Minimax A- and D-optimal integer-valued wavelet designs for estimation

Alwell J. OYET

Key words and phrases: A-optimality; D-optimality; Daubechies wavelet; Gasser–Müller estimator; Haar wavelet; multiwavelet; simulated annealing; weighted least squares.

MSC 2000: Primary 62K05, 62G35; secondary 62G08.

Abstract: The author discusses integer-valued designs for wavelet estimation of nonparametric response curves in the possible presence of heteroscedastic noise using a modified wavelet version of the Gasser–Müller kernel estimator or weighted least squares estimation. The designs are constructed using a minimax treatment and the simulated annealing algorithm. The author presents designs for three case studies in experiments for investigating Gompertz's theory on mortality rates, nitrite utilization in bush beans and the impact of crash helmets in motorcycle accidents.

1. INTRODUCTION

In recent years, there has been a growing interest among statisticians in the application of wavelets to statistical problems. Exciting statistical applications have already been made in the analysis of time series, outlier detection, nonparametric curve estimation and in the construction of classical and robust designs. The success of these applications can be attributed to the flexible nature of wavelet systems that allow them to adapt to the structure of the underlying process governing the behaviour of several systems. Several examples can be found in Abramovich & Silverman (1998), Härdle, Kerkyacharian, Picard & Tsybakov (1998), Antoniadis, Gijbels & Grégoire (1997) and Wang (1995). Herzberg & Traves (1994) were probably the first to discuss classical designs for wavelet regression models, in which the Haar wavelets are the regressors.

The design problem for regression models with a prespecified parametric form for the mean response function \( \eta(x) = \mathbb{E}(Y|x) \) has been discussed in great details by several authors. On the other hand, studies of design construction for nonparametric models have been limited by the uncertainty in the “true” structure of the mean response function. Chan (1991), Mitchell, Sacks & Ylvisaker (1994) and Bandemer, Nather & Pilz (1987) adopted a Bayesian approach and made several assumptions in order to impose some structure on \( \eta(x) \). Some authors simply assigned a convenient approximation to \( \eta(x) \) based on experience and prior information. In such cases, it is assumed that the deviation of the approximate model from the true model is negligible.

As an example of such an approximation, consider an experiment on the utilization of nitrite in bush beans described in Bates & Watts (1988). In the experiment, portions of primary leaves from three 16-day-old bean plants were subjected to eight levels of light intensity \( x \in [0, 170] \) \( (\mu E/m^2 s) \), and the nitrite utilization \( Y \) (nmol/g hr) was measured. The experiment was performed on two different days, resulting in two sets of 24 observations at 8 levels of light intensity. Since the experimenters have no knowledge of the true model for representing the response, Bates &
Watts (1988) proposed that the Michaelis–Menten model and the simple exponential rise model should be considered good approximations. In a report for the researchers, the Michaelis–Menten model

$$\eta(x) = E(Y \mid x) = \frac{\theta_0 x}{\theta_1 + x + \theta_2 x^2}$$

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. However, it is well known that any deviation of the assumed Michaelis–Menten model from the true model will completely wipe out any advantages that the minimum variance designs possess. The approach we adopt in this paper, which combines the flexible and adaptive properties of wavelets for estimation and a minimax treatment in constructing designs for nonparametric models, would then seem appropriate.

Our approach is to transform the nonparametric regression model into a wavelet regression model by expressing the mean response function $\eta(x)$ as a linear combination of components of a wavelet system. We note that only a finite number of parameters in the wavelet model can be estimated. The remaining terms will then cause the estimated parameters to be biased. This implies that the classical designs which minimize variance alone will no longer be “optimal” due to the bias. We therefore adopt a minimax approach and optimality criteria, or loss functions, which are convex functions of bias and variance. More specifically, our design criteria are the determinant and the trace of the mean squared error matrix (MSE) of the estimated parameters. We will refer to the trace of the MSE as the A-optimality criterion. The determinant of the MSE will be called the D-optimality criterion. The A- and D-optimality criteria have been used by several authors to construct classical and robust minimax designs. Several examples, including the results in this paper, show that using two or more criteria to construct designs for a particular experiment or model does not necessarily lead to the same design; see, e.g., Pesotchinsky (1982, p. 512), Pukelsheim (1993, pp. 212–237) and Wiens (1992, pp. 365–371).

We consider a design space which is finite but arbitrarily large. This eliminates any practical restrictions due to finiteness and enables us to use the simulated annealing algorithm to search for exact integer-valued design points. The result of this approach is a considerable simplification in the mathematical problem, both analytically and numerically. Fang & Wiens (2000) used this approach to construct exact designs for polynomial regression models based on the average mean squared error loss. The simulated annealing algorithm has also been put to use by other authors including Bohachevsky, Johnson & Stein (1986), Haines (1987) and Zhou (2001). Previous work has generally resulted in robust minimax designs, which are possibly continuous probability functions $\xi(x)$ on a continuous design space (e.g., Box & Draper 1959; Huber 1975; and Wiens 1998). These experimenters then hoped that an $n$-observation design whose associated probability distribution approximates $\xi(x)$ would be close to optimal.

Oyet & Wiens (2000b) have studied the construction of integer-valued wavelet designs for nonparametric estimation based on the average mean squared error (AMSE) loss function in the presence of (a) homoscedastic errors under ordinary least squares estimation (OLS); (b) heteroscedastic errors under OLS; and (c) heteroscedastic errors under weighted least squares estimation (WLS). The results of this approach in (c) are approximate minimax weights and designs for estimation. By contrast, we specify a weight function and construct exact, rather than approximate, designs that can be used along with those weights.

We construct a second set of designs with the assumption that a modified wavelet version of the Gasser–Müller kernel estimator can be used for estimation instead of weighted least squares. More specifically, we construct integer-valued wavelet designs for nonparametric estimation in the presence of (a) homoscedastic errors under the modified Gasser–Müller kernel estimator; (b) homoscedastic errors under weighted least squares estimation (WLS); (c) heteroscedastic errors under the modified Gasser–Müller kernel estimator; and (d) heteroscedastic errors under
weighted least squares estimation (WLS). The Gasser–Müller kernel estimator is a very useful method in the nonparametric literature for response curve estimation.

This paper is organized as follows. In Section 2, we discuss the weighted least squares method and the wavelet version of the modified Gasser–Müller kernel estimator. We also discuss our two loss functions for judging designs. The minimax design problem is developed in Section 3, where Theorems 1 and 2 outline solutions to the maximization component of the problem. Corollary 1 deals with A-optimal designs for heteroscedastic error models. The simulated annealing algorithm is implemented in Section 3.1 and Section 3.2 to construct designs which minimize the maximum loss in Theorems 1 and 2 based on the Daubechies and multiwavelet systems. In Section 4, we apply the methods developed in Sections 2 and 3 to a biological experiment on mortality rate of Mediterranean fruit flies, the nitrite utilization experiment and a motorcycle impact experiment.

2. PRELIMINARIES
We begin this section by considering a situation where the wavelet model

$$Y_{ij} = q_m(x_i)\beta + f(x_i) + \varepsilon_{ij}, \quad i = 1, \ldots, N, j = 1, \ldots, n_i; \sum_{i=1}^{N} n_i = n, x_i \in [0, 1], \quad (1)$$

arises from representing the mean response function $\eta(x)$ in a nonparametric regression model by a finite $m$th order wavelet expansion. In (1), $q_m(x)$ is a $2^{m+1} \times 1$ vector consisting of a system of dilated and translated versions $\psi^{-\ell,k}(x) = 2^{\ell/2} \psi(2^{\ell} x - k), (\ell = 0, \ldots, m; k = 0, \ldots, 2^\ell - 1)$ of a primary wavelet $\psi(x)$ and a scaling function $\phi(x)$. The term $f(x)$, which represents components of the wavelet system not used in the approximation, accounts for the uncertainty in the true structure of the response function. The presence of this term automatically introduces bias in the estimates of the response $\eta(x)$. To control the magnitude of the bias, we impose a bound on $f(x)$,

$$\frac{1}{N} \sum_{i=1}^{N} f^2(x_i) \leq \tau^2, \quad (2)$$

for a known constant $\tau$. We note that in actual computations, only the first term in (1) can be estimated. The error terms $\varepsilon_{ij}$ are assumed to be uncorrelated with zero mean, and possibly nonconstant variance $\sigma_i^2 = r(x_i)\sigma^2$, where $N^{-1}\sum_{i=1}^{N} r^2(x_i) \leq 1$. This idea is similar to the local variance of Kovac & Silverman (2000). Cai, Hurvich & Tsai (1998) have also used the same assumption on the variance to study tests for heteroscedasticity in wavelet regression.

The scaling function $\phi(x)$ and the primary wavelet $\psi(x)$ of a wavelet system are constructed by a procedure commonly referred to as a multiresolution analysis of the space of square integrable functions. The scaling function $\phi(x)$ is constructed to satisfy the dilation equation $\phi(x) = \sum_k h_k \phi(2x - k)$, where $h_k = \int \phi(s) \phi(2s - k) \, ds$ and $\sum_k |h_k|^2 = 1$. Once $\phi(x)$ has been constructed, the primary wavelet $\psi(x)$ can be obtained through the relation $\psi(x) = \sum_k (-1)^k h_{1-k} \phi(2x - k)$, where $k$ is an integer. Readers who are interested in more details can refer to Vidakovic (1999) and Härdele, Kerkyacharian, Picard & Tsybakov (1998). A simple example of a scaling function and a primary wavelet are the Haar scaling function defined by $\phi(x) = h_{[0,1]}(x)$ and the Haar primary wavelet defined by $\psi(x) = \phi(2x) - \phi(2x - 1)$.

Other examples of orthogonal wavelet systems are the Daubechies (1992) wavelets and the multiwavelets. The Daubechies primary wavelet has no closed form but can be constructed numerically. We have used an algorithm from Strang (1989) to construct the primary wavelets in Figures 1(c) and 1(d). These are commonly represented as $\psi_M(x)$ in order to emphasize the number $2M$ of non-zero coefficients $h_k$ used in the numerical construction of the wavelet. The multiwavelet system is generated by more than one scaling function and primary wavelet. Closed
forms of the scaling functions and primary wavelets are in Alpert (1992). We use the notation $\psi_n(x)$ in Figures 1(a) and 1(b) to indicate that it is the first primary wavelet of a multiwavelet system generated by $L$ scaling functions and $L$ primary wavelets. For example, $\psi_1(x)$ is the second primary wavelet of a multiwavelet system generated by two primary wavelets and two scaling functions. When the multiwavelet is used in the representation (1), $q_n(x)$ becomes a $L2^{m+1} \times 1$ vector. For the Daubechies wavelet system $L = 1$.

An important feature of the Daubechies primary wavelets in Figures 1(c) and (d) is their smoothness whereas in Figure 1(a), (b) the primary multiwavelets have cusps and jumps. At the end of this section, we use some examples to illustrate the fact that response curves approximated by these wavelets usually inherit the features of their primary wavelets. The choice of a wavelet system to be used will therefore depend on whether the experimenter expects the response to be a smooth function, contain discontinuities, or be a step function.

![Wavelet plots]

**Figure 1**: Panels a and b show the two primary wavelets of a multiwavelet system with $L = 2$: (a) $\psi_0(x)$ (b) $\psi_1(x)$. Panels c and d show two primary Daubechies wavelets: (c) $\psi(x)$, (d) $\psi(x)$.

The number of coefficients in (1) is determined by the value of $m$. In practical situations, the choice of a value for $m$ will depend on the complexity of the graph or structure of the response function for a particular experiment. This information is usually available to experimenters with previous experience in similar or related experiments. For example, specific parametric nonlinear models are usually proposed for studying an experiment based on such information. In situations where such information is not available, we suggest that the experimenter should conduct a pilot experiment based on arbitrarily chosen design points. A plot of the data from the pilot experiment can then be used to choose a value for $m$. For example, due to its simple structure (see Figure 5(b)) the response function for the nitrite utilization experiment can be adequately represented by a wavelet model with $m = 2$. In this case,

$$q_m(x) = (\phi(x), \psi(x), \psi^{-1,0}(x), \psi^{-1,1}(x), \psi^{-2,0}(x), \psi^{-2,1}(x), \psi^{-2,2}(x), \psi^{-2,3}(x)).$$

More complicated responses (e.g., motorcycle impact response) may require a higher value of $m$, say, $m = 3$ or 4. Note that as $m$ increases, the number of terms in $q_m$ doubles with each unit increase in $m$. Thus the larger the value of $m$ the higher the number of observations required to uniquely estimate the response. The consequence of choosing unreasonably large values for $m$ is that the wavelet estimate of the response will include wiggles and spikes that are not features of the experiment (see Figure 5(c), (d)). In the words of Härdle, Kerkyacharian, Picard & Tsybakov (1998), "with increasing level the linear wavelet estimates become rougher". Antoniadis,
Grégoire & McKeague (1994) state that “for sample sizes between 100 and 200, we have found that it suffices to examine only \( m = 3, 4 \) and 5”. We note that there are several problems associated with this suggestion. First, most experimenters may not be able to make 100 observations due to resource constraints, especially if the experiment involves destroying the experimental unit. If the response curve for the experiment is simple but the experimenter is able to make more than 100 observations, this approach will lead to an estimated curve with spikes and wiggles that are not features of the experiment. The choice of \( m \) should not depend on the number of observations but on the structure of the response curve.

Let \( n_i \) be the number of observations made at the point \( x_i \). Note that the “design space” \( \mathcal{S} \) consists of \( N \) locations \( x_1, \ldots, x_N \), where \( N \) may be greater or less than \( n \). Put \( p_i = n_i / n \). The wavelet version of the Gasser–Müller kernel estimator \( \hat{\eta}(x) \) discussed by Antoniadis, Grégoire & McKeague (1994) requires the partitioning of the \([0, 1]\) interval into \( N \) subintervals \([A_{i-1}, A_i]\) such that \( x_i \in [A_{i-1}, A_i] \) and \( \bigcup_{i=1}^{N} [A_{i-1}, A_i] = [0, 1] \). In that estimator, the wavelet kernel can be represented as \( K_m(x, s) = n q_m^\prime(x) q_m(s) \). We use instead \( K_m(x, s) = q_m^\prime(x) B^{-1} q_m(s) \), where \( B = Z' P Q = \sum_{i=1}^{N} p_i \zeta_i q_m^\prime(x_i) \) and \( \zeta_i = \int_{A_{i-1}}^{A_i} q_m(s) \, ds \). The rows of the \( N \times \mathfrak{r} \) matrix \( Q \) consist of vectors \( q_m^\prime(x_i) \), \( i = 1, \ldots, N \); \( Z = N \times \mathfrak{r} \) is the matrix with rows \( \zeta_i \), \( i = 1, \ldots, N \) and \( P = \text{diag}(p_1, \ldots, p_N) \). We will construct designs for estimating \( \eta(x) \) by the modified kernel estimator

\[
\hat{\eta}_M(x) = \frac{1}{n} \sum_{i=1}^{N} n_i \sum_{j=1}^{n_i} Y_{ij} \int_{A_{i-1}}^{A_i} K_m(x, s) \, ds.
\]

This estimator can also be written in terms of an estimator of the wavelet coefficients as \( \hat{\eta}(x) = q_m^\prime(x) \hat{\beta}_M \), where \( \hat{\beta}_M = n^{-1} \sum_{i=1}^{N} \sum_{j=1}^{n_i} Y_{ij} B^{-1} \zeta_i \).

The modified kernel is motivated by a desire to obtain an exact expression for the bias of \( \hat{\beta}_M(x) \). We use the matrices \( f = (f(x_1), \ldots, f(x_N))^\prime \), \( \mathbf{b} = Z' P f = \sum_{i=1}^{N} p_i \mathbf{z}_i f(x_i) \), \( \mathbf{R} = \text{diag}(r_1, \ldots, r_N) \), \( \mathbf{A} = Z' P \mathbf{R} Z = \sum_{i=1}^{N} p_i \mathbf{z}_i \mathbf{z}_i' \) and \( \mathbf{H} = B' A^{-1} B \) in the expressions for the bias and covariance. It is straightforward to verify that \( \mathbf{d} = \text{bias}(\hat{\beta}_M) = B^{-1} \mathbf{b} \) and \( \mathbf{C} = \sigma^2 \mathbf{H}^{-1} / n \) is the covariance of \( \hat{\beta}_M(x) \). It follows that the mean squared error of \( \hat{\beta}_M \) is

\[
\text{MSE}(\hat{\beta}_M) = B^{-1} \mathbf{b} (B')^{-1} + \sigma^2 \mathbf{H}^{-1} / n.
\]

Oyet & Wiens (2000b) note that regression weights are useful in increasing the accuracy of wavelet estimation. Thus, the experimenter may consider weighted least squares estimation of \( \beta \):

\[
\hat{\beta}_{\text{WLS}} = \left\{ \frac{1}{n} \sum_{i=1}^{N} n_i w_i q_m(x_i) q_m^\prime(x_i) \right\}^{-1} \frac{1}{n} \sum_{i=1}^{N} n_i \sum_{j=1}^{n_i} w_i q_m(x_i) Y_{ij},
\]

where \( w_i = w(x_i) = \int_{A_{i-1}}^{A_i} ||q_m(s)|| \, ds / ||q_m(x_i)|| \). Then \( \hat{\eta}_{\text{WLS}}(x) = q_m^\prime(x) \hat{\beta}_{\text{WLS}} \). Oyet & Wiens (2000a) have shown that \( w_i \) are minimum variance unbiased weights for weighted least squares estimation of the parameters in wavelet regression models.

Define \( m_i = p_i w_i \), \( \mathbf{M} = \text{diag}(m_1, \ldots, m_N) \) and \( \mathbf{W} = \text{diag}(w_1, \ldots, w_N) \). Then, in terms of \( \mathbf{B}_* = Q' \mathbf{M} Q \), \( \mathbf{b}_* = Q' \mathbf{M} \mathbf{f} \), \( \mathbf{D}_* = Q' \mathbf{M} \mathbf{W} \mathbf{Q} \) and \( \mathbf{H}_* = \mathbf{B}_* \mathbf{D}_*^{-1} \mathbf{B}_* \), we find that \( \mathbf{d}_* = \text{bias}(\hat{\beta}_{\text{WLS}}) = B_*^{-1} \mathbf{b}_* \) and \( \mathbf{C}_* = \text{var}(\hat{\beta}_{\text{WLS}}) = \sigma^2 \mathbf{H}_*^{-1} / n \). Thus, \( \text{MSE}(\hat{\beta}_{\text{WLS}}) = B_*^{-1} \mathbf{b}_* (B_*')^{-1} + \sigma^2 \mathbf{H}_*^{-1} / n \).

If the components of the wavelet system used in (1) are orthogonal then \( \int_{0}^{1} q_m(s) f(s) \, ds = 0 \). This in turn implies that \( \beta \) in the model (1) is identifiable, provided the matrix \( \int_{0}^{1} q_m(s) q_m^\prime(s) \, ds \) is invertible. In Section 3, we have used the discrete analogue of the identifiability condition

\[
\frac{1}{N} \sum_{i=1}^{N} q_m(x_i) f(x_i) = 0,
\]

(4)
to obtain an expression for the contamination term $f(x)$. The minimax designs we construct will be robust against deviations in the class

$$\mathcal{F} = \left\{ f : \frac{1}{N} \sum_{i=1}^{N} q_m(x_i) f(x_i) = 0, \frac{1}{N} \sum_{i=1}^{N} f^2(x_i) \leq \tau^2 \right\}.$$  

Pesotchinsky (1982) constructed D-, A- and E-optimal minimax continuous designs for a multiple linear regression model with contamination $\psi(x)$ which are robust against deviations in the class $\left\{ \psi : |\psi(x)| \leq \phi(x) \right\}$ for some convex function $\phi(x)$.

The design criteria (or loss functions we shall use to determine the number of observations $n_i$ that should be made at each point $x_i$ in $S$) are defined as follows:

1. Determinant Loss (D-optimality): 
   $$\text{det} \{ \text{MSE}(\hat{\beta}_M) \} = \left(\frac{\sigma^2}{n}\right) \frac{1}{|H|} \left(1 + n b^t A^{-1} b / \sigma^2 \right).$$
   The determinant of $\text{MSE}(\hat{\beta}_{WLS})$ is obtained by replacing $A$ by $D_*$, $b$ by $b_*$, and $H$ by $H_*$ in the expression for $|\text{MSE}(\hat{\beta}_M)|$.

2. The trace (A-optimality): 
   $$\text{tr} \begin{bmatrix} \text{MSE}(\hat{\beta}_M) \end{bmatrix} = b^t (B')^{-1} b + \frac{\sigma^2}{n} \text{tr} (H^{-1}).$$
   The expression for $\text{tr} \begin{bmatrix} \text{MSE}(\hat{\beta}_{WLS}) \end{bmatrix}$ is obtained by replacing $b$, $B$, $H$ by $b_*$, $B_*$, and $H_*$ respectively. We note that $B_*$ is symmetric.

Figure 2: A scatter plot of data generated from true response (solid line) and (a) the $m = 2$ Haar wavelet fit $\hat{h}_{WLS}(x)$ (broken line) and Daubechies wavelet fit $\hat{h}_M(x)$ (dotted line); (b) the $m = 1$ multiwavelet fit $\hat{h}_{WLS}(x)$ (broken line) and the $m = 2$ Daubechies wavelet fit $\hat{h}_{WLS}(x)$ (dotted line); (c) the $m = 2$ multiwavelet fit $\hat{h}_{WLS}(x)$ (broken line) and the $m = 3$ Daubechies wavelet fit $\hat{h}_{WLS}(x)$ (dotted line).

Although functions of bias and variance are good optimality criteria, they do not tell the whole story in curve estimation or smoothing problems. An estimator could do well by these criteria and yet never look much like the underlying response in qualitative terms. Other important criteria are whether the estimator displays or blurs discontinuity features, whether the locations and magnitude of such features are approximately right, and whether the estimate has roughly the same curvature and preserves jumps. The weighted least squares and the modified Gasser–Müller estimators will only preserve these features of a response if combined with appropriate wavelet systems. This underlines the importance of choosing an appropriate wavelet system for constructing designs for estimating a specific response. For instance, in Figure 2(a) we observe that the Haar wavelet is suitable for estimating responses with a step pattern. Figures 2(b) and 2(c) illustrate that the Daubechies and the multiwavelet systems are more suitable...
for smooth response curves and curves with jumps and cusps respectively. Figure 2 was obtained by applying (3) and the weighted least squares estimator to data generated from the curves
\[
\eta_1(x) = 0.5I_{[0,1]}(x) + 0.3I_{[0.4,0.5]}(x) + 0.2I_{[0.6,0.8]}(x),
\]
\[
\eta_2(x) = 1 - x + e^{-200(x-1/2)^2},
\]
\[
\eta_3(x) = 2 - 2|\varepsilon - 0.26|^{1/5}I_{[0, 0.26]} - 2|x - 0.26|^{1/5}I_{[0.26, 1]} + I_{[0.78, 1]},
\]
where \( I \) is the indicator function. We assumed that \( \sigma = 0.2 \) for \( \eta_1(x) \) and \( \eta_3(x) \), \( \sigma = 1 \) for \( \eta_2(x) \) and that \( \varepsilon_i \sim N(0, \sigma^2) \).

3. MINIMAX INTEGER-VALUED DESIGNS

Assume that the \( N \times r \) matrices \( Q \) and \( Z \) are of full rank, and define the singular value decomposition of \( Q \) by \( Q = U_N \Sigma R_N \), where \( U'U = V'V = I_r \) and \( \Sigma \) is a diagonal matrix of singular values \( \lambda_1(Q), \ldots, \lambda_r(Q) \) of \( Q \). Similarly, define \( Z = U_N \Sigma R_N \), where \( U'U = V'V = I_N \), and any vector \( f \) satisfying (2) and (4) is representable as \( f = \alpha \tilde{U}c \), where \( ||e|| = 1 \) and \( \alpha \) is a normalizing constant. The choice \( \alpha = \tau \sqrt{N} \) ensures equality in (2). That is, \( f = \tau \sqrt{N} \tilde{U}c \). Our problem then is to solve \( \min_{\beta} \max_{f \in F} L\{\text{MSE}(\beta)\} \), where \( L \) is either the trace or determinant loss function. Define

\[
P_1 = U'PU; \quad P = U'PU'P; \quad \tilde{P}_1 = P_1; \quad P_j = U'P_j'U; \quad M_j = U'M_jU; \quad M^* = U'MWRU \quad \text{and} \quad \tilde{M}_1 = M_1M_1^{-1}M_1.
\]

A sketch of the proofs of Theorems 1 and 2 is given in the Appendix.

**Theorem 1.** Let \( \lambda_{\text{max}}(G) \) be the maximum eigenvalue of \( G = P^* - (P^*_1 - P_1) \) and let \( \lambda_1(P_1), \ldots, \lambda_r(P_1) \) be the eigenvalues of \( P_1 \). Then, for fixed \( \nu = \sigma^2/(n\tau^2) \) and \( r(x_i) \),

\[
\max_{f \in F} \text{MSE}(\beta_f) = \left( \frac{\sigma^2}{n} \right)^{r-1} + \nu^{r-1}N\lambda_{\text{max}}(G)\lambda_1^2(Q)\lambda_1(P_1),
\]

\( r(x_i) \) is the indicator function. We assumed that \( \sigma = 0.2 \) for \( \eta_1(x) \) and \( \eta_3(x) \), \( \sigma = 1 \) for \( \eta_2(x) \) and that \( \varepsilon_i \sim N(0, \sigma^2) \).

**Theorem 2.** Let \( \lambda_{\text{max}}(G_1) \) be the maximum eigenvalue of \( G_1 = I_r - P_1^{-1}(P_1^* - P_1) \) and \( \lambda_1(P_1), \ldots, \lambda_r(P_1) \) be the diagonal elements of \( P_1^{-1} \). Then, for fixed \( \nu = \sigma^2/(n\tau^2) \) and \( r(x_i) \),

\[
\max_{f \in F} \text{tr} \{\text{MSE}(\beta_f)\} = \tau^2 \left( N\lambda_{\text{max}}(G_1) + \nu \sum_{i=1}^r \lambda_i^2(Q)\cdot \ell_i \right),
\]

\( r(x_i) \) is the indicator function. We assumed that \( \sigma = 0.2 \) for \( \eta_1(x) \) and \( \eta_3(x) \), \( \sigma = 1 \) for \( \eta_2(x) \) and that \( \varepsilon_i \sim N(0, \sigma^2) \).

Observe that as \( \nu \to \infty \), the variance components of (5) and (6) become dominant. The biases become dominant in the maximum loss functions as \( \nu \to 0 \). The parameter \( \nu \) can be viewed by the experimenter as a measure of the relative importance of bias versus variance. The maximum loss for the homoscedastic error model is obtained by setting \( r(x_i) = 1 \) for all values of \( i \) in (5) and (6). When \( r(x_i) \) is unknown, the experimenter will need to determine the least favourable \( r(x) \) by maximizing (5) and (6) with respect to \( r(x) \). We note that the matrices \( G \) and \( P_1 \) in (5) are functions of \( r(x) \). Thus constructing D-optimal designs for heteroscedastic error
models becomes difficult unless the structure of \( r(x) \) is known. Authors such as Cai, Hurvich & Tsai (1998) constructed score tests for heteroscedasticity in wavelet regression models under the assumption that the form of \( r(x) \) is known. On the other hand, A-optimal designs can be constructed even when the form of \( r(x) \) is unknown. Expression (7) in Corollary 1 is obtained by applying the Cauchy–Schwarz inequality to the second term in (6).

**Corollary 1.** Let \( E = = A^{-1} P_1^{-1} U' \) and let \( h_i \) be the diagonal elements of the matrix \( E'E \). Let \( G_1 \) be as defined in Theorem 2 for both modified Gasser–Müller and weighted least squares estimation. Then, for fixed \( \nu = \sigma^2/(n \tau^2) \),

\[
\max_{r(x)} \max_{f \in \mathcal{F}} \text{tr} \left\{ \text{MSE}(\hat{\beta}_M) \right\} = \tau^2 \left( N \lambda_{\text{max}}(G_1) + \nu \sqrt{N} \left( \sum_{i=1}^{N} h_i^2 \rho_i^2 \right)^{1/2} \right). \tag{7}
\]

Replace \( P_1 \) by \( M_1 \), \( U \) by \( U \) and \( \rho_i \) by \( m_i w_i \) in (7) to get \( \max_{r(x)} \max_{f} \text{tr} \left\{ \text{MSE}(\hat{\beta}_{\text{WLS}}) \right\} \).

We note that since the expressions (5), (6) and (7) depend on the vector \( q_m \) through the bias and variance components, the designs to be constructed will also depend on the choice of \( m \) and the wavelet system used in constructing the designs. In Section 2, we noted the consequence of choosing arbitrarily large values for \( m \) and that the estimated response will inherit the smoothness properties of the wavelet system used. The case study on the motorcycle impact experiment in Section 4.3 is an example of how \( m \) affects the design.

### 3.1. Integer-valued designs.

In this section, we use the simulated annealing algorithm to determine the distribution \( p = (p_1, \ldots, p_N) \) which minimizes the maximum loss in Theorems 1 and 2. The algorithm begins with the experimenter specifying an initial configuration for \( p \). A set of procedures and a criterion which depends on a parameter, say \( T \), for generating subsequent and acceptable new configurations are then implemented iteratively, until the minimum is found. The main idea behind the algorithm is that the path to a minimum may sometimes go uphill as well as downhill; but the lower the value of \( T \), the less likely is any significant uphill excursion. Thus a progressive reduction in \( T \) is necessary in order to strike a minimum. Details of the procedure are in Fang & Wiens (2000).

In our examples, we consider model (1) and a design space \( S \) with equally spaced design points \( x_i = (2i - 1)/(2N), i = 1, \ldots, N \). The iterative procedure was implemented in three ways in order to ensure that an extensive search for the minimum has been conducted: (a) changing the starting value of \( T \); (b) restricting to symmetric designs; and (c) without restriction to symmetry. Designs for \( m = 1 \) and \( m = 2 \) are constructed for the Daubechies and multiwavelet systems. The algorithm can be used for constructing designs for \( m > 2 \) as shown in Section 4.

### 3.2. Designs for multiwavelet models.

Here, we present integer-valued designs which minimize (5) and (6) when the model (1) is based on the multiwavelet system and the observations are to be estimated by the modified Gasser–Müller estimator. For the purpose of illustration, we take \( n = 48, N = 192, \nu = 5 \) and \( m = 1 \). The integer-valued D-optimal design shown in Figure 3(a) assigns \( n = 2 \) observations to the point 0.0026 and \( n = 1 \) observations to each of the points 0.0078, 0.0130, 0.0182 and 0.0234 (rounded to four decimal places). In all, the D-optimal design assigned a total of 12 observations to points in the interval \([0, 0.25]\). The points we have not listed can be generated by using the fact that the D-optimal design in this example is symmetric within \([0, 0.25]\) and periodic with period \( x = 0.25 \).
that we were adding the same random errors to the predicted values. The values of \( n, N, \nu \) and \( m \) defined above were used in constructing the A-optimal design shown in Figure 3(b) for the purpose of comparing the structure of the A-optimal and D-optimal designs. The A- and D-optimal designs in this particular example do not coincide. We observe that the only feature common to both designs is symmetry about \( x = 1/2 \).

3.3. Designs for Daubechies models.

In contrast to the designs based on the multiwavelet system, the optimal designs for the Daubechies model are neither symmetric nor periodic; see Figures 3(c,d). Here, we construct a D-optimal design when \( n = 64, \ N = 32, \nu = 5 \) and \( m = 2 \) under weighted least squares estimation. The D-optimal design assigned three observations to each of the points 0.078, 0.203, 0.453, 0.828 and 0.953; 4 observations to each of the points 0.297, 0.328, 0.547, 0.578, 0.672 and 0.703; and five observations to each of 0.047, 0.172, 0.422, 0.797 and 0.922 (rounded to three decimal places). The A-optimal design in Figure 3(d) was also constructed under weighted least squares estimation with \( n = 24, N = 96 \) and \( \nu = 10 \).

4. CASE STUDIES

In this section, we construct integer-valued minimax designs for estimating the response in three experiments. The first is an experiment for investigating the theory of Gompertz (1825) on mortality rates. The second is the nitrite utilization experiment as described in Section 1 and the third is a simulated motorcycle impact experiment. In Section 2, we noted two important factors which should influence the choice of designs for estimating a response curve. These are: (a) minimum loss and (b) the estimated response curve should preserve the important features of the experiment. The simulated annealing algorithm is used to construct designs with minimum loss. Since observations at the optimal design points may not be available, we generated data at the design points by fitting a LOESS smoother on S-PLUS to original data from the experiments. Predicted values \( \hat{y}_i^t \) of the smoother were then obtained at each of the design points and a random error \( \varepsilon_i \) added to these values. The random errors were independently and normally distributed with mean 0 and a standard deviation equal to that of the estimated standard error of \( \hat{y}_i^t \). We set the seed at 5 for each combination of estimating method and design criterion, in order to ensure that we were adding the same random errors to the predicted values.
In Section 2, we noted that the choice of a wavelet system for constructing the designs and estimation will depend on the features of the response, such as jumps, discontinuities and so on. We have used the Daubechies wavelet generated by $\psi(x)$ in estimation and construction of the designs since the response curves were all smooth curves with no jumps or discontinuities.

4.1. The theory of Gompertz.

Carey, Liedo, Orozco & Vaupel (1992) examined the theory of Gompertz (1825) in an experiment involving 1,203,646 Mediterranean fruit flies. The theory states that mortality rate increases at an exponential rate as age increases. They recorded the number of flies found dead each day until the last fly died. The natural response variable is “mortality rate” calculated by counting the number of dead flies on a given day and dividing that by the number of flies alive at the previous count. Carey, Liedo, Orozco and Vaupel found that examining the number of dead flies on a daily basis was a waste of resources since flies do not die daily leading to zero mortality rate for some days, and in some cases for several days in a row. A new experiment is to be performed and resource constraints would allow the experimenter to observe the mortality rate for only 43 days out of the 172 days during which the experiment is to be performed. On what days should the mortality rate be observed in order to estimate mortality rate with minimum error? If Gompertz’s theory were reasonable, the log of mortality rate should be a linear curve. A plot of log(Mortality Rate) versus day, however, shows that the exponential model is not reasonable. So, we apply the flexible wavelet approach as discussed in this paper to construct the design.

We are led by the structure of the response curve for the natural logarithm of mortality rate to use the 3 Daubechies wavelet model to construct A- and D-optimal designs for estimating the mortality rate under the modified Gasser–Müller (MGM) and weighted least squares (WLS) approach. Since there is no indication that the variance is not constant, we use the simulated annealing algorithm to construct the design that minimizes expressions (5) and (6), with $r(x) = 1$. The initial configuration used in the algorithm for constructing the designs for this experiment is the almost uniform design: day 1, 5, 9, 13, ..., 73, 77, 81, 86, 92, 96, 100, ..., 172. The losses for this design under the A-optimality criterion are 335.246 and 343.2462 for the MGM and WLS estimation, respectively. In order to express the trace and the determinant losses in the same units, we compute the determinant loss as $\left\{ \det (\cdots) \right\}^{1/16}$. For the D-optimality criterion, the losses are 1.603 and 1.402 for the MGM and WLS estimation, respectively. Notice that uniform designs can be constructed from the design space starting at different days (1 or 2 or 3 or 4). The values of the A and D loss functions for these uniform designs are shown in Table 1. The values of the loss functions for the optimal designs shown in Figure 4(a)–(d) are 305.5594, 310.3861, 1.340, and 1.158, respectively. In the last column of Table 1, we present the percentage reduction in loss when we compare the optimal design to the uniform design.

Following the approach described at the beginning of this section, we used the D-optimal design to generate data and obtained the modified Gasser–Müller estimate of the response curve shown in Figure 5(a). Suppose that we denote the A-optimal design combined with the MGM approach by A(MGM); we found that the A(MGM), A(WLS) and D(WLS) fits could not adequately estimate the extremes of the curve. The problem was most severe for the A(MGM) and A(WLS) fits. The D-optimal design combined with the MGM approach therefore seems to be more appropriate for this experiment. The D(MGM) optimal design shown in Figure 4(c) recommends taking the first reading on day 4 and the last reading on day 167. The D(WLS) optimal design also starts on day 4 and ends on day 167. Both designs have 41 points that coincide and 2 points that are different. Thus, the poor performance of the D(WLS) estimated curve at the extremes may be due to the weights associated with these points. The A-optimal designs have 13 points that are different and 30 points that coincide. The designs start on day 5 and end on day 172. There are also points that are common to all four optimal designs. This example illustrates the point that a design may be optimal with respect to a design criterion in the minimax sense, but may fail to capture some important features of the underlying response.
Figure 4: A(MGM), A(WLS), D(MGM) and D(WLS) designs for: (a)–(d) \( m = 3 \) Daubechies model for the medfly experiment \( (\nu = 5, n = 43, N = 172) \); (e)–(h) \( m = 2 \) Daubechies model for the nitrite experiment \( (\nu = 0.05, n = 16, N = 608) \). A(MGM) and A(WLS) designs for: (c) and (j) \( m = 3 \); (k) and (l) \( m = 4 \) Daubechies model, respectively, for the motorcycle impact experiment.

4.2. Nitrite utilization.

In Section 1, we described an experiment on nitrite utilization in bush beans conducted by Elliot and Peirson. The experimenters were seeking a model for relating nitrite utilization \( Y \) in bush beans to light intensity \( x \in [0, 170] \). Preliminary analysis of the data from a previous experiment by Bates & Watts (1988, p. 110) showed that “the variance is effectively constant”. Thus we construct A- and D-optimal Daubechies wavelet designs with \( m = 2 \), and \( r(x) = 1 \) for estimating the response from the experiment. Using the simulated annealing algorithm, we select \( n = 16 \) optimal light intensities taken from \( N = 608 \) points that minimize expressions (5) and (6). The initial configuration for the algorithm was near uniform with losses 1109.09, 1061.09, 1.996 and 4.048 for A(MGM), A(WLS), D(MGM) and D(WLS), respectively. The values of the loss function under the uniform design are 1012.98, 978.62, 1.961 and 3.990 for A(MGM), A(WLS), D(MGM) and D(WLS), respectively. The optimal designs shown in Table 2 have minimum losses 798.7, 742.8, 1.573 and 3.152, respectively. A plot of the optimal designs is shown in Figure 4(e)–(h). All points but one in the D-optimal designs coincide. The A-optimal designs are, however, completely different for both methods of estimation. The recommended light intensities by the A-optimal design shown in Table 2 range from about 18 to 170 units, whereas the D-optimal designs starts with light intensities of about 9.6 and ends at approximately 160 units. Plots of the estimated response at the optimal design points show that the A(WLS) fit has spikes that are not features of the response, whereas the D(WLS) fit overestimates the response when light intensity is about 9.6 to 45. The A(MGM) and D(MGM) fits appear to have done very well in capturing the locations of curvatures and in estimating the extremes of the response. Thus, any one of the A-optimal and D-optimal designs combined with the MGM approach, should be
suitable for the nitrite utilization experiment. A plot of the modified Gasser–Müller estimated response based on the A-optimal design is shown in Figure 5(b).

### Table 1: Values of loss functions (5) and (6) for uniform designs for the medfly experiment.

<table>
<thead>
<tr>
<th>Method of Starting</th>
<th>Criterion</th>
<th>% Loss reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>Loss</td>
<td>Optimal vs Uniform</td>
</tr>
<tr>
<td>MGM 1</td>
<td>344.94</td>
<td>11.4</td>
</tr>
<tr>
<td>2</td>
<td>342.28</td>
<td>10.7</td>
</tr>
<tr>
<td>3</td>
<td>336.64</td>
<td>9.2</td>
</tr>
<tr>
<td>4</td>
<td>328.91</td>
<td>7.1</td>
</tr>
<tr>
<td>WLS 1</td>
<td>348.73</td>
<td>10.99</td>
</tr>
<tr>
<td>2</td>
<td>346.12</td>
<td>10.3</td>
</tr>
<tr>
<td>3</td>
<td>340.99</td>
<td>8.97</td>
</tr>
<tr>
<td>4</td>
<td>336.89</td>
<td>7.9</td>
</tr>
<tr>
<td>MGM 1</td>
<td>1.581</td>
<td>15.24</td>
</tr>
<tr>
<td>2</td>
<td>1.578</td>
<td>15.02</td>
</tr>
<tr>
<td>3</td>
<td>1.578</td>
<td>15.02</td>
</tr>
<tr>
<td>4</td>
<td>1.586</td>
<td>15.46</td>
</tr>
<tr>
<td>WLS 1</td>
<td>1.379</td>
<td>16.04</td>
</tr>
<tr>
<td>2</td>
<td>1.378</td>
<td>15.95</td>
</tr>
<tr>
<td>3</td>
<td>1.440</td>
<td>19.52</td>
</tr>
<tr>
<td>4</td>
<td>1.380</td>
<td>16.04</td>
</tr>
</tbody>
</table>

4.3. Motorcycle impact experiment.

The motorcycle impact experiment, a simulated motorcycle crash, is an experiment designed to measure the efficacy of crash helmets. Observations from a previous experiment consist of 133 accelerometer readings taken in time from 2.4 to 57.6 seconds. For various reasons, multiple observations were taken at some time points and the objective of the experiment is to discern the general shape of the underlying acceleration curve and to draw inferences about its minimum and maximum values. Silverman (1985) used the spline smoothing approach to estimate the underlying curve and estimate the minimum and maximum values. Hárdle (1990) discussed the regressogram smooth, the running median and the \( k \cdot N \) nonparametric estimates of the acceleration curve. The shortcoming of these methods is that none of them resulted in an explicit mathematical expression for relating acceleration to time. Also, one may ask whether the experimenters really need 133 observations in order to achieve their objectives. For future experiments, we exploit the techniques discussed in this paper to select 64 A-optimal time points, at which to measure the accelerations in order to estimate the underlying curve and the minimum and maximum values. We design to minimize (7) under an A-optimal Daubechies wavelet model because the variance of the data from previous experiments is not constant; see Silverman (1985, p. 8).

For \( m = 3 \), the A(MGM) design shown in Figure 4(i) suggests measuring one observation at 2.51 seconds and ending at 57.49 seconds. The A(WLS) design in Figure 4(j) also starts at 2.51 seconds but ends at 57.6 seconds. We find that most of the points in the A(MGM) and the A(WLS) designs are distinct. When \( m = 4 \), the A(MGM) optimal design in Figure 4(k) places six observations at 2.4 seconds whereas the A(WLS) design in Figure 4(l) places three observations at the same point. Two other observations are placed at 3.59 seconds by the A(MGM) design and one observation each at 56 other points. Thus the
A(MGM) design places observations at 58 distinct points. Apart from the three observations at 2.4 seconds, the A(WLS) design also recommends taking three other observations at 2.51 seconds and one at each of 58 other points. So, the A(WLS) design recommends 60 distinct points at which to conduct the experiment.

Following the procedure described at the beginning of this section, we generated the data used in constructing the Daubechies wavelet fits shown in Figure 5(c), (d). In Section 1, we noted that unreasonably high values of \( m \) introduce spikes and wiggles in the estimate that are not features of the response. Figure 5(d) constructed with \( m = 4 \) is a good illustration of this fact. The wiggles seen in the estimated acceleration curve between 35 and 55 seconds are not features of the curve. The \( m = 4 \) model also requires estimating twice as many parameters as the \( m = 3 \) wavelet model. The recommendation to take six observations at a single point under the \( m = 4 \) model may be unacceptable to an experimenter. The estimated wavelet response in Figure 5(c) for \( m = 3 \) does not have the extra wiggles seen in Figure 5(d), and also does well in estimating the extremes of the response. The \( m = 3 \) A(MGM) and A(WLS) optimal designs therefore seem to be more suitable for the purpose of discerning the underlying acceleration curve and to draw inferences about its minimum and maximum values. The values of the loss function for the A(MGM) and A(WLS) optimal designs are 334.22 and 334.54 respectively. Under the uniform design, the losses are 475.22 and 609.82 respectively, corresponding to approximately 30% and 46% increases respectively, over the loss for the optimal designs.

**Table 2: A- and D-optimal design for nitrite utilization experiment.**

<table>
<thead>
<tr>
<th></th>
<th>A-optimal designs</th>
<th>D-optimal designs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( x_i(GM) )</td>
<td>( x_i(WLS) )</td>
</tr>
<tr>
<td>1</td>
<td>17.99</td>
<td>18.28</td>
</tr>
<tr>
<td>2</td>
<td>20.79</td>
<td>22.19</td>
</tr>
<tr>
<td>3</td>
<td>24.15</td>
<td>44.31</td>
</tr>
<tr>
<td>4</td>
<td>43.47</td>
<td>44.59</td>
</tr>
<tr>
<td>5</td>
<td>47.95</td>
<td>48.51</td>
</tr>
<tr>
<td>6</td>
<td>49.91</td>
<td>79.02</td>
</tr>
<tr>
<td>7</td>
<td>68.94</td>
<td>82.94</td>
</tr>
<tr>
<td>8</td>
<td>71.74</td>
<td>87.69</td>
</tr>
<tr>
<td>9</td>
<td>73.70</td>
<td>88.26</td>
</tr>
<tr>
<td>10</td>
<td>91.06</td>
<td>122.13</td>
</tr>
<tr>
<td>11</td>
<td>96.38</td>
<td>125.49</td>
</tr>
<tr>
<td>12</td>
<td>103.09</td>
<td>130.81</td>
</tr>
<tr>
<td>13</td>
<td>123.53</td>
<td>157.12</td>
</tr>
<tr>
<td>14</td>
<td>128.85</td>
<td>169.44</td>
</tr>
<tr>
<td>15</td>
<td>130.81</td>
<td>169.72</td>
</tr>
<tr>
<td>16</td>
<td>170.00</td>
<td>170.00</td>
</tr>
</tbody>
</table>

\( x_i(GM) \) — Design based on modified Gasser–Müller estimator

\( x_i(WLS) \) — Design based on Weighted Least Squares estimator
Note that we can write 
\[ /BP /CU /CQ \]
Using the results in Section 2, we have

**Proof of Theorem 1**

**APPENDIX: DERIVATIONS**

Figure 4: Estimated response curves for (a) medfly experiment based on the design in Figure 4(c); (b) nitrite experiment based on the design in Figure 4(e); (c), (d) motorcycle experiment based on the designs in Figures 4(i) and 4(k), respectively.

5. CONCLUDING REMARKS

The results in this study show that by combining bias- and variance-based optimality criteria and an appropriate method of estimation, designs for wavelet estimation of a wide variety of response curves can be constructed. Since no restriction was imposed on the structure of the response, the approach we have discussed can be applied to a wide variety of situations including parametric nonlinear regression experiments. When applied to such experiments, however, the parameters in the wavelet model will have a different interpretation from the parameters in the parametric nonlinear model. We find that the curves estimated with designs constructed under the modified Gasser–Müller procedure are better at estimating the extremes of the curve than those under the weighted least squares procedure. These designs will not only help in locating the important features of the response but are also useful when resources are limited.

ACKNOWLEDGEMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada and by the Memorial University of Newfoundland. The author is grateful for comments and suggestions for improvements from an Associate Editor and two referees.

APPENDIX: DERIVATIONS

**Proof of Theorem 1.** Using the results in Section 2, we have

\[ |H| = |VA\hat{P}_1AV'| = |A|^2 \cdot |\hat{P}_1| = \Pi_{i=1}^r \lambda_i^2(Q) \cdot \lambda_i(\hat{P}_1) \]

and \[ b' A^{-1} b = f' P U^* P_{-1}^* U^* f' \]. Define \( G = U^* P U^* P_{-1}^* U^* P \) and use the fact that \( f = \tau \sqrt{N} U \)c, where \( c \) is the eigenvector corresponding to \( \lambda_{\text{max}}(G) \) to obtain

\[ \max_f b' A^{-1} b = \max_f \tau^2 N c' G c = \tau^2 N \lambda_{\text{max}}(G). \]

Note that we can write

\[ \lambda_{\text{max}}(G) = \lambda_{\text{max}}(P_{-1/2}^* U^* P U^* P_{-1/2}^*). \]
Expression (5) is then obtained by using the definition of D-optimality in Section 2 and the result $\mathbf{U}^T \mathbf{U}' = \mathbf{I} - \mathbf{U}^T \mathbf{U}'$ to show that $d_{\text{max}}(\mathbf{G}) = \lambda_{\text{max}} \left\{ \mathbf{P}^{-1}_i (\mathbf{P}^*_i - \mathbf{P}^i) \right\}$. Similar steps lead to the expression for $\max_f [\text{MSE}(\beta_{\text{WLS}})]$.

**Proof of Theorem 2.** Again, we use the results in Section 2 to obtain

$$\text{tr} \mathbf{H}^{-1} = \text{tr} \left( \mathbf{V} \mathbf{A} \mathbf{P} \mathbf{A}^T \mathbf{V}' \right)^{-1} = \text{tr} \left( \mathbf{A}^{-2} \cdot \mathbf{P}^{-1} \right) = \sum_{i=1}^r \lambda_i^{-2} (\mathbf{Q}) \cdot \ell_i,$$

where $\ell_i$ are the diagonal elements of $\mathbf{P}_i^{-1}$. Now, we write

$$b(\mathbf{B})^{-1} \mathbf{b} = f^T \mathbf{P} \mathbf{U}^T (\mathbf{P}^*_i)^{-1} \mathbf{A}^{-2} \mathbf{P}^{-1}_i \mathbf{U}' \mathbf{P}' \mathbf{f}$$

and define $\mathbf{G}_1 = \mathbf{U}^T \mathbf{P} \mathbf{U}^T (\mathbf{P}^*_i)^{-1} \mathbf{A}^{-2} \mathbf{P}^{-1}_i \mathbf{U}' \mathbf{P}'$. Then,

$$\max_f b(\mathbf{B})^{-1} \mathbf{b} = \max_f \tau^2 \mathbf{N} \mathbf{c}^T \mathbf{G}_1 \mathbf{c} = \tau^2 \mathbf{N} \lambda_{\text{max}} (\mathbf{G}_1).$$

The result is obtained by following the steps in Theorem 1.

**REFERENCES**


Received 5 July 2000

Alwell J. OYET: aoyet@math.mun.ca

Accepted 4 April 2002

Department of Mathematics and Statistics

Memorial University of Newfoundland

Saint John's, Newfoundland, Canada A1C 5S7