ELSEVIER

Statistics & Probability Letters 61 (2003) 111–121

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

On exact minimax wavelet designs obtained by simulated annealing

Alwell J. Oyet[a], Douglas P. Wiens[b,*]

[a] *Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Ffld., Canada, AIC 5S7*

[b] *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alta., Canada, T6G 2G1*

Received October 2001; received in revised form September 2002

Abstract

We construct minimax robust designs for estimating wavelet regression models. Such models arise from approximating an unknown nonparametric response by a wavelet expansion. The designs are robust against errors in such an approximation, and against heteroscedasticity. We aim for *exact*, rather than approximate, designs; this is facilitated by our use of simulated annealing. The relative simplicity of annealing allows for a much more complete treatment of some hitherto intractable problems initially addressed in Oyet and Wiens (J. Nonparametric Stat. 12 (2000) 837). Thus, we are able to exhibit integer-valued designs for estimating higher order wavelet approximations of nonparametric curves. The exact designs constructed for multiwavelet approximations of various orders are found to be symmetric and periodic, as anticipated in Oyet and Wiens (J. Nonparametric Stat. 12 (2000) 837). We also construct integer-valued designs based on the Daubechies wavelet system with a wavelet number of 5.

Keywords: Daubechies wavelet; Heteroscedastic; Multiwavelet; Nonparametric regression; Rounding; Weighted least squares

1. Introduction

In this article we consider the construction of regression designs, following an approximation of a regression response by an initial segment of its wavelet series expansion. Such designs should be

* Corresponding author. Tel.: +1-780-492-4406; fax: +1-780-492-4826. Research of both authors supported by the Natural Sciences and Engineering Research Council of Canada.

E-mail address: doug.wiens@ualberta.ca (D.P. Wiens).

0167-7152/03/$ - see front matter
PII: S0167-7152(02)00361-9

robust against the bias engendered by model misspecification as well as from natural, and possibly heterogeneous, variation.

In designing an experimental study, researchers typically must represent the relationship between an experimental outcome and the design, or input variables, by an approximate response function. For such purposes it makes sense to search for functions which are flexible and which can adapt to the structure of the underlying processes governing the behaviour of several alternate systems. Here the adaptive structure of wavelets becomes appealing. Like Fourier series, wavelet series have the ability to approximate arbitrarily closely any square integrable regression response. Thus wavelet analysis complements, and since the early 1980s has to a certain extent replaced, traditional Fourier analysis as a standard tool of the applied mathematician or statistician.

We note that wavelet representations inherit the smoothness properties of the particular wavelet system used in the approximation. For instance, a Haar wavelet representation will follow the general structure of the response but show up as a step function. Thus, an important step is choosing an appropriate wavelet system for a particular response. Our examples in this work are based on the Daubechies wavelets (Daubechies, 1992) and the multiwavelet systems (Alpert, 1992) due to their flexibility in representing curves of varying shapes. A response curve with jumps and cusps can be adequately represented using the multiwavelet system. The Daubechies wavelet system is however best suited for representing smooth curves. Several examples of the estimation of response curves involving the Haar and the Daubechies wavelet systems in a nonparametric framework can be found in Härdle et al. (1998).

Our approach is to represent the regression response as a series of components of a particular wavelet system. In this way, a nonparametric problem is transformed into an approximate linear regression problem in which finitely many such components serve as regressors. It turns out that regression weights can be helpful in increasing the accuracy of the wavelet approximation; several examples can be found in Härdle et al. (1998), see also Antoniadis et al. (1994).

The problem of constructing designs involves determining the points within the input or design space at which measurements should be taken in order to minimize the error in estimation or prediction. Such design points should control both for the bias resulting from the truncation of the wavelet series in the formation of the linear regression model, and for variation. See for instance, Box and Draper (1959) or Wiens (1992) for developments of these notions in general regression models. Particular applications to wavelet models have been studied by Herzberg and Traves (1994), Xie (1998), Oyet and Wiens (2000) and Oyet (2002).

A common approach in optimal design theory is to exhibit "approximate designs" that are, possibly continuous, probability functions $\xi(\mathbf{x})$ on the design space. The number of observations, out of a total of n, allocated to a particular design point $\mathbf{x}_i$ is then $n\xi(\mathbf{x}_i)$, with the experimenter left to approximate this quantity when it is not an integer. In this article our focus is instead on constructing exact—i.e. integer-valued—rather than approximate designs for estimating nonparametric response curves based on wavelet approximations. For this we employ simulated annealing algorithms, which have previously been put to this purpose by Bohachevsky et al. (1986), Haines (1987), Fang and Wiens (2000) and Zhou (2001). Our use of these algorithms has made it possible to exhibit exact designs for multiwavelet models of higher orders than those discussed in Oyet and Wiens (2000). We are also able to construct designs based on the Daubechies wavelet approximation without imposing a condition of unbiasedness as was required in Oyet (2002).

Of particular interest is the fact that certain structural features—symmetry and periodicity—of the exact designs, found here to be optimal for multiwavelet models, agree with those imposed on the continuous designs of Oyet and Wiens (2000) even though the *restriction* to symmetric, periodic designs has been dropped in the present work.

We outline the general theory underlying our study in Section 2. Section 3 discusses the minimax problems of interest and the annealing algorithm. We study several examples of the application of the algorithm in constructing designs based on the multiwavelet and Daubechies wavelet systems.

2. Preliminaries

We consider the situation in which the square integrable response function $E[Y|x]=\eta(x)$ can be decomposed into 'details' $\eta^{(j)}$ at various levels j such that

$$\eta(x)=\eta^{\phi}+\eta^{(0)}+\eta^{(1)}+\cdots+\eta^{(\alpha)}+\cdots=\eta(x;\alpha)+f(x). \tag{1}$$

In (1), η^{ϕ} is a scalar multiple $d\cdot\phi(x)$ of a 'scaling' function $\phi(x)$ and $\eta^{(j)}$ is a linear combination of 2^j dilated and translated versions of a 'primary wavelet' $\psi(x)$. Thus, $\eta^{(j)}=\sum_{k=0}^{2^j-1} c_{jk}\psi^{-j,k}(x)$, where $\psi^{-j,k}(x)=2^{j/2}\psi(2^jx-k)$. For the multiwavelet system, this representation will involve $L\geqslant 1$ scaling functions and a corresponding number of primary wavelets. See Fig. 1 for examples.

We denote by $f(x)$ the remainder after an approximation of $\eta(x)$ by a wavelet decomposition $\eta(x;\alpha)=\eta^{\phi}+\eta^{(0)}+\eta^{(1)}+\cdots+\eta^{(\alpha)}$. Thus, $f(\cdot)$ accounts for the uncertainty in the "true" structure of $\eta(\cdot)$. It turns out that we can write $\eta(x;\alpha)=\mathbf{q}^{\mathrm{T}}(x)\boldsymbol{\beta}$, where the $r=L\cdot 2^{\alpha+1}$ elements of the vector $\mathbf{q}(x)$ are the components of the wavelet system used in the decomposition and $\boldsymbol{\beta}$ is made up of the filter coefficients $\{d_l, c_{jkl}\}$ (the subscript $l\in\{0,\ldots,L-1\}$ is needed when dealing with multiwavelets). The response can then be written as

$$Y=E[Y|x]+\varepsilon(x)=\mathbf{q}^{\mathrm{T}}(x)\boldsymbol{\beta}+f(x)+\varepsilon(x).$$

Example 2.1. *Multiwavelet systems*: For $x\in[0,1)$ the multiwavelet systems employ scaling functions $\phi_l(x)=\sqrt{2l+1}P_l(2x-1)$, where $P_l(y)$ is the lth degree Legendre polynomial with the standard normalization $\int_{-1}^{1} P_l^2(y)\,\mathrm{d}y=(l+0.5)^{-1}$. The primary wavelets have somewhat more involved definitions. We give them here only for $L=1$ and 2, which are the values of interest in our examples. For $L=1$ ("Haar wavelets") the primary wavelet is $\psi_{1,0}(x)=I_{[0,1/2)}(x)-I_{[1/2,1)}(x)$. For $L=2$ the primary wavelets are

$$\psi_{2,0}(x)=\sqrt{3}(4|x-1/2|-1)\cdot I_{[0,1)}(x),$$

$$\psi_{2,1}(x)=2(1-3|x-1/2|)\cdot(I_{[0,1/2)}(x)-I_{[1/2,1)}(x)).$$

See Fig. 1(a) and (b). The vector $\mathbf{q}(x)$ contains the $L\cdot 2^{\alpha+1}$ regressors $\{\phi_l(x), \psi_{L,l}^{-j,k}(x)\,|\,j=0,\ldots,\alpha,\ k=0,\ldots,2^j-1,\ l=0,\ldots,L\}$ in some order.

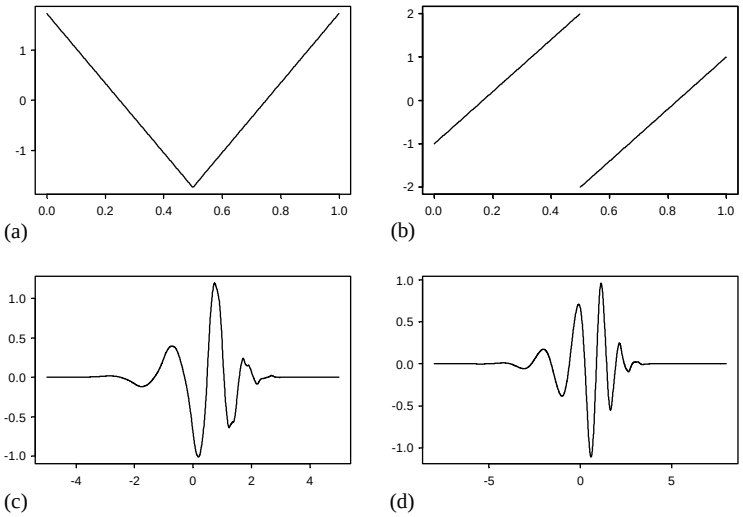

Fig. 1. Primary multiwavelets: (a) ${}_2\psi_0(x)$, (b) ${}_2\psi_1(x)$. Daubechies primary wavelets: (c) ${}_5\psi(x)$, (d) ${}_8\psi(x)$.

Example 2.2. *Daubechies wavelets*: The scaling functions and primary wavelets of the Daubechies wavelet systems, commonly denoted ${}_N\phi(x)$ and ${}_N\psi(x)$, respectively, have no closed form expressions. They are constructed numerically for different values of the "wavelet number" N which identifies the number of nonvanishing coefficients in the "dilation equation" ${}_N\phi(x) = \sum_k c_k\, {}_N\phi(2x-k)$ used in the construction. The choice $N=1$ again yields the Haar wavelets. Once ${}_N\phi(x)$ has been constructed, the corresponding primary wavelet is obtained from ${}_N\psi(x) = \sum_k (-1)^k c_{1-k}\, {}_N\phi(2x-k)$. The functions ${}_N\phi(x)$ and ${}_N\psi(x)$ have compact support with vanishing moments of orders 1 to N. This property, commonly referred to as a moment condition, guarantees good approximation properties of the corresponding wavelet expansion of a response function $\eta(x)$ in ${}_N\phi(x)$ and ${}_N\psi(x)$. That is, it determines how quickly the wavelet expansion will converge to the true response $\eta(x)$. See Härdle et al. (1998) for details. A useful recursive algorithm we have used in constructing the examples in Fig. 1(c) and (d) can be found in Strang (1989).

The development in the rest of this section is analogous to that in Fang and Wiens (2000) and so we give only a survey. The design points are to be chosen from a finite but otherwise arbitrary design space $\mathscr{S} = \{x_i\}_{i=1}^N$. We assume that $\mathrm{Var}(\varepsilon(x)) = \sigma^2 g(x)$ for some $g(\cdot)$ satisfying $N^{-1}\sum_{i=1}^N g^2(x_i) \leqslant 1$. The problem is that of determining the "optimal" number $n_i \geqslant 0$ of observations to be taken at x_i, where $\sum_{i=1}^N n_i = n$ is specified in advance.

We define the regression parameter by $\boldsymbol{\beta} = \arg\min_{\boldsymbol{\omega}} N^{-1}\sum_{i=1}^N \{\eta(x_i) - \mathbf{q}^{\mathrm{T}}(x_i)\boldsymbol{\omega}\}^2$. The condition $N^{-1}\sum_{i=1}^N \mathbf{q}(x_i) f(x_i) = \mathbf{0}$ arising from this definition ensures the identifiability of $\boldsymbol{\beta}$. To ensure a balance between the contributions of bias and variation to the overall error, we place a bound on the magnitude of the remainder: $N^{-1}\sum_{i=1}^N f^2(x_i) \leqslant \tau^2$ for a constant τ.

Our results depend on the values of σ^2 and τ^2 only through that of $v := \sigma^2/(n\tau^2)$; this quantity can in turn be chosen by the experimenter to reflect his view of the relative importance of variance versus bias. We give a more complete discussion of this point immediately preceding Section 3.1.

Upon making uncorrelated observations Y_{ij} $(j=1,\dots,n_i)$ at x_i $(i=1,\dots,N)$ the experimenter estimates $\boldsymbol{\beta}$ by (possibly weighted) least squares:

$$\hat{\boldsymbol{\beta}} = \left[\frac{1}{n}\sum_{i=1}^{N} n_i w_i \mathbf{q}(x_i)\mathbf{q}^{\mathrm{T}}(x_i)\right]^{-1} \frac{1}{n}\sum_{i=1}^{N}\sum_{j=1}^{n_i} w_i \mathbf{q}(x_i) Y_{ij}.$$

Let $\{p_i = n_i/n\}$ be the integer-valued design on $\mathscr{S}$ and define $m_i = p_i w_i$. There is no restriction in assuming that the weights average to 1 on $\mathscr{S}$, i.e. that $\sum_{i=1}^{N} m_i = 1$. Then we are seeking a probability distribution $\{m_i\}$ and possibly weights $\{w_i\}$ which, subject to the constraint $\sum_{i=1}^{N}(m_i/w_i) = 1$, minimize the maximum, over f and g, value of the loss function.

Let $\mathbf{Q}$ be the matrix with rows $\mathbf{q}^{\mathrm{T}}(x_1),\dots,\mathbf{q}^{\mathrm{T}}(x_N)$ and define diagonal matrices $\mathbf{M}=\mathrm{diag}(m_1,\dots,m_N)$, $\mathbf{W} = \mathrm{diag}(w_1,\dots,w_N)$, $\mathbf{G} = \mathrm{diag}(g(x_1),\dots,g(x_N))$ and a vector $\mathbf{f} = (f(x_1),\dots,f(x_N))^{\mathrm{T}}$. Define $\mathbf{b} = \mathbf{Q}^{\mathrm{T}}\mathbf{M}\mathbf{f}$, $\mathbf{B} = \mathbf{Q}^{\mathrm{T}}\mathbf{M}\mathbf{Q}$, $\mathbf{D} = \mathbf{Q}^{\mathrm{T}}\mathbf{M}\mathbf{W}\mathbf{G}\mathbf{Q}$. Then the bias vector and covariance matrix of $\hat{\boldsymbol{\beta}}$ are $\mathbf{d} = \mathrm{bias}(\hat{\boldsymbol{\beta}}) = \mathbf{B}^{-1}\mathbf{b}$ and $\mathbf{C} = \mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\sigma^2/n)\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}$, respectively.

The design criterion used here is the minimization of the average mean squared Error (AMSE) $A = A(f,g,w,m)$ of $\hat{\eta}(x;\alpha) = \mathbf{q}^{\mathrm{T}}(x)\hat{\boldsymbol{\beta}}$ as an estimate of $\eta(x)$. This measure of loss is the discrete version of integrated mean squared error (IMSE) and is defined by

$$A = \frac{1}{N}\sum_{i=1}^{N} E[(\hat{\eta}(x_i;\alpha) - \eta(x_i))^2] = \mathrm{AV} + \mathrm{ASB},$$

where the average variance (AV) and the average squared Bias (ASB) are

$$\mathrm{AV} = \frac{1}{N}\sum_{i=1}^{N} E[\{\hat{\eta}(x_i;\alpha) - E[\hat{\eta}(x_i;\alpha)]\}^2] = \frac{1}{N}\,\mathrm{tr}[\mathbf{Q}\mathbf{C}\mathbf{Q}^{\mathrm{T}}],$$

$$\mathrm{ASB} = \frac{1}{N}\sum_{i=1}^{N} \{E[\hat{\eta}(x_i;\alpha)] - \eta(x_i)\}^2 = \frac{1}{N}\,\|\mathbf{Q}\mathbf{d}\|^2 + \frac{1}{N}\,\|\mathbf{f}\|^2.$$

We assume that $\mathbf{Q}_{N\times r}$ is of full rank and define the singular value decomposition of $\mathbf{Q}$ by $\mathbf{Q} = \mathbf{U}_{N\times r}\boldsymbol{\Lambda}_{r\times r}\mathbf{V}^{\mathrm{T}}_{r\times r}$, where $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}_r$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of the square roots of the eigenvalues of $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$. Then with $\mathbf{M}_j := \mathbf{U}^{\mathrm{T}}\mathbf{M}^j\mathbf{U}$ and h_i the ith diagonal element of $\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^{\mathrm{T}}$:

$$\mathrm{AV} = \frac{\sigma^2}{nN}\,\mathrm{tr}[\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^{\mathrm{T}}\mathbf{M}\mathbf{W}\mathbf{G}] = \tau^2\frac{v}{N}\sum_{i=1}^{N} m_i w_i g_i h_i,$$

$$\mathrm{ASB} = \frac{1}{N}\,\mathbf{f}^{\mathrm{T}}[\mathbf{I} + \mathbf{M}\mathbf{U}\mathbf{M}_1^{-2}\mathbf{U}^{\mathrm{T}}\mathbf{M}]\mathbf{f}.$$

We note that the number of terms in $\hat{\boldsymbol{\beta}}$ doubles with each unit increase in α. Thus, one may consider shrinking $\hat{\boldsymbol{\beta}}$ by wavelet thresholding in order to eliminate the components of $\hat{\boldsymbol{\beta}}$ that are "small" when compared to a fixed threshold parameter $\lambda \geqslant 0$. In other words, components of $\hat{\boldsymbol{\beta}}$ that contribute little to estimating $\eta(x)$ are set to zero before the AMSE of $\hat{\eta}(x;\alpha)$ is computed. Let $\mathbf{J}_{r\times r} = \mathrm{diag}(\mathbf{1}(|\hat{\beta}_1| > \lambda),\dots,\mathbf{1}(|\hat{\beta}_r| > \lambda))$. Two common threshold estimators (see Vidakovic, 1999)

correspond to "hard thresholding":

$$\boldsymbol{\delta}^h(\hat{\boldsymbol{\beta}}, \lambda) = \mathbf{J}\hat{\boldsymbol{\beta}}$$

and "soft thresholding":

$$\boldsymbol{\delta}^s(\hat{\boldsymbol{\beta}}, \lambda) = \mathbf{J}(\hat{\boldsymbol{\beta}} - \operatorname{sgn}(\hat{\boldsymbol{\beta}}) \cdot \lambda).$$

In either case, $\hat{\eta}(x;\alpha) = \mathbf{q}^{\mathrm{T}}(x)\boldsymbol{\delta}(\hat{\boldsymbol{\beta}}, \lambda)$ where $\boldsymbol{\delta}(\hat{\boldsymbol{\beta}}, \lambda)$ is a nonlinear function of $\hat{\boldsymbol{\beta}}$. Consequently, the expressions for AV and ASB become more involved, leading to further complications in computing the least favourable function $f(x)$ in Section 3. We do not consider threshold estimators in this article.

3. Minimax wavelet designs

In this section we discuss solutions to the minimax design problem

$$\min_{\{w_i, m_i\}_{i=1}^N} \max_{f,g} A(f, g, w, m),$$

with f and g satisfying the conditions of Section 2. This problem is denoted by P3. In addition to P3, we shall discuss solutions to two other special cases:

P1. Set $g = w = 1$ for all values of x, corresponding to ordinary least squares in the presence of homoscedastic errors. Then $\{m_i = p_i\}$ is to minimize $\max_f A(f, \mathbf{1}, \mathbf{1}, m)$.

P2. Set $w = 1$ for all values of x, corresponding to ordinary least squares in the presence of heteroscedastic errors. Then $\{m_i = p_i\}$ is to minimize $\max_{f,g} A(f, g, \mathbf{1}, m)$.

Oyet and Wiens (2000) constructed approximate continuous designs for P1 based on the multiwavelet system. For first-order ($\alpha = 0$) multiwavelet approximations these were—partly for mathematical convenience—restricted to periodic and symmetric designs. Oyet (2002) then extended the results to P2 and P3 under the same restrictions. Our use of annealing allows us to obtain solutions for higher order approximations without these restrictions.

We observe that general analytic solutions can be found for the problems of maximizing $A(f, g, w, m)$ with respect to f and g and minimizing with respect to the weights w. The minimization with respect to m is then carried out numerically using the simulated annealing algorithm. From Fang and Wiens (2000) we have the following results for P1, P2 and P3, respectively:

$$\mathscr{A}_1(\mathbf{m}) := \max_f A(f, \mathbf{1}, \mathbf{1}, m) = \tau^2 \left[\lambda_1(\mathbf{m}) + \frac{v}{N} \sum_{i=1}^N m_i h_i \right], \tag{2}$$

$$\mathscr{A}_2(\mathbf{m}) := \max_{f,g} A(f, g, \mathbf{1}, m) = \tau^2 \left[\lambda_1(\mathbf{m}) + \frac{v}{\sqrt{N}} \left(\sum_{i=1}^N m_i^2 h_i^2 \right)^{1/2} \right], \tag{3}$$

$$\mathscr{A}_3(\mathbf{m}) := \min_w \max_{f,g} A(f, g, w, m) = \tau^2 \left[\lambda_1(\mathbf{m}) + \frac{v}{\sqrt{N}} \left(\sum_{i=1}^N m_i^{4/3} h_i^{2/3} \right)^{3/2} \right], \tag{4}$$

where $\mathbf{m} = (m_1, \ldots, m_N)$ and $\lambda_1(\mathbf{m})$ is the largest eigenvalue of $\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{M}_1^{-1}$. In (2) and (3) the design is obtained directly from m: $p_i = m_i = n_i/n$. The minimizing weights in (4) are $w_i m_i^{-1/3} h_i^{-2/3}$, where the constant of proportionality is $\sum_{i=1}^{N} m_i^{4/3} h_i^{2/3}$. In this case $p_i = m_i w_i$ and the search for optimal *integers* n_i becomes more involved.

Note that in (2)–(4), τ^2 appears only multiplicatively and so can be ignored. Only v need be specified; it can be viewed as a weighting factor supplied by the experimenter and reflecting the relative importance to him of loss due to variation versus that due to bias.

3.1. Integer-valued designs for P1 and P2

The simulated annealing algorithm is used to select the numbers n_i of observations (equivalently the design $\{p_i = n_i/n\}$) minimizing the maximum loss (2) or (3). We restrict our discussion to the case where one of (n, N) is a multiple of the other. We also assume that the experiment is to be carried out on the interval [0,1) and assume $\mathscr{S}$ to have equally spaced design points:

$$x_i = \frac{2i-1}{2N}, \quad i = 1, 2, \ldots, N.$$

The algorithm begins with the experimenter specifying an initial configuration $\mathbf{n} = (n_1, \ldots, n_N)$ for the system. A set of procedures and a criterion for generating subsequent and acceptable new configurations is then implemented, iteratively, until the minimum is found.

The algorithm used imposes symmetry around $\frac{1}{2}$ on the designs; we did this for computational efficiency only after noting that the designs obtained without this restriction were still invariably symmetric. The initial configuration, and method of randomly generating new configurations, is as used by Fang and Wiens (2000). A new configuration $\tilde{\mathbf{m}}$ is accepted, and iterations continue, if the difference in loss $\Delta\mathscr{A} = \mathscr{A}(\tilde{\mathbf{m}}) - \mathscr{A}(\mathbf{m})$ is negative. Otherwise, $\tilde{\mathbf{m}}$ is accepted if the value $\exp\{-\Delta\mathscr{A}/T\}$ of the Boltzmann acceptance probability exceeds 0.5. The initial value of the 'temperature' T is chosen in such a way that at least 50% of the new configurations are accepted with the value reduced by a factor of 0.9 after each 100 iterations; see Haines (1987) and Press et al. (1989). In the terminology of thermodynamics, the main idea is that the path to a global minimum might go through points of high energy states. The lower the temperature, the less likely are any significant excursions into such states. Thus, the progressive reduction in T is necessary in order to strike a minimum.

Example 3.1.1. *Multiwavelet systems*: We exhibit designs for the multiwavelet approximation obtained by implementing the annealing algorithm described above for $v = 0.5$, 5 and 10. Fig. 2(a) illustrates a typical example of the progress of the algorithm. The orders of approximation considered are $\alpha = 0$, 1 and 2. See Fig. 2(b)–(g). Designs for $\alpha \geqslant 3$ can be obtained in a similar manner.

For the case illustrated in Fig. 2(c) we find that the exact designs are the same for $v = 5$ and 10. In all cases the structure of the exact designs constructed here, and the approximate continuous designs of Oyet and Wiens (2000), are similar in the sense that they are periodic, with period $2^{-(\alpha+1)}$, and symmetric within each of the $2^{\alpha+1}$ subintervals $[k \cdot 2^{-(\alpha+1)}, (k+1) \cdot 2^{-(\alpha+1)}]$.

Example 3.1.2. *Daubechies wavelets*: We use the Daubechies wavelet ${}_5\psi(x)$ in our examples. This choice is based on the fact that it appears to perform better than the others in approximating curves

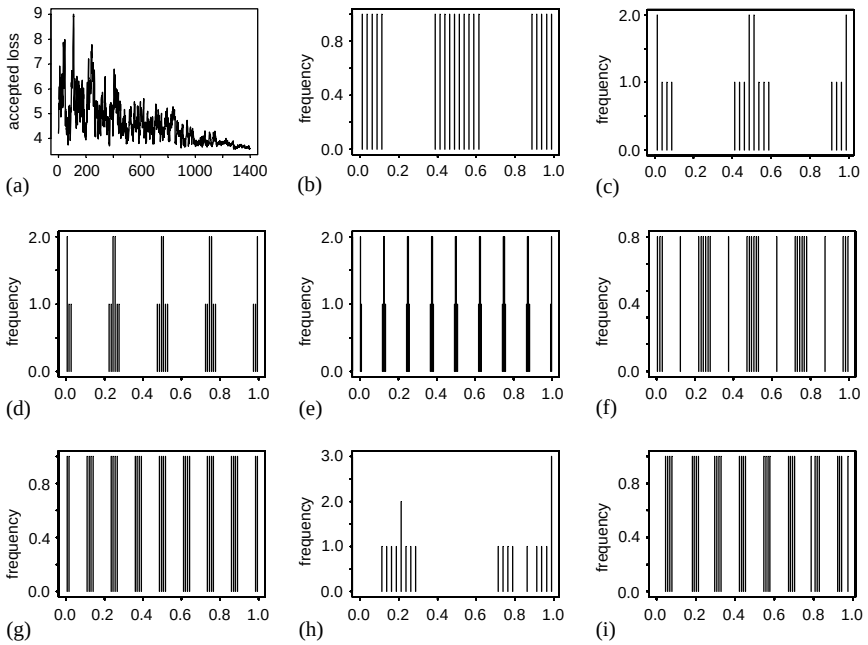

Fig. 2. Exact integer-valued designs for P1 and P2. Figs. (a)–(e) are for the $L = 2$ multiwavelet system and P1. (a) Accepted loss versus iteration number; $(N, n, \alpha, v) = (40, 20, 0, 0.5)$. (b)–(e) Design points and frequencies. Values of (N, n, α, v) are: $(40, 20, 0, 0.5)$, $(40, 20, 0, 5)$, $(96, 32, 1, 10)$, $(192, 48, 2, 10)$, resp. (f), (g) $L = 2$ multiwavelet system and P2; $(N, n, \alpha, v) = (84, 28, 1, 5)$ and $(96, 32, 2, 10)$, resp. (h),(i) Daubechies wavelet designs for P1 and P2 using $(N, n, \alpha, v) = (40, 20, 0, 10)$ and $(96, 32, 2, 5)$, resp.

based on weighted least squares; see Oyet (2002). As in Fig. 2(h) and (i) the optimal designs are nonperiodic and nonsymmetric. As α increases we find that the designs become periodic and symmetric around $x = \frac{1}{2}$.

3.2. Approximate designs for P3

For P1 and P2, it was not difficult to construct exact integer-valued designs using the annealing algorithm because $m_i = p_i = n_i/n$. The situation is not the same for P3. Here, $n_i = nm_i/w_i$ is not necessarily an integer, even if the nm_i are integers. Thus, the n_i must be rounded to integers which sum to n. The rounding techniques we have applied are the *quota* method of Kiefer (1971) and the *efficient rounding* or *multiplier* method of Pukelsheim and Rieder (1992). Both are discussed in Pukelsheim (1993).

Example 3.2.1. Examples involving the multiwavelet and the Daubechies wavelet systems are shown in Fig. 3. One feature which can be seen in Figs. 3(b)–(d) is symmetry about $x = \frac{1}{2}$. This property is absent in the plots for the Daubechies wavelet in Figs. 3(f)–(h). In these particular examples, the quota and the multiplier methods yielded the same designs. In certain other examples (not shown) this was not the case.

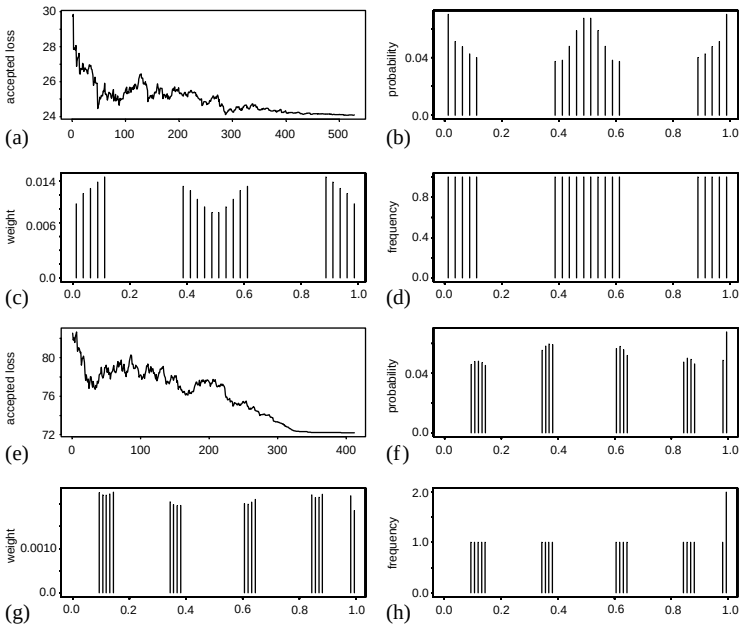

Fig. 3. Approximate integer-valued designs for P3. $L=2$ multiwavelets with $(N,n,\alpha,v)=(40,20,0,5)$. (a) accepted loss, (b) probabilities p_i, (c) weights w_i, (d) design points and frequencies. Daubechies wavelets with $(N,n,\alpha,v)=(80,20,1,10)$: (e) accepted loss, (f) probabilities p_i, (g) weights w_i and (h) design points and frequencies.

3.3. Case study: nitrite utilization

Bates and Watts (1988) describes an experiment by Elliott and Peirson on the utilization of nitrite in bush beans. In the experiment, portions of primary leaves from three 16-day-old bean plants were subjected to eight levels of light intensity $x\in[0,170]$ (μE/m^{2}s), and the nitrite utilization Y (nmol/g h) was measured. The experiment was performed on two different days, resulting in two sets of 24 observations at eight levels of light intensity. Since the experimenters did not have a theoretical model for explaining the behaviour of nitrite utilization at various light intensities, Bates and Watts (1988) considered the Michaelis–Menten model and the simple exponential rise model based on the behaviour of the data from the experiment. In a report prepared for the researchers, a three-parameter Michaelis–Menten model given by $Y(x)=\theta_0 x/(\theta_1+x+\theta_2 x^2)$ was recommended with a suggestion that "additional experiments be run, especially at higher light intensities". In view of this, Bates and Watts discussed the construction of minimum variance D-optimal designs based on the assumption that the model is correctly specified.

It is well known that slight deviations from the assumed model can completely eliminate advantages that minimum variance designs possess. The approach we adopt in this article, which combines the flexible and adaptive properties of wavelets for estimation and a minimax treatment in constructing designs, would then seem appropriate. The only requirement is for the experimenter to choose a specific wavelet system, the order m of approximation and a specific method of estimation.

We chose $n=16$ levels of light intensities from $N=608$ equally spaced points in [0,170]. Our proposed model uses the Daubechies wavelets $({}_5\phi(x),{}_5\psi(x),{}_5\psi^{-1,0}(x),\ldots,{}_5\psi^{-2,3}(x))$ as regressors.

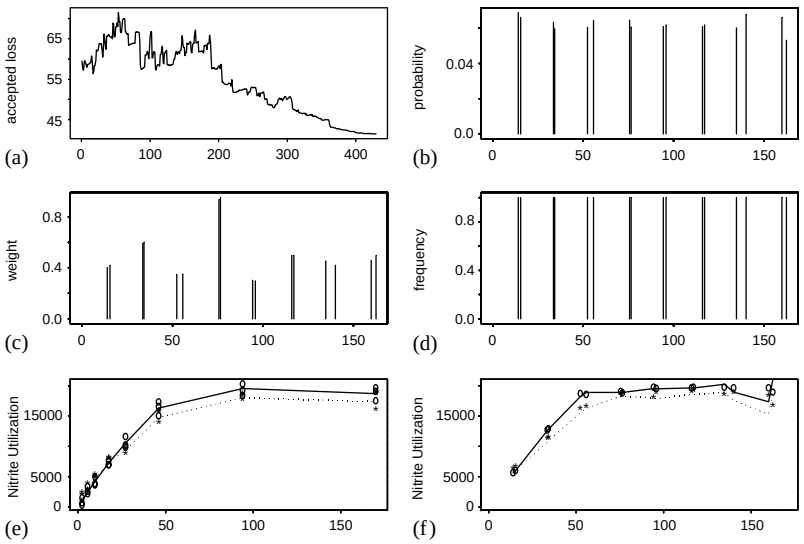

Fig. 4. Design for nitrite utilization experiment. Daubechies wavelets with $(N, n, \alpha, v) = (608, 16, 2, 0.05)$: (a) accepted loss, (b) probabilities p_i, (c) weights w_i, (d) design points and frequencies, (e) *loess* fit of experimental data (day 1—solid line, day 2—dotted line), (f) wavelet fit to data generated from *loess* fit for optimal design points (day 1—solid line, day 2—dotted line).

Designs for this model, meeting the requirements of problem P3, are illustrated in Fig. 4. We applied a nonparametric *loess* fit to the experimental data and predicted observations at each of the optimal design points. We then fitted the Daubechies wavelet model, using weighted least squares with the optimal weights, to the new data. See Fig. 4 for the design, weights and fits.

4. Summary

We have illustrated the use of simulated annealing as a method of constructing minimax robust regression designs, upon approximating the response function by the first few terms of a wavelet expansion. Three formulations of optimality, differing in their methods of handling variance heterogeneity, have been considered. Annealing has proven to be an attractive manner of addressing the analytic difficulties associated with the typically intractable problem of wavelet design. In particular, it has allowed us to construct integer-valued designs in situations in which only continuous designs could previously be considered.

References

Alpert, B.K., 1992. Wavelets and other bases for fast numerical linear algebra. In: Chui, C.K. (Ed.), Wavelets: A Tutorial in Theory and Applications. Academic Press, Boston, pp. 181–216.

Antoniadis, A., Gregoire, G., McKeague, I.W., 1994. Wavelet methods for curve estimation. J. Amer. Statist. Assoc. 89, 1340–1352.

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and its Applications. Wiley, New York.

Bohachevsky, I.O., Johnson, M.E., Stein, M.L., 1986. Generalized simulated annealing for function optimization. Technometrics 28, 209–217.

Box, G.E.P., Draper, N.R., 1959. A basis for the selection of a response surface design. J. Amer. Statist. Assoc. 54, 622–654.

Daubechies, I., 1992. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, Philadelphia.

Fang, Z., Wiens, D.P., 2000. Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. J. Amer. Statist. Assoc. 95, 807–818.

Haines, L.M., 1987. The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models. Technometrics 29, 439–447.

Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A., 1998. Wavelets, Approximation, and Statistical Applications. Springer, New York.

Herzberg, A.M., Traves, W.N., 1994. An optimal experimental design for the haar regression model. Canad. J. Statist. 22, 357–364.

Kiefer, J.C., 1971. The role of symmetry and approximation in exact design optimality. In: Gupta, S.S., Yackel, J. (Eds.), Statistical Decision Theory and Related Topics, Proceedings of a Symposium, Purdue University. Academic Press, New York, pp. 109–118.

Oyet, A.J., 2002. Wavelet designs for estimating nonparametric regression models with heteroscedastic error. J. Statist. Plann. Inference, to appear.

Oyet, A.J., Wiens, D.P., 2000. Robust designs for wavelet approximations of regression models. J. Nonparametric Statist. 12, 837–859.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1989. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, Cambridge.

Pukelsheim, F., 1993. Optimal Design of Experiments. Wiley, Toronto.

Pukelsheim, F., Rieder, S., 1992. Efficient rounding of approximate designs. Biometrika 79, 763–770.

Strang, G., 1989. Wavelets and dilation equations: a brief introduction. SIAM Rev. 31, 614–627.

Vidakovic, B., 1999. Statistical Modelling by Wavelets. Wiley, New York.

Wiens, D.P., 1992. Minimax designs for approximately linear regression. J. Statist. Plann. Inference 31, 353–371.

Xie, M.Y., 1998. Some optimalities of uniform designs and projection uniform designs under multi-factor models, Ph.D. Thesis, Hong Kong Baptist University, Department of Mathematics, unpublished.

Zhou, J., 2001. Integer-valued, minimax robust designs for approximately linear models with correlated errors. Commun. Statist: Theory Methods 30, 21–39.