Smoothing Data with Roughness Penalties

Why do we use
Defining smoothness
Penalized least
Spline Smoothing
Choosing smoothing
A simulation study
Confidence limits
Summary
Home Page
Title Page
•• ••
< ▶
Page 1 of 100
Go Back
Eull Scroop
Full Screen
Close
Quit

1. Why do we use roughness penalties?



Why do we use . . . Defining smoothness Penalized least . . .

- Controlling smoothness by limiting the number of basis functions is discontinuous; roughness penalties allow continuous control over smoothness.
- We want to be able to define "smooth" in ways that are appropriate to our problems.
 - We may want a smooth derivative rather than just a smooth function.
 - What is smooth in one situation is not smooth in another. Smoothness has to be defined differently for periodic functions, for example.
- We find that roughness penalty smoothing gives better results.
- Roughness penalties are connected to fitting data by a differential equation; they are models for process dynamics.



2. Defining roughness

Why do we use	
Defining smoothness	
Penalized least	
Spline Smoothing	
Choosing smoothing	
A simulation study	
Confidence limits	
Summary	
Home Page	
Title Page	
•• ••	
• •	
Page 4 of 100	
Go Back	
Full Screen	
Close	
01036	

We have two competing objectives:

- 1. Fit the data well; keep bias low.
- 2. Keep the fit smooth so as to
 - filter out noise
 - get better estimates of derivatives

Mean squared error = $Bias^2 + Sampling Variance$

We can often greatly reduce MSE by trading a little bias off against a lot of sampling variance.

Defining smoothness
Penalized least
Spline Smoothing
Choosing smoothing
A simulation study
Confidence limits
Summary
Home Page
Title Page
•• ••
• •
▲ ▶ Page 5 of 100
 ✓ ✓ Page 5 of 100
Page 5 of 100 Go Back
Page 5 of 100 Go Back
Page 5 of 100 Go Back Full Screen
Page 5 of 100 Go Back Full Screen Close
Page 5 of 100 Go Back Full Screen Close
Page 5 of 100 Go Back Full Screen Close Quit
Page 5 of 100 Go Back Full Screen Close Quit

Quantifying roughness

• The classic: curvature in the function

$$\texttt{PEN}_2(x) = \int [D^2 x(s)]^2 \, ds$$

 $[D^2x(s)]^2$ measures the *squared curvature* in x at s. This penalty measures squared total curvature.

• Curvature in acceleration:

$$\operatorname{pen}_4(x) = \int [D^4 x(s)]^2 \, ds$$

• These two penalties also define what we mean by "smooth"; any function that has zero penalty is "hypersmooth." A straight line for the classic, a cubic polynomial for the acceleration penalty.



Harmonic acceleration

- If the process is periodic, it is natural to think of a constant + sinusoid as "hyper-smooth".
- This suggests that we use

$$\operatorname{PEN}_{H}(x) = \int [D^{3}x(s) + \omega^{2}Dx(s)]^{2} ds$$

where $2\pi/\omega$ is the period.

- The functions $1, sin(\omega t)$, and $cos(\omega t)$ all have zero penalties, as does any linear combination of them.
- Writing

$$Lx(s) = D^3x(s) + \omega^2 Dx(s)$$

we have

$$\operatorname{PEN}_H(x) = \int [Lx(s)]^2 \, ds$$

Some questions to think about

- Can we think of other *differential operators L* that might be useful?
- If we have a small number of "hyper-smooth" functions in mind, can we find a differential operator *L* that will assign zero penalty to them?
- Can use the data themselves to tell us something about the right differential operator *L*?

Why do we u	se	
Defining smoothness		
Penalized least		
Spline Smoothing		
Choosing smoothing		
A simulation study		
Confidence li	Confidence limits	
Summary		
Ноте	Page	
Title	Page	
44 >>		
••	••	
••	••	
••	>>	
	>>>	
▲▲ Page &	••• • 8 of 100	
Image €	••• • of 100	
Image €	Image: Solution of the solution	
↓ ↓ Page & Go ↓	••• • of 100 Back	
Image 8 Go a Full 5	B of 100 Back Gcreen	
Image € Go Full \$	B of 100 Back Ccreen	
Image 6 Image 6 Go Full 5	Image: bit with the second	
↓ ↓ Page § Go I Full 5	a of 100 Back Creen Dse	
Image and the second secon	Creen	

3. Penalized least squares estimation

Why do we u	
ing do we u	se
Defining smo	othness
Penalized lea	st
Spline Smoot	thing
Choosing sm	oothing
A simulation :	study
Confidence li	mits
Summary	
Home	Page
Title	Page
••	••
•	•
Page 9	of 100
Gol	Back
Go I	Back
Go I Full S	Back Ccreen
Go I Full S	Back Creen
Go L Full S Clo	Back Creen
Go L Full S Cla	Back Creen
Go I Full S Cla	Back Creen

- - y is the *n*-vector of data y_j to be smoothed.
 - **t** is the *n*-vector of values of t_j .
 - W is a symmetric positive definite weight matrix.
 - $x(\mathbf{t})$ is the n-vector of fitted values, and x(t) has the basis function expansion

$$x(t) = \sum_{k}^{K} c_k \phi_k(t) = \mathbf{c}' \boldsymbol{\phi}(t)$$

• The penalized least squares criterion is

$$\label{eq:pensse} \mathsf{Pensse}_{\lambda}(x|\mathbf{y}) = [\mathbf{y} - x(\mathbf{t})]' \mathbf{W}[\mathbf{y} - x(\mathbf{t})] + \lambda \, \mathsf{pen}(x) \; ,$$

why do we l	lse
Defining smo	oothness
Penalized least	
Spline Smoothing	
Choosing sn	noothing
A simulation	study
Confidence I	limits
Summary	
Home	e Page Page
••	••
••	••
••	++
•	>>
Image 1	••• • 0 of 100
↓ ↓ Page 1	•• • 0 of 100
↓ ↓ Page 1 Go	••• • 0 of 100 Back
 ▲ Page 1 Go 	••• • 0 of 100 Back
↓ ↓ Page 1 Go Full 5	A of 100 Back Screen
↓ ↓ Page 1 Go Full 5	C of 100 Back Screen
Image 1 Image 1 Go Full \$ Cl	O of 100 Back Screen
Image 1 Image 1 Go Full 5	Image: the second s
Image 1 Image 1 Go Full 5 Cl	C of 100 Back Screen Cose Duit

How the smoothing parameter works

Smoothing parameter λ controls the amount of roughness.

- As $\lambda \to 0$, roughness matters less and less, and x(t) fits the data better and better.
- As $\lambda \to \infty$, roughness matters more and more, and x(t) becomes more and more "hyper–smooth."
- Our job is to find the right value where we trade enough bias off against sampling variance to minimize mean squared error.

W	hy do we u	se
De	fining smo	othness
Pe	nalized lea	ast
Sp	line Smoo	thing
Ch	noosing sm	oothing
A s	simulation	study
Сс	onfidence li	imits
Su	Immary	
	Ноте	e Page
	Title	Page
	••	>>
	••	••
	••	>>
	••	>>
	 ▲ Page 1 	••• • 1 of 100
	▲ ▲ Page 1	••• • 1 of 100
	↓ Page 1 Go 1	•• • 1 of 100 Back
	↓ Page 1 Go 1	•• • 1 of 100 Back
	↓ ↓ Page 1 Go I Full S	I of 100 Back Screen
	↓ ↓ Page 1 Go I Full S	1 of 100 Back Creen
	A A Page 1 Go I Full S Clo	I of 100 Back Ccreen Dse
	A A Page 1 Go I Full S Clo	I of 100 Back Creen Dse
	↓ ↓ Page 1 Go I Full S Cla	I of 100 Back Screen ose uit

The roughness penalty matrix

• For the classic penalty,

$$PEN_2(x) = \int [D^2 \mathbf{c}' \boldsymbol{\phi}(t)]^2 dt$$

= $\mathbf{c}' \int [D^2 \boldsymbol{\phi}(t)] [D^2 \boldsymbol{\phi}'(t)] dt \mathbf{c}$
= $\mathbf{c}' \mathbf{R} \mathbf{c}$

• The order K roughness penalty matrix ${f R}$ is

$$\mathbf{R} = \int [D^2 \boldsymbol{\phi}(t)] [D^2 \boldsymbol{\phi}'(t)] dt = \int (D^2 \boldsymbol{\phi}) (D^2 \boldsymbol{\phi}')$$

• substitute L for D^2 for more general roughness penalties.

(1)

The roughness penalized estimates for c and y

- Φ is the *n* by *K* matrix of basis function values $\phi_k(t_j)$.
- The penalized least squares criterion becomes

 $\mathtt{PENSSE}(y|c) = (\mathbf{y} - \mathbf{\Phi c})' \mathbf{W} (\mathbf{y} - \mathbf{\Phi c}) + \lambda \mathbf{c}' \mathbf{Rc}$.

 \bullet This is quadratic in c, and is minimized by

 $\hat{\mathbf{c}} = (\mathbf{\Phi}' \mathbf{W} \mathbf{\Phi} + \lambda \mathbf{R})^{-1} \mathbf{\Phi}' \mathbf{W} \mathbf{y}$.

Why do we use
Defining smoothness
Penalized least
Spline Smoothing
Choosing smoothing
A simulation study
Confidence limits
Summary
Home Page
Title Page
•• ••
Page 13 of 100
Page 13 of 100
Page 13 of 100
Page 13 of 100 Go Back
Page 13 of 100 Go Back
Page 13 of 100 Go Back Full Screen
Page 13 of 100 Go Back Full Screen
Page 13 of 100 Go Back Full Screen Close
Page 13 of 100 Go Back Full Screen Close

The smoothing matrix $\mathbf{S}_{\phi,\lambda}$

• The data-fitting vector $\hat{\mathbf{y}} = x(\mathbf{t})$ is

 $\hat{\mathbf{y}} = \mathbf{\Phi} (\mathbf{\Phi}' \mathbf{W} \mathbf{\Phi} + \lambda \mathbf{R})^{-1} \mathbf{\Phi}' \mathbf{W} \mathbf{y} \; ,$

• Smoothing matrix

 $\mathbf{S}_{\phi,\lambda} = \mathbf{\Phi} (\mathbf{\Phi}' \mathbf{W} \mathbf{\Phi} + \lambda \mathbf{R})^{-1} \mathbf{\Phi}' \mathbf{W}$

maps the data into the fit, and has many useful applications.



Equivalent degrees of freedom $df(\lambda)$

- It is useful to compare a fit using a roughness penalty to one using a fixed number of basis functions.
- A measure of the "degrees of freedom" in a roughness penalized fit is

 $df(\lambda) = {
m trace}\, {f S}_{\phi,\lambda}$

• This corresponds to the number of basis functions *K* in an un–penalized fit.



4. Spline Smoothing

Why do we use Defining smoothness Penalized least Spline Smoothing Choosing smoothing A simulation study Confidence limits Summary Home Page Title Page Title Page Title Cage Colose		
Defining smoothness Penalized least Spline Smoothing Choosing smoothing A simulation study Confidence limits Summary Home Page Title Page Title Page Title Page A D D D D A D D D D A D D D A D D D A D D D A D A	Why do we u	ise
Penalized least Spline Smoothing Choosing smoothing A simulation study Confidence limits Summary Home Page Title Page Title Page ↓ Page 16 of 100 Go Back Full Screen Close	Defining smo	oothness
Spline Smoothing Choosing smoothing A simulation study Confidence limits Summary Home Page Title Page Title Page Title Page O Bage 16 of 100 Go Back Full Screen	Penalized lea	ast
Choosing smoothing A simulation study Confidence limits Summary Home Page Title Page Title Page A Page 16 of 100 Go Back Full Screen Close	Spline Smoo	othing
A simulation study Confidence limits Summary Home Page Title Page Title Page A Page 16 of 100 Go Back Full Screen Close	Choosing sn	noothing
Confidence limits Summary Home Page Title Page () () () () () () () () () ()	A simulation	study
Summary Home Page Title Page I I I Page Page 16 of 100 Go Back Full Screen Close	Confidence I	limits
Home Page Title Page Title Page Page 16 of 100 Go Back Full Screen Close	Summary	
Home Page Title Page		
Title Page Image	Home	e Page
Title Page ▲ ▶ Page 16 of 100 Go Back Full Screen Close		
Image 16 of 100 Go Back Full Screen Close	Title	Page
 ↓ Page 16 of 100 Go Back Full Screen Close 		
 ✓ ✓		
 ✓ Page 16 of 100 Go Back Full Screen Close 	••	••
Page 16 of 100 Go Back Full Screen Close	••	**
Page 16 of 100 Go Back Full Screen Close	••	>>
Go Back Full Screen Close	••	•
Go Back Full Screen Close	Image 1	•••
Full Screen Close	Image 1	••• • 6 of 100
Full Screen Close	↓ ↓ Page 1 Go	
Close	↓ ↓ Page 1 Go	••• 6 of 100 Back
Close	↓ ↓ Page 1 Go	6 of 100 Back
Close	Image 1 Go Full 5	6 of 100 Back
	↓ ↓ Page 1 Go Full 5	6 of 100 Back
	↓ ↓ Page 1 Go Full 5	6 of 100 Back Screen

- The term "smoothing spline" has come to mean the following procedure:
 - Use natural or B-spline basis functions.
 - Place a knot at each data point t_j .
 - Use a penalty on D^2x .
- However, we find that
 - We can often achieve the same results by just using a number K of basis functions that is "large" relative to the resolution of the data.
 - We certainly want to be able to play with alternative roughness penalties.
 - Other basis functions systems are also desirable.



Two estimates of an acceleration curve.



Why do we use . . . Defining smoothness

5. Choosing smoothing parameter λ

Why do we use	
Defining smoothness	
Penalized least	
Spline Smoothing	
Choosing sm	oothing
A simulation study	
Confidence li	mits
Summary	
Home	Page
Title	Page
••	••
•	**
Image 19	•• • • of 100
Image: 1st Go E	••• • • of 100 Back
Image: 12 Image: 12 Go Image: 12 Full S	A of 100 Back creen
↓↓ ↓ Page 19 Go B Full S Cloc	►► For the second s

$\begin{array}{c} \textbf{Cross-validation for choosing the} \\ \textbf{smoothing parameter } \lambda \end{array}$

- In cross-validation,we
 - set aside a subset of data, the validation sample
 - call the balance of the data the training sample
 - fit the model to the training sample
 - assess fit to the validation sample
 - choose the λ value that gives the best fit



- \bullet We can also, for a sequence of values of $\lambda,$
 - set aside each observation (t_j, y_j) in turn
 - fit the data with the rest of the sample,
 - sum fits to the left out values to get a *cross–validated* error sum of squares $CV(\lambda)$.
 - select the λ value that minimizes $CV(\lambda)$.

	se
Defining smo	othness
Penalized lea	st
Spline Smoot	thing
Choosing sm	oothing
A simulation :	study
Confidence li	mits
Summary	
Home	Page
Title	Page
••	>>
•	•
Page 2	1 of 100
Page 2	1 of 100
Page 2 Go I	1 of 100 Back
Page 2 Go I	1 of 100 Back
Page 2 Go I Full S	1 of 100 Back creen
Page 2 Go I Full S	1 of 100 Back Icreen
Page 2 Go I Full S	t of 100 Back Icreen
Page 2 Go I Full S Clo	t of 100 Back Icreen
Page 2 Go I Full S Clo	t of 100 Back Icreen Dise

Generalized cross–validation for choosing the smoothing parameter λ

- Cross-validation is time-consuming, and tends too often to under-smooth the data.
- The generalized cross-validation criterion is

$$GCV(\lambda) = (\frac{n}{n-df(\lambda)})(\frac{\texttt{SSE}}{n-df(\lambda)})$$

where df is the equivalent degrees of freedom of the smoothing operator.

- The right factor is just the unbiassed estimate s_e^2 of residual variance familiar in regression analysis.
- The left factor further "discounts" this measure further to allow for the influence of optimizing with respect to λ .

6. A simulation study

Ch	noosing sm	oothing
A s	simulation	study
Сс	onfidence li	mits
Su	ımmary	
	Home	Page
	Title	Page
	••	•••
	•	
	•	•
	✓ Page 2.	• 3 of 100
	◀ Page 2.	► 3 of 100
	A Page 2 Go I	► 3 of 100 Back
	Page 2. Go I	▶ 3 of 100 Back
	Page 2. Go I Full S	► 3 of 100 Back Creen
	Page 2. Go I Full S	► 3 of 100 Back Correen
	 Page 2. Go I Full S Close 	► 3 of 100 Back Creen DSE
	Age 2 Go I Go I Go I Clo	► 3 of 100 Back Screen Dse

Why do we use ... Defining smoothness Penalized least ... Spline Smoothing

- How does GCV work in a simulated data example?
- A parametric growth model by Pierre Jolicoeur at the Université de Montréal offers a nice test problem.
- We simulate 1000 samples, each observation being a random sample from realistic Jolicoeur models plus realistic error.
- We smooth using a range of values of λ, and note the value giving the best value of GCV.
- How well do we estimate the Jolicoeur acceleration curves?

Defining smoothness		
Penalized least		
Spline Smoothing		
Choosing smoothing		
A simulation study		
Confidence limits		
Summary		
Home Page		
Title Page		
•• ••		
Page 24 of 100		
Page 24 of 100		
Page 24 of 100 Go Back		
Page 24 of 100 Go Back		
Page 24 of 100 Go Back Full Screen		
Page 24 of 100 Go Back Full Screen		
Page 24 of 100 Go Back Full Screen		
Page 24 of 100 Go Back Full Screen Close		
Page 24 of 100 Go Back Full Screen Close		

20 Jolicoeur acceleration curves



Why do we use . . . Defining smoothness

GCV and Root-Mean-Squared-Error



Why do we use . . . Defining smoothness

What we see

- In the top panel, GCV favors $\lambda = 0.1$.
- This is about right for optimal MSE for ages 8 and 16, but less smoothing would be better for age 12, in the middle of the pubertal growth spurt.
- One smoothing parameter value does not work best for all ages, but
- The value chosen by GCV certainly does a fine job.

Why do we use
Defining smoothness
Penalized least
Spline Smoothing
Choosing smoothing
A simulation study
Confidence limits
Summary
Home Page
Title Page
•• ••
Page 27 of 100
Go Back
Full Screen
Close
0#

RMSE, Bias, and Standard Error



Why do we use . . . Defining smoothness

What we see

- The performance of the spline smoothing estimate deteriorates badly at the extremes.
- The sharp curvature at the pubertal growth spurt also causes some problems.
- Except at the extremes and PGS, the bias is negligible.
- The standard error is about the same as RMSE.
- Would we do better at the extremes if the smooth respected monotonicity?

WI	Why do we use		
De	fining sm	oothness	
Pe	nalized le	ast	
Sp	line Smoo	othing	
Ch	loosing sr	moothing	
A simulation study			
Со	onfidence	limits	
Su	mmary		
	Hom	e Page	
	Title	Page	
		<u> </u>	
	44		
	Page 2	29 01 100	
	Go	Back	
	Full	Screen	
	~	lose	
	6		
	(Quit	

7. Confidence limits

Why do we use
Defining smoothness
Penalized least
Spline Smoothing
Choosing smoothing
A simulation study
Confidence limits
Summary
Home Page
Title Page
44 >>
• •
• •
▲ ▶ Page 30 of 100
Page 30 of 100
 ✓ ✓ Page 30 of 100 Go Back
Page 30 of 100 Go Back
 ✓ Page 30 of 100 Go Back Full Screen
 ▲ Page 30 of 100 Go Back Full Screen
 ✓ ✓ Page 30 of 100 Go Back Full Screen Close
 ✓ Page 30 of 100 Go Back Full Screen Close

- Because the mapping from data y to the coefficient vector c is linear, it is a simple matter to work out the standard error of any linear functional of a curve defined by c.
- The variance of a quantity $\rho(x)$ associated with linear mapping ${\bf M}$ from $\hat{{\bf c}}$ to $\hat{\rho}(x)$ is

 $\operatorname{Var}[\hat{\rho}(x)] = \mathbf{MS}_{\phi,\lambda} \boldsymbol{\Sigma}_e \mathbf{S}_{\phi,\lambda} \mathbf{M}'$

• Simple, that is, if we can get a good estimate of the variance-covariance matrix Σ_e of the residual vector.

Defining smoothness		
Penalized least		
Spline Smoothing		
Choosing smoothing		
A simulation study		
Confidence limits		
Summary		
Home Page Title Page		
•• ••		
• •		
Page 31 of 100		
Go Back		
Full Screen		
Close		

95% point–wise confidence limits for growth acceleration





Why do we use ... Defining smoothness Penalized least... Spline Smoothing Choosing smoothing... A simulation study Confidence limits Summary Home Page Title Page 44 •• Page 33 of 100 Go Back Full Screen Close Quit

8. Summary

- Roughness penalization, also called *regularization*, is a flexible and effective way to ensure that an estimated function is "smooth."
- We can tailor the definition of "smooth" to our needs.
- The roughness penalty idea extends to any type of *functional parameter* that we want to estimate from the data.
- Roughness penalties are one of the main ways in which we exploit the smoothness that we assume in the process generating the data.
- "Roughness" is like *energy* in physics; roughness requires energy to produce, and smoothness implies limited energy.
- Where we imagine that the amount of energy behind the data is limited, it is natural to assume smoothness.