CHAPTER 9

Further Nonlinear Generalizations

9.1 Partial spline models in nonlinear regression.

Consider the nonlinear partial spline model

$$y_i = \psi(t_i, \theta) + f(t_i) + \epsilon_i, \quad i = 1, \ldots, n \tag{9.1.1}$$

where $\theta = (\theta_1, \ldots, \theta_q)$ is unknown, $\psi(t_i, \theta)$ is given, $f \in \mathcal{H}_R$, and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)' \sim \mathcal{N}(0, \sigma^2 I)$.

We can fit this data by finding $\theta \in E^q$, $f \in \mathcal{H}$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \psi(t_i, \theta) - f(t_i))^2 + \lambda\|f\|^2. \tag{9.1.2}$$

Note that we have used a norm rather than a seminorm in the penalty functional in (9.1.2). Here any part of the "signal" for which there is to be no penalty should be built into ψ, to avoid hard-to-analyze aliasing when (9.1.2) is minimized using iterative methods. In most applications, f would be a smooth nuisance parameter, and testing whether or not it is zero would be a way of testing whether or not the model $\psi(t, \theta)$ is adequate.

It is easy to see that the minimizer f_λ of (9.1.2) must be of the form

$$f = \sum_{i=1}^{n} c_i R_{t_i}$$

where R_{t_i} is the representer of evaluation at t_i, in $\mathcal{H}_R$. Letting $\psi(\theta) = (\psi(t_1, \theta), \ldots, \psi(t_n, \theta))'$ and Σ be the $n \times n$ matrix with ijth entry $R(t_i, t_j)$, we have that the minimization problem of (9.1.2) becomes the following. Find $\theta \in E^q$ and $c \in E^n$ to minimize

$$\frac{1}{n}\|y - \psi(\theta) - \Sigma c\|^2 + \lambda c'\Sigma c. \tag{9.1.3}$$

For any fixed θ, we have (assuming Σ is of full rank), $c = c(\theta)$ satisfies

$$(\Sigma + n\lambda I)c(\theta) = y - \psi(\theta). \tag{9.1.4}$$

Substituting (9.1.4) into (9.1.3), (9.1.3) becomes the following. Find θ to minimize

$$\lambda(y-\psi(\theta))'(\Sigma+n\lambda I)^{-1}(y-\psi(\theta)). \tag{9.1.5}$$

The Gauss–Newton iteration goes as follows.

Let $T^{(l)}$ be the $n\times q$ matrix with $i\nu$th entry $\partial\psi(t_i,\theta)/\partial\theta_\nu|_{\theta=\theta^{(l)}}$, where $\theta^{(l)}$ is the lth iterate of θ. Expanding $\psi(\theta)$ in a first-order Taylor series gives

$$\psi(\theta)\simeq\psi(\theta^{(l)})-T^{(l)}\theta^{(l)}+T^{(l)}\theta.$$

Letting

$$z^{(l)}=y-\psi(\theta^{(l)})+T^{(l)}\theta^{(l)},$$

we have that $\theta^{(l+1)}$ is the minimizer of

$$(z^{(l)}-T^{(l)}\theta)'(\Sigma+n\lambda I)^{-1}(z^{(l)}-T^{(l)}\theta).$$

That is, $\theta^{(l+1)}$ satisfies

$$T^{(l)'}(\Sigma+n\lambda I)^{-1}T^{(l)}\theta^{(l+1)}=T^{(l)'}(\Sigma+n\lambda I)^{-1}z^{(l)}.$$

Letting

$$\Sigma=UDU'$$

gives

$$T^{(l)'}U(D+n\lambda I)^{-1}U'T^{(l)}\theta^{(l+1)}=T^{(l)}U(D+n\lambda I)^{-1}U'z^{(l)},$$

so that the same $n\times n$ matrix decomposition can be used for all iterations and values of λ.

For fixed λ, the iteration is carried to convergence, $l=L=L(\lambda)$, say, and the solution $(\theta_\lambda,c_\lambda)$ is the solution to the linearized problem

$$\frac{1}{n}\|z^{(L)}-T^{(L)}\theta-\Sigma c\|^2+\lambda c'\Sigma c, \tag{9.1.6}$$

for which the influence matrix $A(\lambda)=A^{(L)}(\lambda)$ is given by the familiar formula

$$I-A^{(L)}(\lambda)=n\lambda Q_2^{(L)}(Q_2^{(L)'}\Sigma Q_2^{(L)}+n\lambda I)^{-1}Q_2^{(L)'} \tag{9.1.7}$$

where

$$T^{(L)}=(Q_1^{(L)}:Q_2^{(L)})\begin{pmatrix}R^{(L)}\\0\end{pmatrix}.$$

One has $n\lambda c=(I-A^{(L)})z^{(L)}$, and the GCV function can be defined as

$$V(\lambda)=\frac{\frac{1}{n}\|(I-A^{(L)}(\lambda))z^{(L)}\|^2}{(\frac{1}{n}\mathrm{Tr}\,(I-A^{(L)}(\lambda)))^2}.$$

The matrix $A^{(L)}$ has q eigenvalues that are one, thus this formula is assigning q degrees of freedom for signal to the estimation of $(\theta_1,\ldots,\theta_q)$. It has been noted by many authors (see the comments in Friedman and Silverman (1989)) that when q parameters enter nonlinearly q may not be the real equivalent degrees of freedom. It is an open question whether a correction needs to be made here, and what modifications, if any, to the hypothesis testing procedure in Section 6 when the null space is not a linear space. See the recent book by Bates and Watts (1988) for more on nonlinear regression.

9.2 Penalized GLIM models.

Suppose

$$y_i \sim \text{ Binomial } (1, p(t_i)), \qquad i = 1, \ldots, n, \tag{9.2.1}$$

and let the logit $f(t)$ be $f(t) = \log[p(t)/(1-p(t))]$, where f is assumed to be in $\mathcal{H}_R$, and it is desired to estimate f. The negative log-likelihood $\mathcal{L}(y)$ of the data is

$$\mathcal{L}(y) = -\sum_{i=1}^{n}(y_i \log p(t_i) + (1-y_i)\log(1-p(t_i))) \tag{9.2.2}$$

and, since $p = e^f/(1+e^f)$,

$$\mathcal{L}(y) = \sum_{i=1}^{n}(\log(1+e^{f(t_i)}) - y_i f(t_i)) = Q(y,f), \text{ say.} \tag{9.2.3}$$

McCullagh and Nelder (1983) in their book on GLIM (generalized linear models) suggest assuming that f is a parametric function (for example, $f(t) = \theta_1 + \theta_2 t$), and estimating the parameters by minimizing $Q(y,f) = Q(y,\theta)$. O'Sullivan (1983) and O'Sullivan, Yandell, and Raynor (1986) considered the estimation of f by supposing that $f \in \mathcal{H}_R$ and finding f to minimize the penalized log-likelihood

$$I_\lambda(y,f) = Q(y,f) + \lambda\|P_1 f\|^2, \tag{9.2.4}$$

and extended GCV to this setup.

Penalized likelihood estimates with various penalty functionals have been proposed by a number of authors. (See the references in O'Sullivan (1983); we note only the work of Silverman (1978) and Tapia and Thompson (1978) where the penalty functionals are seminorms in reproducing kernel spaces. See also Leonard (1982).)

If (9.2.1) is replaced by

$$y_i \sim \text{ Poisson } (\Lambda(t_i))$$

and $f(t) = \log \Lambda(t)$, we have

$$Q(y,f) = \sum_{i=1}^{n}\left\{e^{f(t_i)} - y_i f(t_i) + \log(y_i!)\right\}. \tag{9.2.5}$$

Of course $y_i \sim \mathcal{N}(f(t_i), \sigma^2)$ is the model we have been considering all along, and the setup we are discussing works whenever y_i has an exponential density of the form

$$p(y_i) = e^{-[\{b(f(t_i)) - y_i f(t_i)\}/a_i] + c(y_i)} \tag{9.2.6}$$

where a_i, b, and c are given.

Here we note for further reference that $Ey_i = b'(f(t_i))$ and var $y_i = b''(f(t_i))a_i$, and below we will let $a_i = 1$.

Approximating f by a suitable basis function representation in $\mathcal{H}_R$, we have

$$f \simeq \sum_{k=1}^{N} c_k B_k$$

and we need to find $c = (c_1, \ldots, c_N)'$ to minimize

$$I_\lambda(c) = \sum_{i=1}^{n} b\left(\sum_{k=1}^{N} c_k B_k(t_i)\right) - y_i\left(\sum_{k=1}^{N} c_k B_k(t_i)\right)$$

$$+\lambda \sum_{k,k'=1}^{N} c_k c_{k'} < P_1 B_k, P_1 B_{k'} > . \tag{9.2.7}$$

Using a Newton–Raphson iteration this problem can be solved iteratively, and at the last step of the iteration one can obtain an approximating quadratic problem, from which one can extract a GCV function.

The second-order Taylor expansion of $I_\lambda(c)$ for the lth iterate $c^{(l)}$ is

$$I_\lambda(c) \simeq I_\lambda(c^{(l)}) + \nabla I_\lambda(c - c^{(l)}) + \frac{1}{2}(c - c^{(l)})' \nabla^2 I_\lambda(c - c^{(l)}) \tag{9.2.8}$$

where the gradient ∇I_λ is given by

$$\nabla I_\lambda = \left(\frac{\partial I_\lambda}{\partial c_1}, \ldots, \frac{\partial I_\lambda}{\partial c_N}\right)\bigg|_{c=c^{(l)}} \tag{9.2.9}$$

and the Hessian $\nabla^2 I_\lambda$ is the $N \times N$ matrix with jkth entry

$$\{\nabla^2 I_\lambda\}_{jk} = \frac{\partial^2 I_\lambda}{\partial c_j \partial c_k}\bigg|_{c=c^{(l)}}. \tag{9.2.10}$$

Then $c = c^{(l+1)}$, the minimizer of (9.2.8), is given by

$$c^{(l+1)} = c^{(l)} - (\nabla^2 I_\lambda)^{-1} \nabla I'_\lambda. \tag{9.2.11}$$

Letting X be the $n \times N$ matrix with ikth entry $B_k(t_i)$ and Σ be the $N \times N$ matrix with kk'th entry $< P_1 B_k, P_1 B_{k'} >$, we have that (9.2.7) becomes

$$I_\lambda(c) = \sum_{i=1}^{n} b((Xc)_i) - y'Xc + \lambda c'\Sigma c,$$

where $(Xc)_i$ is the ith entry of the vector Xc. We have

$$\nabla I'_\lambda = X'(\mu(c) - y) + 2\lambda\Sigma c \tag{9.2.12}$$

where $\mu(c) = (\mu_1(c), \ldots, \mu_n(c))'$ with $\mu_i(c) = b'(f(t_i)) = b'((Xc)_i)$; and

$$\nabla^2 I_\lambda = X'D(c)X + 2\lambda\Sigma \tag{9.2.13}$$

where $D(c)$ is the $n \times n$ diagonal matrix with iith entry $b''(f(t_i)) = b''((Xc)_i)$.

Substituting (9.2.12) and (9.2.13) into (9.2.11) gives the Newton–Raphson update

$$c^{(l+1)} = c^{(l)} - (X'D(c^{(l)})X + 2\lambda\Sigma)^{-1}(-X'(y - \mu(c^{(l)})) + 2\lambda\Sigma c^{(l)}) \quad (9.2.14)$$

$$= (X'D(c^{(l)})X + 2\lambda\Sigma)^{-1}X'D^{1/2}(c^{(l)})z^{(l)} \quad (9.2.15)$$

where the pseudodata $z^{(l)}$ is

$$z^{(l)} = D^{-1/2}(c^{(l)})(y - \mu(c^{(l)})) + D^{1/2}(c^{(l)})Xc^{(l)}. \quad (9.2.16)$$

Then $c^{(l+1)}$ is the minimizer of

$$\frac{1}{2}\|z^{(l)} - D^{1/2}(c^{(l)})Xc\|^2 + \lambda c'\Sigma c. \quad (9.2.17)$$

The predicted value $\hat{z}^{(l)} = D^{1/2}Xc$ of $z^{(l)}$ is related to $z^{(l)}$ by

$$\hat{z}^{(l)} = A(\lambda)z^{(l)}$$

where

$$A(\lambda) = D^{1/2}(c^{(l)})X(X'D(c^{(l)})X + \lambda\Sigma)^{-1}X'D^{1/2}(c^{(l)}). \quad (9.2.18)$$

Running (9.2.14) to convergence, (9.2.17) at convergence becomes

$$\frac{1}{2}\|D^{-1/2}(c)(y - \mu(c))\|^2 + \lambda c'\Sigma c \quad (9.2.19)$$

and letting $w = D^{-1/2}(c)y$ and $\hat{w} = D^{-1/2}\mu(c)$, it is seen that

$$\frac{\partial \hat{w}_i}{\partial w_j} \simeq (A(\lambda))_{ij}$$

resulting in the GCV function

$$V(\lambda) = \frac{\|D^{-1/2}(c)(y - \mu(c))\|^2}{(\text{Tr}\,(I - A(\lambda)))^2}$$

evaluated at the converged value of c.

Properties of these estimates are discussed in O'Sullivan (1983), Cox and O'Sullivan (1989, 1990), and Gu (1990). The method has been extended to the Cox proportional hazards model and other applications by O'Sullivan (1986b, 1988b).

9.3 Estimation of the log-likelihood ratio.

Suppose one is going to draw a sample of n_1 observations from population 1 with density $h_1(t)$, $t \in \mathcal{T}$, and a sample of n_2 observations from population 2 with density $h_2(t)$, and it is desired to estimate $f(t) = \log\,(h_1(t)/h_2(t))$, the log-likelihood ratio. Without loss of generality we will suppose $n_1 = n_2 = n/2$. (Details for removing this limitation may be found in Villalobos and Wahba (1987).) Suppose the n observations are labeled $t_1, \ldots, t_n$, and with each observation is attached a tag y_i, $i = 1, \ldots, n$ that is 1 if t_i came from population 1, and 0 if t_i came from population 2.

Given that an observation has value t_i, the conditional probability that its tag y_i is 1, is $h_1(t_i)/(h_1(t_i) + h_2(t_i)) = p(t_i)$, and we have that $f(t) = \log(p(t)/(1 - p(t)) = \log(h_1(t)/h_2(t))$. f can be estimated by minimizing $Q(y, f) + \lambda\|P_1 f\|^2$, where Q is given by (9.2.3). This way of looking at log likelihood estimation is due to Silverman (1978).

Note that if h_1 and h_2 are d-variate normal densities, ($t = (x_1, \ldots, x_d)$), then f is quadratic in $x_1, \cdots, x_d$ and will be in the null space of the thin plate penalty functional for $m = 3$ (provided $6 - d > 0$; see Section 2.4). Thus, if h_1 and h_2 are believed to be "close" to multivariate normal, then this penalty functional is a natural one (see Silverman (1982)).

9.4 Linear inequality constraints.

Suppose we observe

$$y_i = L_i f + \epsilon_i, \tag{9.4.1}$$

and it is known a priori that $f \in \mathcal{C} \subset \mathcal{H}_R$ where $\mathcal{C}$ is a closed convex set.

We want to find $f \in \mathcal{H}_R$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - L_i f)^2 + \lambda\|P_1 f\|^2 \tag{9.4.2}$$

subject to $f \in \mathcal{C}$. Since any closed convex set can be characterized as the intersection of a family of half planes, we can write

$$\mathcal{C} = \{f : < \chi_s, f > \geq \alpha(s), \quad s \in \mathcal{S}\},$$

for some family $\{\chi_s, \ s \in \mathcal{S}\}$. Frequently, we can approximate $\mathcal{C}$ by $\mathcal{C}_L$,

$$\mathcal{C}_L = \{f : < \chi_s, f > \geq \alpha(s), \ s = s_1, \ldots, s_L\},$$

where $\chi_{s_1}, \ldots, \chi_{s_L}$ is a discrete approximation to $\{\chi_s, \ s \in \mathcal{S}\}$. For example, if $\mathcal{C} = \{f : \ f(t) \geq 0, \ t \in \mathcal{T}\}$, then we have $\{\chi_s, \ s \in \mathcal{S}\} = \{R_s, \ s \in \mathcal{T}\}$, and if $\mathcal{T} = [0, 1]$, we may approximate $\mathcal{C}$ by $\mathcal{C}_L = \{R_{1/L}, R_{2/L}, \ldots, R_{L/L}\}$. If $\mathcal{C}_L$ is a good approximation to $\mathcal{C}$, one may frequently find after minimizing (9.4.2) subject to $f \in \mathcal{C}_L$ that the result is actually in $\mathcal{C}$. Letting η_i be the representer of L_i and $\xi_i = P_1\eta_i$, and letting $\rho_j = P_1\chi_{s_j}$, it is known from Kimeldorf and Wahba (1971), that if

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - L_i f)^2$$

has a unique minimizer in $\mathcal{H}_0$, then (9.4.2) has a unique minimizer in $\mathcal{C}_L$, and it must have a representation

$$\sum_{i=1}^{n} c_i \xi_i + \sum_{j=1}^{L} b_j \rho_j + \sum_{\nu=1}^{M} d_\nu \phi_\nu$$

for some coefficient vectors $a = (c' : b')'$, and d. The coefficients a and d are found by solving the following quadratic programming problem. Find $a \in E^{n+L}$, $d \in E^d$ to minimize

$$\|\Sigma_1 a + T_1 d - y\|^2 + n\lambda a' \Sigma a \tag{9.4.3}$$

subject to

$$\Sigma_2 a + T_2 d \geq \alpha \tag{9.4.4}$$

where

$$\Sigma_1 = (\Sigma_{11} : \Sigma_{12}), \tag{9.4.5}$$

$$\Sigma_2 = (\Sigma_{21} : \Sigma_{22}), \tag{9.4.6}$$

$$\Sigma = \begin{pmatrix} \Sigma_1 \\ \cdots \\ \Sigma_2 \end{pmatrix} \tag{9.4.7}$$

and the Σ_{ij} and T_i are given in Table 9.1 ($\Sigma_{12} = \Sigma'_{21}$), and $\alpha = (\alpha(s_1), \ldots, \alpha(s_L))'$.

TABLE 9.1
Definitions of Σ_{ij} and T_i.

Matrix	Dimension	ijth entry
Σ_{11}	$n \times n$	$< \xi_i, \xi_j >$
Σ_{12}	$n \times L$	$< \xi_i, \rho_j >$
Σ_{22}	$L \times L$	$< \rho_i, \rho_j >$
T_1	$n \times M$	$< \eta_i, \phi_j >$
T_2	$L \times M$	$< \chi_{s_i}, \phi_j >$

A GCV function can be obtained for constrained problems via the "leaving-out-one" lemma of Section 4.2.

Let $f_\lambda^{[k]}$ be the minimizer in C_L of

$$\frac{1}{n} \sum_{\substack{i=1 \\ i \neq k}}^{n} (y_i - L_i f)^2 + \lambda \|P_1 f\|^2$$

(supposed unique) and let

$$a_{kk}^*(\lambda, \delta) = \frac{L_k f_\lambda[k, \delta] - L_k f_\lambda}{\delta} \tag{9.4.8}$$

where $f_\lambda[k,\delta]$ is the minimizer of (9.4.2) in C_L with y_k replaced by $y_k+\delta$. If there are no active constraints, then L_kf_λ is linear in the components of y and

$$a^*_{kk}(\lambda,\delta)=\frac{\partial L_kf_\lambda}{\partial y_k}=a_{kk}(\lambda), \tag{9.4.9}$$

where $a_{kk}(\lambda)$ is the kkth entry of the influence matrix of (1.3.23). From Theorem 4.2.1, we have that the ordinary cross-validation function $V_0(\lambda)$ satisfies

$$V_0(\lambda)\equiv\frac{1}{n}\sum_{i=1}^{n}(y_k-L_kf_\lambda^{[k]})^2\equiv\frac{1}{n}\sum_{i=1}^{n}(y_k-L_kf_\lambda)^2/(1-a^*_{kk}(\lambda,\delta_k)) \tag{9.4.10}$$

where $\delta_k=L_kf_\lambda^{[k]}-y_k$. By analogy with the linear case, the GCV function is defined as

$$V(\lambda)=\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-L_if_\lambda)^2}{(1-\frac{1}{n}\sum_{k=1}^{n}a^*_{kk}(\lambda,\delta_k))^2}. \tag{9.4.11}$$

To evaluate $V(\lambda)$ for a single value of λ we need to solve n quadratic programming problems in $n+L-M$ variables. To avoid this it is suggested in Wahba (1982c) that the approximate GCV function

$$V_{\text{app}}(\lambda)=\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-L_if_\lambda)^2}{(1-\frac{1}{n}\sum_{k=1}^{n}a_{kk}(\lambda))^2} \tag{9.4.12}$$

where

$$a_{kk}(\lambda)=\left.\frac{\partial L_kf_\lambda}{\partial y_k}\right|_y \tag{9.4.13}$$

be used. The right-hand side of (9.4.13) is well defined and continuous in λ except at boundaries when a constraint changes from active to inactive or vice versa.

We can obtain $\partial L_kf_\lambda/\partial y_k$ for the constrained problem by examining an approximating quadratic minimization problem. It is not hard to see that the approximating problem is found as follows. Fix λ, and solve the quadratic programming problem of (9.4.3) and (9.4.4). Find all the active constraints (suppose there are l). Now let $\tilde\Sigma_1,\tilde\Sigma_2,\tilde\Sigma,\tilde T_1$, and $\tilde T_2$, and $\tilde\alpha$, be defined as the corresponding elements in (9.4.5), (9.4.6), (9.4.7), and Table 9.1 with all rows and/or columns corresponding to inactive constraints deleted. Then c,d, and the nonzero values of b in the solution to quadratic programming problems of (9.4.3) and (9.4.4) are given by the solution to the following equality constrained minimization problem. Find $\tilde a,\tilde d$ to minimize

$$\|\tilde\Sigma_1\tilde a+\tilde T_1\tilde d-y\|^2+n\lambda\tilde a'\tilde\Sigma\tilde a \tag{9.4.14}$$

subject to

$$\tilde\Sigma_2\tilde a+T_2\tilde d=\tilde\alpha. \tag{9.4.15}$$

To write the solution to this minimization problem quickly, let

$$W_\xi = \begin{pmatrix} I_{n\times n} & O_{n\times l} \\ O_{l\times n} & \xi I_{l\times l} \end{pmatrix}$$

and consider the following minimization problem. Find $\tilde{a}$ and d to minimize

$$(\tilde{\Sigma}\tilde{a} + \tilde{T}d - \tilde{y})' W_\xi^{-1} (\tilde{\Sigma}\tilde{a} + \tilde{T}d - \tilde{y}) + n\lambda \tilde{a}'\tilde{\Sigma}\tilde{a}, \qquad (9.4.16)$$

where

$$\tilde{y} = \begin{pmatrix} y \\ \tilde{\alpha} \end{pmatrix}.$$

It is not hard to see that the minimizer of (9.4.16) satisfies

$$(\tilde{\Sigma} + n\lambda W_\xi)\tilde{a} + \tilde{T}d = \tilde{y}, \qquad (9.4.17)$$

$$\tilde{T}'\tilde{a} = 0, \qquad (9.4.18)$$

for any W_ξ with $\xi > 0$, and if we let $\xi \to 0$ we get the minimizer of (9.4.14) subject to (9.4.15).

Let the QR decomposition of $\tilde{T}$ be

$$\tilde{T} = (\tilde{Q}_1 : \tilde{Q}_2)\begin{pmatrix} \tilde{R} \\ 0 \end{pmatrix}$$

and letting $W = W_0$, we can derive, following the arguments leading to (1.3.19),

$$\tilde{a} = \tilde{Q}_2(\tilde{Q}_2'(\tilde{\Sigma} + n\lambda W)\tilde{Q}_2)^{-1}\tilde{Q}_2'\tilde{y}. \qquad (9.4.19)$$

Now

$$\hat{\tilde{y}} \equiv \begin{pmatrix} \hat{y} \\ \tilde{\alpha} \end{pmatrix} = \tilde{\Sigma}\tilde{a} + \tilde{T}d, \qquad (9.4.20)$$

and subtracting (9.4.20) from (9.4.17) gives

$$\begin{pmatrix} y - \hat{y} \\ 0 \end{pmatrix} = n\lambda W\tilde{a} = n\lambda W\tilde{Q}_2(\tilde{Q}_2'(\tilde{\Sigma} + n\lambda W)\tilde{Q}_2)^{-1}\tilde{Q}_2'\tilde{y}, \qquad (9.4.21)$$

so for $j = 1, \ldots, n$, the jjth entry in the matrix on the right of (9.4.21) is $1 - (\partial L_j f_\lambda / \partial y_j)|_y$. Thus

$$\sum_{j=1}^{n} a_{jj}(\lambda) = n - n\lambda \ \mathrm{tr}\ \Delta\,(\Phi + n\lambda\Delta)^{-1}$$

where

$$\Delta = \tilde{Q}_2' W \tilde{Q}_2$$

and

$$\Phi = \tilde{Q}_2'\tilde{\Sigma}\tilde{Q}_2;$$

furthermore,

$$\text{tr } \Delta (\Phi + n\lambda\Delta)^{-1} = \sum_{j=1}^{n+l-M} \frac{w_j}{1+n\lambda w_j}$$

where $w_1, \ldots, w_{n+l-M}$ are the eigenvalues of the real symmetric eigenvalue problem

$$\Delta u_j = w_j \Phi u_j, \quad j = 1, \ldots, n+l-M.$$

These arguments are from Villalobos and Wahba (1987). (The derivations there are modified to apply directly to thin-plate splines.) A numerical strategy for carrying out the computation of $V_{\text{app}}(\lambda)$ is given there. It uses an "active set" algorithm of Gill et al. (1982) for the quadratic optimization problem. This type of algorithm is known to converge rapidly when a good starting guess is available for the active constraint set. If the set of active constraints changes slowly with λ, then a good guess for the active set for an updated λ is the active set for the preceding λ. The unconstrained problem is solved to obtain a starting guess. Discontinuities in $V_{\text{app}}(\lambda)$ as the active constraint set changed were evident in the examples tried in Villalobos and Wahba (1987), but were not a practical problem. Another recent work on inequality constraints is Elfving and Andersson (1988).

9.5 Inequality constraints in ill-posed problems.

In the solution of Fredholm integral equations of the first kind, if the number of linearly independent pieces of information in the data (see Section 8.2) is small, the imposition of known linear inequality constraints may add crucial missing information. A numerical experiment that illustrates this fact was performed in Wahba (1982c) and here we show two examples from that study. Data were generated according to the model

$$y_i = \int_0^1 k(\frac{i}{n} - u) f(u) du + \epsilon_i, \quad i = 1, \cdots, n \tag{9.5.1}$$

with $n = 64$, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and $\sigma = .05$. The periodic convolution kernel k is shown in Figure 9.1. The two example f's from that paper are the solid black lines in Figures 9.2(a) and 9.3(a). (The explicit formulae may be found in Wahba (1982c).) The dashed curves in Figures 9.2(a) and 9.3(a) are $g(x) = \int_0^1 k(x-u) f(u) du$, and the circles are the data y_i. f was estimated as the minimizer of

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - \int_0^1 k\left(\frac{i}{n} - u\right) f(u) du)^2 + \lambda \int_0^1 (f''(u))^2 du \tag{9.5.2}$$

in a 64-dimensional space of sines and cosines, using GCV to estimate λ. The numerical problem here is much simplified over that in Section 9.4 due to the periodic nature of the problem and the equally spaced data. The estimates are given as the finely dashed curves marked $f_{\hat{\lambda}}$ in Figures 9.2(b) and 9.3(b).

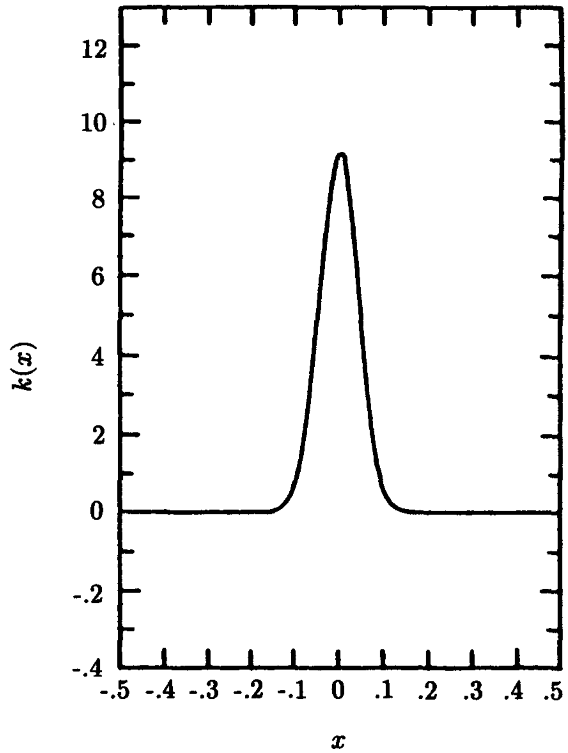

FIG. 9.1. *The convolution kernel.*

Then (9.5.2) was minimized in the same space of sines and cosines, subject to $f(\frac{i}{n}) \geq 0$, $i = 1, \ldots, n$, and V_{app} of (9.4.12) used to choose λ. The result is the coarsely dashed curves in the same figures, marked $f^c_{\hat{\lambda}_c}$. It can be seen that the imposition of the positivity constraints reduces the erroneous side lobes in the solution as well as improves the resolution of the peaks. We remark that although in theory there are 64 strictly positive eigenvalues, in this problem the ratio $(\lambda_{42}/\lambda_1)^{1/2}$ was 10^{-7}.

9.6 Constrained nonlinear optimization with basis functions.

Let

$$y_i = N_i f + \epsilon_i \tag{9.6.1}$$

where N_i is a nonlinear functional, and suppose it is known that

$$< \chi_s, f > \geq \alpha(s), \qquad s \in \mathcal{S}. \tag{9.6.2}$$

Approximating f by

$$f \simeq \sum_{k=1}^{N} c_k B_k \tag{9.6.3}$$

and $\mathcal{S}$ by $\{s_1, \ldots, s_J\}$, we seek to find c to minimize

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - N_i(c))^2 + \lambda c' \Sigma c \tag{9.6.4}$$

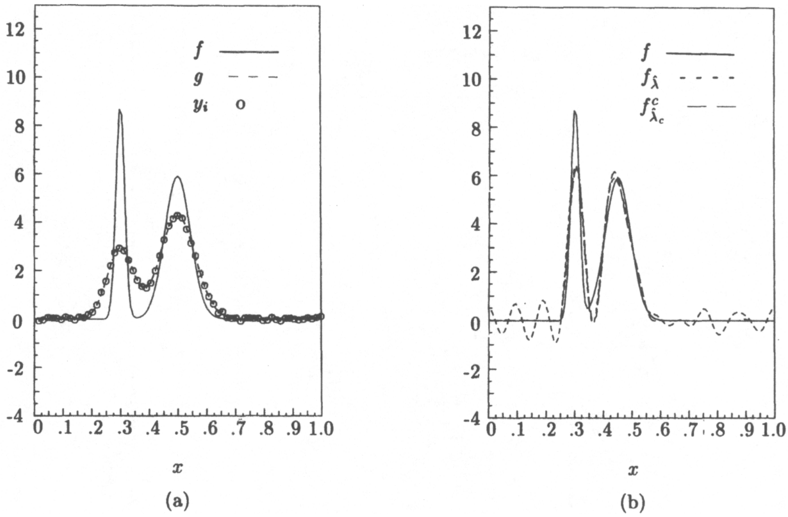

FIG. 9.2. *Example* 1.

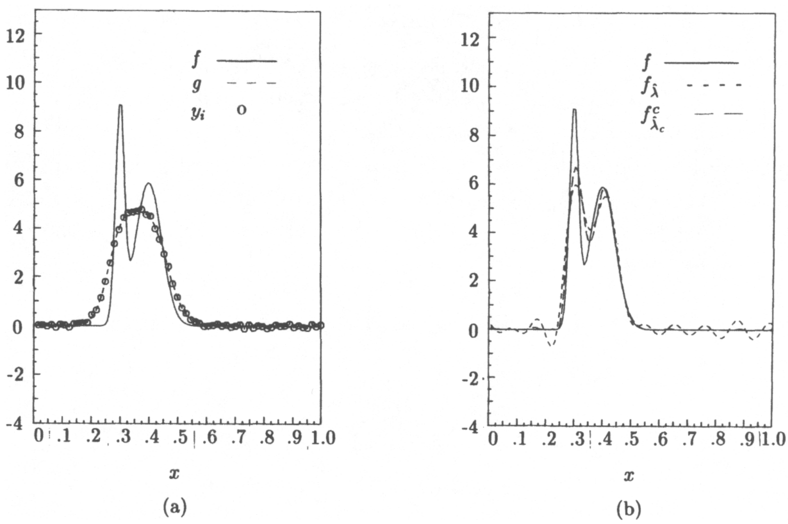

FIG. 9.3. *Example* 2.

subject to

$$\sum_{k=1}^{N} c_k < \chi_{s_r}, B_k > \geq \alpha(s_r), \qquad r = 1, \ldots, J. \tag{9.6.5}$$

Here, Σ is the $N \times N$ matrix with ijth entry $< P_1 B_i, P_1 B_j >$, as in Section 7.1. Letting

$$N_i(c) \simeq N_i(c^{(l)}) + \sum_{k=1}^{N} \frac{\partial N_i}{\partial c_k}(c_k - c_k^{(l)})$$

as in (8.3.5), and letting $X^{(l)}$ be the $n \times N$ matrix with i kth entry $\partial N_i / \partial c_k |_{c=c^{(l)}}$, and letting

$$y^{(l)} = y - \begin{pmatrix} N_1(c^{(l)}) \\ \vdots \\ N_n(c^{(l)}) \end{pmatrix} + X^{(l)} c^{(l)}$$

be as in (8.3.5)–(8.3.7), at the lth step of an iterative solution we have that the problem is to find c to minimize

$$\frac{1}{n} \| y^{(l)} - X^{(l)} c \|^2 + \lambda c' \Sigma c$$

subject to

$$Cc \geq \alpha$$

where C is the $J \times N$ matrix with rkth entry $< \chi_{s_r}, B_k >$, and $\alpha = (\alpha(s_1), \ldots, \alpha(s_J))'$. In principle at least, this problem can be iterated to convergence for fixed λ, and V_{app} for constrained functions can be evaluated. Here the influence matrix for V_{app} is

$$A^{(L)}(\lambda) = X^{(L)} F' (F X^{(L)'} X^{(L)} F' + n\lambda F \Sigma F')^{-1} F X^{(L)'}$$

where $X^{(L)}$ is the converged value of $X^{(l)}$, and, if there are J' active constraints and $C^{(L)}$ is the $J' \times N$ submatrix of C corresponding to these J' constraints, then F is any $N - J' \times N$ matrix with $F'F = I_{N-J'}$ and $FC^{(L)} = 0_{N-J' \times J}$.

9.7 System identification.

The key idea in this section (equation (9.7.14)), which allows the use of GCV in the system identification problem, has been adapted from O'Sullivan (1986a). Kravaris and Seinfeld (1985) have proposed the method of (9.7.4) below that, adopting the nomenclature of the field, might be called the penalized output least squares method. Another important recent reference is O'Sullivan (1987b), where convergence properties of the method are discussed.

The dynamic flow of fluid through a porous medium is modeled by a diffusion equation

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} - \frac{\partial}{\partial \mathbf{x}} \left\{ \rho(\mathbf{x}) \frac{\partial}{\partial \mathbf{x}} u(\mathbf{x}, t) \right\} = q(\mathbf{x}, t), \ \mathbf{x} \in \Omega, t \in [t_{\min}, t_{\max}] \tag{9.7.1}$$

subject to prescribed initial and boundary conditions, for example, $u(\mathbf{x},0) = u_0(\mathbf{x})$ (initial condition) and $\partial u/\partial w = 0$ where w is the direction normal to the boundary. Here, if $\mathbf{x} = (x_1, \ldots, x_d)$ then $\partial/\partial\mathbf{x} = \sum_{\alpha=1}^{d} \partial/\partial x_\alpha$. Here u is, say, pressure, q represents a forcing function (injection of fluid into the region), and ρ is the transmittivity or permeability of the medium. If u_0 and q are known exactly, then for fixed ρ in some appropriate class ($\rho(\mathbf{x}) > 0$, in particular), u is determined (implicitly) as a function of ρ. Typically, ρ must be nonnegative to be physically meaningful, and such that there is measurable flow.

The practical problem is, given measurements

$$y_{ij} = u(\mathbf{x}(i), t_j, \rho) + \epsilon_{ij} \tag{9.7.2}$$

on u, the initial boundary functions, and q, estimate ρ.

We remark that if $\partial/\partial\mathbf{x}u(\mathbf{x},t)$ is zero for $\mathbf{x}$ in some region $\Omega_0 \subset \Omega$, all t, then there is no information in the experiment concerning $\rho(\mathbf{x})$ for $x \in \Omega_0$. Although the algorithm below may provide an estimate for $\rho(\mathbf{x})$ for $x \in \Omega_0$, in this case the information is coming from the prior, and not the experiment.

This is an extremely important practical problem; see, e.g., the references in O'Sullivan (1986a) and Kravaris and Seinfeld (1985). Deveaux and Steele (1989) study a somewhat different but related inverse problem.

The problem will be solved approximately in the span of a suitable set of N basis functions

$$\rho(\mathbf{x}) = \sum_{k=1}^{N} c_k B_k(\mathbf{x}),$$

and since ρ must be nonnegative, we put a sufficiently large number of linear inequality constraints on $c = (c_1, \ldots, c_N)$, that is,

$$\sum_{k=1}^{N} c_k B_k(\mathbf{x}) \geq 0 \tag{9.7.3}$$

for $\mathbf{x}$ in some finite set, so that the estimate is positive. If stronger information than just positivity is known, then it should be used. We seek to find c subject to (9.7.3) to minimize

$$\sum_{ij} (y_{ij} - u(\mathbf{x}(i), t_j, c))^2 + \lambda c'\Sigma c, \tag{9.7.4}$$

where $c'\Sigma c = \|P_1\rho\|^2$. For the moment we suppose that u_0 and q are known exactly. Then

$$u(\mathbf{x}(i), t_j, \rho) \simeq u(\mathbf{x}(i), t_j; c)$$

is a nonlinear functional of c, but only defined implicitly. If $u(\mathbf{x}(i), t_j; c)$ could be linearized about some reasonable starting guess

$$\rho_0(\mathbf{x}) = \sum_{k=1}^{N} c_k^{(0)} B_k(\mathbf{x})$$

then the methods of Section 9.6 could be used to numerically find the minimizing c_λ and to choose λ by GCV.

Given a guess $c^{(l)}$ for c, we would like to be able to linearize about $c^{(l)}$,

$$u(\mathbf{x}(i), t_j; c) \simeq u(\mathbf{x}(i), t_j; c^{(l)}) + \sum_k X_{ijk}(c_k - c_k^{(l)}), \tag{9.7.5}$$

where

$$X_{ijk} = \frac{\partial u}{\partial c_k}(\mathbf{x}(i), t_j; c)\bigg|_{c=c^{(l)}} \tag{9.7.6}$$

If this could be done, then c and λ could be determined, at least in principle, via the constrained Gauss–Newton iteration and the GCV procedure described in Section 9.6.

Let

$$L_c = \frac{\partial}{\partial t} - \frac{\partial}{\partial \mathbf{x}}\left\{\sum_{k=1}^{N} c_k B_k(\mathbf{x})\frac{\partial}{\partial \mathbf{x}}\right\}, \tag{9.7.7}$$

let

$$\mathcal{B} = \{u : \ u \text{ satisfies the given initial and boundary conditions,}\}$$

$$\mathcal{B}_0 = \{u : \ u \text{ satisfies homogeneous initial and boundary conditions,}\}$$

and let

$$\delta_k = (0, \cdots, 0, \delta, 0, \cdots, 0), \ \delta \text{ in the } k\text{th position.}$$

Let u_c be the solution to

$$L_c u_c = q, \qquad u_c \in \mathcal{B}, \tag{9.7.8}$$

let $u_{c+\delta_k}$ be the solution to

$$L_{c+\delta_k} u_{c+\delta_k} = q, \qquad u_{c+\delta_k} \in \mathcal{B}, \tag{9.7.9}$$

and let

$$h_{c,k}(\delta) = \frac{u_{c+\delta_k} - u_c}{\delta}. \tag{9.7.10}$$

Observe that

$$L_{c+\delta_k} = L_c - \delta\frac{\partial}{\partial \mathbf{x}}B_k(\mathbf{x})\frac{\partial}{\partial \mathbf{x}}; \tag{9.7.11}$$

then substituting (9.7.9) into (9.7.10) gives

$$\left(L_c - \delta\frac{\partial}{\partial \mathbf{x}}B_k(\mathbf{x})\frac{\partial}{\partial \mathbf{x}}\right)(u_c + \delta h_{c,k}(\delta)) = q, \tag{9.7.12}$$

$$u_c + \delta h_{c,k}(\delta) \in \mathcal{B}.$$

Subtracting (9.7.8) from (9.7.12) and dividing through by δ gives

$$L_c h_{c,k}(\delta) = \frac{\partial}{\partial \mathbf{x}}B_k(\mathbf{x})\frac{\partial}{\partial \mathbf{x}}(u_c + \delta h_{c,k}(\delta)). \tag{9.7.13}$$

Assuming that we can take limits as $\delta \to 0$, and letting $\lim_{\delta\to 0} h_{c,k}(\delta) = h_{c,k}$, this gives that $h_{c,k}$ is the solution to the problem

$$L_c h = \frac{\partial}{\partial \mathbf{x}} B_k(\mathbf{x}) \frac{\partial}{\partial \mathbf{x}} u_c$$

$$h \in \mathcal{B}_0.$$

Thus if everything is sufficiently "nice," $X_{ijk}^{(l)}$ can be obtained by solving

$$L_{c^{(l)}} h = \frac{\partial}{\partial \mathbf{x}} B_k(\mathbf{x}) \frac{\partial}{\partial \mathbf{x}} u_{c^{(l)}} \tag{9.7.14}$$

and evaluating the solution at $\mathbf{x}(i), t_j$.

O'Sullivan (1988a) has carried out this program on a one-dimensional example.

We emphasize that this is a nonlinear ill-posed problem complicated by the fact that the degree of nonlinearity as well as the degree of ill-posedness can depend fairly strongly on the unknown solution. To see more clearly some of the issues involved, let us examine a problem sitting in Euclidean n-space that has many of the features of the system identification problem. Let $X_1, \ldots, X_N$ be $N \le n$ matrices each of dimension $(n-M) \times n$, let B be an $M \times n$ matrix of rank M, and let $u \in E^n$, $q \in E^{n-M}$, and $b \in E^M$ be related by

$$\left(\sum_{k=1}^{N} c_k X_k\right) u = q \tag{9.7.15}$$

$$Bu = b.$$

Think of c, q, and b, respectively, as stand-ins for ρ, the forcing function, and the initial/boundary conditions.

Suppose q and b are known exactly and it is known a priori that $c_k \ge \alpha_k > 0$, $k = 1, \ldots, N$, and that this condition on the c_k's ensures that the matrix $\begin{pmatrix} \sum_k c_k X_k \\ B \end{pmatrix}$ is invertible. Suppose that one observes

$$y_i = u_i + \epsilon_i, \quad i = 1, \ldots, n$$

where u_i is the ith component of u. Letting $\Psi_{ij}(c)$ be the ijth entry of $\begin{pmatrix} \sum_k c_k X_k \\ B \end{pmatrix}^{-1}$, we may estimate c as the minimizer of

$$\sum_{i}^{n} \left(y_i - \sum_{j=1}^{n-M} \Psi_{ij}(c) q_j - \sum_{j=n-M+1}^{n} \Psi_{ij}(c) b_{j-(n-M)} \right)^2 + \lambda c' \Sigma c, \tag{9.7.16}$$

subject to $c_k \ge \alpha_k$. The ability to estimate the c's can be expected to be quite sensitive to the true values of c as well as q and b.

Returning to the original system identification problem, we now consider the case where the boundary conditions are not completely known. If (as in a one-dimensional, steady-state problem) there are only $M << n$ unknowns in the initial/boundary values, then the analogue of (9.7.16) could (in principle) be minimized with respect to c and $b = (b_1, \ldots, b_M)$.

More generally, suppose that the forcing function q and the boundary conditions $\partial u/\partial w = 0$ are known exactly, but the initial conditions $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ are observed with error, that is

$$z_i = u_0(\mathbf{x}(i)) + \epsilon_i.$$

Modeling $u_0(\mathbf{x})$ as

$$u_0(\mathbf{x}) \simeq \sum_{\nu=1}^{M} b_\nu \tilde{B}_\nu(\mathbf{x}) \tag{9.7.17}$$

where the $\tilde{B}_\nu$ are appropriate basis functions (not necessarily the same as before) and letting $b = (b_1, \ldots, b_M)$, we have

$$u \simeq u(\mathbf{x}, t; c, b)$$

and we want to choose b and c, subject to appropriate constraints, to minimize

$$\frac{1}{n}\left\{\sum_{ij}(y_{ij} - u(\mathbf{x}(i), t_j; c, b))^2 + \sum_i \left(z_i - \sum_\nu b_\nu \tilde{B}_\nu(\mathbf{x}(i))\right)^2\right\}$$

$$+ \lambda_1 c'\Sigma c + \lambda_2 b'\tilde{\Sigma} b \tag{9.7.18}$$

where $b'\tilde{\Sigma}b$ is an appropriate penalty on u_0. The penalty functionals $c'\Sigma c$ and $b'\tilde{\Sigma}b$ may be quite different, since the first contains prior information about the permeability and the second about the field. This expression assumes that all the measurement errors have the same variance.

For fixed λ_1 and λ_2 this minimization can, in principle, be done as before, provided we have a means of calculating

$$z_{ij\nu} = \frac{\partial u}{\partial b_\nu}(\mathbf{x}(i), t_j; c; b). \tag{9.7.19}$$

The $z_{ij\nu}$ can be found by the same method used for the X_{ijk}. Let $u_{c,b}$ be the solution to the problem

$$L_c u_{c,b} = q, \quad \frac{\partial u_{c,b}}{\partial w} = 0, \qquad u_{c,b}(\mathbf{x}, 0) = \Sigma b_\nu \tilde{B}_\nu(\mathbf{x}). \tag{9.7.20}$$

Let $\delta_\nu = (0, \ldots, \delta, \ldots, 0), \delta$ in the νth position, and let $u_{c,b+\delta_\nu}$ be the solution to

$$L_c u_{c,b+\delta_\nu} = q, \quad \frac{\partial u_{c,b+\delta_\nu}}{\partial w} = 0, \quad u_{c,b+\delta_\nu}(\mathbf{x}, 0) = \delta \tilde{B}_\nu + \left(\sum_\mu b_\mu \tilde{B}_\mu(\mathbf{x})\right) \tag{9.7.21}$$

and let

$$\tilde{h}_{c,b,\nu}(\delta) = \frac{u_{c,b+\delta_\nu} - u_{c,b}}{\delta}. \tag{9.7.22}$$

Then, subtracting (9.7.21) from (9.7.20) as before, we see that $\tilde{h}_{c,b,\nu}(\delta) = \tilde{h}_{c,b,\nu}(0)$ is the solution to the problem

$$L_c u = 0, \ \frac{\partial u}{\partial w} = 0, u(\mathbf{x}, 0) = \tilde{B}_\nu(\mathbf{x}). \tag{9.7.23}$$

$V(\lambda_1, \lambda_2)$ can be minimized, at least in principle, to estimate good values of λ_1 and λ_2 by GCV.