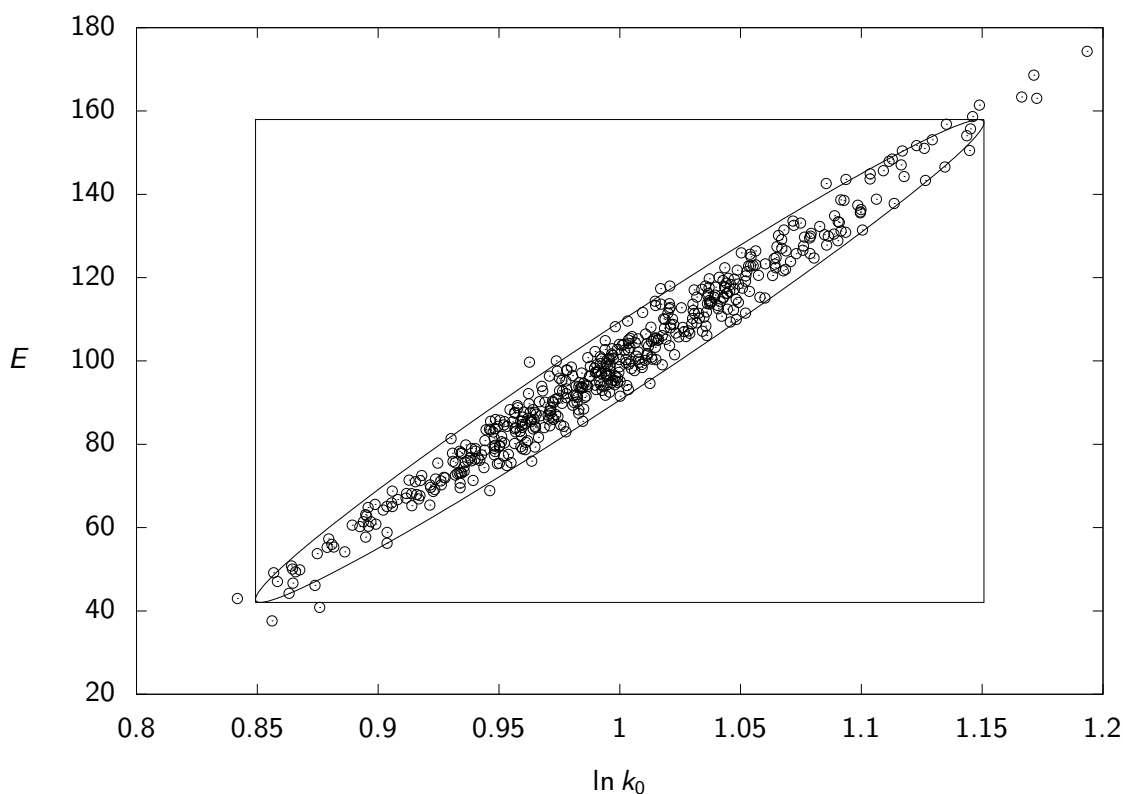
$$\begin{bmatrix} \ln k_0 \\ E \end{bmatrix} = \begin{bmatrix} 1 \\ 100 \end{bmatrix} \pm \begin{bmatrix} 0.15 \\ 60 \end{bmatrix}$$

- Notice these limits are misleading. The rectangle in the figure does *not* indicate the strong correlation between the parameters. The ellipse is much more informative in this case.

32 / 116



- The box does not show the strong parameter correlation



33 / 116

## Reparameterization of the Rate Constant

- Next we show how a simple reparameterization of the rate constant can reduce the parameter correlation.
- Consider the mean of the temperatures at which data were collected, and reparameterize the rate constant as in Chapter 6

$$k = k_m \exp(-E(1/T - 1/T_m)) \quad (9.13)$$

- We may wish to consider the mean of the temperature or the inverse temperature for  $T_m$ . Here we let  $T_m$  be the mean temperature

$$T_m = \frac{300 + 500}{2} = 400 \text{ K}$$

- Both Equations 9.11 and 9.13 are two parameter models

$$k = k_0 \exp(-E/T) \quad k = k_m \exp(-E(1/T - 1/T_m))$$

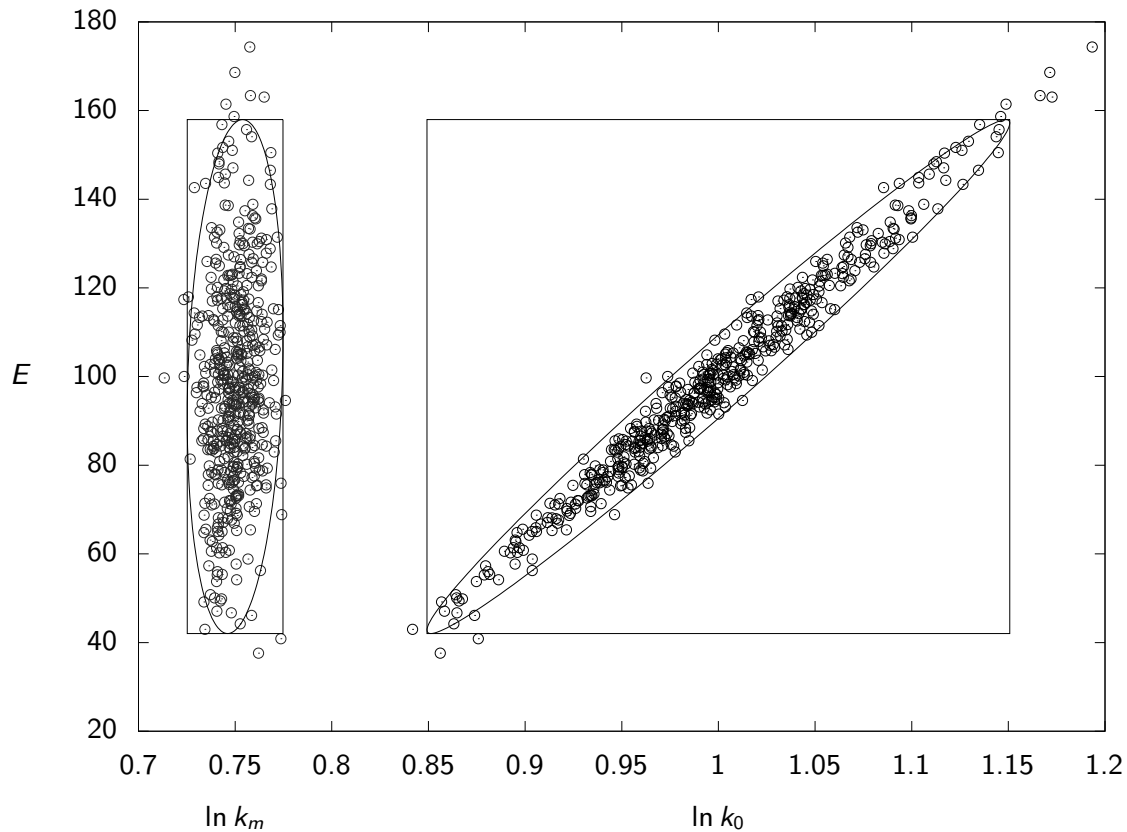
- We can convert between them using

$$k_0 = k_m \exp(E/T_m)$$

- But in Equation 9.13 we estimate the rate constant at the mean temperature in contrast to infinite temperature ( $1/T = 0$ ) in Equation 9.11.

34 / 116

If we estimate  $E$  and  $k_m$  in place of  $k_0$  we obtain the results shown in the figure.



35 / 116

- Notice that the correlation between  $k_m$  and  $E$  is much reduced compared to  $k_0$  and  $E$ .
- In fact, reporting confidence limits

$$\begin{bmatrix} \ln k_m \\ E \end{bmatrix} = \begin{bmatrix} 0.75 \\ 100 \end{bmatrix} \pm \begin{bmatrix} 0.05 \\ 60 \end{bmatrix}$$

is an accurate representation of the true 95% confidence interval ellipse.

36 / 116

- In the previous problem, we assumed the variance in the error was known to us. It is often the case that we do not know the measurement error variance, but must estimate it also from the data.
- In this case, it can be shown that the distribution of parameter estimates is a multivariate  $t$ -distribution (instead of a normal distribution) using the same least-squares estimate as before

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (9.14)$$

37 / 116

## Unknown Variance — Confidence Intervals

- For the confidence intervals, we merely compute the size of the ellipse containing a given probability of the multivariate  $t$ -distribution.
- That can be shown to be an  $F$  probability function [3].
- Given the number of estimated parameters,  $n_p$ , the confidence level,  $\alpha$ , and the number of data points,  $n_d$ ,

$$\frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{s^2} \leq n_p F(n_p, n_d - n_p, \alpha) \quad (9.15)$$

defines the confidence interval ellipse.

- The  $F$  distribution is also tabulated in statistics handbooks [2] and available in computing environments.

38 / 116

- The sample variance

$$s^2 = \frac{1}{n_d - n_p} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}}) \quad (9.16)$$

is the estimate of the unknown error variance.

- Notice the number of data points,  $n_d$ , shows up in the confidence interval when the error variance is unknown.

39 / 116

## Large Number of Data Points

- In the limit of large  $n_d$ , the  $F$  distribution converges to the  $\chi^2$

$$\lim_{n_d \rightarrow \infty} n_p F(n_p, n_d - n_p, \alpha) = \chi^2(n_p, \alpha)$$

and the confidence intervals given in Equations 9.10 and 9.15 are the same.

- The sample variance also converges to the error variance in the limit of large number of data points.

$$\frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{\sigma^2} \leq \chi^2(n_p, \alpha)$$
$$\frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{s^2} \leq n_p F(n_p, n_d - n_p, \alpha)$$

40 / 116

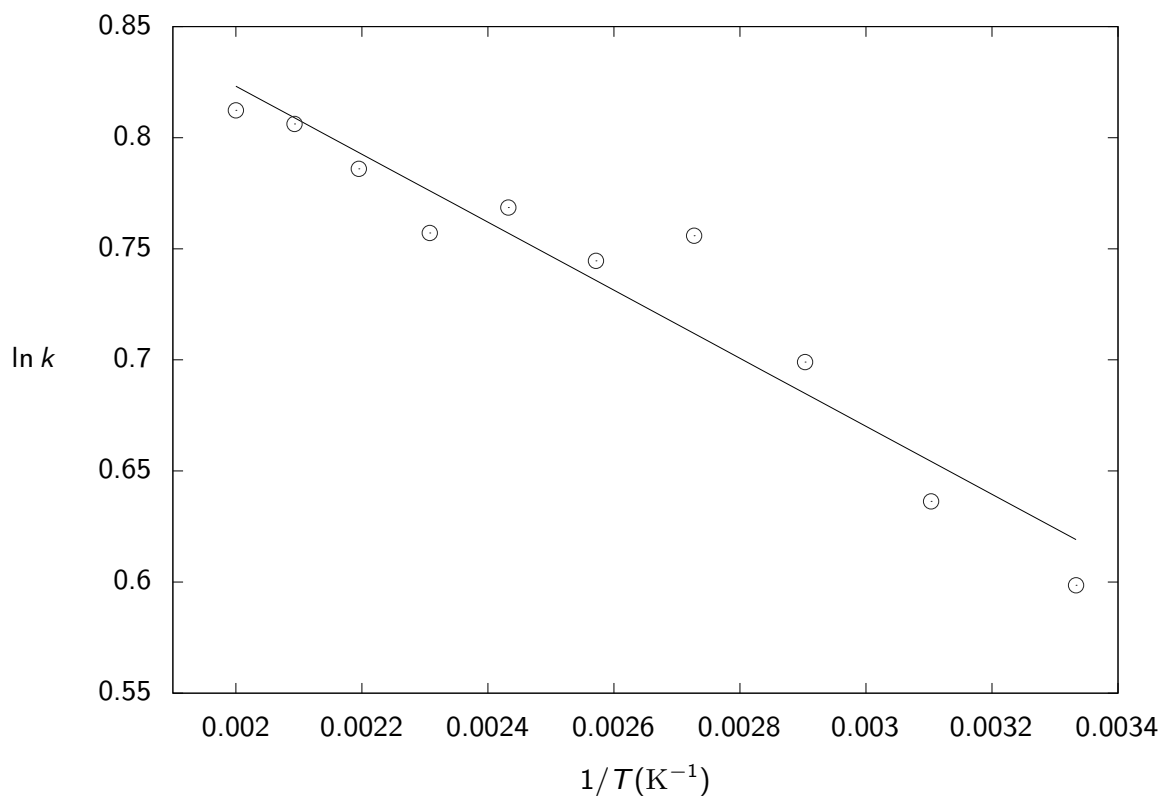
## Example 9.2: Unknown measurement variance and few data points

- Consider the data shown in the next figure with two unknown parameters and only 10 data points.
- The measurement errors are drawn from a normal distributed with zero mean and variance  $\sigma^2 = 10^{-3}$ .
- Compute the best estimates of activation energy and mean rate constant and the 95% confidence intervals for the cases of known and unknown measurement variance.

□

41 / 116

## The Data



42 / 116

- We calculate  $\hat{\theta}$  from Equation 9.14 and the sample variance from Equation 9.16 and obtain

$$\hat{\theta} = \begin{bmatrix} 0.747 \\ 153 \end{bmatrix}, \quad s^2 = 0.000454$$

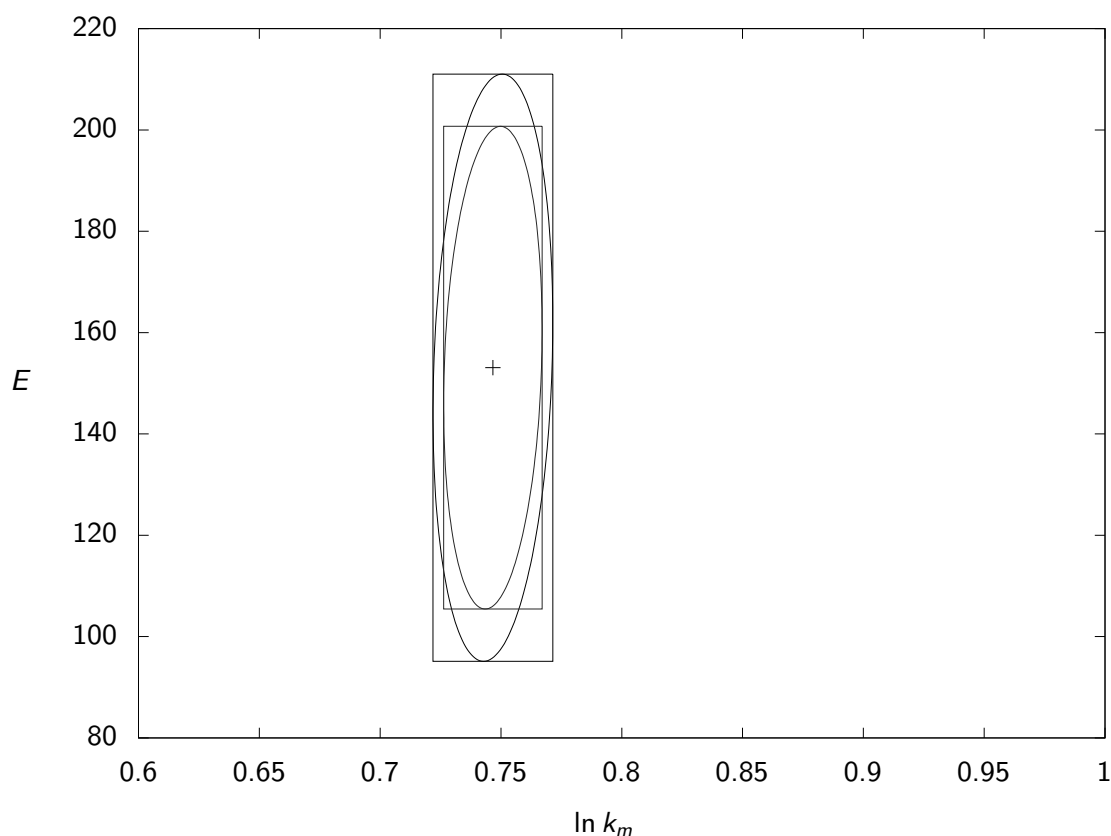
- We construct the 95% confidence interval from Equations 9.10 and 9.15. We obtain from a statistics table  $F(2, 8, 0.95) = 4.46$  so the two confidence intervals are given by

$$(\theta - \hat{\theta})^T \mathbf{X}^T \mathbf{X} (\theta - \hat{\theta}) \leq (0.001)(5.99), \quad \text{known variance}$$

$$(\theta - \hat{\theta})^T \mathbf{X}^T \mathbf{X} (\theta - \hat{\theta}) \leq (0.000454)(2)(4.46), \quad \text{unknown variance}$$

- The two 95% confidence ellipses are shown in the next figure.

### Known (solid line) and Unknown (dashed line) error variance



- Notice that although  $n_p F$  is 50% *larger* than  $\chi^2$ , our 95% confidence interval for the unknown measurement variance case is *smaller* than the known measurement variance case.
- Can you resolve this apparent contradiction?
- What experiment can you propose in which we would find the confidence interval with estimated measurement variance to be about 50% larger than known measurement variance in agreement with the notion that we lose information by estimating the variance as well as parameters.

## Nonlinear Least Squares

- In the previous two examples we transformed the model by taking logarithms to obtain a linear estimation problem.
- In many situations we do not want to make such a transformation, or such transformations simply do not exist.
- If we decide to treat the estimation problem using the nonlinear model, the problem becomes more challenging.
- As we will see, the parameter estimation becomes a nonlinear optimization that must be solved numerically instead of a linear matrix inversion that can be solved analytically
- Moreover, the confidence intervals become more difficult to compute, and they lose their strict probabilistic interpretation as  $\alpha$ -level confidence regions.
- As we will see, however, the “approximate” confidence intervals remain very useful in nonlinear problems.

- Consider the nonlinear model

$$y_i = h(x_i, \boldsymbol{\theta}), \quad i = 1, \dots, n_d \quad (9.17)$$

and the least-squares objective

$$\Phi(\boldsymbol{\theta}) = \sum_{i=1}^{n_d} (\tilde{y}_i - h(x_i, \boldsymbol{\theta}))^2$$

- We obtain the parameter estimates by solving the optimization problem

$$\min_{\boldsymbol{\theta}} \Phi(\boldsymbol{\theta})$$

subject to Equation 9.17.

- Call the solution to this problem  $\hat{\boldsymbol{\theta}}$ . We now consider the function  $\Phi(\boldsymbol{\theta})$  near the optimum.

47 / 116

## Taylor Series at the Optimum

- Expanding in a second-order multivariable Taylor series gives

$$\Phi(\boldsymbol{\theta}) \approx \Phi(\hat{\boldsymbol{\theta}}) + (\nabla \Phi)^T \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \quad (9.18)$$

- The gradient of the objective function is the vector of first derivatives of  $\Phi$  with respect to the model parameters

$$(\nabla \Phi)_j = \frac{\partial \Phi}{\partial \theta_j}$$

- The Hessian of the objective function is the matrix of second derivatives

$$H_{kj} = \frac{\partial^2 \Phi}{\partial \theta_k \partial \theta_j}$$

48 / 116



- The gradient is zero at the optimum, which allows us to rearrange Equation 9.18 to give

$$\Phi(\boldsymbol{\theta}) - \Phi(\hat{\boldsymbol{\theta}}) \approx \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{H} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

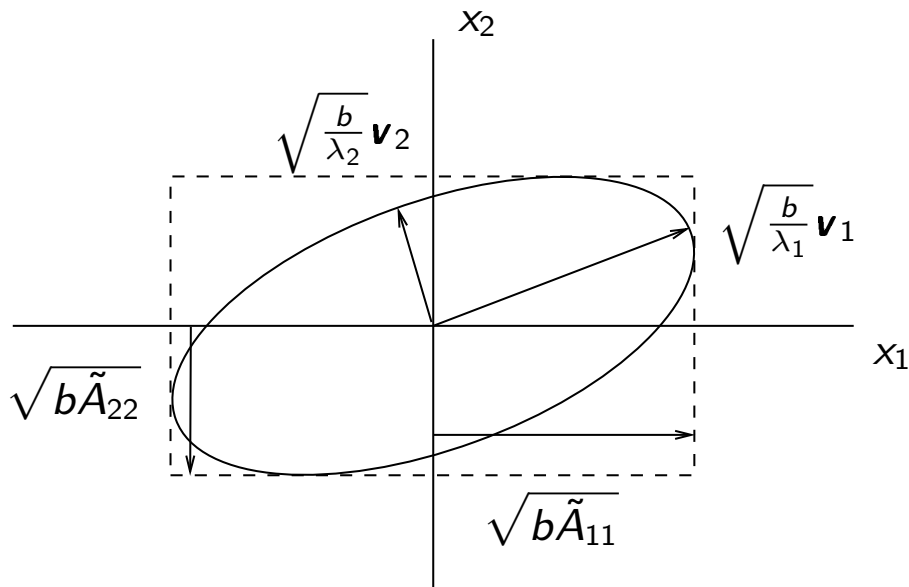
- Lines of constant objective function are approximately ellipses in which we use  $\mathbf{H}$  as the  $\mathbf{A}$  matrix.

## Quadratic Approximation

Use  $\mathbf{H}$  for  $\mathbf{A}$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = b$$

$$\mathbf{A} \mathbf{v}_i = \lambda_i \mathbf{v}_i$$



- We also can obtain order-of-magnitude confidence intervals using the relation

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq 2s^2 n_p F(n_p, n_d - n_p, \alpha) \quad (9.19)$$

- $s^2$  is again the sample variance

$$s^2 = \frac{1}{n_d - n_p} \sum_{i=1}^{n_d} \left( \tilde{y}_i - h(x_i, \hat{\boldsymbol{\theta}}) \right)^2 = \frac{\Phi(\hat{\boldsymbol{\theta}})}{n_d - n_p}$$

- These confidence intervals are correct only if the model is linear, in which case  $\mathbf{H} = 2\mathbf{X}^T \mathbf{X}$ .
- The intervals should be checked occasionally with Monte Carlo simulations when the model is nonlinear. We illustrate this check next.

51 / 116

## Fitting single and multiple adsorption experiments

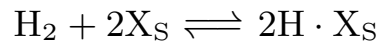
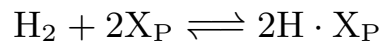
### Example 9.3: Fitting single and multiple adsorption experiments

- To illustrate the use of nonlinear models, we study an adsorption experiment.
- The system studied was the adsorption of  $\text{H}_2$  on a Pd catalyst with  $\text{SiO}_2$  support [8, 7].

□

52 / 116

- The adsorption is assumed dissociative.
- Both the catalyst and the support adsorb  $H_2$  so the adsorption isotherm model is



in which  $X_P$  represents a Pd vacant site and  $X_S$  represents a  $SiO_2$  vacant site.

- We can apply the methods of Section 5.6 to derive the surface coverage of H atoms

$$\bar{c}_H = \frac{\bar{c}_{mP} \sqrt{K_P} \sqrt{c_{H_2}}}{1 + \sqrt{K_P} \sqrt{c_{H_2}}} + \frac{\bar{c}_{mS} \sqrt{K_S} \sqrt{c_{H_2}}}{1 + \sqrt{K_S} \sqrt{c_{H_2}}}$$

53 / 116

## Parameter estimation

- It is known that the adsorption constant on the Pd is large, so we may reduce the model to

$$\bar{c}_H = \bar{c}_{mP} + \frac{\bar{c}_{mS} \sqrt{K_S} \sqrt{c_{H_2}}}{1 + \sqrt{K_S} \sqrt{c_{H_2}}} \quad (9.20)$$

- We have three parameters to estimate from data,

$$\theta^T = [ \sqrt{K_S} \quad \bar{c}_{mS} \quad \bar{c}_{mP} ]$$

- As the measure of fit to the data we choose a least-squares objective,

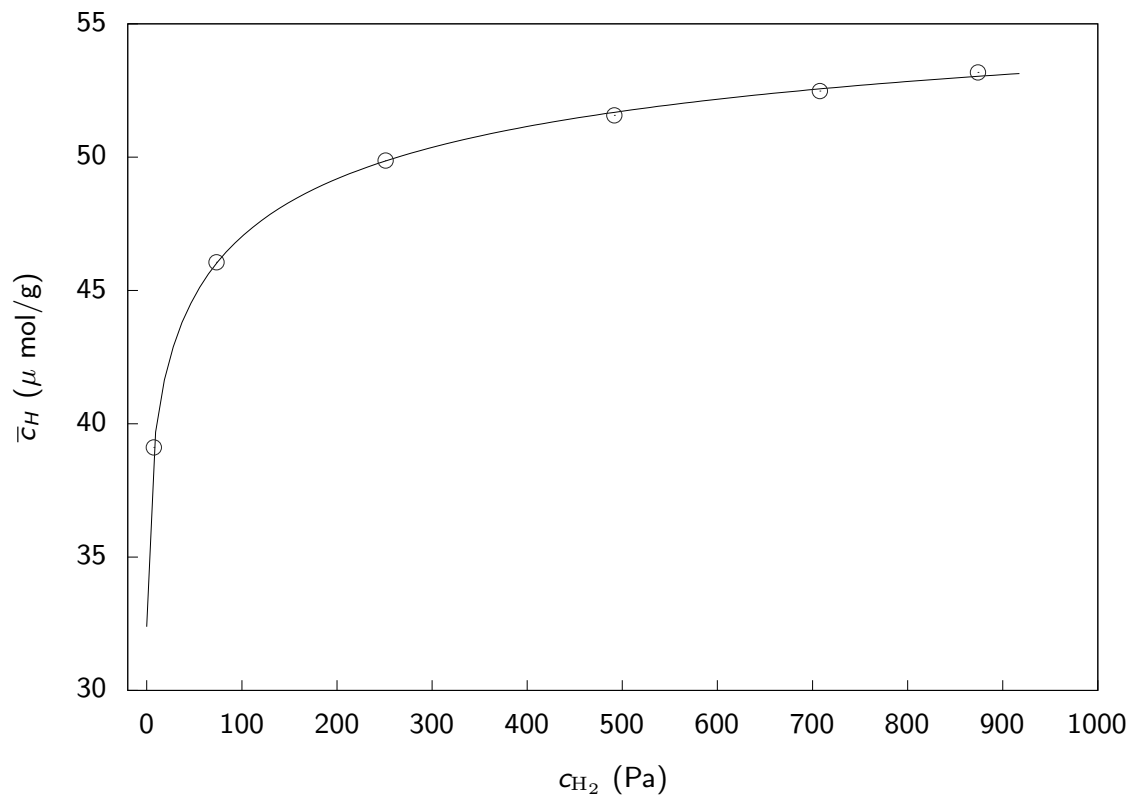
$$\Phi = \sum_i \left( \tilde{c}_{Hi} - \bar{c}_{Hi} \right)^2$$

- The parameter estimation problem

$$\min_{\theta} \Phi(\theta)$$

54 / 116

This figure shows the data and the best fit for a single adsorption experiment.



55 / 116

## Parameter estimates and confidence intervals

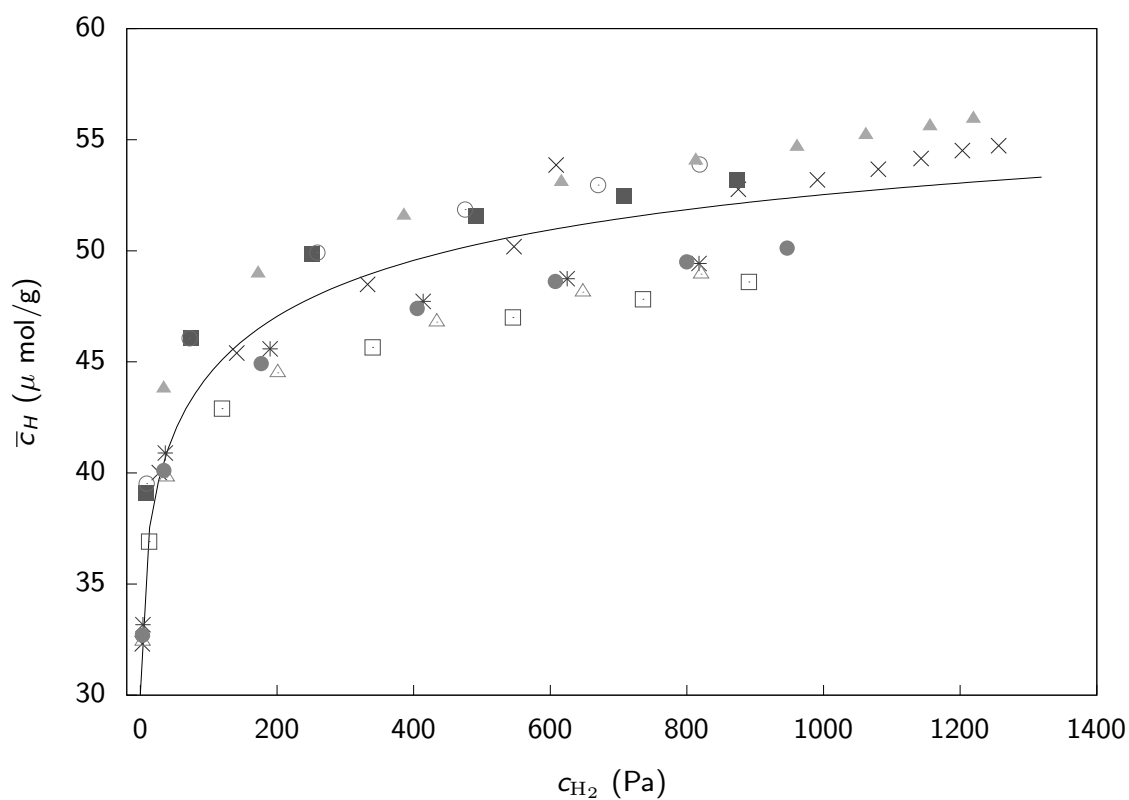
- The parameters and the 95% confidence intervals are given by

$$\hat{\theta} = \begin{bmatrix} \sqrt{K_S} \\ \bar{c}_{mS} \\ \bar{c}_{mP} \end{bmatrix} = \begin{bmatrix} 0.127 \\ 26.1 \\ 32.4 \end{bmatrix} \pm \begin{bmatrix} 0.048 \\ 1.5 \\ 2.4 \end{bmatrix}$$

- We can see that the model fit to the data is excellent, and the parameters are determined with fairly tight confidence intervals.

56 / 116

- For this catalyst sample, additional adsorption experiments are available.
- After the first adsorption experiment, reaction studies were performed with the catalyst sample. Then the catalyst was regenerated with heat treatment, and a second adsorption experiment was performed, followed by an additional reaction study, and so on.
- The additional adsorption data are shown in the next figure.



- Notice these are not replicate experiments, because the adsorption experiments are performed after different reaction studies have been performed, and the catalyst regeneration step does not necessarily return the catalyst to exactly the same condition.
- Estimating the parameters by fitting all of these data simultaneously produces the solid line in Figure 9.16 and the following parameter values and confidence intervals

$$\hat{\theta} = \begin{bmatrix} \sqrt{K_S} \\ \bar{c}_{mS} \\ \bar{c}_{mP} \end{bmatrix} = \begin{bmatrix} 0.0895 \\ 30.4 \\ 30.1 \end{bmatrix} \pm \begin{bmatrix} 0.081 \\ 4.7 \\ 5.4 \end{bmatrix}$$

- The total coverage estimated here,  $60.5 \mu\text{mol/g}$ , compares favorably to the apparent saturation limit or total uptake value of  $55 \mu\text{mol/g}$  reported by Natal-Santiago et al. [7, p.157].

## Interpretation

- Notice because the reaction studies have introduced significant variability in the data, the confidence intervals for the parameter values are significantly larger than those for the single adsorption experiment.
- Both sets of estimated parameters are valid, but they have different meanings.
- If the model is intended to represent the catalyst in its native state, one might use the estimated values from a single adsorption experiment on a freshly prepared sample.
- If the model is intended to represent the “average behavior” of the catalyst during the period it is used and regenerated, one would naturally use the estimated values from the many adsorption experiments.

Please read section 9.2.6 on your own.

## Replicate experiments

- We next take a quick look at the impact of replicating experiments.
- The main reason one replicates an experiment is to obtain a direct measure of the experimental reproducibility.
- By replication we are confronted with the simple experimental fact that no matter how much care we take and how much effort we exert, the results of two experiments are never exactly the same.

- The differences among replicated experiments give us a direct quantitative measure of the experimental error or noise.
- These errors are partly due to our inability to measure accurately the outputs of interest, but they are also due to our inability to control completely the experimental environment.
- The irreproducibility of the experiment is one of the key motivations for the random variable description of  $\mathbf{e}$  in Equation 9.7.

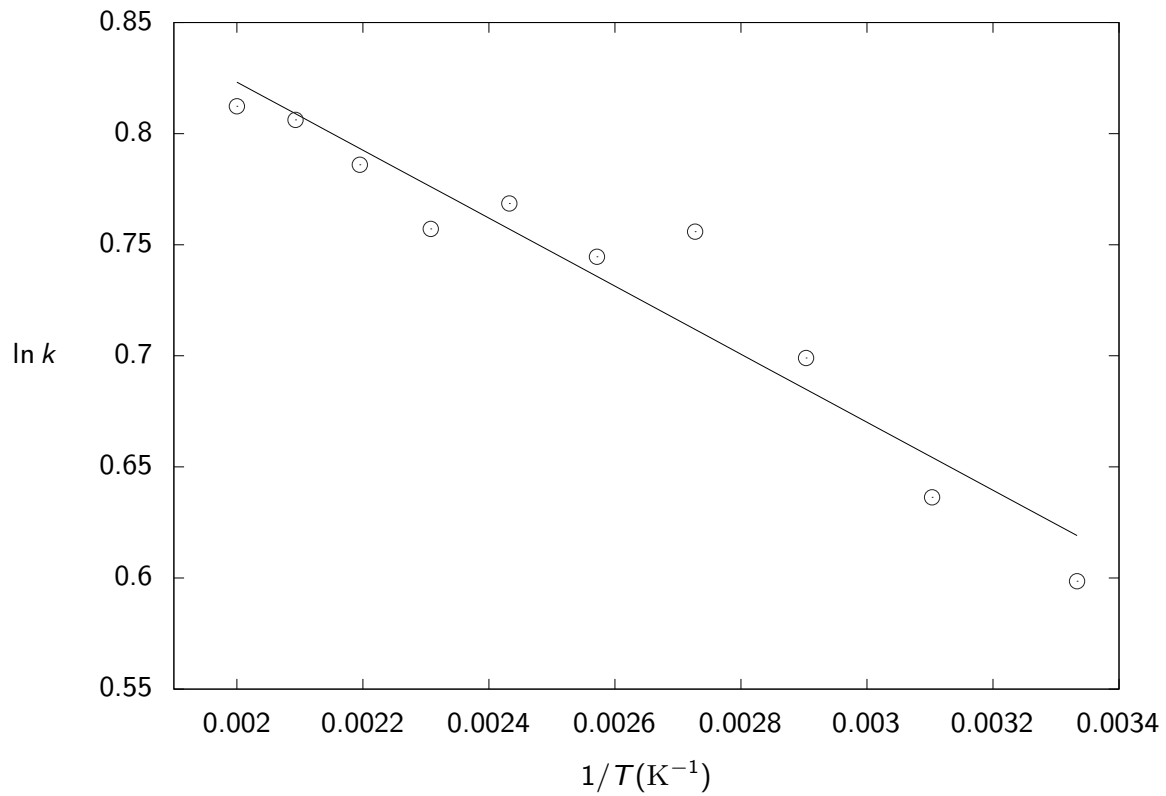
$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$$

## Replication and Confidence

- But replicating an experiment has another interesting and expected benefit. It increases our confidence in our conclusions.
- To see this benefit, we examine again the least-squares problem for estimating  $k_m, E$ .
- Assume for one experiment, we simply make measurements at 10 evenly spaced temperatures as shown in the next figure.



## One experiment

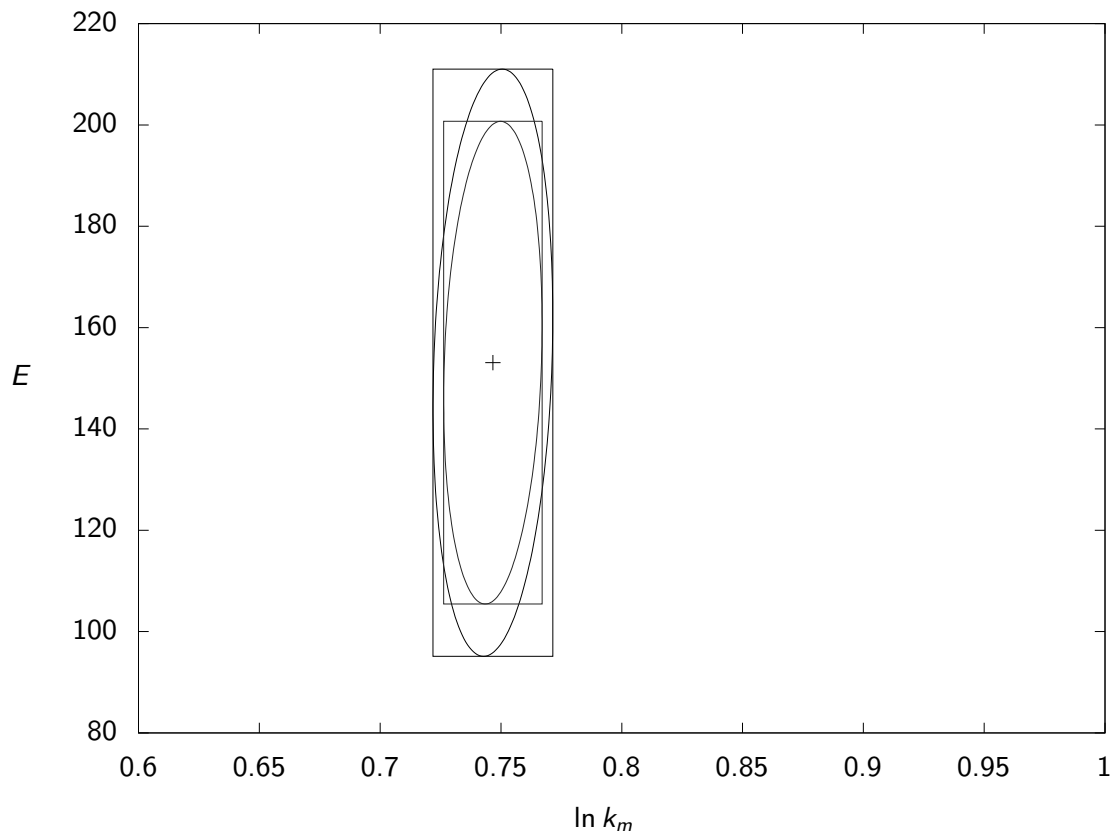


65 / 116

## Replication and Confidence

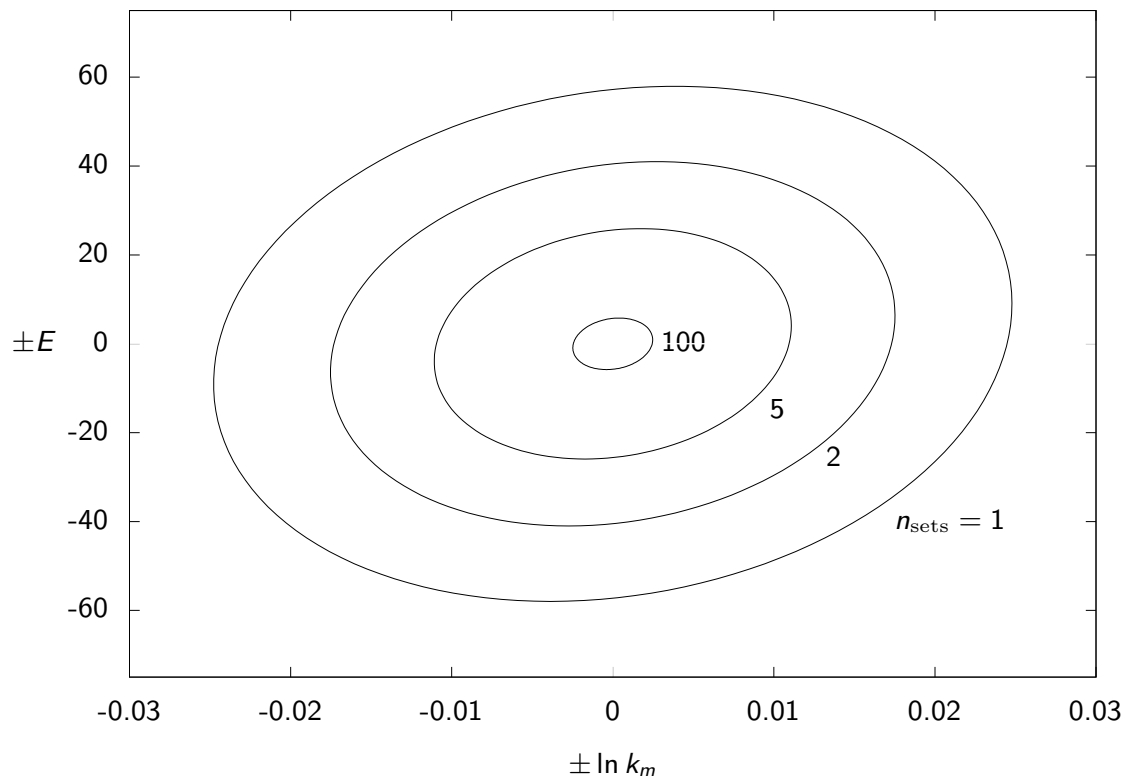
- If we perform one set of 10 measurements we achieve the confidence intervals we have shown previously

66 / 116



## Replication and Confidence

- What happens if we simply repeat the experiment and perform two sets of 10 measurements? The confidence region becomes smaller, as shown in the next figure.
- We also show the result of using 5 and 100 replicates of the experiment. The confidence intervals continue to shrink as we replicate the experiments.
- So replicating experiments provides another general avenue for increasing the confidence in the parameter estimates.
- Be aware, however, as discussed further in Exercise 9.8, that this “brute force” approach may be time consuming and expensive.



## Parameter Estimation with Differential Equation Models

- Now we turn to the single most important parameter estimation problem in chemical reactor modeling: **determining reaction-rate constants given dynamic concentration measurements.**
- We devote the rest of the chapter to developing methods for this problem.

- To get started, we consider a simple reactor model consisting of a single differential equation with a single experimentally measured quantity

$$\frac{dx}{dt} = f(x; \theta) \quad (9.21)$$

$$x(0) = g(x_0; \theta) \quad (9.22)$$

$$y = h(x) \quad (9.23)$$

in which  $x$  is the single material balance of interest,  $\theta$  are the unknown model parameters,  $x_0$  is the initial condition, and  $y$  is the experimentally measurable quantity.

- For simplicity let us assume here that  $x$  itself is measured, in which case  $h(x) = x$ .
- As we see in several of the examples, it may be necessary to include some of the initial conditions also as unknown parameters.
- Often  $t$  is time, but in steady-state tubular PFRs, reactor volume or length can take the place of time without changing the structure of the parameter estimation problem.

71 / 116

## Least Squares Problem

- As before, we define a least-squares objective to measure our fit to the data

$$\Phi(\theta) = \sum_i (\tilde{x}_i - x_i)^2 \quad (9.24)$$

- $\tilde{x}_i$  is the experimental measurement at time  $t_i$ ,
- $x_i$  is the solution to the model at time  $t_i$ ,  $x_i = x(t_i; \theta)$ . Note  $x_i$  is the only part of the objective function that depends on the model parameters.
- Again, we minimize this objective function to obtain our parameter estimates

$$\min_{\theta} \Phi(\theta) \quad (9.25)$$

subject to Equations 9.21–9.23.

72 / 116

- The major change is that the model constraint consists of nonlinear differential equations rather than linear or nonlinear algebraic equations as in the previous sections.
- The differential equations make it much more expensive to evaluate the constraints while solving the optimization problem.
- We can increase the efficiency of the optimizer if we provide an accurate gradient of the objective function.

73 / 116

## Gradient and Hessian of the Parameter Estimation Problem

$$\Phi(\theta) = \sum_i (\tilde{x}_i - x_i)^2$$

- Recall the gradient of the objective function is the vector of first derivatives of  $\Phi$  with respect to the model parameters.

Differentiating this equation gives

$$\frac{\partial \Phi}{\partial \theta_j} = -2 \sum_i (\tilde{x}_i - x_i) \frac{\partial x_i}{\partial \theta_j} \quad (9.26)$$

- Differentiating a second time gives the Hessian of the objective function, which we again use to construct approximate confidence intervals

$$\frac{\partial^2 \Phi}{\partial \theta_k \partial \theta_j} = 2 \sum_i \left[ \frac{\partial x_i}{\partial \theta_k} \frac{\partial x_i}{\partial \theta_j} - (\tilde{x}_i - x_i) \frac{\partial^2 x_i}{\partial \theta_k \partial \theta_j} \right] \quad (9.27)$$

74 / 116

- In the Gauss-Newton approximation of the Hessian, we neglect the second term in Equation 9.27 to yield,

$$\frac{\partial^2 \Phi}{\partial \theta_k \partial \theta_j} \approx 2 \sum_i \frac{\partial x_i}{\partial \theta_k} \frac{\partial x_i}{\partial \theta_j} \quad (9.28)$$

- Two arguments support the Gauss-Newton approximation. If the model fits the data well at the optimal value of parameters, the residuals are small in magnitude and of different signs. The sum in the second term is then small.
- Alternatively, the second derivative of the model may be small compared to the first derivative.
- If the model is linear in the parameters, for example, the second derivatives are identically zero and the Gauss-Newton approximation is exact.

## Some restrictions on using Gauss-Newton

- The Gauss-Newton approximation is not valid if the model solution is a highly nonlinear function of the parameters, or if the residuals are large and not randomly distributed about zero at the optimal value of parameters.
- In the latter case one should question the model structure because the model does not well represent the data.
- For the highly nonlinear case, one may try numerical finite difference formulas to compute the Hessian.
- Computing a reliable finite difference approximation for a *second* derivative is not a trivial matter either, however, and the step size should be carefully chosen to avoid amplifying the errors introduced by the finite numerical precision.

- The first derivatives of the model solution with respect to the model parameters are known as the model sensitivities,

$$S_{ij} = \frac{\partial x_i}{\partial \theta_j}$$

- The sensitivities also can be described as the solution to a set of differential equations (see also Appendix A).

## Sensitivities — Evolution Equation

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) & \mathbf{x}(0) &= \mathbf{g}_0(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \boldsymbol{\theta}^T} \frac{d\mathbf{x}}{dt} &= \frac{\partial}{\partial \boldsymbol{\theta}^T} \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \\ \frac{d}{dt} \left( \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^T} \right) &= \frac{\partial \mathbf{f}}{\partial \mathbf{x}^T} \left( \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}^T} \right) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}^T} \end{aligned}$$

$$\boxed{\frac{d\mathbf{S}}{dt} = \mathbf{f}_x \mathbf{S} + \mathbf{f}_\theta \quad \mathbf{S}(0) = \mathbf{g}_\theta}$$

- This fact allows us to solve the model and sensitivity equations simultaneously with an ODE solver, rather than use finite difference formulas to obtain the sensitivities.
- Bard provides a readable account for further study on these issues [1].
- Caracotsios, Stewart and Sørensen developed this approach and produced an influential software code (GREG) for parameter estimation problems in chemical reaction engineering [4, 9, 5].

## From Sensitivities to Gradients and Hessians

- From the sensitivities and model solution, we can then calculate the gradient of the objective function and the Gauss-Newton approximation of the Hessian matrix.
- Reliable and robust numerical optimization programs are available to find the optimal values of the parameters.
- These programs are generally more efficient if we provide the gradient in addition to the objective function.
- The Hessian is normally needed only to calculate the confidence intervals after the optimal parameters are determined.



- If we define  $\mathbf{e}$  to be the residual vector

$$e_i = \tilde{x}_i - x_i$$

We can express the gradient in matrix notation as

$$\begin{aligned} \frac{\partial \Phi}{\partial \theta_j} &= -2 \sum_i (\tilde{x}_i - x_i) \frac{\partial x_i}{\partial \theta_j} \\ \nabla \Phi &= -2 \mathbf{S}^T \mathbf{e} \end{aligned} \tag{9.29}$$

- In terms of the sensitivities, we can express the Gauss-Newton approximation of the Hessian as

$$\begin{aligned} \frac{\partial^2 \Phi}{\partial \theta_k \partial \theta_j} &\approx 2 \sum_i \frac{\partial x_i}{\partial \theta_k} \frac{\partial x_i}{\partial \theta_j} \\ H_{kj} &= 2 \sum_i S_{ik} S_{ij} \\ \mathbf{H} &= 2 \mathbf{S}^T \mathbf{S} \end{aligned} \tag{9.30}$$

- Given these expressions for the gradient and Hessian, we can construct a fairly efficient parameter estimation method for differential equation models using standard software for solving nonlinear optimization and solving differential equations with sensitivities.

- ① Guess initial parameter values.
- ② Using an appropriate ODE solver, solve the model and sensitivity equations simultaneously given the current parameter values. Compute  $x_i$  and  $S_{ij}$ .
- ③ Evaluate  $\Phi$  and  $\nabla\Phi$  using Equations 9.24 and 9.29.
- ④ Update parameter values to minimize  $\Phi$ . This step and the next are usually controlled by an optimization package.
- ⑤ Check convergence criteria. If not converged, go to ②.
- ⑥ On convergence, set  $\hat{\theta}$  to current parameter values. Calculate  $H$  using Equation 9.30. Calculate confidence intervals using Equation 9.19.

## Example: Fitting reaction-rate constant and order

### Example 9.4: Fitting reaction-rate constant and order

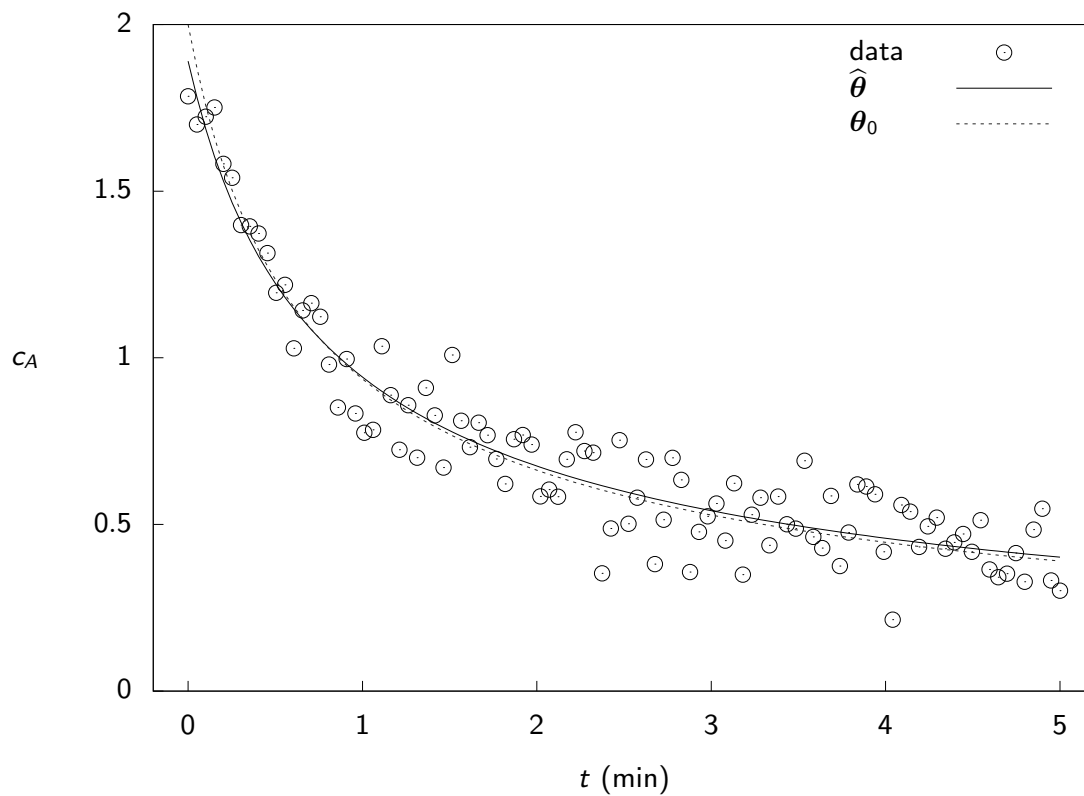
- We illustrate these methods on a classic reactor modeling problem: finding the rate constant and reaction order from isothermal concentration measurements in a batch reactor.
- Consider an irreversible,  $n$ th order reaction



taking place in a liquid-phase batch reactor containing a large excess of reactant B.

- Given the measured concentration of component A shown in Figure 9.21, determine the best values of the model parameters.

□



85 / 116

## Solution

- The material balance for species A is

$$\begin{aligned} \frac{dc_A}{dt} &= -kc_A^n \\ c_A(0) &= c_{A0} \end{aligned} \quad (9.32)$$

- Given the experimental data, it does not seem reasonable to assume we know  $c_{A0}$  any more accurately than the other measurements, so we include it as a parameter to be estimated.
- The model therefore contains three unknown parameters

$$\boldsymbol{\theta}^T = [ k \quad c_{A0} \quad n ]$$

86 / 116

- We can generate a reasonable initial parameter set by guessing values and solving the model until the model simulation is at least on the same scale as the measurements.
- We provide this as the starting point, and then solve the nonlinear optimization problem in Equation 9.25 using the least-squares objective as shown in Equation 9.24.
- We then compute the approximate confidence intervals using Equation 9.19 with Equation 9.30 for  $\mathbf{H}$ . The solution to the optimization problem and the approximate confidence intervals are given in Equation 9.33.

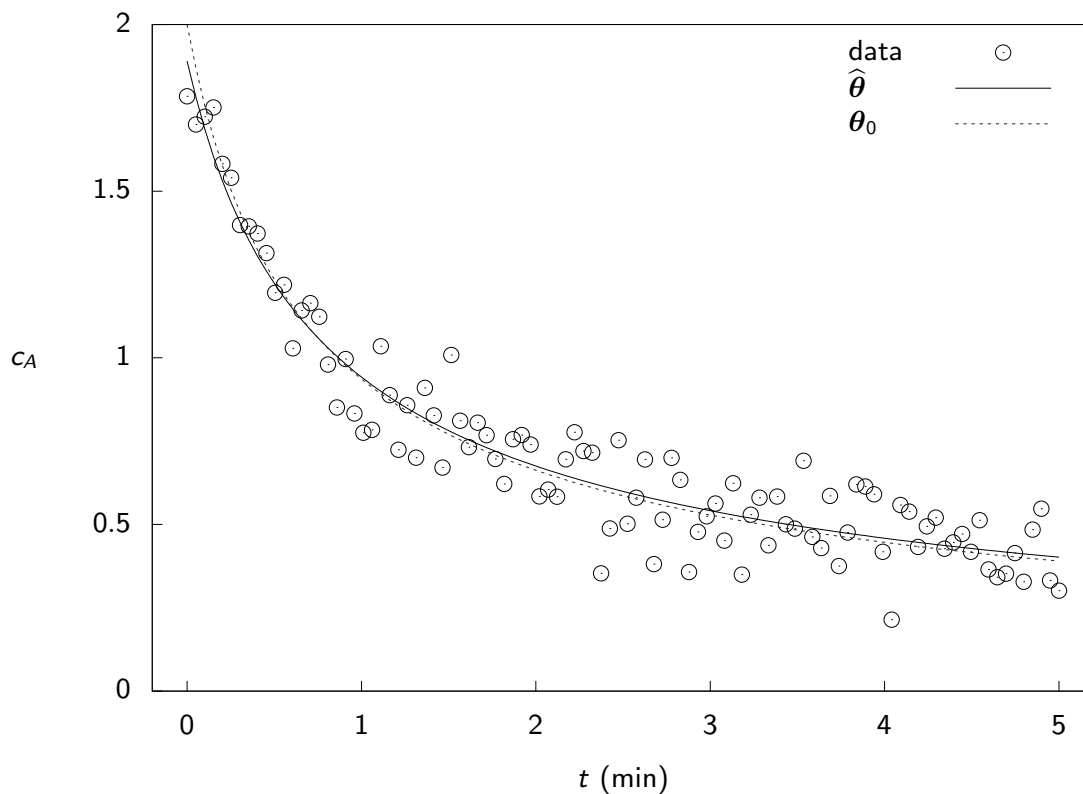
87 / 116

## Results

$$\boldsymbol{\theta}_0 = \begin{bmatrix} k \\ c_{A0} \\ n \end{bmatrix} = \begin{bmatrix} 0.5 \\ 2.0 \\ 2.5 \end{bmatrix} \quad \hat{\boldsymbol{\theta}} = \begin{bmatrix} 0.47 \\ 1.89 \\ 2.50 \end{bmatrix} \pm \begin{bmatrix} 0.052 \\ 0.18 \\ 0.42 \end{bmatrix} \quad (9.33)$$

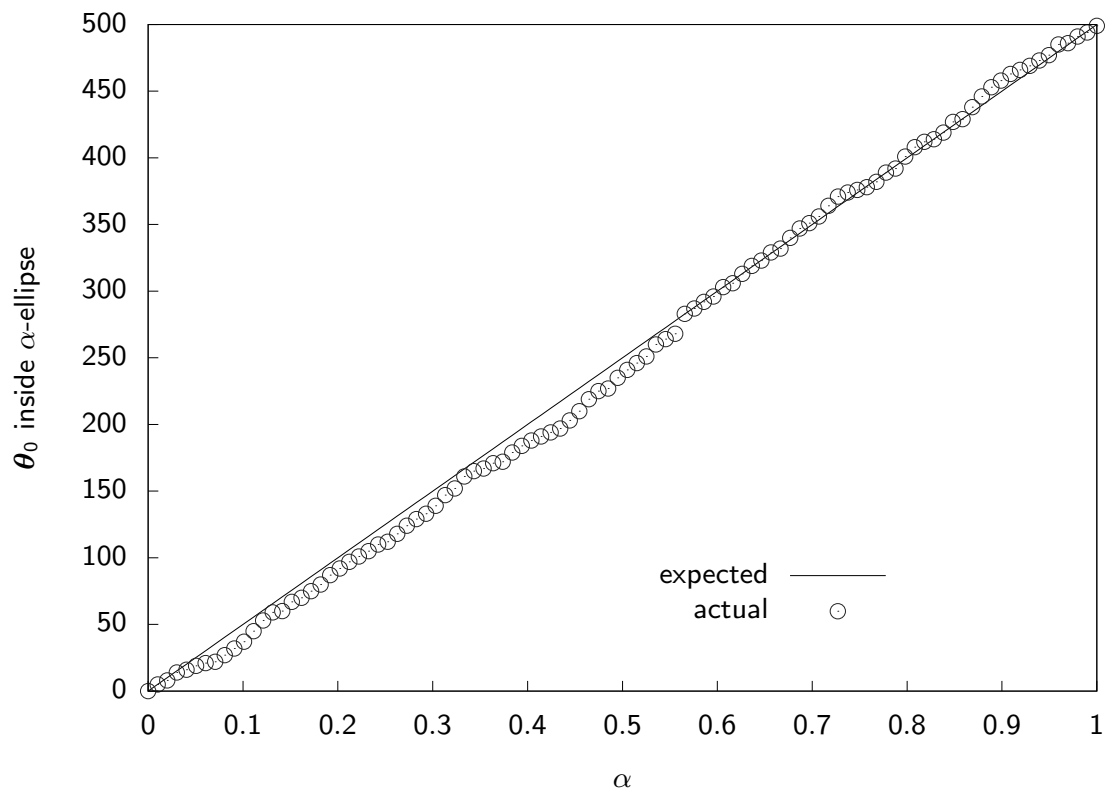
- The parameters that we used to generate the data also are given in Equation 9.33.
- Notice the estimates are close to the correct values, and we have fairly tight approximate confidence intervals.

88 / 116



## Validating the Confidence Intervals

- Next we examine the quality of these approximate confidence intervals for this problem.
- In this study we generate 500 datasets by adding zero-mean measurement noise with variance  $\sigma^2 = 0.01$  to the model solution with the correct parameters.
- For each of these 500 datasets, we solve the optimization problem to obtain the parameter estimates.
- We also produce a value for the Hessian for each dataset, and we use the mean of these for  $\mathbf{H}$ .
- Finally we calculate what fraction of these 500 estimates fall within each  $\alpha$ -confidence region as defined by Equation 9.19 as we vary  $\alpha$ .
- This result is plotted as the points in the next figure. Notice for all  $\alpha$  values, approximately the correct number of parameter estimates fall within the corresponding ellipse, which indicates the approximate confidence intervals given in Equation 9.19 are fairly reliable for this problem.

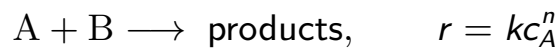


## Validating the Confidence Intervals

- Be aware that this conclusion may not be true for other nonlinear problems, and should be checked. This computational check is fairly expensive; note that we had to solve 500 optimization problems to produce the figure.
- But given the dramatic increase in computational speed, a few Monte Carlo simulations in the final stages of a modeling study are well justified.

# Breaking a Bad Habit

- In your undergraduate training, did any of you learn the following method for estimating reaction orders and rate constants?
- Differentiate the concentration  $c(t)$  data and obtain the rate directly from the data



$$r(t_i) \approx -\frac{c_A(t_{i+1}) - c_A(t_i)}{t_{i+1} - t_i}$$

- Transform variables by taking the logarithm of the rate expression

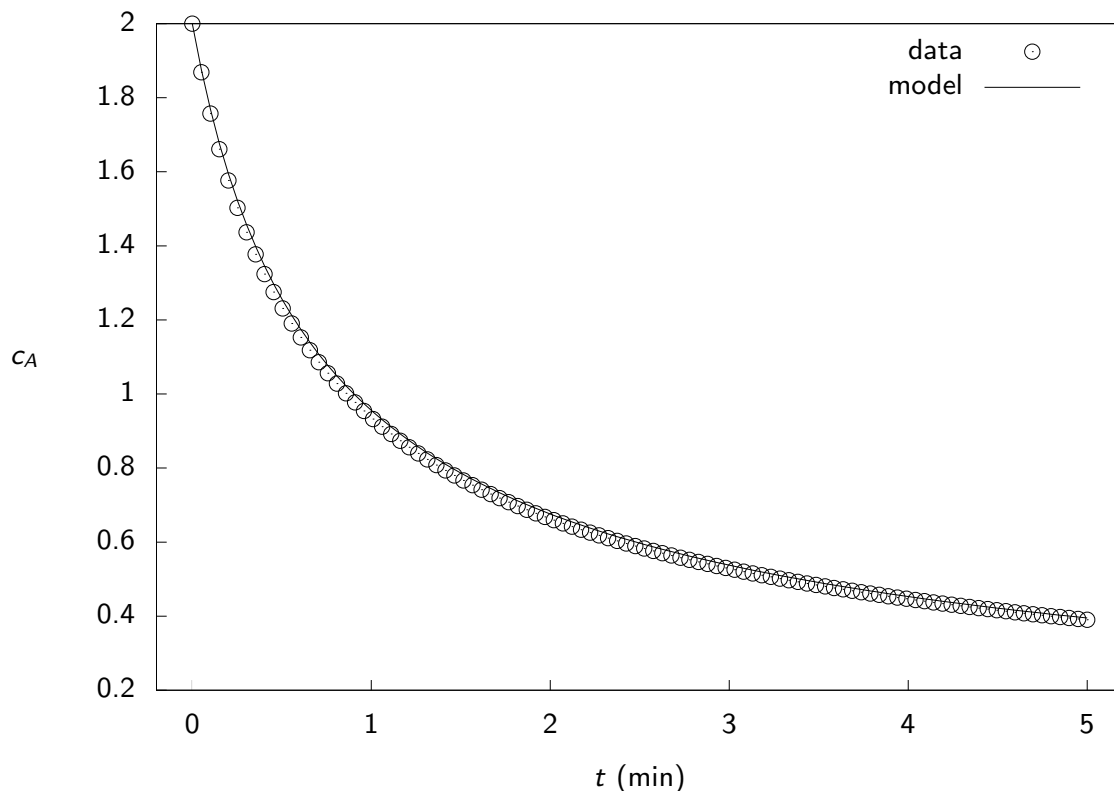
$$\ln(r_i) = \ln(k) + n \ln(c_{A_i}), \quad i = 1, \dots, n_d$$

- Set up the least squares problem

$$\mathbf{y} = \begin{bmatrix} \ln(r_1) \\ \ln(r_2) \\ \vdots \\ \ln(r_{n_d}) \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & \ln(c_{A1}) \\ 1 & \ln(c_{A2}) \\ \vdots & \vdots \\ 1 & \ln(c_{A_{n_d}}) \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} \ln(k) \\ n \end{bmatrix}$$
$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}$$

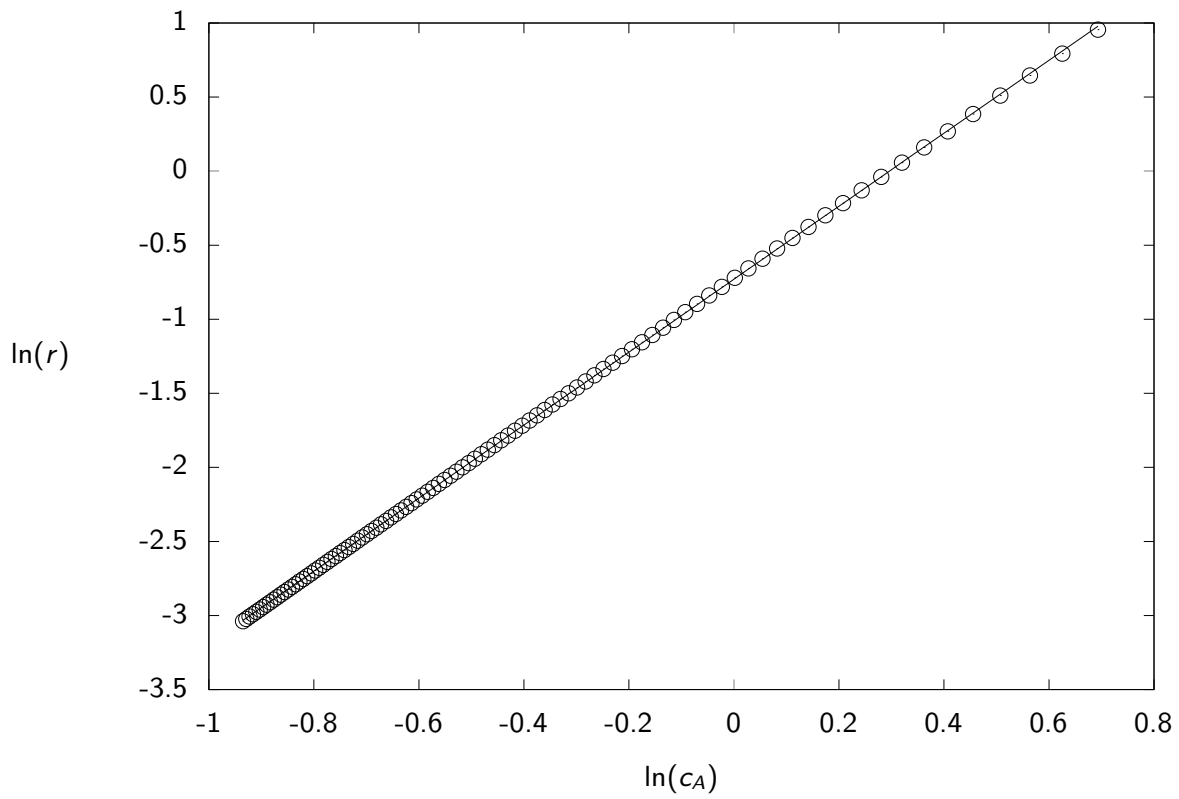
93 / 116

Why is this approach a bad idea? No noise. I



94 / 116

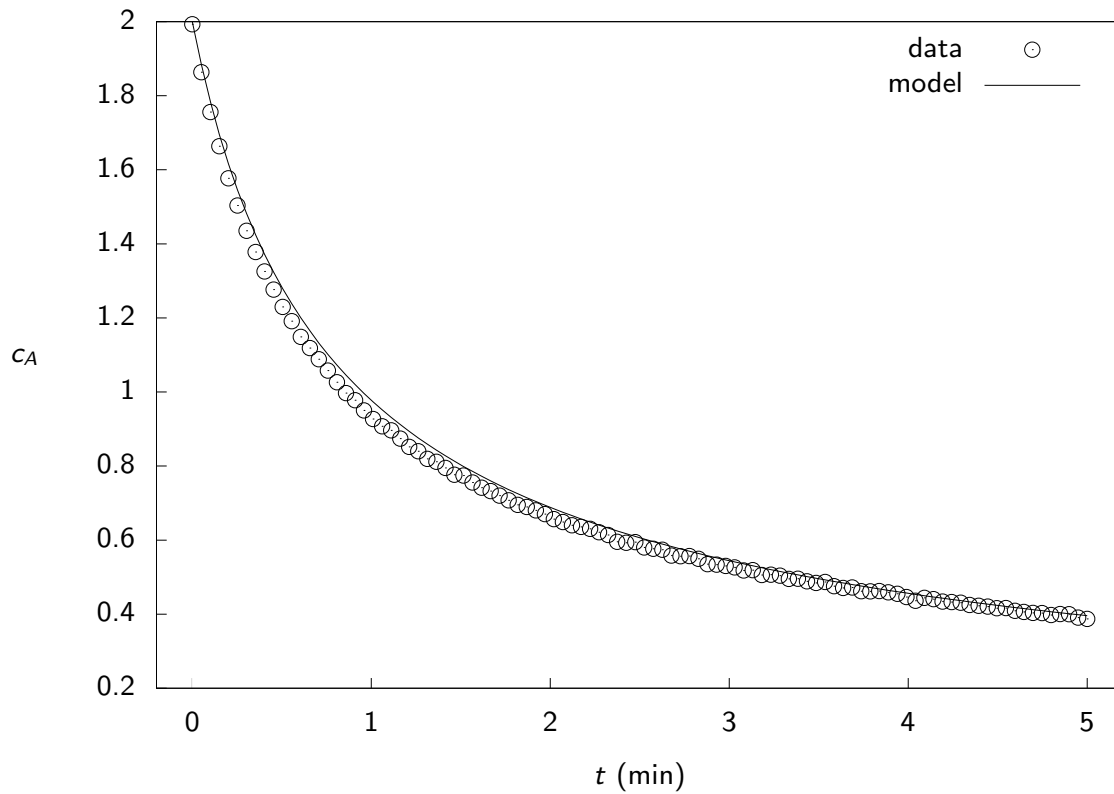
# Why is this approach a bad idea? No noise. II



$$\begin{bmatrix} \ln(\hat{k}) \\ \hat{n} \end{bmatrix} = \begin{bmatrix} -0.729 \\ 2.46 \end{bmatrix} \pm \begin{bmatrix} 0.0018 \\ 0.0031 \end{bmatrix} \quad \begin{bmatrix} \ln(k) \\ n \end{bmatrix} = \begin{bmatrix} -0.693 \\ 2.5 \end{bmatrix}$$

95 / 116

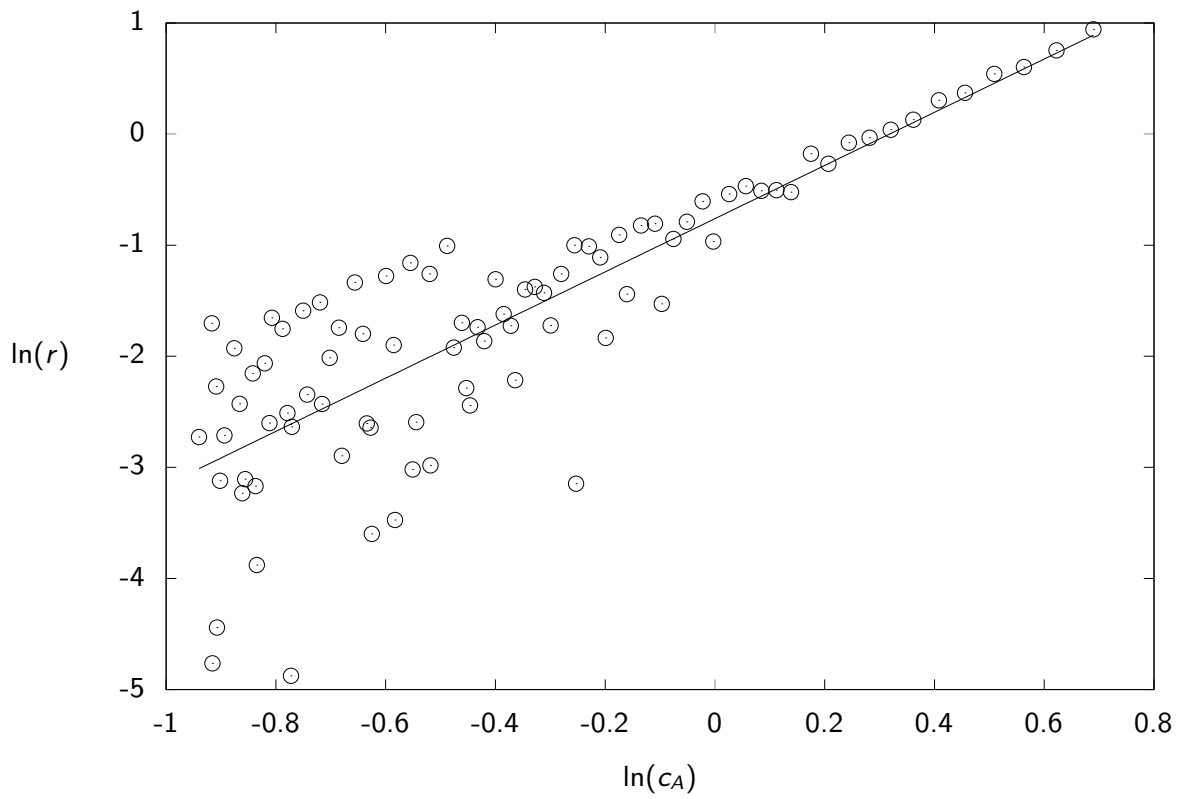
# So far so good. Small noise. I



96 / 116



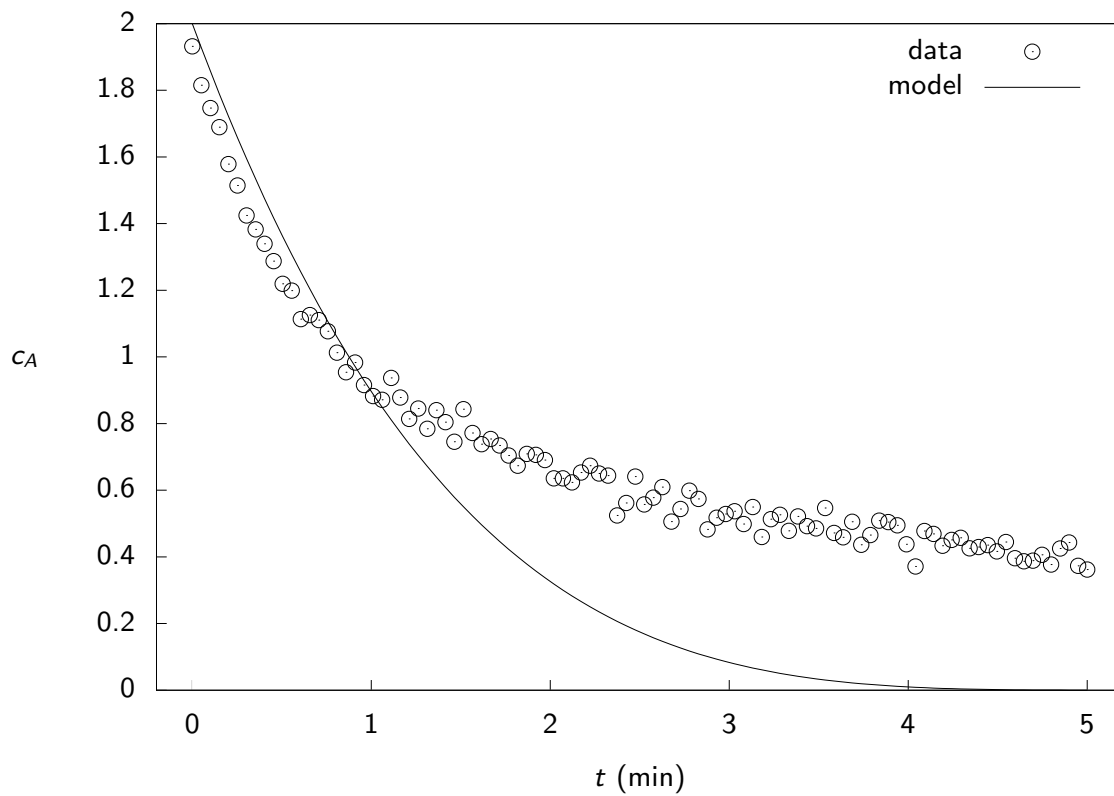
# So far so good. Small noise. II



$$\begin{bmatrix} \ln(\hat{k}) \\ \hat{n} \end{bmatrix} = \begin{bmatrix} -0.761 \\ 2.39 \end{bmatrix} \pm \begin{bmatrix} 0.23 \\ 0.41 \end{bmatrix} \quad \begin{bmatrix} \ln(k) \\ n \end{bmatrix} = \begin{bmatrix} -0.693 \\ 2.5 \end{bmatrix}$$

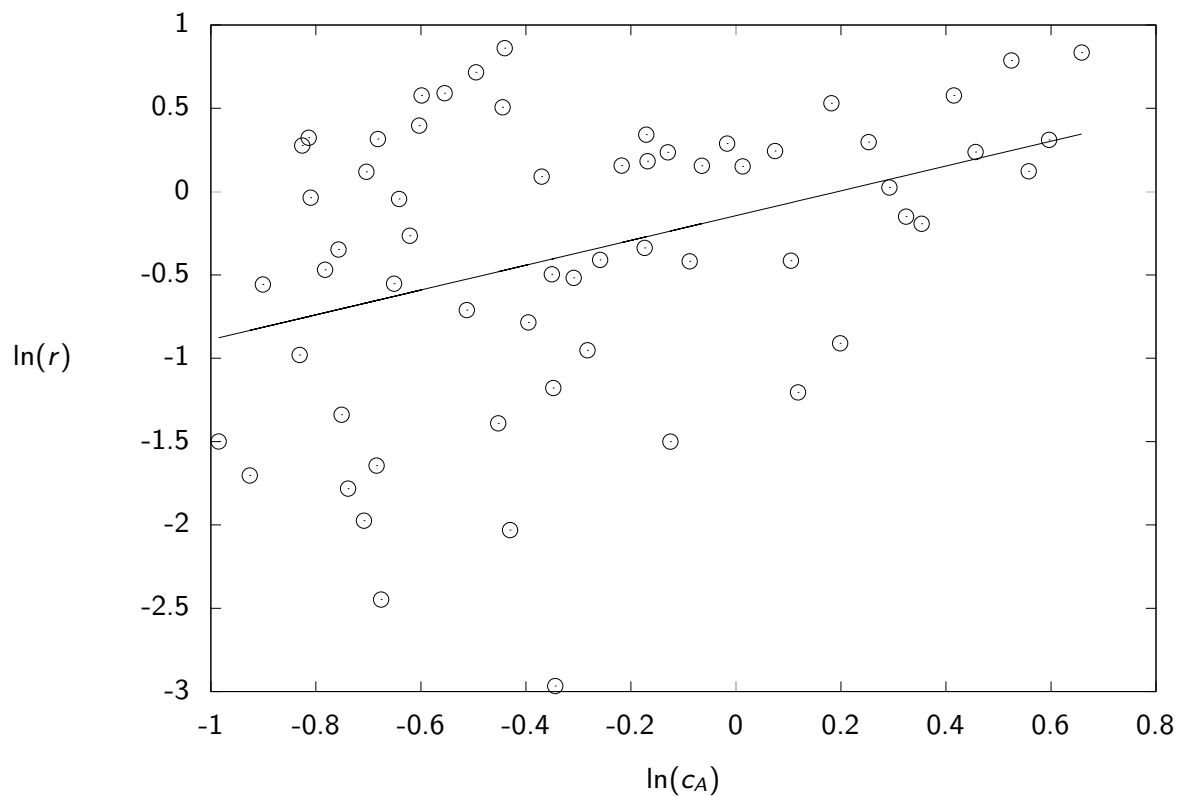
97 / 116

# Hmmm... not even large noise and... I



98 / 116

## Hmmm... not even large noise and... II



$$\begin{bmatrix} \ln(\hat{k}) \\ \hat{n} \end{bmatrix} = \begin{bmatrix} -0.144 \\ 0.744 \end{bmatrix} \pm \begin{bmatrix} 0.32 \\ 0.60 \end{bmatrix} \quad \begin{bmatrix} \ln(k) \\ n \end{bmatrix} = \begin{bmatrix} -0.693 \\ 2.5 \end{bmatrix}$$

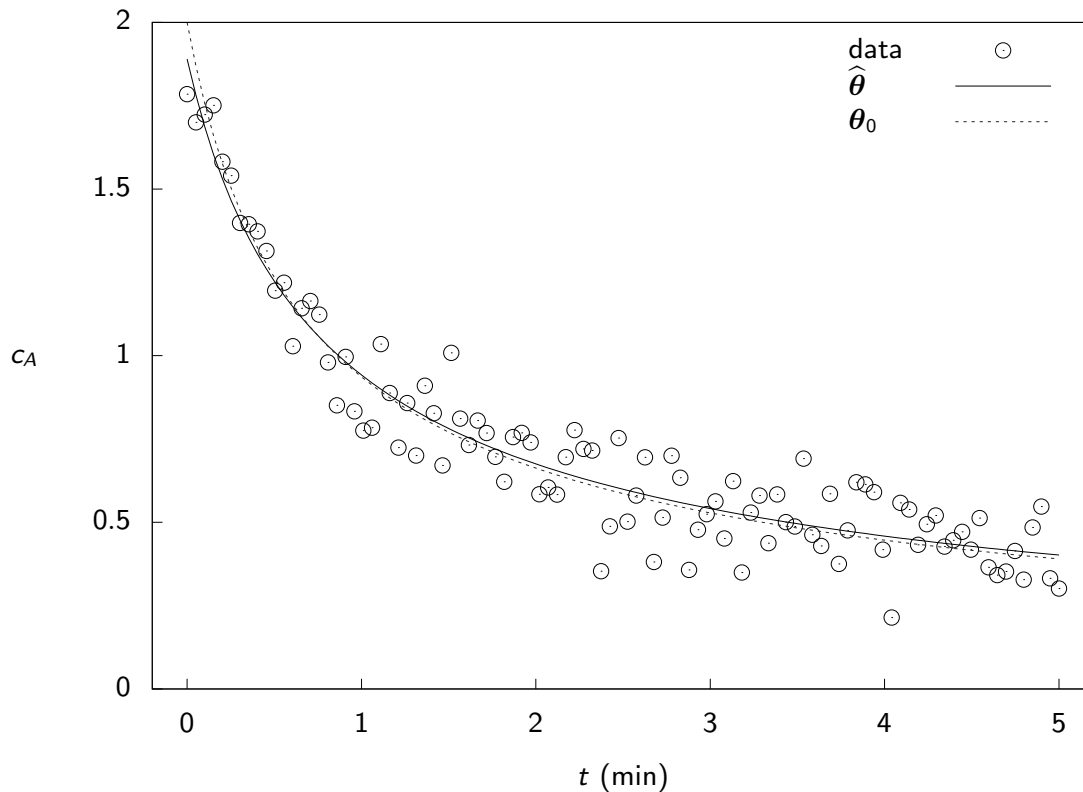
99 / 116

## Hmmm... not even large noise and... III

Note. Some of the rates are *negative*, and the logarithm is complex valued! I deleted those points (39/100) so the least squares estimate would not be complex valued.

100 / 116

## And the results again without differentiation



$$\theta_0 = \begin{bmatrix} k \\ c_{A0} \\ n \end{bmatrix} = \begin{bmatrix} 0.5 \\ 2.0 \\ 2.5 \end{bmatrix} \quad \hat{\theta} = \begin{bmatrix} 0.47 \\ 1.89 \\ 2.50 \end{bmatrix} \pm \begin{bmatrix} 0.052 \\ 0.18 \\ 0.42 \end{bmatrix}$$

101 / 116

## Summary

- If you have noise in your data, do not differentiate the data. Differentiation greatly amplifies noise.
- Rather than first smoothing the data and then differentiating, it is simpler to just optimize over the rate constants to fit the concentration measurement directly.
- The instantaneous rate is not needed to estimate the parameters, only the concentration measurement is needed. In fact, one good estimate of the instantaneous rate comes from solving the model with the best parameters. That produces an estimate of the rate without differentiating the data.
- If you have a complicated model with many species and many reactions with many rate constants, parameter estimation with optimization is a general approach. Transformations are not necessary and straight lines are not necessary.

102 / 116

- Most reactor models of interest contain balances for several species and therefore consist of several differential equations.
- To determine the parameters in such models, usually the concentrations of several species are measured as well.

- To treat this case, we consider the general differential equation model

$$\frac{dx}{dt} = f(x; \theta) \quad (9.34)$$

$$x(0) = g(x_0; \theta) \quad (9.35)$$

$$y = h(x) \quad (9.36)$$

in which  $x$  is the vector of states that defines the reactor model,  $\theta$  are the unknown model parameters,  $x_0$  are the initial conditions, and  $y$  are the experimentally measurable quantities.

- In this case, we again define a scalar objective function that measures our fit to the data. When we have different measured quantities, however, it often does not make sense to sum the squares of the residuals.
- The measured variables may differ in size from each other by orders of magnitude. The influence a measurement has on the objective also would be influenced by the arbitrary choice of the units of the measurement, which is obviously undesirable.
- The simplest way to address this issue is to employ *weighted* least squares. The reader should be aware that more general procedures are available for the multiple measurement case, including the maximum likelihood method [9, 10, 11].
- For simplicity of presentation, we focus here on weighted least squares.

- In weighted least squares, we combine the different measurements by forming the *weighted* sum of the residuals.
- Let  $\mathbf{e}_i$  be the residual vector at the  $i$ th sample time

$$\mathbf{e}_i = \tilde{\mathbf{y}}_i - \mathbf{h}(\mathbf{x}(t_i; \boldsymbol{\theta}))$$

The objective function is then

$$\Phi(\boldsymbol{\theta}) = \sum_i \mathbf{e}_i^T \mathbf{W} \mathbf{e}_i \quad (9.37)$$

in which  $\mathbf{W}$  is a symmetric, positive-definite weighting matrix.

- Usually  $\mathbf{W}$  is chosen to be a diagonal matrix. The elements on the diagonal are the weights assigned to each measurement type.
- To estimate the parameters we now solve

$$\min_{\boldsymbol{\theta}} \Phi(\boldsymbol{\theta}) \quad (9.38)$$

subject to Equations 9.34–9.36.

## Gradient and Hessian Approximation

- The sensitivities are now a time-varying matrix,

$$S_{jk}(t_i) = \frac{\partial x_j(t_i; \boldsymbol{\theta})}{\partial \theta_k}$$

- We can compute the gradient as before

$$\nabla \Phi = -2 \sum_i \mathbf{s}_i^T \frac{\partial \mathbf{h}_i^T}{\partial \mathbf{x}_i} \mathbf{W} \mathbf{e}_i$$

- The Gauss-Newton approximation of the Hessian is

$$\mathbf{H} = 2 \sum_i \mathbf{s}_i^T \frac{\partial \mathbf{h}_i^T}{\partial \mathbf{x}_i} \mathbf{W} \frac{\partial \mathbf{h}_i}{\partial \mathbf{x}_i^T} \mathbf{s}_i$$

in which  $\mathbf{S}_i = \mathbf{S}(t_i)$ ,  $\mathbf{x}_i = \mathbf{x}(t_i)$ , and  $\mathbf{h}_i = \mathbf{h}(\mathbf{x}(t_i))$ .

Please read Example 9.5 on your own.

This example illustrates the following issues.

- The model consists of six reactions and six unknown rate constants. We assume three species concentrations can be measured.
- The six parameters cannot be identified from the three measurements without very large confidence intervals.
- We show how to reduce the model to a four parameter model whose four parameters can be determined accurately from the measurements.
- This model reduction step is usually necessary with complex models and limited measurements (the usual industrial situation).

## An Industrial Case Study

Please read Section 9.3, an industrial case study, on your own.

In this case study we illustrate the following steps:

- Proposing an initial model given a rough idea of the reaction chemistry.
- Estimating the parameters for the full model given composition measurements.
- Reducing the model based on the parameter confidence intervals.
- Designing new experiments based on the reduced model's confidence intervals.
- Using the final model to find new operating policies to double the production rate.

- In analyzing data, we used probability and random variables to model the irreproducible part of the experiment.
- For models that are linear in the parameters with normally distributed measurement error, we can perform parameter estimation and construct exact confidence intervals analytically. Thank you, Messrs. Gauss and Laplace.
- For models that are nonlinear in the parameters, we compute parameter estimates and construct approximate confidence intervals using nonlinear optimization methods.
- We then covered the important topic of estimating parameters for differential-equation models.
- We employed computational methods for solving differential equations and sensitivities, and solving nonlinear optimization problems in order to tackle this challenging problem.
- The linear, elliptical confidence intervals are approximate, but we can check their validity with Monte Carlo sampling.

## Closing

- I hope the examples and methods in this chapter serve to inspire students and practicing engineers to build models as part of understanding new processes and chemistries of interest.
- The modeling experience often leads to deeper process understanding and produces a compact summary of current knowledge that is easily and efficiently communicated to other colleagues and team members.
- Process understanding coupled with creativity often leads to process improvement and new discovery.



## Notation I

$A$	absorbance
$b$	path length for the absorbance measurement
$c$	concentration of absorbing species
$c_j$	concentration of component $j$
$\hat{c}_j(t_i)$	measured concentration of component $j$ at sampling time $t_i$
$D_p$	catalyst particle diameter
$D_t$	fixed-bed reactor tube diameter
$\mathbf{e}$	residual vector
$\mathbf{e}$	measurement error vector
$H_{kj}$	Hessian matrix, $H_{kj} = \partial^2 \Phi / \partial \theta_k \partial \theta_j$
$L$	fixed-bed reactor tube length
$m$	mass velocity
$n_d$	number of data points
$n_p$	number of model parameters
$N_j$	molar flow of component $j$
$p(\mathbf{x})$	probability density function of random variable $\mathbf{x}$
$Q$	volumetric flowrate
$Q_f$	feed volumetric flowrate
$r$	reaction rate of (single) reaction
$r_i$	reaction rate for $i$ th reaction
$\mathbf{r}$	reaction-rate vector








113 / 116

## Notation II

$R_j$	production rate for $j$ th species
$\mathbf{R}$	production-rate vector
$S$	open tube area for flow
$S_{jk}$	sensitivity of state $x_j$ with respect to parameter $\theta_k$
$t_i$	sampling time
$V_R$	reactor volume
$\mathbf{x}$	state vector
$\mathbf{y}$	vector of measured responses
$\epsilon$	molar absorptivity
$\epsilon_B$	bed porosity
$\boldsymbol{\theta}$	parameter vector
$\mu$	fluid viscosity
$\nu_{ij}$	stoichiometric number for the $j$ th species in the $i$ th reaction
$\phi$	transformed parameter vector
$\Phi$	objective function

114 / 116

## References I

-  Y. Bard.  
*Nonlinear Parameter Estimation.*  
Academic Press, New York, 1974.
-  G. E. P. Box, W. G. Hunter, and J. S. Hunter.  
*Statistics for Experimenters.*  
John Wiley & Sons, New York, 1978.
-  G. E. P. Box and G. C. Tiao.  
*Bayesian Inference in Statistical Analysis.*  
John Wiley & Sons, New York, 1973.
-  M. Caracotsios and W. E. Stewart.  
Sensitivity analysis of initial value problems with mixed ODEs and algebraic equations.  
*Comput. Chem. Eng.*, 9(4):359–365, 1985.
-  M. Caracotsios and W. E. Stewart.  
Sensitivity analysis of initial-boundary-value problems with mixed PDEs and algebraic equations.  
*Comput. Chem. Eng.*, 19(9):1019–1030, 1995.
-  C. F. Gauss.  
*Theory of the combination of observations least subject to errors: part one, part two, supplement.*  
SIAM, Philadelphia, 1995.  
Translated by G. W. Stewart.
-  M. A. Natal-Santiago, S. G. Podkolzin, R. D. Cortright, and J. A. Dumesic.  
Microcalorimetric studies of interactions of ethene, isobutene, and isobutane with silica-supported Pd, Pt, and PtSn.  
*Catal. Lett.*, 45:155–163, 1997.

115 / 116

## References II

-  S. G. Podkolzin and J. A. Dumesic.  
Adsorption data for H<sub>2</sub> on Pd.  
Personal communication, 2001.
-  W. E. Stewart, M. Caracotsios, and J. P. Sørensen.  
Parameter estimation from multiresponse data.  
*AIChE J.*, 38(5):641–650, 1992.
-  W. E. Stewart, T. L. Henson, and G. E. P. Box.  
Model discrimination and criticism with single-response data.  
*AIChE J.*, 42(11):3055–3062, 1996.
-  W. E. Stewart, Y. Shon, and G. E. P. Box.  
Discrimination and goodness of fit of multiresponse mechanistic models.  
*AIChE J.*, 44(6):1404–1412, 1998.

116 / 116