Manny Parzen and Reproducing Kernel Hilbert Spaces

Grace Wahba

At the session on

Parzen's Legacy on Modern Nonparametric Statistics

JSM 2009 Washington, DC, August 3

These and other slides at

`http://www.stat.wisc.edu/~wahba/` → TALKS

Preprint collection at

`http://www.stat.wisc.edu/~wahba/` − > TRLIST

I learned about Reproducing Kernel Hilbert Spaces from Manny Parzen, my thesis advisor, somewhere around 1962-64, and these wonderful objects have essentially formed the foundation of my career. Randy Eubank has already told you a lot of modern things about them so I am going to revisit ill-posed inverse problems, the RKHS setting rephrased from Manny's work, and some computational and convergence results from 1990 and earlier.

References

[1] E. Parzen. Regression analysis of continuous parameter time series. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 469–489, Berkeley, California, 1960. University of California Press.

[2] E. Parzen. An approach to time series analysis. *Ann. Math. Statist.*, 32:951–989, 1962.

[3] E. Parzen. A new approach to the synthesis of optimal smoothing and prediction systems. In R. Bellman, editor, *Mathematical Optimization Techniques*, pages 75–108. Univ. of California Press, 1963.

[4] E. Parzen. Probability density functionals and reproducing kernel Hilbert spaces. In M. Rosenblatt, editor, *Proceedings of the Symposium on Time Series Analysis*, pages 155–169. Wiley, 1963.

[5] E. Parzen. Statistical inference on time series by RKHS methods. In R. Pyke, editor, *Proceedings 12th Biennial Seminar*, Montreal, 1970. Canadian Mathematical Congress. 1-37.

Figure 1: *Manny Parzen's 60th Birthday Party, 1989.*

l. to r. Don Ylvisaker, Grace Wahba, Joe Newton, Marcello Pagano, Randy Eubank, Manny Parzen, Will Alexander, Marvin Zelen, Scott Grimshaw

Ill-Posed Inverse Problems and (Tihonov) Regularization

1. I'll review how one can solve ill-posed inverse problems in an RKHS setting. These result are already in Manny's papers, and, in any case, I learned them at his feet at Stanford in the 60's.

2. I'll make some remarks as to how to actually compute the solutions in a stable way (Nychka et al 1989)

3. I'll note some old convergence results and discuss the question "When is the optimal regularization parameter review insensitive to the choice of the loss function?" (Wahba and Wang 1990)

Searching google for "Ill Posed Inverse Problems" gives "about 27,800 hits". and adding "convergence" to the "all of these words" box gives "about 12,000 hits - my guess is that these kinds of convergence results are being rediscovered often.

Some Early Applications:

- G. Wahba. Constrained regularization for ill posed linear operator equations, with applications in meteorology and medicine, in "Statistical Decision Theory and Related Topics III", Vol. 2, S.S. Gupta and J.O. Berger, eds., 383-418, Academic Press (1982).

- D. Nychka, G. Wahba, S. Goldfarb, and T. Pugh. Cross-validated spline methods for the estimation of three dimensional tumor size distributions from observations on two dimensional cross sections. *J. Am. Stat. Assoc.*, 79:832–846, 1984.

- F. O'Sullivan and G. Wahba. A cross validated Bayesian retrieval algorithm for non-linear remote sensing. *J. Comput. Physics*, 59:441–455, 1985.

Solving Ill Posed Integral Equations in an RKHS Setting

Given data

$$y_i = \int_\Omega G(u_i, s) f(s) ds + \epsilon_i, i = 1, \cdots, n \tag{1}$$

with $\epsilon \approx \mathcal{N}(0, \sigma^2)$, σ^2 unknown. Suppose $f \in \mathcal{H}_K$, an RKHS with RK K. Find $f_\lambda \in \mathcal{H}_K$ to minimize

$$\sum_{i=1}^{n} (y_i - \int_\Omega G(u_i, s) f(s) ds)^2 + \lambda \|f\|^2_{\mathcal{H}_K}. \tag{2}$$

Solving Ill Posed Integral Equations in an RKHS Setting, continued

Assume $\int_{\mathcal{H}_K} G(u_i, s) f(s) ds \equiv L_i f$ is a bounded linear functional in $\mathcal{H}_K$. Then by the Riesz Representation theorem there exist some η_i in $\mathcal{H}_K$ such that

$$L_i f \equiv < \eta_i, f >_{\mathcal{H}_K} \tag{3}$$

where $< \cdot, \cdot >_{\mathcal{H}_K}$ is the inner product in $\mathcal{H}_K$. What is η_i? Recall that $K_t(\cdot) \equiv K(\cdot, t)$ is the representer of evaluation in $\mathcal{H}_K$. This means that for any fixed t, $\eta_i(t) =< \eta_i, K_t >_{\mathcal{H}_K} \equiv L_i K_t$. So,

$$\eta_i(t) =< \eta_i, K_t >_{\mathcal{H}_K} \equiv L_i K_t \equiv \int_\Omega G(u_i, s) K_t(s) ds \equiv \int_\Omega G(u_i, s) K(s, t) ds. \tag{4}$$

Solving Ill Posed Integral Equations in an RKHS Setting, continued

Theorem: (Special case of Kimeldorf and Wahba, 1971)

The minimizer of $\sum_{i=1}^{n}(y_i - \int_\Omega G(u_i, s)f(s)ds)^2 + n\lambda\|f\|^2_{\mathcal{H}_K}$ has a representation of the form

$$f_\lambda = \sum_{j=1}^{n} c_j \eta_j . \tag{5}$$

A closed form solution is now at hand, using

$$\begin{aligned} \int_\Omega G(u_i, s)f(s)ds &= L_i f = L_i \sum_{j=1}^{n} c_j \eta_j =< \eta_i, \sum_{j=1}^{n} c_j \eta_j > \\ \|f_\lambda\|_{\mathcal{H}_K} &= \sum_{i,j=1^n} c_i c_j < \eta_i, \eta_j > . \end{aligned}$$

Solving Ill Posed Integral Equations in an RKHS Setting, continued

We're almost done. We just have to know what $< \eta_i, \eta_j >_{\mathcal{H}_K}$ is. Let $L_{(i(t)}$ mean L_i applied to what follows, considered as a function of t. Then

$$\begin{aligned} < \eta_i, \eta_j >= L_i\eta_j &= L_{i(t)} \int_\Omega G(u_j, s)K(s,t)ds \\ &= \int_\Omega \int_\Omega G(u_j, s)K(s,t)G(u_i, t) \\ &= R_{ij}, [say] \end{aligned}$$

Almost!

Solving Ill Posed Integral Equations in an RKHS Setting, continued

Almost! Let R be the nxn matrix with ij entry

$$R_{ij} = \int_\Omega \int_\Omega G(u_j, s) K(s,t) G(u_i, t)$$

Returning to the original optimization problem and writing it in terms of the η_i: Find f in $\mathcal{H}_K$ to minimize

$$\sum_{i=1}^{n} (y_i - \int_\Omega G(u_i, s) f(s) ds)^2 + n\lambda \|f\|^2_{\mathcal{H}_K}$$

becomes: Find $c \in E^n$ to minimize

$$\|(y - Rc)\|^2 + \lambda c' Rc.$$

and $f_\lambda = \sum_j c_j \eta_j$ with

$$c = (R + \lambda I)^{-1} y. \tag{6}$$

Computing

- If n is large the solution may be computed in the span of a subset of the basis functions (the η_i's).
- Do not attempt to compute the $R_{ij} = \int_\Omega \int_\Omega G(u_j, s)K(s,t)G(u_i,t)$ by quadrature. It is better to approximate the η_i by a quadrature approximation $\tilde{\eta}_i = \sum_{r=1}^N b_{ir} K_{t_r}$ with the t_r playing the role of quadrature points. Then $< \tilde{\eta}_i, \tilde{\eta}_j >= \sum_{r,r'} b_{ir} b_{jr'} K(t_r, t'_r)$
- If $\|f\|^2$ is replaced by $\int (f^{(m)}(s))^2 ds$ then the K_t are splines, and B-spline basis functions are convenient.

Three Early Convergence Rate Papers:

1. G. Wahba. Convergence rates of certain approximate solutions to Fredholm integral equations of the first kind. *J. Approx. Theory*, 7:167–185, 1973.

2. D. Nychka and D. Cox. Convergence rates for regularized solutions of integral equations from discrete noisy data. *Ann. Statist.*, 17:556–572, 1989.

3. G. Wahba and Y. Wang. When is the optimal regularization parameter insensitive to the choice of the loss function? *Commun. Statist.-Theory Meth.*, 19:1685–1700, 1990.

When is the optimal regularization parameter insensitive to the choice of the loss function?

Simple case to get closed form answers, periodic convolution equation and periodic solution function f..

$$y_i = \int_0^1 h(\frac{i}{n} - s) f(s) ds + \epsilon_i, i = 1, 2, \cdots, n.$$

- $\|f\|^2 = \int_0^1 (f^{(m)})^2(s) ds.$
- Fourier coefficients of f go to 0 at the rate $n^{-\alpha}$.
- Fourier coefficients of h go to 0 at the rate $n^{-\beta}$.
- $MSE(solution) \equiv MSE(S) = \int_0^1 (f_\lambda(s) - f(s))^2 ds$
- $MSE(prediction) \equiv MSE(P) = \int_0^1 (g_\lambda(s) - g(s))^2 ds, g = h * f.$

When is the optimal λ for $MSE(P)$ the same as the optimal λ for $MSE(S)$ (ratewise)? The answer depends on α, β and m.

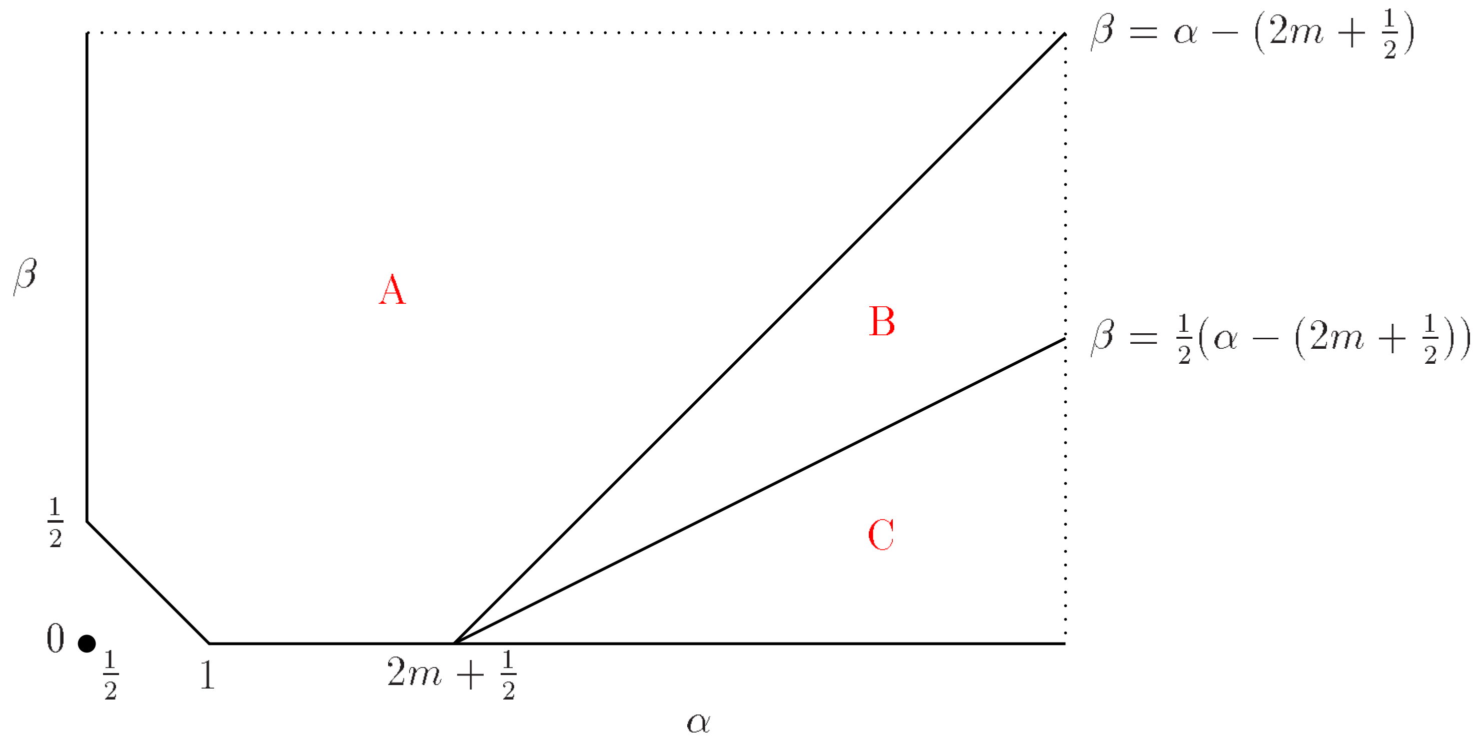

Figure 2: *Optimal λ and $MSE(S)$ as a function of α, β and m.*

- Region A, $\lambda_S \sim n^{-\frac{m+\beta}{\alpha+\beta}}$, $MSE(S) \sim n^{-\frac{\alpha-1/2}{\alpha+\beta}}$, $\lambda_P \sim \lambda_S$
- Region B, $\lambda_S \sim n^{-\frac{m+\beta}{\alpha+\beta}}$, $MSE(S) \sim n^{-\frac{\alpha-1/2}{\alpha+\beta}}$, $\lambda_P = o(\lambda_S)$
- Region C,
  $\lambda_S \sim n^{-\frac{2(m+\beta)}{(4m+6\beta+1)}}$, $MSE(S) \sim n^{-\frac{4(m+\beta)}{(4m+6\beta+1)}}$, $\lambda_P = o(\lambda_S)$

Conclusions

So, it all started for me back in the 60's, with fond memories of sitting on the grass in front of the old Sequoia Hall at Stanford hearing abut these wonderful objects, and now "Kernel Methods" (meaning "Reproducing Kernel Hilbert Space Methods") are everywhere. Of course in density estimation we have the Parzen Kernel, so, Manny is the father of both reproducing kernel methods and Parzen kernel methods!