PARAMETER ESTIMATION CONCEPTS

James V. Beck, beck@msu.edu
Prof. Em., Dept. of Mech. Eng.
Michigan State University

This is an introduction to the topic.
See book, ñParameter Estimation in Engineering and Scienceò
by J.V. Beck and K.N. Arnold, Wiley, 1977

GENERAL TOPICS:

MEASUREMENT ERRORS

SCALED SENSITIVITY COEFFICIENTS

ESTIMATION METHODS

EXAMPLE

Parameter estimation is a bridge between experimentation and analysis/ modeling /computation

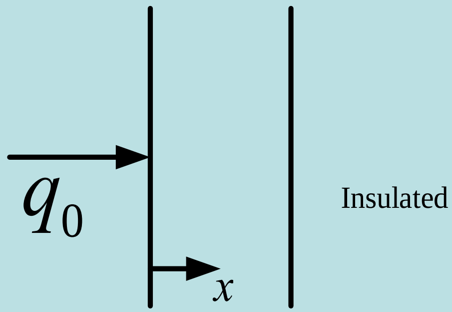
Parameter Estimation vs Function Estimation

Parameter Estimation

- Small number of parameters, 2, 3, 4,5
- Nonlinear even if ode or pde are linear
- Not ill-posed
- Parameters may be a physical property, e.g., thermal conductivity

Function Estimation

Large number of parameters, 50, 100 or even 1000
Might be linear if ode or pde are linear
Ill-posed
Parameters not a physical property, e.g., heat flux



Heat Conduction Example

$$k\frac{\partial^2 T}{\partial x^2} = C\frac{\partial T}{\partial t},\ \ 0 < x < L,\ \ t > 0$$

$$-k\frac{\partial T}{\partial x}(0,t) = q_0,\ \ \frac{\partial T}{\partial x}(L,t) = 0$$

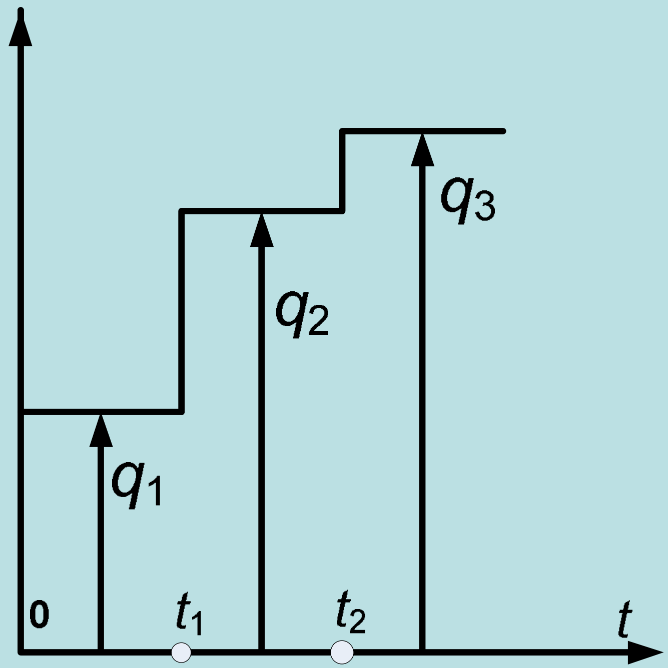
$$T(x,0) = 0$$

Parameter estimation problem: Estimate *k* and *C* given measurements of *T*(*x,t*) and q_0

k **= thermal conductivity, physical characteristic of a material, W/(m C)**

C **= volumetric heat capacity, physical characteristic of a material, J/(m^3 C)**

A FUNCTION ESTIMATION problem can be formed if *k* and C are known and T-measurements are available but $q_0 = q_0(t)$ is an unknown function



Where are **the measurement errors?**
How can they be described?

Eight Standard Statistical Assumptions

1. Additive errors,

$$Y_i = T_i + \varepsilon_i$$

2. Zero mean errors

$$E(\varepsilon_i) = 0$$

3. Constant standard deviation,

$$\sigma_i = \sigma = \text{constant}$$

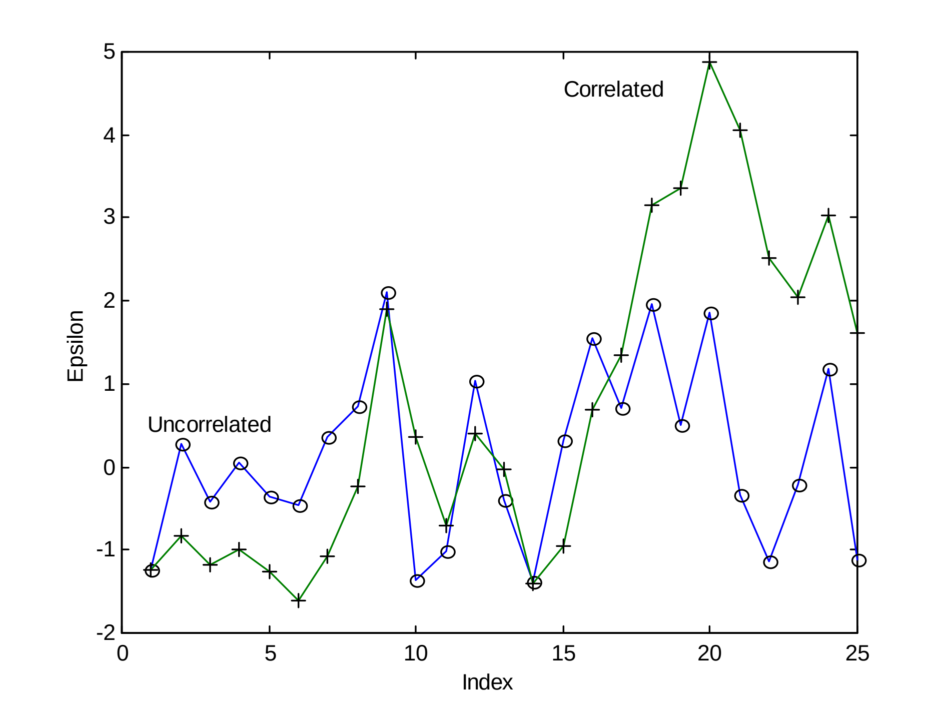
4. Uncorrelated errors

5. Normal distribution of errors

6. Known statistical parameters

7. Errors only in the dependent variable

8. No prior information



Uncorrelated, $\sigma = 1$

13 runs

Correlated, $\sigma = 1$

$\rho = 0.9$, 6 runs

$$\varepsilon_i = \rho \varepsilon_{i-1} + u_i$$

randnum

A PRINCIPLE OF PARAMETER ESTIMATION

We want to design experiments so that the scaled sensitivity coefficients are large and uncorrelated.

***i*th Sensitivity Coefficient,** $\frac{\partial T}{\partial \beta_i}$

***i*th Scaled Sensitivity Coefficient,** $\beta_i \frac{\partial T}{\partial \beta_i}$

Scaled Sensitivity Coefficients are important. Why??

COMPUTATION OF SCALED SENSITIVITY COEFFICIENTS

1. **ANALYTICAL DIFFERENTIATION, GOOD BUT ERROR-PRONE**
2. **FINITE DIFFERERCE, (Intrinsic Verification)**

$$\beta_j \frac{\partial T}{\partial \beta_j}$$

$$\approx \beta_j \frac{T(x,t,\beta_1,\cdots,\beta_j(1+\delta),\cdots,\beta_p) - T(x,t,\beta_1,\cdots,\beta_p)}{\beta_j(1+\delta)-\beta_j}$$

$$\approx \frac{T(x,t,\beta_1,\cdots,\beta_j(1+\delta),\cdots,\beta_p) - T(x,t,\beta_1,\cdots,\beta_p)}{\delta}$$

$$\delta = 0.0001 \text{ or } 0.00001$$

TAYLOR SERIES ON PARAMETER VALUES

$$Let\ \boldsymbol{\beta}_0 = (\beta_{10},\ \beta_{20},\ \beta_{30}, \cdots, \beta_{p0})$$

$$T(x,t,\boldsymbol{\beta}) \approx T(x,t,\boldsymbol{\beta}_0) + \left.\frac{\partial T}{\partial \beta_1}\right|_{\boldsymbol{\beta}_0} (\beta_1 - \beta_{10}) + \cdots + \left.\frac{\partial T}{\partial \beta_p}\right|_{\boldsymbol{\beta}_0} (\beta_p - \beta_{p0})$$

$$\approx T(x,t,\boldsymbol{\beta}_0) + \left(\beta_1 \left.\frac{\partial T}{\partial \beta_1}\right|_{\boldsymbol{\beta}_0} \right) \frac{\beta_1 - \beta_{10}}{\beta_1} + \cdots + \left(\beta_p \left.\frac{\partial T}{\partial \beta_p}\right|_{\boldsymbol{\beta}_0} \right) \frac{\beta_p - \beta_{p0}}{\beta_p}$$

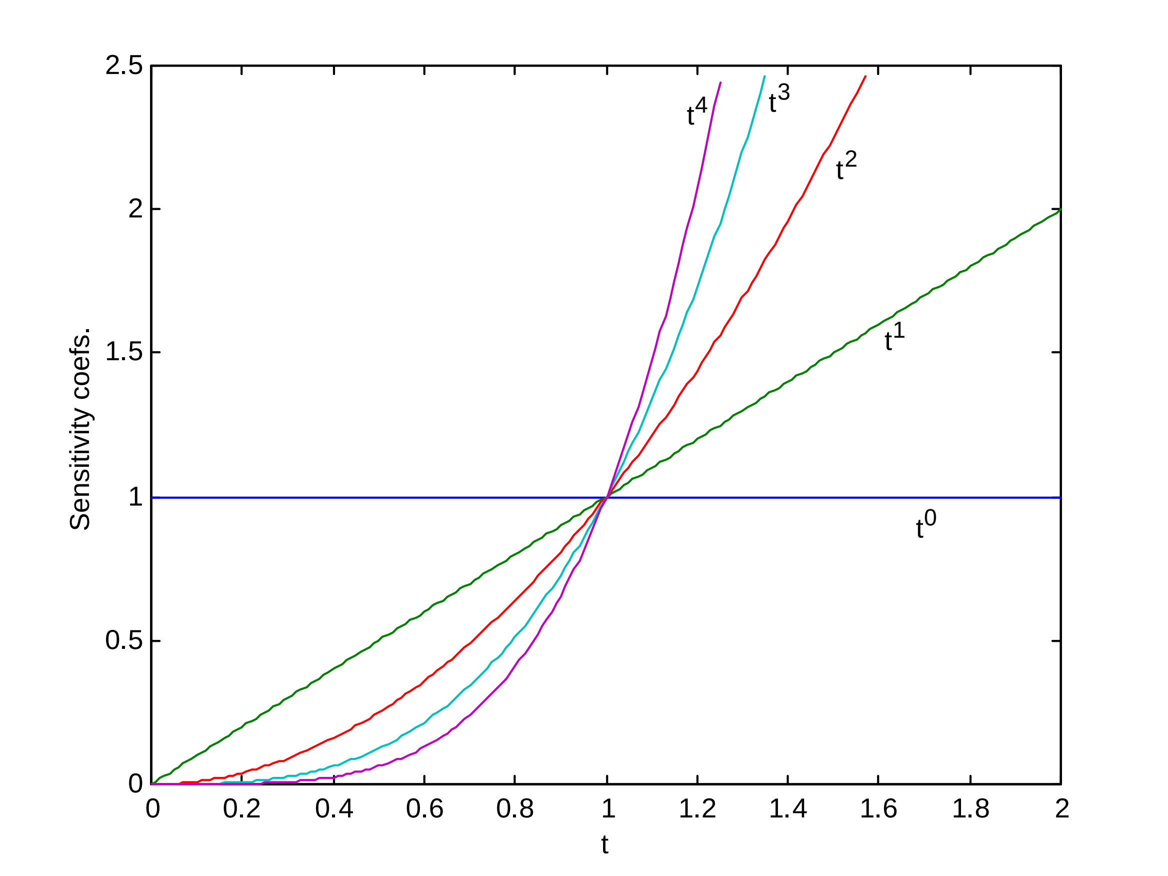
1st scaled sensitivity

***p*th scaled sensitivity**

Want sensitivity coefficients to be "large" compared to *T* and uncorrelated with each other

Polynomial model: $\eta_i = \beta_1 + \beta_2 t_i \; + \beta_3 t_i^2 + \beta_4 t_i^3 + \beta_3 t_5^4$

$$\frac{\partial \eta_i}{\partial \beta_1} = 1, \;\; \frac{\partial \eta_i}{\partial \beta_2} = t_i, \;\; \frac{\partial \eta_i}{\partial \beta_3} = t_i^2, \;\; \frac{\partial \eta_i}{\partial \beta_4} = t_i^3, \;\; \frac{\partial \eta_i}{\partial \beta_5} = t_i^4$$



t^3 and t^4 sens. coef. becoming more correlated.

IDENTIFIABILITY CONDITION:

To estimate parameters, must <u>not</u> have:

$$A_1\beta_1 \frac{\partial T}{\partial \beta_1} + A_2\beta_2 \frac{\partial T}{\partial \beta_2} + \cdots + A_p\beta_p \frac{\partial T}{\partial \beta_p} = 0$$

for not all the A_i values equal to zero..

T **= temperature/dependent variable**

β_i **= *i*th parameter**

p **= no. of parameters**

A_i = dimensionless constant, like 0, +1, -1

If we can show that

$$\beta_1 \frac{\partial T}{\partial \beta_1} + \beta_2 \frac{\partial T}{\partial \beta_2} + \cdots + \beta_p \frac{\partial T}{\partial \beta_p} \neq 0$$

it may be possible to estimate all the parameters simultaneously. Can we show that? Consider the same transient heat conduction example,

$$k \frac{\partial^2 T}{\partial x^2} = C \frac{\partial T}{\partial t}, \quad T(x,0) = 0$$

$$-k \frac{\partial T}{\partial x}(0,t) = q_0, \quad \frac{\partial T}{\partial x}(L,t) = 0$$

DERIVATION OF SUM OF SCALED SENSITIVITY COEFFICIENTS RELATIONSHIP

Consider the same problem in dimensionless form:

$$\tilde{T} = \frac{T}{\frac{q_0 L}{k}}, \; \tilde{x} = \frac{x}{L}, \; \tilde{t} = \frac{kt}{CL^2}, \quad T = \frac{q_0 L}{k}\tilde{T}\left(\tilde{x}, \tilde{t}(k, C, L, t)\right)$$

Find *k* sensitivity coefficient using chain rule

$$\left.\frac{\partial T}{\partial k}\right|_{x,t,C\ fixed} = -\frac{q_0 L}{k^2}\tilde{T} + \frac{q_0 L}{k}\left[\left.\frac{\partial \tilde{T}}{\partial \tilde{t}}\right|_{\tilde{x}} \frac{\partial \tilde{t}}{\partial k}\right]$$

$$= -\frac{q_0 L}{k^2}\tilde{T} + \frac{q_0 L}{k}\left[\left.\frac{\partial \tilde{T}}{\partial \tilde{t}}\right|_{\tilde{x}} \frac{t}{CL^2}\right]$$

Multiply by *k* to get

$$k\frac{\partial T}{\partial k}\bigg|_{x,t,C\ fixed} = -\frac{q_0 L}{k}\tilde{T} + \frac{q_0 L}{k}\tilde{t}\,\frac{\partial \tilde{T}}{\partial \tilde{t}}\bigg|_{\tilde{x}} = -T + \frac{q_0 L}{k}\tilde{t}\,\frac{\partial \tilde{T}}{\partial \tilde{t}}\bigg|_{\tilde{x}}$$

Repeat for *C* sensitivity coefficient

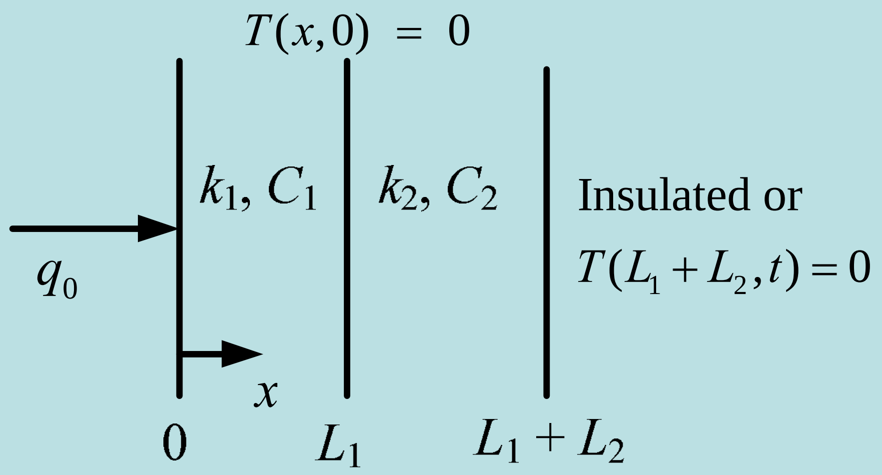
$$\frac{\partial T}{\partial C}\bigg|_{x,t,k\ fixed} = \frac{q_0 L}{k}\frac{\partial \tilde{T}}{\partial \tilde{t}}\bigg|_{\tilde{x}}\frac{\partial \tilde{t}}{\partial C} = \frac{q_0 L}{k}\frac{\partial \tilde{T}}{\partial \tilde{t}}\bigg|_{\tilde{x}}\left(-\frac{kt}{C^2 L^2}\right)$$

$$C\frac{\partial T}{\partial C}\bigg|_{x,t,k\ fixed} = -\frac{q_0 L}{k}\tilde{t}\,\frac{\partial \tilde{T}}{\partial \tilde{t}}\bigg|_{\tilde{x}}$$

Add these two scaled sensitivities to get

$$k\frac{\partial T(x,t)}{\partial k} + C\frac{\partial T(x,t)}{\partial C} = -T(x,t)$$

ANOTHER EXAMPLE, TWO PLATES/MATERIALS



We can derive:

$$k_1 \frac{\partial T(x,t)}{\partial k_1} + C_1 \frac{\partial T(x,t)}{\partial C_1} + k_2 \frac{\partial T(x,t)}{\partial k_2} + C_2 \frac{\partial T(x,t)}{\partial C_2} = -T(x,t)$$

For temperature boundary condition at x = 0:

$$k_1 \frac{\partial T(x,t)}{\partial k_1} + C_1 \frac{\partial T(x,t)}{\partial C_1} + k_2 \frac{\partial T(x,t)}{\partial k_2} + C_2 \frac{\partial T(x,t)}{\partial C_2} = 0$$

Insights from Sensitivities Summation Relation

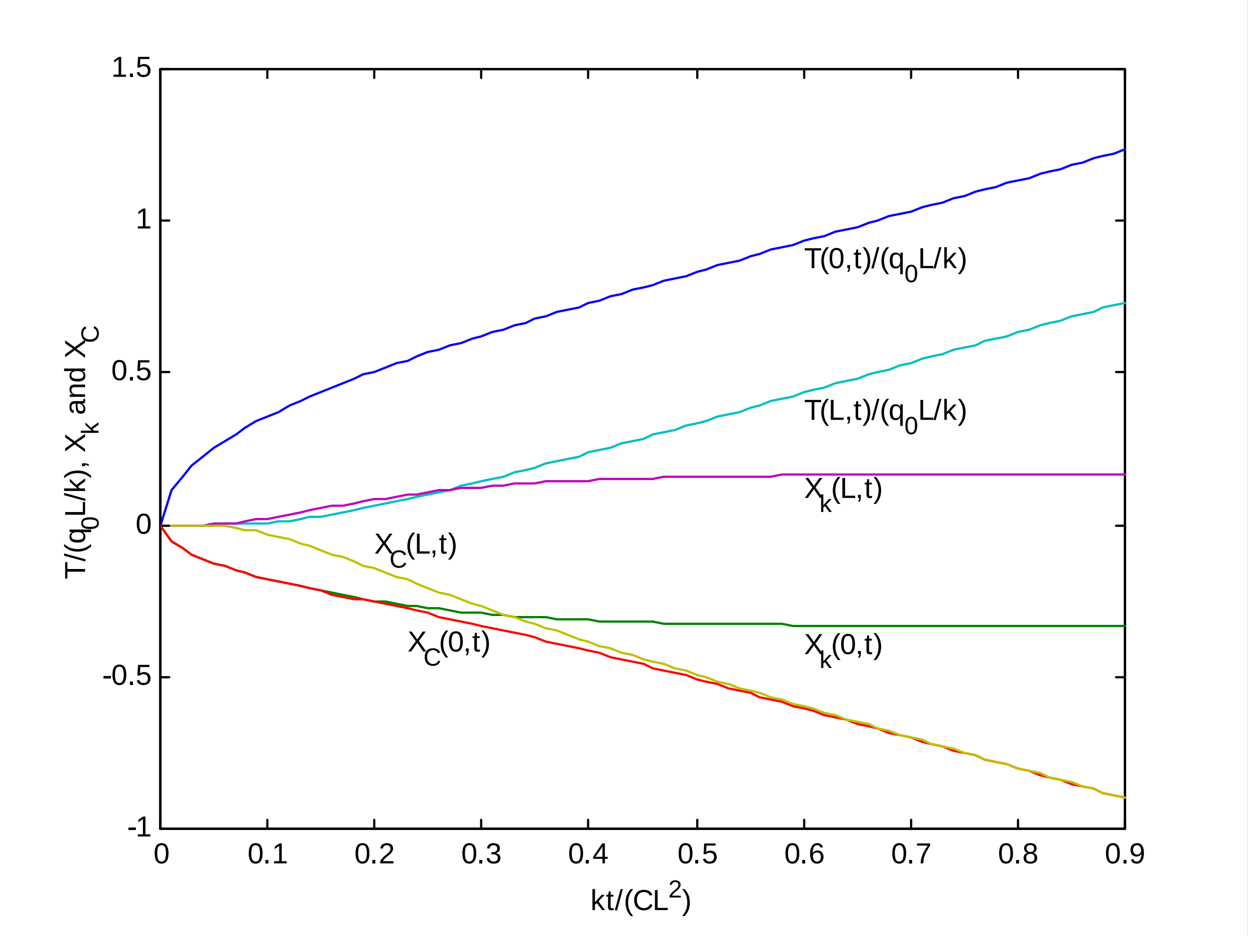
1. If *T* is not zero, scaled sensitivities might be linearly independent. Then *k* and *C* might be simultaneously and independently estimated.

2. For *T* > 0, sum of scaled sensitivities < 0. Sensitivities tend to be negative for *T* > 0. If each sensitivity < 0, each sensitivity is less than abs(*T*)

3. As more parameters are estimated, it becomes more difficult since the sum is still –*T.*

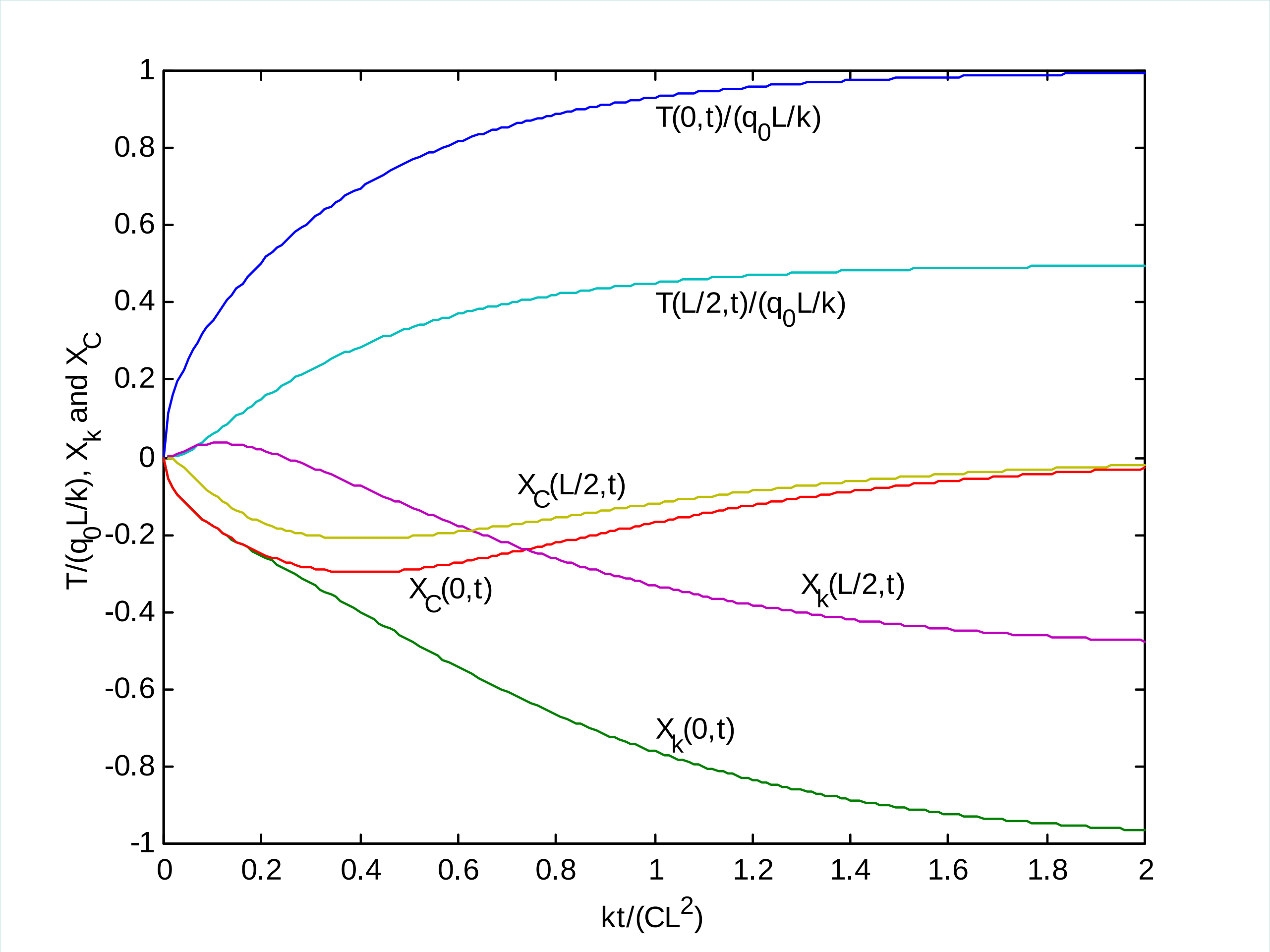
T & scaled sens. coefficients for single material case

$$X_k \equiv \frac{1}{q_0 L / k} k \frac{\partial T}{\partial k}, \; X_C \equiv \frac{1}{q_0 L / k} C \frac{\partial T}{\partial C}$$



Temperature and scaled sensitivity coefficients for X21B10T0 case (const. heat flux @ *x* = 0, *T*(*L*,*t*) = 0)

$$X_k \equiv \frac{1}{q_0 L / k} k \frac{\partial T}{\partial k}, \; X_C \equiv \frac{1}{q_0 L / k} C \frac{\partial T}{\partial C}$$



OBSERVATIONS – SENS. COEF.

1. Sensitivity coefficients tend to be negative.
2. Coefficients tend to be "large."
3. Coefficients have different shapes, that is, are uncorrelated.

However, for measurements just at x = 0 and until dimensionless time until about 0.2

$$X_k = X_C \text{ or } k\frac{\partial T}{\partial k} - C\frac{\partial T}{\partial C} = 0$$

Then, k and C cannot be estimated.

PARAMETER ESTIMATION PROBLEMS ARE NONLINEAR IF THE SENSITIVITY COEFFICIENTS ARE FUNCTIONS OF THE PARAMETERS

Linear model in the parameters:

$$\eta_i = \beta_1 + \beta_2 t_i$$

$\eta_i =$ Dependent variable

$\beta_1,\ \beta_2 =$ Parameters

$t_i =$ Independent variable

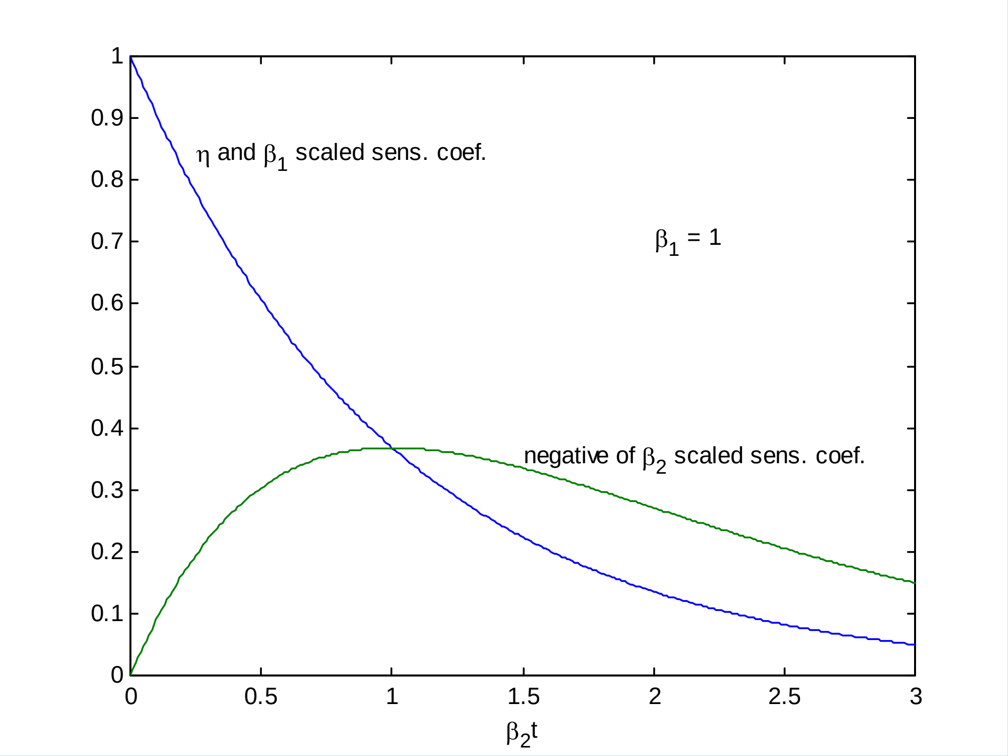
$$\frac{\partial \eta_i}{\partial \beta_1} = 1, \quad \frac{\partial \eta_i}{\partial \beta_2} = t_i,$$

Not functions of parameter, then linear estimation problem

Nonlinear model in the parameters:

$$\eta_i = \beta_1 e^{-\beta_2 t_i}, \qquad \frac{\partial \eta_i}{\partial \beta_1} = e^{-\beta_2 t_i}, \quad \frac{\partial \eta_i}{\partial \beta_2} = -\beta_1 t_i e^{-\beta_2 t_i}$$

Since sens. coef. functions of parameters, the parameter estimation problem is nonlinear..



HEAT CONDUCTION EXAMPLE, *k* and *C* parameters

$$k\frac{\partial^2 T}{\partial x^2} = C\frac{\partial T}{\partial t}, \quad -k\frac{\partial T}{\partial x}(0,t) = q_0, \; \frac{\partial T}{\partial x}(L,t) = 0, \; T(x,0) = 0$$

$$T = T(x,t,k,C,q_0,L), \quad \text{Let } X_C \equiv \left.\frac{\partial T}{\partial C}\right|_{k\ fixed}$$

$$k\frac{\partial^2 X_C}{\partial x^2} = \frac{\partial T}{\partial t} + C\frac{\partial X_C}{\partial t}, \quad -k\frac{\partial X_C}{\partial x}(0,t) = 0, \; \frac{\partial X_C}{\partial x}(L,t) = 0$$

$$X_C(x,0) = 0$$

NONLINEAR P. E. SINCE X_c IS FUNCTION OF PARAMETERS.

ESTIMATION PROCEDURES,

1. ORDINARY LEAST SQUARES

$$S = \sum_{i=1}^{n} \left(Y_i - \eta_i \right)^2 = \left(\mathbf{Y}\text{-}\boldsymbol{\eta} \right)^T \left(\mathbf{Y}\text{-}\boldsymbol{\eta} \right)$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \text{Meas. vector, } \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_n \end{bmatrix} = \text{model vector}$$

For a model linear in terms of the parameters

$$\eta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} \quad \text{or} \quad \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}$$

Sensitivity matrix

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Parameter vector

ORDINARY LINEAR LEAST SQUARES ESTIMATOR

$$\mathbf{b}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Excellent estimator if standard statistical are valid

NOT ALL STANDARD STAT. ASSUMPTIONS VALID:

USE MAXIMUM LIKELIHOOD IF STANDARD DEVIATION IS NOT CONSTANT, ERRORS CORRELATED & NORMAL

Need covariance matrix

$$\boldsymbol{\psi} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

$\sigma_i^2 =$ variance of $Y_i = \eta_i + \varepsilon_i = V(Y_i) = V(\varepsilon_i)$

$\sigma_i =$ standard deviation

$\sigma_{ij} = \operatorname{cov}\left(Y_i, Y_j\right)$

$$S_{ML} = \left(\mathbf{Y}\text{-}\boldsymbol{\eta}\right)^T \boldsymbol{\psi}^{-1} \left(\mathbf{Y}\text{-}\boldsymbol{\eta}\right)$$

For Linear Model: $\mathbf{b}_{ML} = (\mathbf{X}^T \boldsymbol{\psi}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\psi}^{-1} \mathbf{Y}$

IF PRIOR INFORMATION IS AVAILABLE, USE MAXIMUM A POSTERIORI ESTIMATION

$\boldsymbol{\mu}_{\boldsymbol{\beta}}$ = prior parameter values

$\mathbf{V}_{\boldsymbol{\beta}}$ = covariance matrix of the prior parameters

For linear-in-parameters model,

$$\mathbf{b}_{MAP} = \boldsymbol{\mu}_{\boldsymbol{\beta}} + (\mathbf{X}^T \boldsymbol{\psi}^{-1} \mathbf{X} + \mathbf{V}_{\boldsymbol{\beta}}{}^{-1})^{-1} \mathbf{X}^T \boldsymbol{\psi}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\mu}_{\boldsymbol{\beta}})$$

We have given estimators for linear-in-parameter models. Usually non-linear, however.

NONLINEAR-IN-PARAMETERS ESTIMATION

One way is Gauss linearization.

$$S = \left(\mathbf{Y}\text{-}\boldsymbol{\eta}(\boldsymbol{\beta})\right)^T \mathbf{W}\left(\mathbf{Y}\text{-}\boldsymbol{\eta}(\boldsymbol{\beta})\right) + (\boldsymbol{\mu}_{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

$$\mathbf{W} = \text{ Weighting matrix}$$

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \mathbf{P}^{(k)}\left[\mathbf{X}^{T(k)}\mathbf{W}\left(\mathbf{Y}\text{-}\boldsymbol{\eta}^{(k)}\right)\right] + \mathbf{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\mu}_{\boldsymbol{\beta}} - \mathbf{b}^{(k)})$$

$$\mathbf{P}^{-1(k)} = \mathbf{X}^{T(k)}\mathbf{W}\mathbf{X}^{(k)} + \mathbf{V}_{\boldsymbol{\beta}}^{-1}$$

MODIFICATIONS TO GAUSS LINEARIZATION

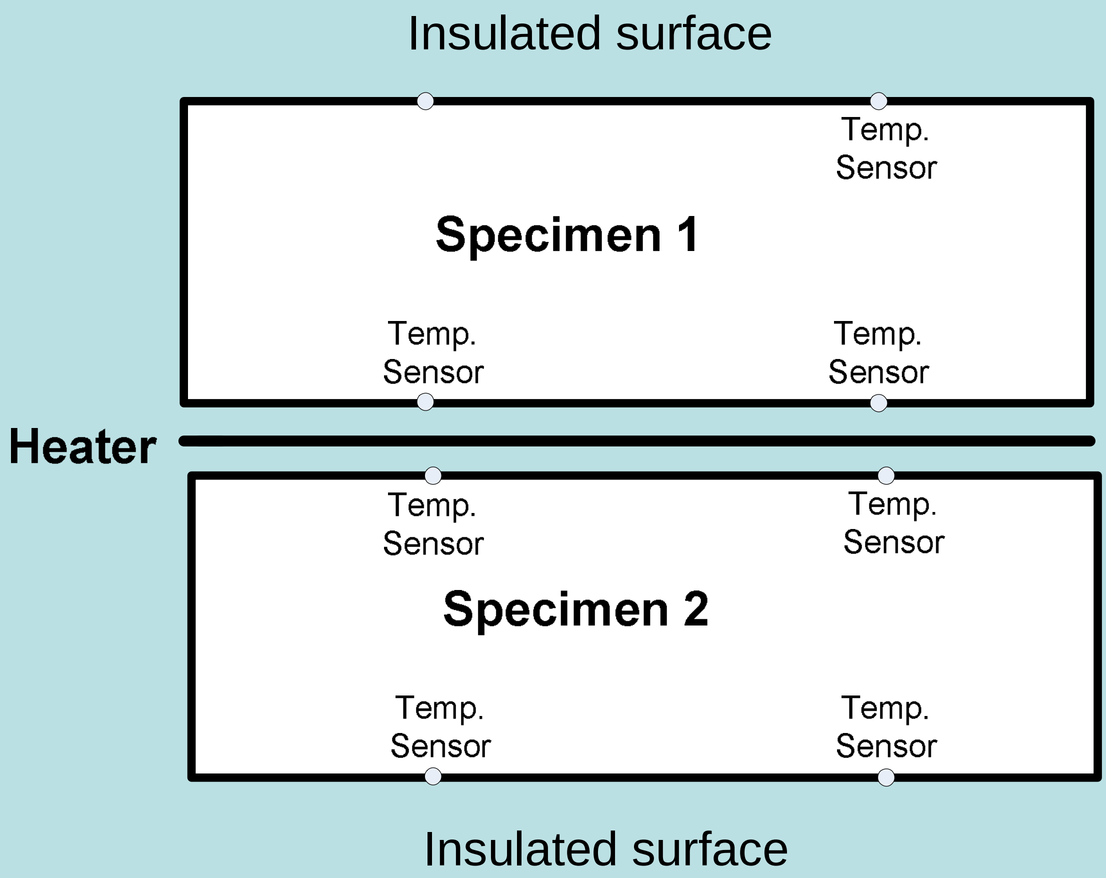
1. Use prior information (maybe non-informative prior)

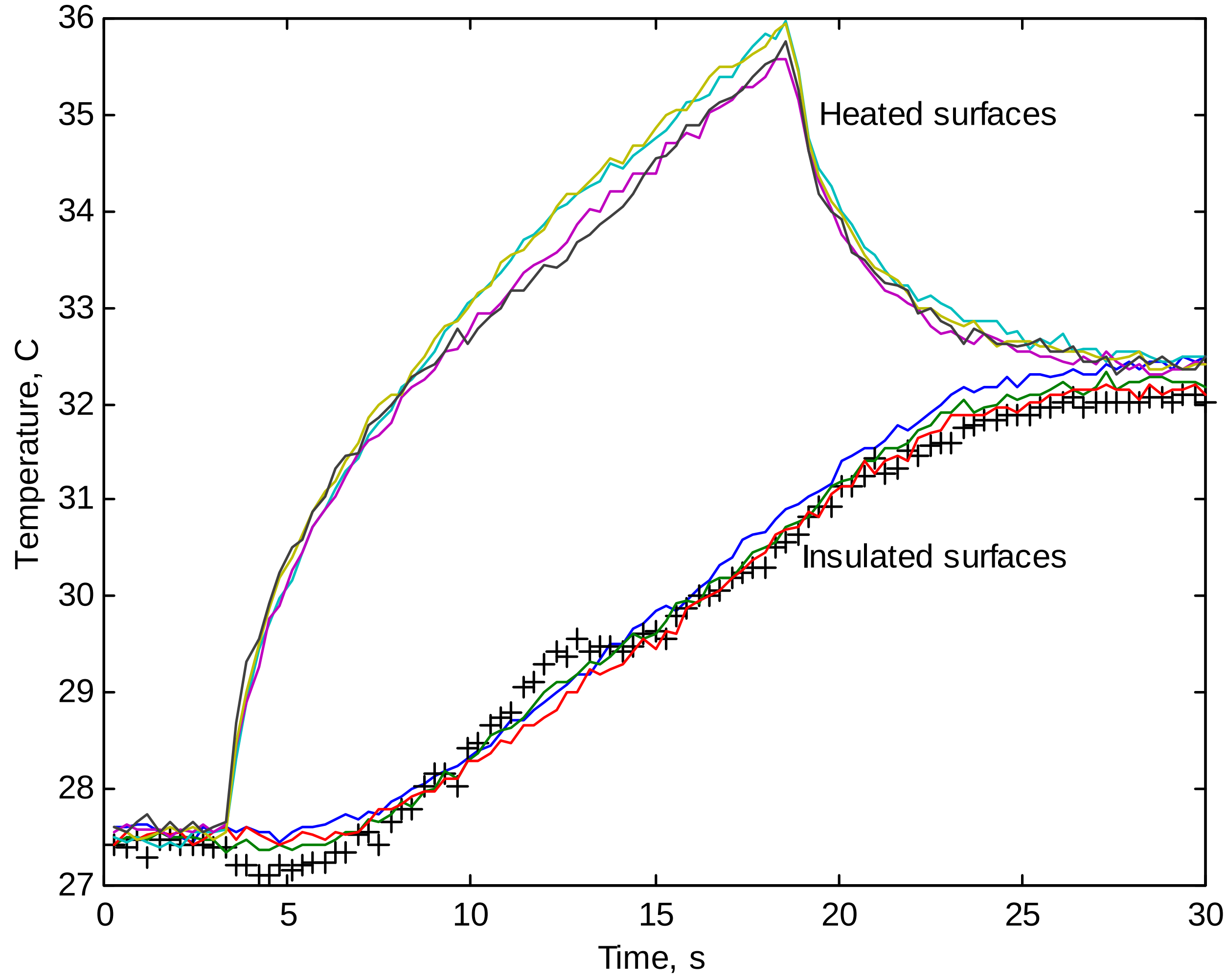
2. Modify step size (Box-Kanemasu)

3. Do sequentially, add one measurement at a time

Transient Heat Conduction Experiment to find *k* and *C*

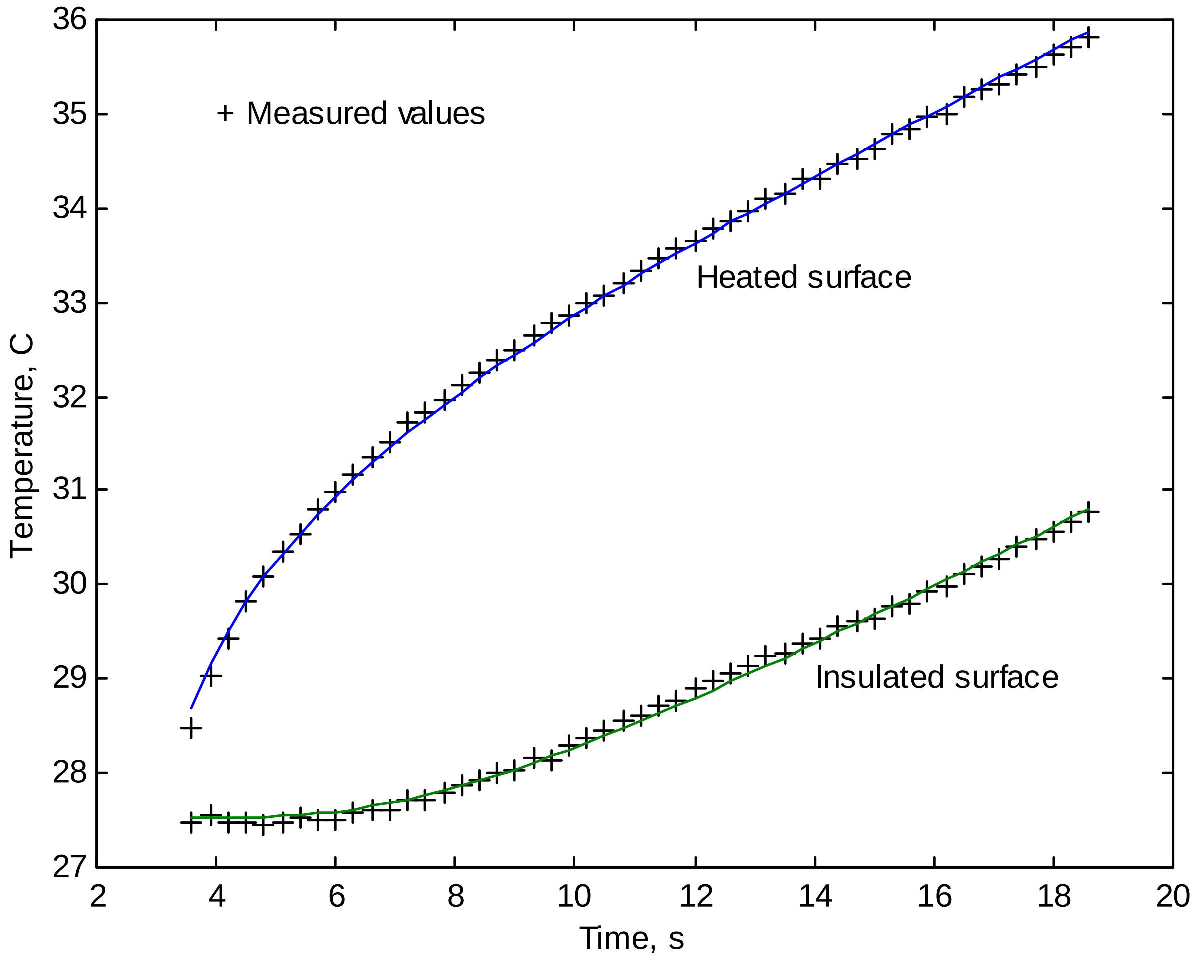
Specimens 0.0254 m thick, 0.076 m diameter, q_0 = 30,300 W/m^2





Measured *T*s. Curve with + sign inaccurate. Drop its data.

data1977kC



Averaged measured & computed *T* values. Good agreement.

data1977kCpl.m Experimant k L

Estimated parameter values of Armco iron

Using PROP1D

Thermal conductivity,   75.4 +- 1.2 W/m-C, +-1.6%

Volumetric heat capacity, 3,700,000 +-56,000 J/m^3-C, +-1.5%

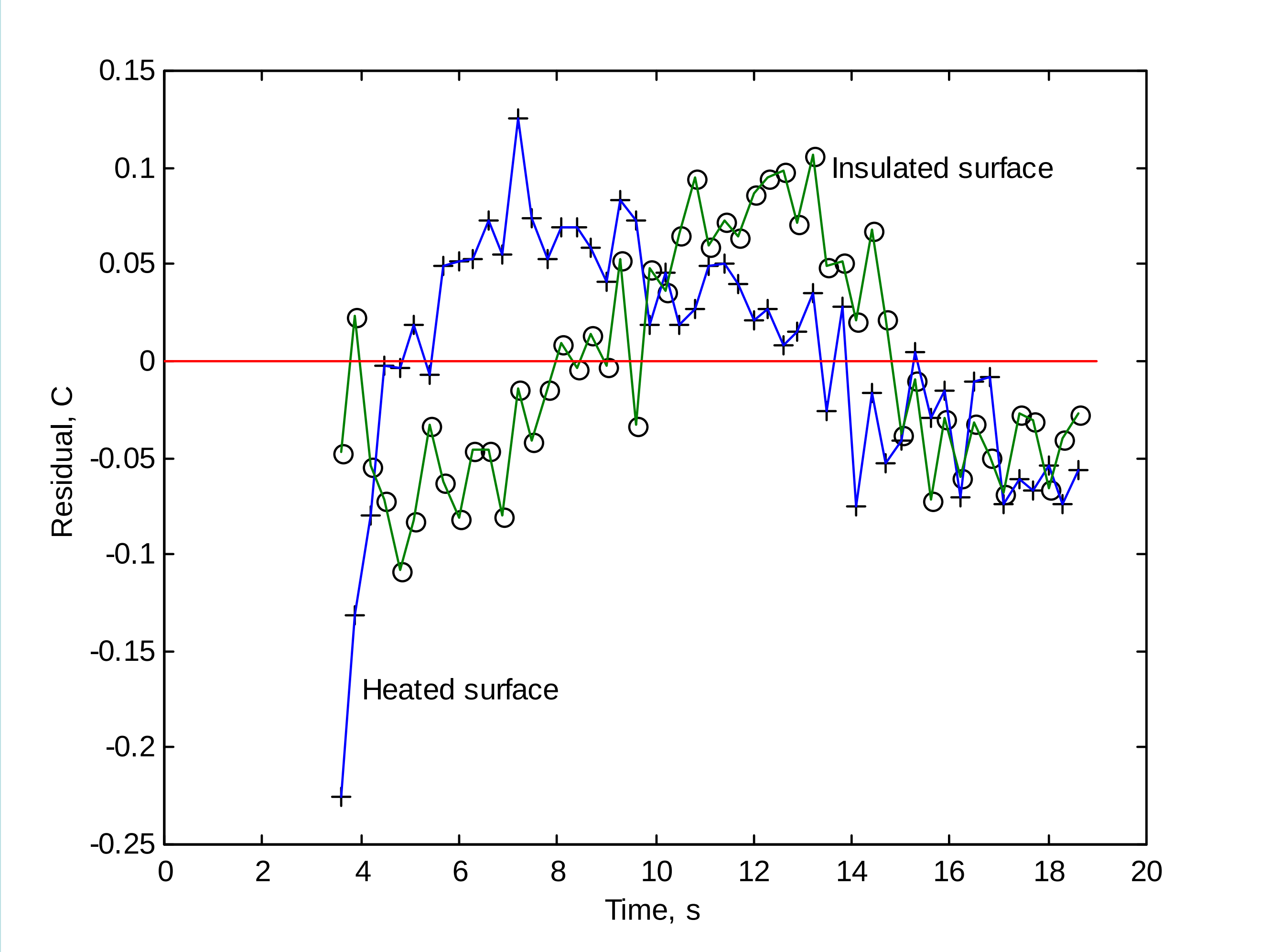
RMS,   0.0599 C

Incropera and DeWitt 99.75 % pure Armco iron at 27 C

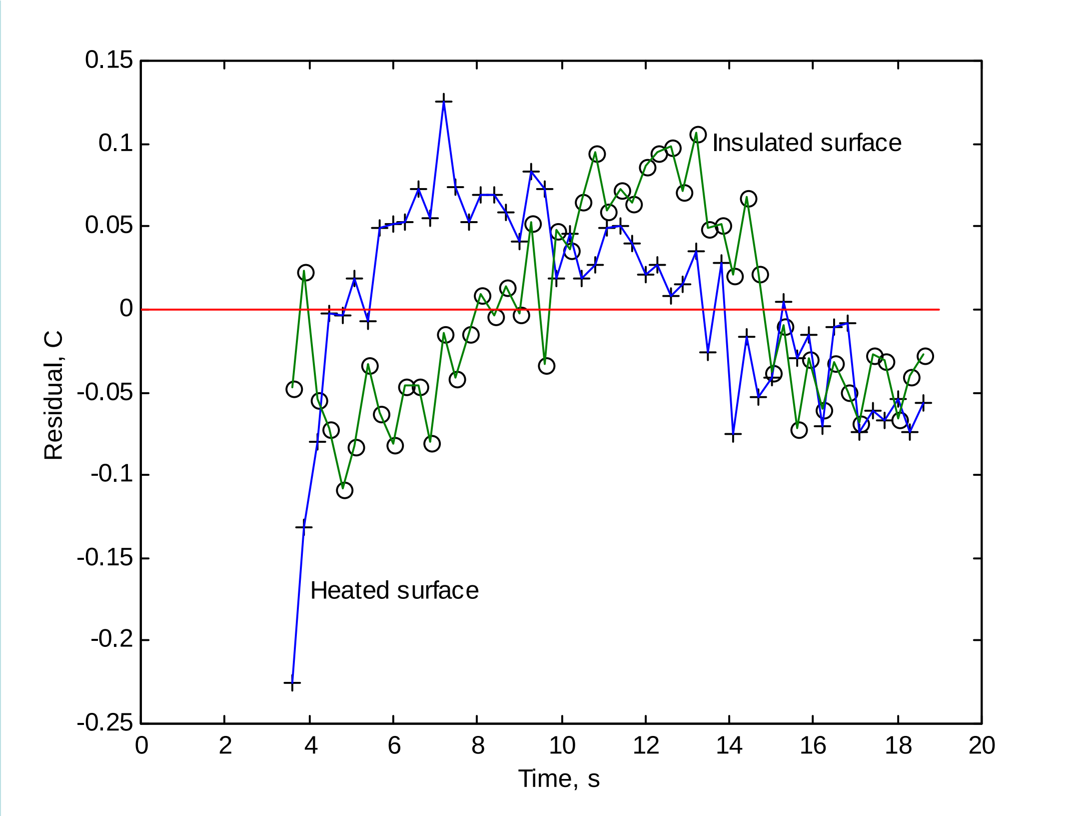
Thermal conductivity 72.7 W/m-C, within 3.6%

Volumetric heat capacity, 3,520,000 J/m3-C, within 4.9%

ITER	RMS	TH COND	VOL HEAT
0		50	.25E+7
1	2.092	66.83	.3314E+07
2	0.522	74.39	.3664E+07
3	0.079	75.35	.3706E+07
4	0.060	75.36	.3706E+07

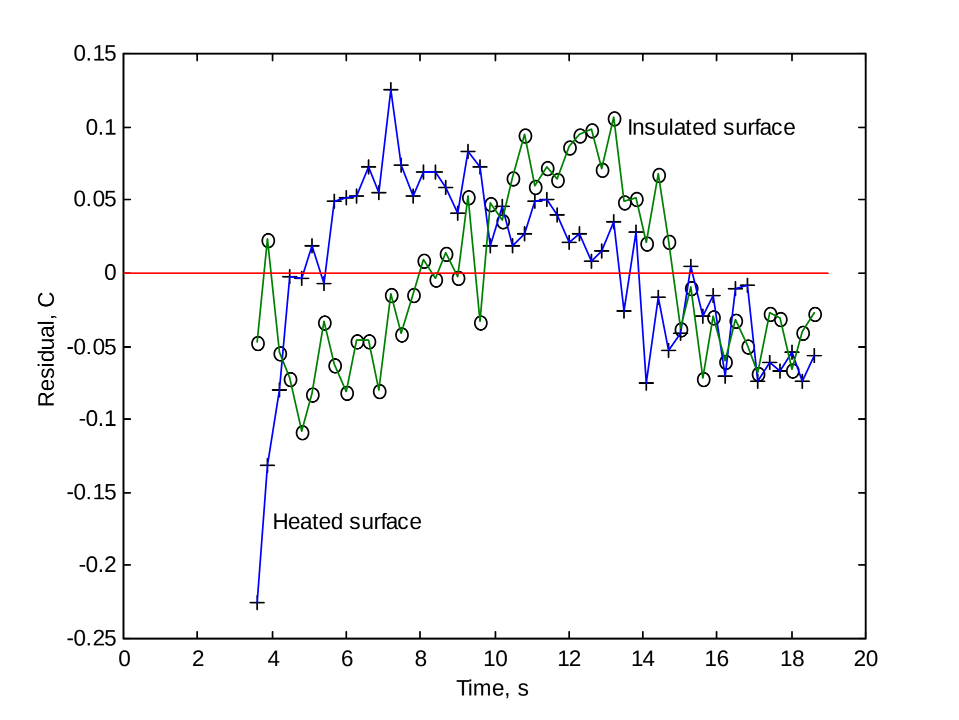


Residuals. What can we learn? Compare with std. assumptions



Comparison with standard statistical assumptions

1. **Additive errors. O.K.**
2. **Zero mean errors. O.K.**
3. **Constant variance. O.K.**
4. **Non-correlated errors. Not valid**
5. **Gaussian errors. O.K.**



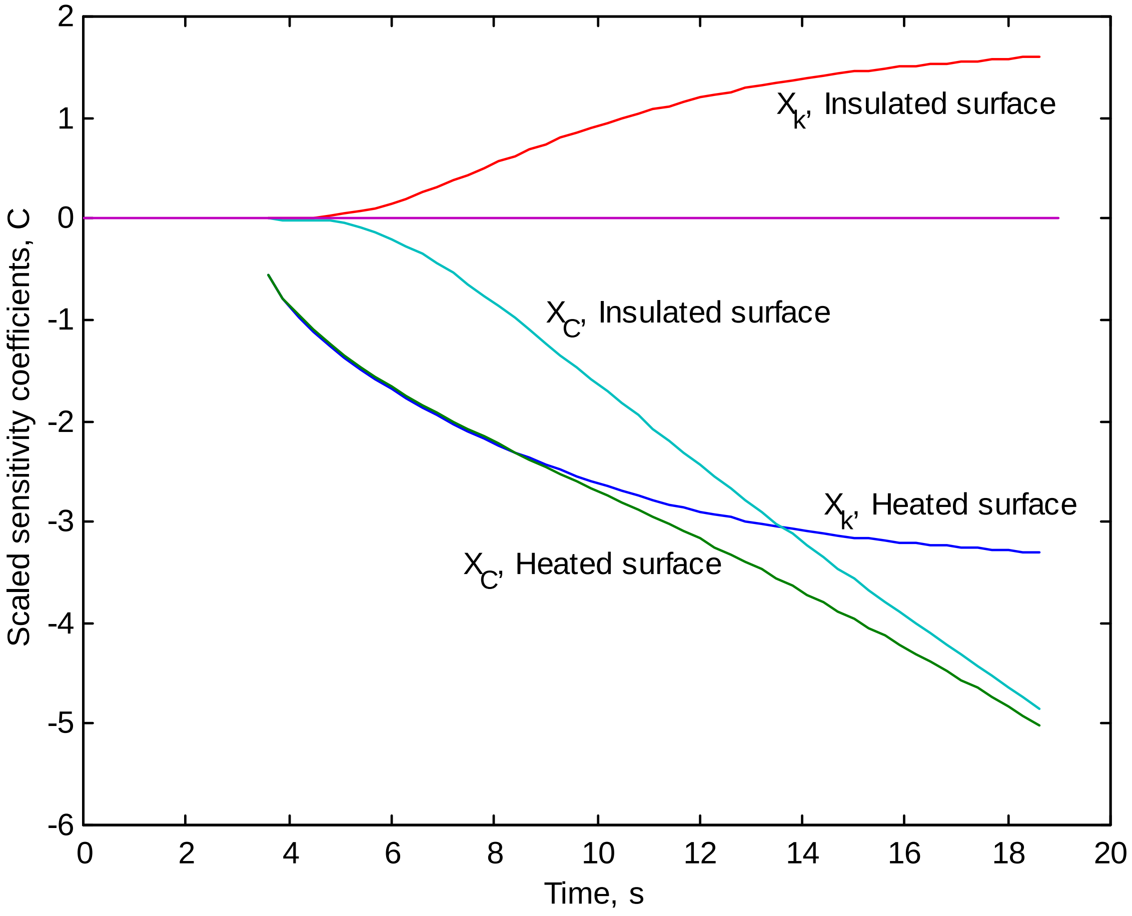
Significant features:

1. **Maximum magnitude residual is at 1st time at heated surf.**

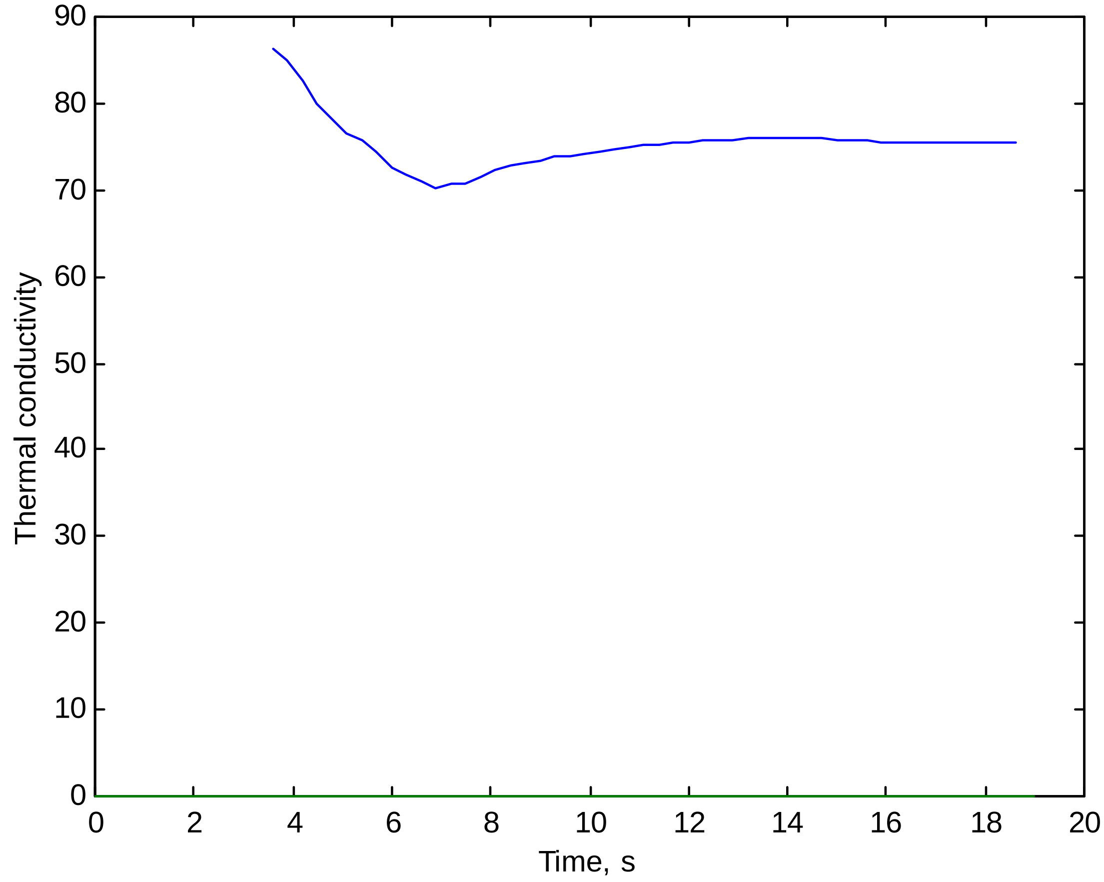
   Heater not modeled. Starting time may not be precisely known

2. **Characteristic "signature" of both $x = 0$ and L residuals are**

   negative at final times. Drift lower. Heat loss not modeled?

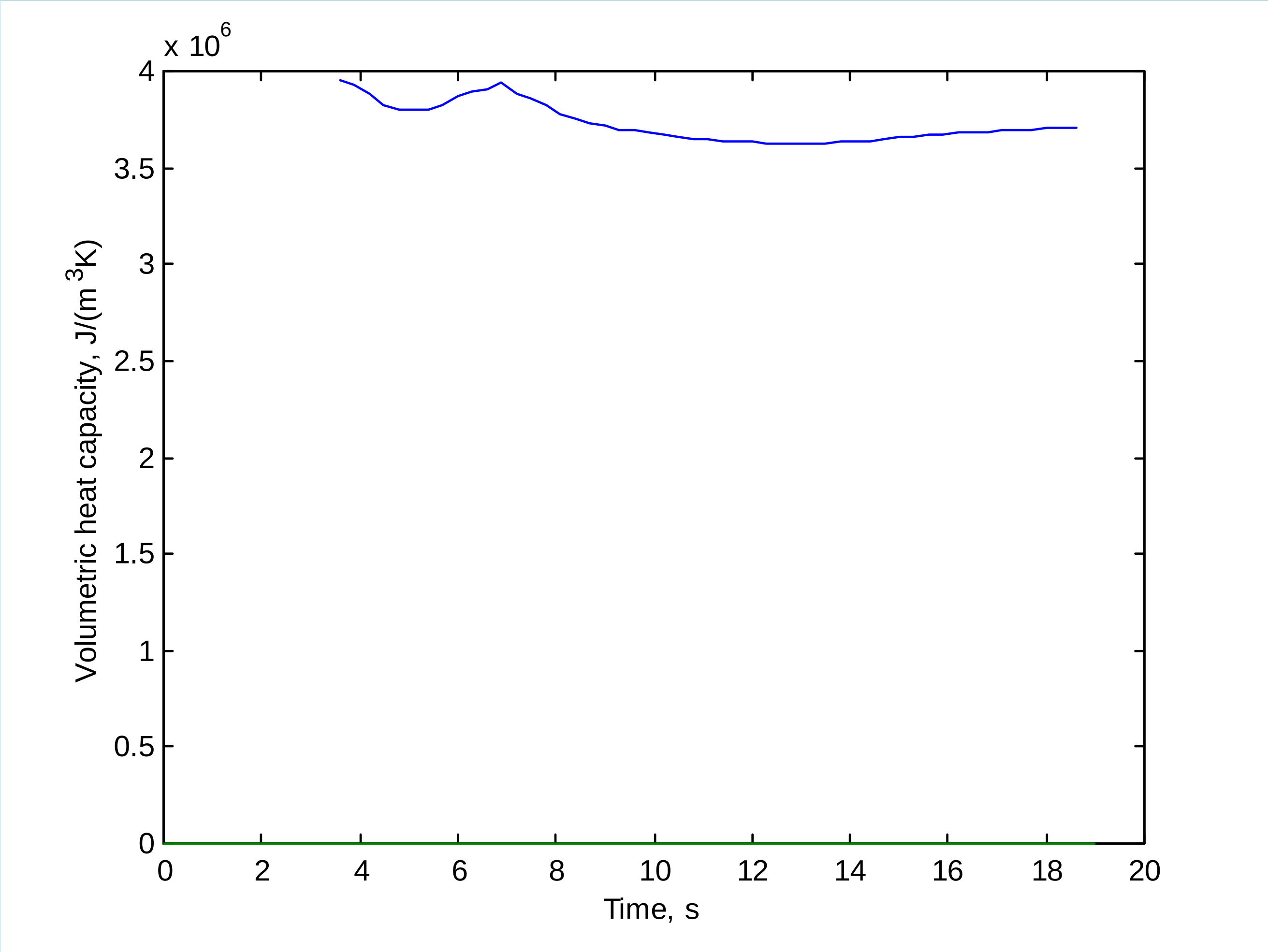


Scaled sensitivity coefficients, C
Time, s
X_k, Insulated surface
X_C, Insulated surface
X_k, Heated surface
X_C, Heated surface
2
1
0
-1
-2
-3
-4
-5
-6
0
2
4
6
8
10
12
14
16
18
20



Time	*k*
9.9	74.20
12.0	75.54
14.1	75.91
16.2	75.55
18.6	75.36
	2.3%

SEQUENTIAL *k* ESTIMATES. Sensitivity coefficients evaluated at converged *k* & *C* values.



Time	$C \times 10^{-6}$
9.9	3.67
12.0	3.63
14.1	3.64
16.2	3.68
18.6	3.71

2.2%

SEQUENTIAL *C* ESTIMATES.

TWO MORE IMPORTANT TOPICS:

1. Confidence regions

2. Optimal experiment design

Both use sensitivity coefficients.

Consider briefly OPTIMAL EXPERIMENTS

Let $\mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}$ **= sensitivity matrix**

Criterion: Maximize $\left|\mathbf{X}^T \mathbf{X}\right|$ **Subject to constraints**

Possible contraints:

1. Given No. of measurement points
2. Given range of dependent variable
3. Cost or whatever

Adjustable conditions

1. Location and no. of sensors
2. Duration of experiment
3. Type of boundary conditions
4. Time variation of boundary conditions

CONCLUSIONS

1. It is important learn about and characterize the measurement errors. Should check assumptions using residuals.
2. The scaled sensitivity coefficients can give important insight into the design of experiments. Also aid in determining which parameters can be estimated.
3. Do not over-parameterize! Adding parameters makes estimation more difficult
4. Some actual experiments are described and related sensitivity coefficients given. Sequential analysis can give insight into results.
5. The residuals are important. Examine to determine if any characteristic signatures are present. Should modedl be improved?