CHAPTER 12
Special Topics

12.1 The notion of "high frequency" in different spaces.

Let $\Psi_\nu(t) = \sqrt{2}\cos 2\pi\nu t$, $\nu = 1, 2, \ldots$ and let $\mathcal{H}$ be the collection of all functions of d variables with a representation

$$f(x_1, \ldots, x_d) = \sum_{\nu_1=1}^{\infty} \cdots \sum_{\nu_d=1}^{\infty} f_{\nu_1,\ldots,\nu_d} \Psi_{\nu_1}(x_1) \ldots \Psi_{\nu_d}(x_d) \tag{12.1.1}$$

such that

$$\sum_{\nu_1,\ldots,\nu_d=1}^{\infty} (2\pi\nu_1)^{2m} \ldots (2\pi\nu_d)^{2m} f^2_{\nu_1,\ldots,\nu_d} < \infty. \tag{12.1.2}$$

We see that the left-hand side of (12.1.2) is then equal to

$$\int_0^1 \cdots \int_0^1 \left(\frac{\partial^{dm}}{\partial x_1^m \ldots \partial x_d^m} f \right)^2 dx_1 \ldots dx_d, \tag{12.1.3}$$

which we may take as the squared norm in $\mathcal{H}$. (We could let the Ψ_ν's be sines as well as cosines and get the same result; to avoid cumbersome notation we will not do this.)

Alternatively, consider the collection of functions of the form (12.1.1) for which

$$\sum_{\nu_1,\ldots,\nu_d=1}^{\infty} [(2\pi\nu_1)^2 + \ldots + (2\pi\nu_d)^2]^m f^2_{\nu_1,\ldots,\nu_d} < \infty. \tag{12.1.4}$$

It is not hard to see that (12.1.4) is equal to

$$\sum_{\mu_1+\ldots+\mu_d=m} \frac{m!}{\mu_1! \ldots \mu_d!} \int_0^1 \cdots \int_0^1 \left(\frac{\partial^m f}{\partial x_1^{\mu_1} \cdots \partial x_d^{\mu_d}} \right)^2 dx_1 \ldots dx_d. \tag{12.1.5}$$

The Hilbert spaces with squared norms (12.1.2) and (12.1.4) are quite different. The eigenvalues of the r.k. for the first space are

$$\lambda_{\nu_1,\ldots,\nu_d} = [(2\pi\nu_1)(2\pi\nu_2) \ldots (2\pi\nu_d)]^{-2m},$$

and if these are lined up in size place, it can be shown that the nth largest is of the order of $\left[((\log n)^{d-1}/n)\right]^{2m}$. To see this, observe that the number of lattice points $(\nu_1, \ldots, \nu_d)$ on the lattice of d tuples of positive integers that satisfies

$$\prod_{\alpha=1}^{d} \nu_\alpha \leq k$$

is, to first order, given by $(1/(d-1)!)k(\log k)^{d-1}(1+o(1))$. This is obtained from the volume approximation given by

$$\int_1^k \cdots \int_1^{k/x_3 \ldots x_{d-1}} \int_1^{k/x_2 \ldots x_{d-1}} \frac{k}{x_1 \ldots x_{d-1}} dx_1 \ldots dx_{d-1} = \frac{k}{(d-1)!}(\log k)^{d-1}.$$

(For a sharper approximation, see Ivic (1985, Chap. 3).) Letting Λ_n be the nth largest eigenvalue, we obtain

$$\Lambda_{[k(\log k)^{d-1}/(d-1)!]} \simeq \frac{1}{(2\pi)^d k^{2m}}.$$

Setting

$$n = [k(\log k)^{d-1}/(d-1)!]$$

gives

$$1/(2\pi)^d k^{2m} = O\left[(\log n)^{d-1}/n\right]^{2m} = \Lambda_n.$$

Similarly, for the space of (12.1.4) the eigenvalues of the r.k. are

$$\lambda_{\nu_1, \ldots, \nu_d} = \left(\sum_{\alpha=1}^{d} (2\pi\nu_\alpha)^2 \right)^{-m},$$

and we use the volume inside a sphere to estimate the number of lattice points for which $\sum_{\alpha=1}^{d} \nu_\alpha^2 \leq k$. The result is the nth largest eigenvalue is of the order of $n^{-2m/d}$.

We have noted that the rate of decay of the eigenvalues of the r.k. plays a role in convergence rates of estimates with noisy data.

Below is a handy theorem giving lower bounds on the r.k. norm of the error, when the data are exact and the estimate is an orthogonal projection.

THEOREM 12.1.1 (Micchelli and Wahba (1981)). *Let $\mathcal{H}_R$ be an* r.k.h.s. *with* r.k. *R and eigenvalues and eigenfunctions (λ_ν, Φ_ν), $\nu = 1, 2, \ldots$. Let V_n be any n-dimensional subspace in $\mathcal{H}_R$ and let $P_{V_n} g$ be the orthogonal projection of g in $\mathcal{H}_R$ onto V_n. Then, for any $p > 1$, there exists $g \in \mathcal{H}_R$ with*

$$g(t) = \int_T R^{p/2}(t, u)\rho(u)\, du \tag{12.1.6}$$

with $\rho \in \mathcal{L}_2[T]$, $\int_T \rho^2(u)\, du = 1$, such that

$$\|g - P_{V_n} g\|^2 \geq \lambda_{n+1}^{p-1}. \tag{12.1.7}$$

To see why this theorem is true, let

$$\mathcal{C}_p = \left\{ g : \; g(t) = \int_T R^{p/2}(t,u)\rho(u)\,du \right\}$$

with $\rho \in \mathcal{L}_2$ and

$$\int \rho^2(s)\,ds = 1.$$

Then

$$g(t) = \int \sum_{\nu=1}^{\infty} \lambda_\nu^{p/2} \Phi_\nu(t)\Phi_\nu(u)\rho(u)\,du,$$

that is,

$$g = \sum_{\nu=1}^{\infty} \lambda_\nu^{p/2} \rho_\nu \Phi_\nu$$

where

$$\rho_\nu = \int \Phi_\nu(u)\rho(u)\,du$$

and

$$\Sigma \rho_\nu^2 \leq 1.$$

Consider

$$\inf_{V_n} \sup_{g \in \mathcal{C}_p} \|g - P_{V_n} g\|^2.$$

We have the following game. You choose V_n to minimize $\|g - P_{V_n} g\|$ and nature chooses g to maximize it. The optimal strategy is to choose $V_n = \text{span}\,(\Phi_1, \ldots, \Phi_n)$. Then nature chooses ρ_ν all zero except $\rho_{n+1} = 1$, then $g = \lambda_{n+1}^{p/2}\, \Phi_{n+1}$ and

$$\|g - P_{V_n} g\|^2 = \|\lambda_{n+1}^{p/2} \Phi_{n+1}\|^2 = \frac{\lambda_{n+1}^p}{\lambda_{n+1}} = \lambda_{n+1}^{p-1}.$$

There are examples (with $\mathcal{T} = [0,1]$) for which $V_n = \text{span}\,\{R_{t_1}, \ldots, R_{t_n}\}$ is also an optimal strategy. The $t_1, \ldots, t_n$ are the n zeros of the $(n+1)$st eigenfunction (see Melkman and Micchelli (1978)).

This is, of course, a theorem giving lower bounds on interpolation error. To see this, let V_n be spanned by $\eta_1, \ldots, \eta_n$ and let $g_i =< g, \eta_i >$, $i = 1, \ldots, n$; then $P_{V_n} g$ is that element $\hat{g}$ in $\mathcal{H}_R$ minimizing $\|g\|$ subject to $< \hat{g}, \eta_i >= g_i$, $i = 1, \ldots, n$. Note that any element in $\mathcal{H}_R$ has a representation as a multiple of an element in $\mathcal{C}_p$ for $p = 1$ (to which this theorem does not apply). As an application, consider W_m, with the r.k. of Chapter 10. Then $\mathcal{C}_2$ is the collection of functions that satisfy $f^{(2m)} \in \mathcal{L}_2$ and certain boundary conditions. Since the eigenvalues of the r.k. decay at the rate ν^{-2m}, this theorem says that if one interpolates to f at the points t_i, $i = 1, \ldots, n$, then $1/n^{2m}$ is a lower bound on the best achievable convergence rate of $\|g - P_{V_n} g\|^2$ for $g \in \mathcal{C}_2$.

Note that if the loss were measured as $\|g - P_{V_n} g\|^2_{\mathcal{L}_2}$, where the subscript indicates the $\mathcal{L}_2$ norm rather than the r. k. norm, the game would be the same, with

$$\inf_{V_n} \sup_{g \in C_p} \|g - P_{V_n} g\|^2_{\mathcal{L}_2} = \lambda^p_{n+1}.$$

12.2 Optimal quadrature and experimental design.

Let

$$f(t) = bg(t) + X(t), \quad t \in \mathcal{T}$$

where X is a zero-mean Gaussian stochastic process with $EX(s)X(t) = R(s,t)$. One will observe $f(t)$ for $t = t_1, \ldots, t_n$, g is known, and it is desired to estimate b. The Gauss–Markov estimate of b is

$$\hat{b} = \frac{f'\Sigma^{-1}g}{g'\Sigma^{-1}g}$$

where $f = (f(t_1) \ldots f(t_n))'$, $g = (g(t_1), \ldots, g(t_n))'$ and Σ is the $n \times n$ matrix with ijth entry $R(t_i, t_j)$. The variance of $\hat{b}$ is $(g'\Sigma^{-1}g)^{-1} = \|P_{V_n} g\|^{-2}$, where P_{V_n} is the orthogonal projection onto $R_{t_1}, \ldots, R_{t_n}$. Letting V_n be span $R_{t_1}, \ldots, R_{t_n}$, we have that this experimental design problem then is equivalent to: Choose $t_1, \ldots, t_n$ to minimize

$$\|g - P_{V_n} g\|^2.$$

Lower bounds on $\|g - P_{V_n} g\|^2$ follow from (12.1.7) in the case $g \in \mathcal{H}_{R^p}$ for $p > 1$. This problem was posed and studied by Sacks and Ylvisaker (1969), and studied by a number of authors (see Wahba (1971) and references there, also Wahba (1976, 1978c) and Athavale and Wahba (1979)). Let

$$Lh = \int \rho(u) h(u)\, du$$

and suppose one wishes to estimate Lh, given data $h(t_i) =< R_{t_i}, h >$, $i = 1, \ldots, n$. Let V_n be span $R_{t_1}, \ldots, R_{t_n}$. Let $P_{V_n} h$ be the minimal norm interpolant to this data; then

$$\widehat{Lh} = \int \rho(u)\ (P_{V_n} h)(u)\, du$$

gives a quadrature formula, that is, a formula of the form

$$\widehat{Lh} = \Sigma w_i h(t_i).$$

Letting h be the representer for L,

$$h(s) = \int R(s,u)\rho(u)\, du, \tag{12.2.1}$$

we have

$$\widehat{Lh} =< h, P_{V_n} h >=< P_{V_n} g, h >$$

and

$$|Lh - \widehat{Lh}| = | < h - P_{V_n} g, h > | \leq \|h - P_{V_n} g\| \|h\|.$$

The optimal quadrature problem then becomes the problem of choosing $t_1, \ldots, t_n$ to minimize $\|g - P_{V_n} g\|^2$. Note that g of (12.2.1) is in $\mathcal{H}_{R^p}$ with $p = 2$.

The major $\mathcal{T} = [0, 1]$ results are loosely described as follows. (For technical details, see the references.) Let the r.k. $R(s, t)$ be a Green's function for a $2m$th order linear differential operator (as in Section 1.2, for example), or equivalent to such an R. Let the characteristic discontinuity of R be

$$\lim_{s \downarrow t} \frac{\partial^{2m-1}}{\partial s^{2m-1}} R(s, t) - \lim_{s \uparrow t} \frac{\partial^{2m-1}}{\partial s^{2m-1}} R(s, t) = (-1)^m \alpha(t)$$

for some $\alpha(t) > 0$. Let $g(s) = \int_0^1 R(s, t) \rho(t)\, dt$ where ρ is strictly positive and has a bounded first derivative on $[0, 1]$. Then an asymptotically optimal design $R_{t_1}, \ldots, R_{t_n}$ for minimizing $\|g - P_{V_n} g\|^2$ is given by $t_1, \ldots, t_n$ satisfying

$$\int_0^{t_i} [\rho^2(u)\alpha(u)]^{1/(2m+1)} = \frac{i}{n} \int_0^1 [\rho^2(u)\alpha(u)]^{1/(2m+1)}\, du, \quad i = 1, \ldots, n. \quad (12.2.2)$$

These results have been used in a sequential procedure that involves starting with a trial design, estimating ρ, and then using (12.2.2) to obtain an additional set of design points, etc., (see Athavale and Wahba (1979)).

Very little is known of optimal designs for $\mathcal{T}$ other than $[0, 1]$. Optimal designs in the tensor product space associated with (12.1.2) can be expected to be different from those for the (thin-plate) space associated with (12.1.4), because the eigenfunctions associated with the largest eigenvalues are different. Some very curious examples for tensor product spaces are given in Wahba (1978c). The designs given there are for evaluation functionals and their span approximates the span of the eigenfunctions with large eigenvalues. These designs are known as blending function designs (see Delvos and Posdorf (1977)). Some recent related work can be found in Donoho and Johnstone (1989).

The noisy data case is of some importance but very little is known. Here let

$$y_i = < \eta_i, f > + \epsilon_i, \quad i = 1, \ldots, n$$

as before with $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \sim \mathcal{N}(0, \sigma^2 I)$ and let f_λ be the minimizer (for simplicity) of

$$\sum_{i=1}^{n} (y_i - < \eta_i, f >)^2 + \lambda \|f\|^2.$$

The problem is to choose $\eta_1, \ldots, \eta_n$ so that some loss function depending on $f - f_\lambda$ is small. One may require that $\eta_i = R_{t_i}$ for some $t_1, \ldots, t_n$, or one may have more freedom to choose the η_i's. In this latter case where the $\|\eta_i\|$ must be bounded to make the problem nontrivial; we set $\|\eta_i\| = 1$. Plaskota (1990) has recently shown that if f is a Gaussian stochastic process with $Ef(s)f(t) = R(s, t)$, then to minimize expected squared $\mathcal{L}_2$-norm of the error (with the expectation taken

over f as well as the ϵ_i's), the optimal η_i's are in the span of a proper subset of the first n eigenfunctions Φ_ν, with replications.

Suppose the design is $\eta_1, \ldots, \eta_N$ with η_i replicated n_i times, $\sum_{i=1}^N n_i = n$. Then the information available is assumed to be equivalent to

$$\tilde{y}_i = < \eta_i, f > + \tilde{\epsilon}_i, \quad i = 1, \ldots, N,$$

where $E\tilde{\epsilon}_i^2 = \sigma^2/n_i$. Here $\tilde{y}_i$ is the average of the n_i observations involving η_i, and f is estimated as the minimizer of

$$\sum_{i=1}^N n_i(\tilde{y}_i - < \eta_i, f >)^2 + n\lambda\|f\|^2. \tag{12.2.3}$$

Let us see what happens if it is assumed that the η_i's are in span $\{\Phi_1, \ldots, \Phi_n\}$. Let $\eta_\nu = \sqrt{\lambda_\nu}\Phi_\nu$: this ensures that $\|\eta_\nu\|_R = 1$. Then $< \eta_\nu, f > = f_\nu/\sqrt{\lambda_\nu}$, where $f_\nu = \int f(t)\Phi_\nu(t)dt$. After some calculations one obtains $f_\lambda = \sum_{\nu=1}^N \hat{f}_\nu \Phi_\nu$, where $\hat{f}_\nu = \sqrt{\lambda_\nu}(n_\nu/(n_\nu + n\lambda))\tilde{y}_\nu$. Further calculation then gives the expected squared $\mathcal{L}_2$-norm of the error as

$$E\|f - f_\lambda\|_{\mathcal{L}_2}^2 = \sum_{\nu=1}^N \left(\frac{n\lambda}{n_\nu + n\lambda}\right)^2 f_\nu^2 + \sigma^2 \sum_{\nu=1}^N \lambda_\nu \frac{n_\nu}{(n_\nu + n\lambda)^2} + \sum_{i=N+1}^\infty f_\nu^2. \tag{12.2.4}$$

Plaskota's assumption that $Ef(s)f(t) = R(s,t)$ entails that $Ef_\nu^2 = \lambda_\nu$ and the optimum $n\lambda$ averaged over sample functions is σ^2. Making these substitutions in (12.2.4), we obtain

$$\begin{aligned} E_f E\|f - f_\lambda\|_{\mathcal{L}_2}^2 &= \sum_{\nu=1}^N \left(\frac{\sigma^2}{n_\nu + \sigma^2}\right)^2 \lambda_\nu \\ &+ \sigma^2 \sum_{\nu=1}^N \frac{\lambda_\nu n_\nu}{(n_\nu + \sigma^2)^2} + \sum_{\nu=N+1}^\infty \lambda_\nu \\ &= \sigma^2 \sum_{\nu=1}^N \frac{\lambda_\nu}{(n_\nu + \sigma^2)} + \sum_{\nu=N+1}^\infty \lambda_\nu. \end{aligned} \tag{12.2.5}$$

Ignoring the requirement that the n_ν be integers we have that $\sum_{\nu=1}^N \lambda_\nu/(n_\nu + \sigma^2)$ is minimized over $n_1, \ldots, n_N$ subject to $\sum_{\nu=1}^N n_\nu = n$ when $(n_\nu + \sigma^2)$ is proportional to $\lambda_\nu^{1/2}$. This gives

$$n_\nu = \left(\frac{n + N\sigma^2}{\sum_{\mu=1}^N \lambda_\mu^{1/2}}\right)\lambda_\nu^{1/2} - \sigma^2$$

and (12.2.5) becomes

$$E_f E\|f_{\sigma^2/n}\|^2 = \sigma^2 \frac{(\sum_{\mu=1}^N \lambda_\mu^{1/2})^2}{n + N\sigma^2} + \sum_{N+1}^\infty \lambda_\mu^2.$$

The optimal N is then the greatest integer for which

$$\sigma^2 \frac{(\sum_{\mu=1}^{N} \lambda_\mu^{1/2})\lambda_N^{1/2}}{n + N\sigma^2} \leq \lambda_N$$

If it is only assumed that $f \in \mathcal{H}_R$, or $f \in \mathcal{C}_\rho$, then an optimal design would depend on the strategy for choosing λ, among other things. It appears plausible that such designs will involve replications of eigenfunctions of the r.k., however.

The nature of optimal designs when $\|f\|^2$ is replaced by $\|P_1 f\|^2$ in (12.2.3) is an open question.