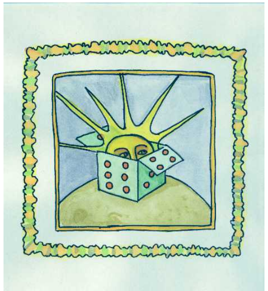
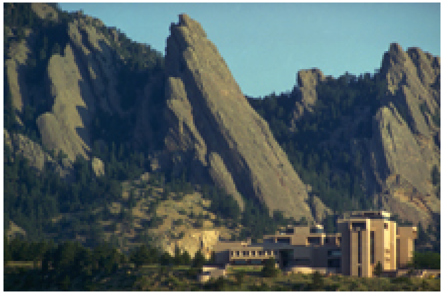
Discussion to *Likelihood Basis Pursuit*

Douglas Nychka Geophysical Statistics Project
National Center for Atmospheric Research

Outline

- Penalized least squares
- Implied priors.
- Something completely different





Supported by the National Science Foundation DMS

Penalized least squares

An abstraction of Grace's model and estimate is

Observe:

$$y_i = f(x_i) + e_i$$

f is smooth and $\{e_i\}$ iid normal

A spline:

$$\hat{f}(x) = \sum_{j=1}^{n} c_j K(x, x_j)$$

where $\boldsymbol{c}$ minimizes

$$||\boldsymbol{y} - K\boldsymbol{c}||^2 + \lambda \boldsymbol{c}^T K \boldsymbol{c}$$

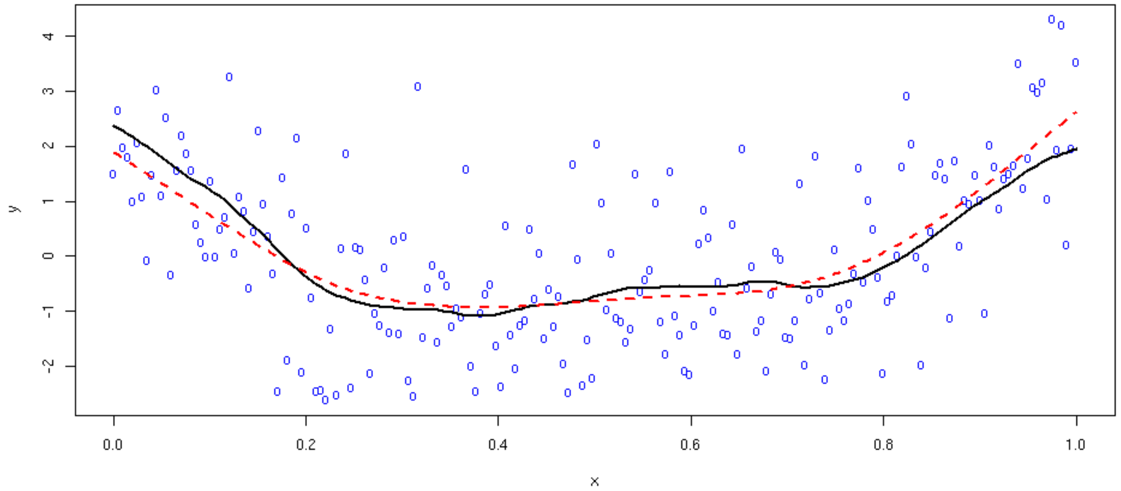
The RKHS:

$K_{i,j} = K(x_i, x_j)$ and K is a reproducing kernel (or a covariance function).

What problem are we solving?

If f is Gaussian process with mean zero and with covariance K

- then $\hat{f}$ is the "Kriged" curve.
- then $\hat{f}$ is the conditional expectation of f given $\boldsymbol{y}$.
- then $\hat{f}$ is the posterior mean/mode with a Gaussian process prior on f



Grace does something different

Instead of

$$||\boldsymbol{y} - K\boldsymbol{c}||^2 + \lambda \boldsymbol{c}^T K \boldsymbol{c}$$

$\boldsymbol{c}$ minimizes

$$||\boldsymbol{y} - K\boldsymbol{c}||^2 + \lambda \sum_{j=1}^{n} |c_j|$$

This forces some components to be identically zero the number being controlled by λ.

What problem are we solving? What about the L_1 penalty?

$$\underbrace{||\boldsymbol{y} - K\boldsymbol{c}||^2}_{-\text{log likelihood}} \quad + \quad \underbrace{\lambda \sum_{j=1}^{n} |c_j|}_{\text{log prior (minus some constants)}}$$

The prior is $\{c_i\}$ are iid double exponential RVs.
i.e.

$$f(c) = \lambda/2 e^{-|c|/\lambda}$$

minimizing this is the same as maximizing the posterior.

Something else

Instead of

$$||\boldsymbol{y} - K\boldsymbol{c}||^2 + \lambda \boldsymbol{c}^T K \boldsymbol{c}$$

Kriging/spatial statistics version

$$||\boldsymbol{y} - K^{1/2}\boldsymbol{\beta}||^2 + \lambda ||\boldsymbol{\beta}||^2$$

$(\boldsymbol{\beta} = K^{-1/2}\boldsymbol{c})$

Version of hard thresholding with wavelets

$$||\boldsymbol{y} - K^{1/2}\boldsymbol{\beta}||^2 + \lambda \sum_{j=1}^{n} |\beta_j|$$

If $K^{1/2}$ has columns that are an orthogonal wavelet basis.

What are the priors for f like?

K generalized covariance related to cubic smoothing spline

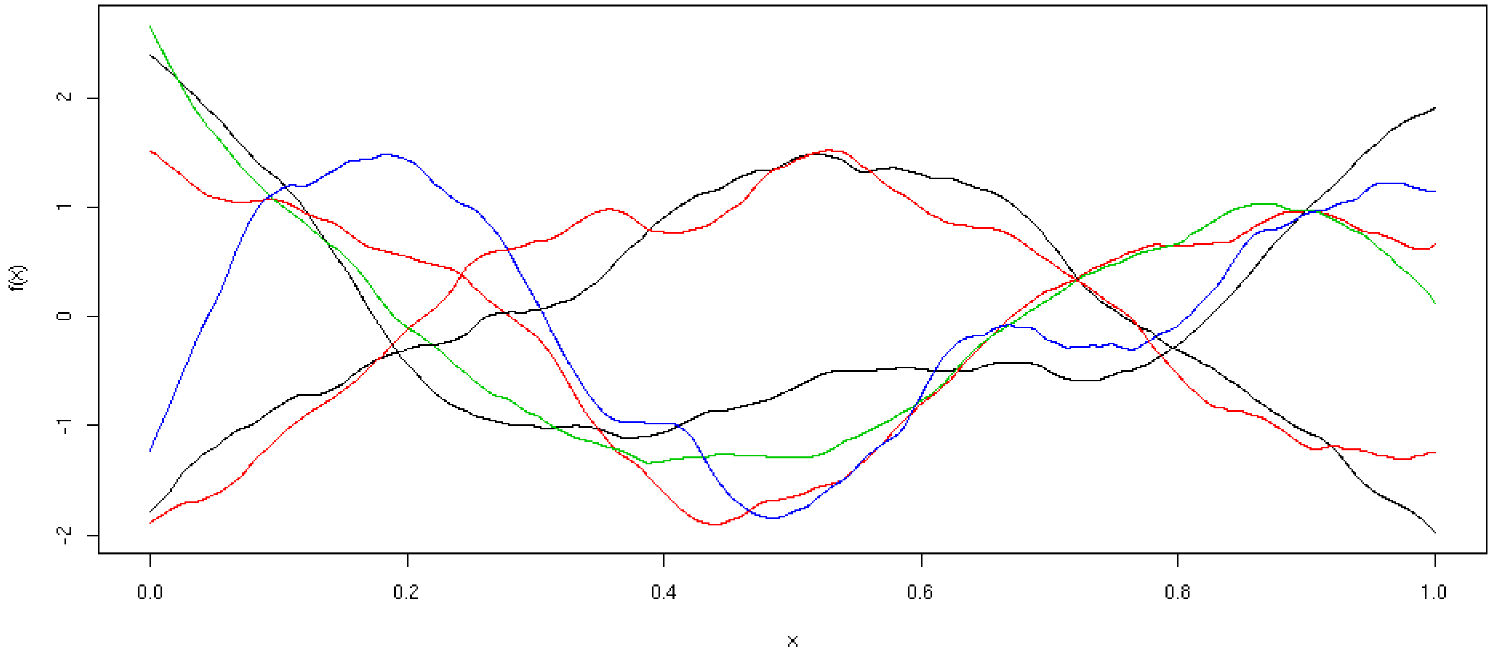
$$f(x) = \sum_j a_j \psi_j(x)$$

Four (2×2) interesting cases

- *Cubic smoothing spline prior*
  a_j iid normal, $\psi_j = K^{1/2}(., x_j)$
- *Wavelet type prior*
  a_j iid double exponential and ψ_j same.
- $\approx$ *Grace's model.*
  $a_j = c_j$ iid double exponential but $\psi_j = K(., x_j)$
- *Gaussian analog*
  $a_j = c_j$ iid normal and $\psi_j = K(., x_j)$

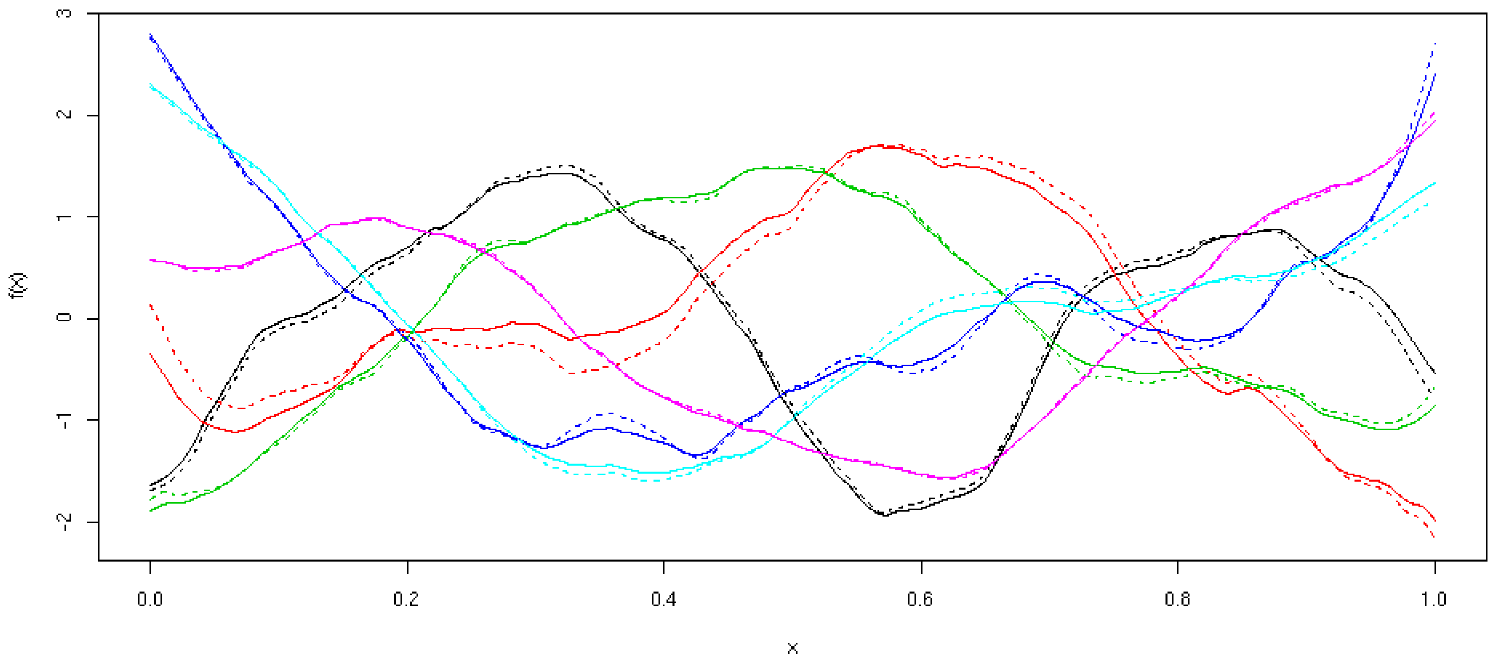
... and also a mixture $\pi\delta_0, (1-\pi)N(0, \sigma^2)$

Sample paths for cubic smoothing spline prior



Adding sample paths with double exponential coefficients

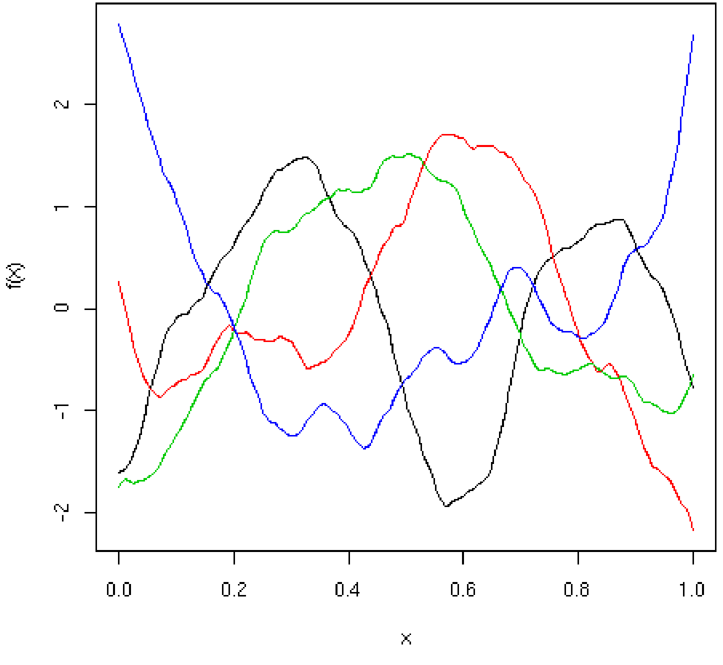
(and using same set of generating uniforms RVs)



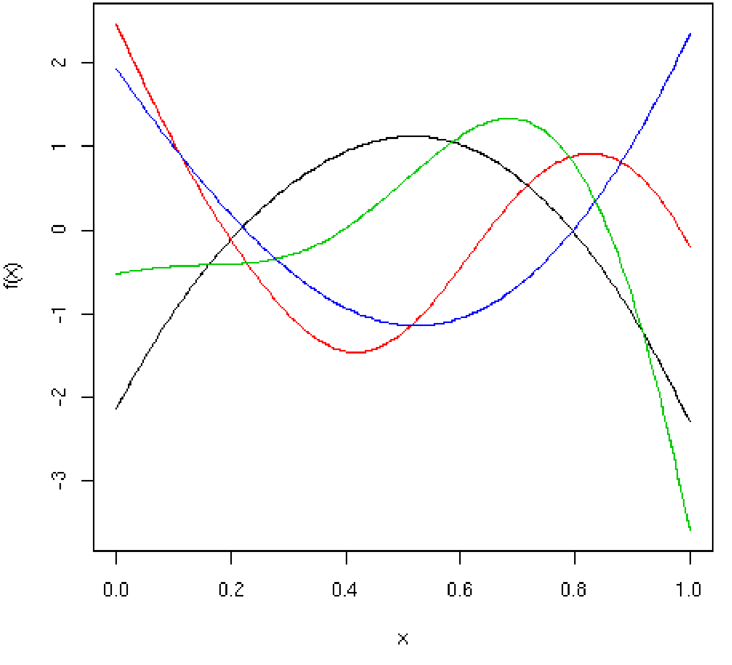
dashed = double exponential

Realizations from prior related to Grace's talk

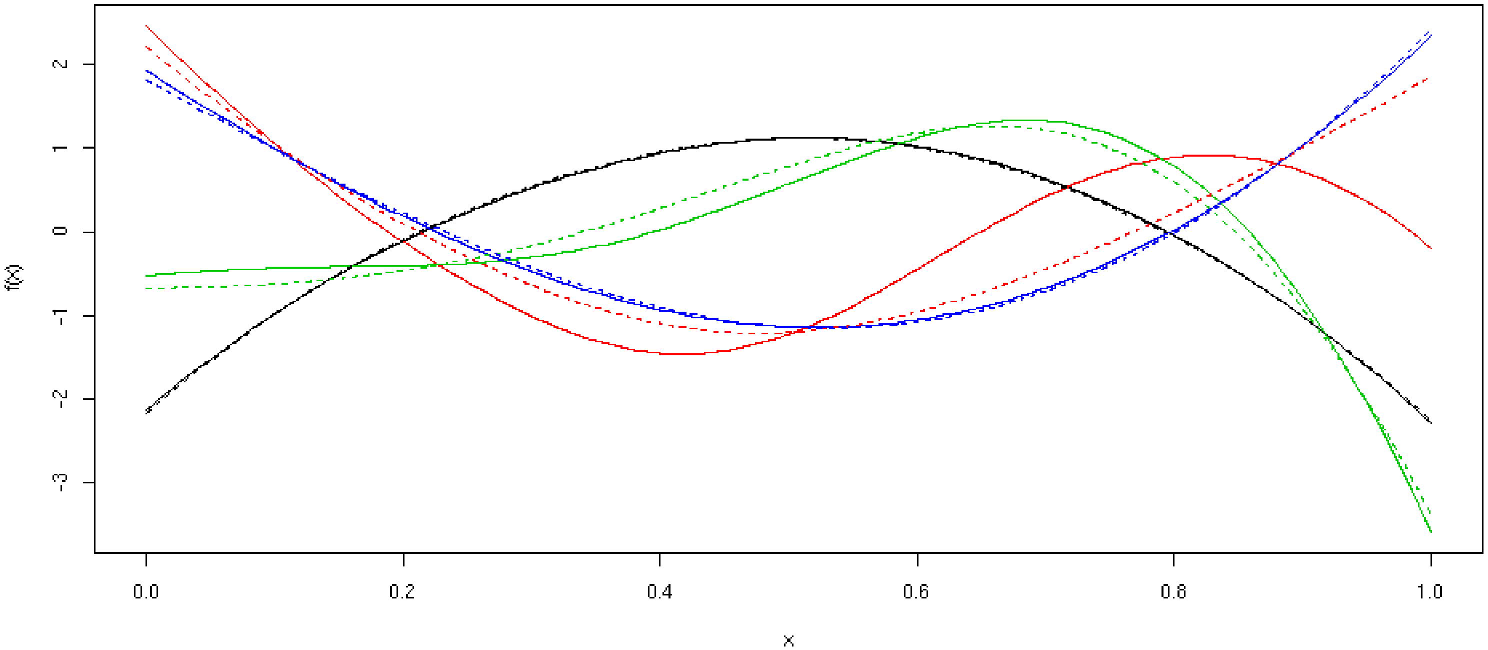
Cubic spline prior



Basis pursuit prior

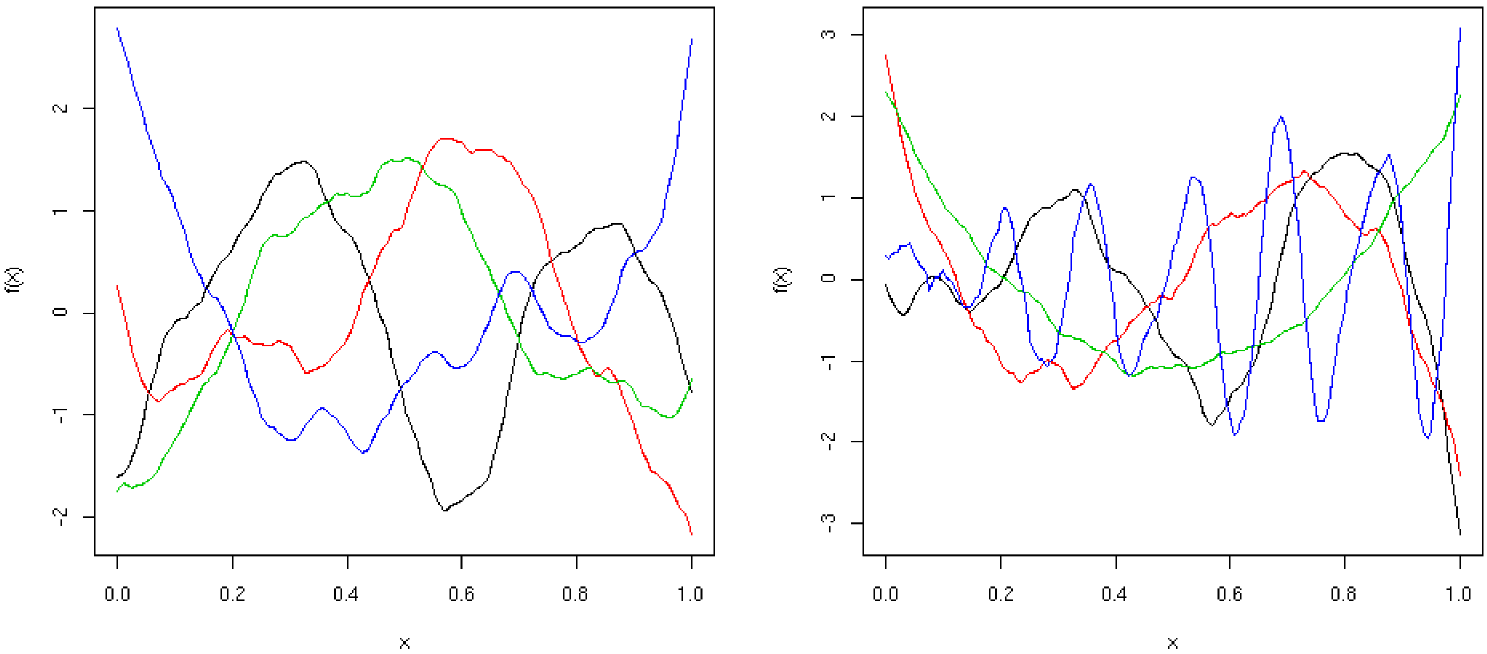


Comparison to Gaussian process



dashed = double exponential

Comparison to mixture ($\pi = .8$)



A project at NCAR

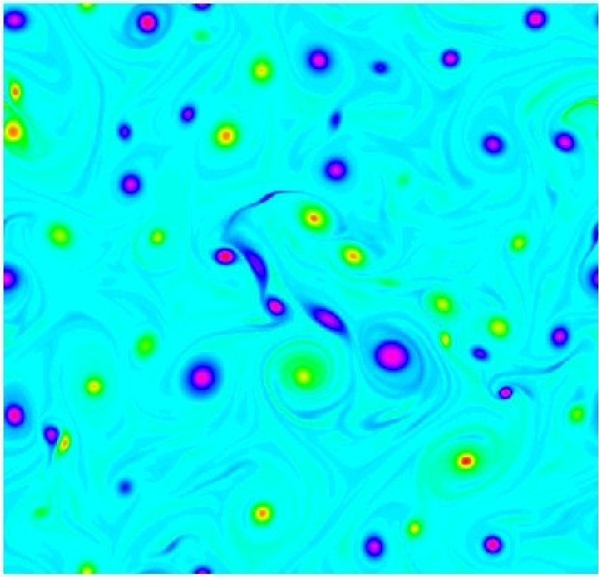
From a numerical simulation of turbulence find individual vortices and expand them in a parsimonious basis.

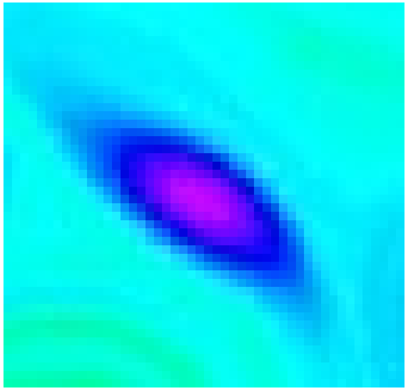
The whole image:

$\zeta(\boldsymbol{x}) = \Sigma_{k=1}^{K} v_k(\boldsymbol{x}) + e(\boldsymbol{x})$

The k^{th} vortex:

$v_k = \Sigma_{i=1}^{3J+1} c_{k,i} \; \psi_j(\boldsymbol{x} - \mu_k)$





Conclusion

Basis pursuit raises issues about what priors are good for functional structure.

Distinguish variable selection from function estimation.