The Annals of Statistics
1996, Vol. 24, No. 5, 2108–2127

OPTIMAL BAYESIAN DESIGNS FOR MODELS WITH PARTIALLY SPECIFIED HETEROSCEDASTIC STRUCTURE

BY HOLGER DETTE[1] AND WENG KEE WONG[2]

Ruhr-Universität Bochum and University of California

We consider the problem of finding a nonsequential optimal design for estimating parameters in a generalized exponential growth model. This problem is solved by first considering polynomial regression models with error variances that depend on the covariate value and unknown parameters. A Bayesian approach is adopted, and optimal Bayesian designs supported on a minimal number of support points for estimating the coefficients in the polynomial model are found analytically. For some criteria, the optimal Bayesian designs depend only on the expectation of the prior, but generally their dependence includes the derivative of the logarithm of the Laplace transform of a measure induced by the prior. The optimal design for the generalized exponential growth model is then determined from these optimal Bayesian designs.

1. Introduction. We begin by considering two design problems.

1.1. *A design problem for a generalized exponential model.* Suppose the statistical model is given by

$$y = x^{v}\exp(-\theta x)\sum_{j=0}^{n-1} a_j x^j + \varepsilon, \qquad x \ge 0,$$

where y is the response and ε is an unobservable error normally distributed with mean 0 and variance σ^2. The model parameters are $a^T = (a_0, a_1, \dots, a_{n-1}) \in R^n$ and $\theta > 0$. If n and v are known, the above model is a more general exponential growth model than those studied in Chaloner (1993) ($n = 1$, $v = 0$ and a_0 known), Mukhopadhyay and Haines (1995) and Dette and Neugebauer (1996a, b) ($n = 1$, $v = 0$) and Dette and Sperlich (1994) ($n = 1$, $v \ge 0$). Since the above model has multiple extrema which decrease exponentially as $x \to +\infty$ and includes the commonly used exponential models, we shall call the above model a generalized exponential growth model.

The main interest here is to find an efficient design for estimating the parameters (a^T, θ) in the generalized exponential growth model. Special

Received August 1994; revised January 1996.
[1] Research partially supported by the Deutsche Forschungsgemeinschaft.
[2] Research partially supported by a UCLA Faculty Career Development Award.
AMS 1991 *subject classifications*. Primary 62K05; secondary 65D30.
Key words and phrases. Approximate designs, design efficiency, efficiency functions, Laplace transform, Bayesian design.

cases of this model like those mentioned above already have numerous applications in the biological and agricultural sciences. For example, when $v = 0$ and $n = 1$, we have the asymptotic regression model which is one of the most widely used curves in applied scientific work. In agricultural research, this is the Mitscherlich growth model frequently used to study the relationship between crop yield and the amount of fertilizer; in fisheries research, this is the Bertalanffy growth curve used for modeling the age and length of fish [Ratkowsky (1983)]. Further applications of related exponential growth models to poultry science and decay growth laws are amply supplied in Box and Lucas (1959). In light of this, studying the design problem for a generalized exponential growth model is useful since this model provides researchers further flexibility in modeling problems.

1.2. *A design problem for a polynomial model with partially known heteroscedastic structure.* Consider the general problem of designing a regression experiment when it is known that the assumption of homoscedasticity is not tenable and a precise description of the heteroscedasticity is problematic. The model of interest here is the general linear model given by

$$y(x) = f^T(x)\kappa + e(x, \theta),$$

where $f(x)$ is a given $k \times 1$ vector of regression functions and $y(x)$ is the response at the x-level of the independent variable x, assumed to lie in a given design space χ. The random error incurred in observing $y(x)$ is $e(x, \theta)$, where θ is an unobservable nuisance parameter. The distribution of $e(x, \theta)$ is assumed to be normal with mean 0 and variance proportional to $1/\lambda(x, \theta)$. The function $\lambda(x, \theta)$ is commonly called the efficiency function in the design literature [Fedorov (1972), page 64]. The parameters in the model are the $(k \times 1)$ vector κ and the vector θ.

The research question here is how to design an efficient experiment to estimate the parameter κ when the functional form of $\lambda(x, \theta)$ is known apart from the values of θ. This is an important design problem to address because (i) almost all previous work in this area assumes θ is known so that the efficiency function is completely determined [see the monographs of Fedorov (1972), Pazman (1986) and Pukelsheim (1993) and the references therein]; (ii) the assumption of a known efficiency function can be unrealistic in practice, particularly if the postulated model has little or no theoretical underpinnings [see Walter and Pronzato (1990) for a further discussion on this issue]; and (iii) the assumption of a known efficiency function is a risky one since a slight misspecification of the function can cause serious problems in estimating parameters that are of interest. To illustrate the third point, consider designing an experiment to gain information about κ using the model $f^T(x) = (1, x, \dots, x^n)$ with $\lambda(x, \theta) = \exp(-\theta x)$, $\theta > 0$, and the design space is $[0, \infty)$. Suppose the true value of θ is θ_2 and the experimenter misspecifies the efficiency function and uses $\exp(-\theta_1 x)$ as the efficiency function instead. If ξ_{θ_1} is the locally D-optimal design for θ_1 [Chernoff (1953)], it can be shown that the D-efficiency of the design ξ_{θ_1} is $\{(\theta_2/\theta_1)\exp(1 - \theta_2/\theta_1)\}^n$ [see Karlin

TABLE 1
D-efficiencies of locally D-optimal designs for polynomial regression of degree n when the value of θ in the efficiency function $\lambda(x, \theta) = \exp(-\theta x)$ is misspecified

θ_2 / θ_1	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**	**2.0**
$n = 1$	0.445	0.729	0.895	0.977	1.000	0.982	0.938	0.878	0.808	0.736
$n = 2$	0.198	0.531	0.801	0.955	1.000	0.965	0.881	0.771	0.654	0.541
$n = 3$	0.088	0.387	0.717	0.933	1.000	0.948	0.826	0.677	0.529	0.398

and Studden (1966), Theorem 3.3, page 330]. This D-efficiency measures the efficiency of a design used for estimating θ and will be more formally defined in the next section.

Table 1 displays the D-efficiencies of ξ_{θ_1} for selected ratios of θ_2 and θ_1. It is evident from the table that misspecification of the efficiency function can result in serious loss of D efficiencies, and becomes increasingly severe as the degree of the polynomial regression function increases. It is therefore crucial to have the efficiency function specified as accurately as possible.

It transpires that the optimal design for the generalized exponential model can be obtained from a general optimal strategy for design problems in models with heteroscedastic errors. We will therefore consider the latter design problem first. Only nonsequential designs are considered here, and so the sequential approach described in Fedorov (1972), Section 4.5, is not applicable. Throughout, we will use a Bayesian approach by incorporating a prior distribution on the parameter θ. This is different from most work in optimal Bayesian designs where a prior distribution is placed on the parameter vector κ and errors are assumed to be homoscedastic; see Chaloner (1984) and Pilz (1991) and the references therein.

Our methodology enables us to find many analytical optimal Bayesian designs supported on a minimal set of support points for a large class of priors in a fairly general setting. This is an interesting situation, since it is generally known that analytical results hardly exist for nonlinear design problems with a nondegenerate prior [Chaloner (1993)]. Our basic technique relies on expansion of the Stieltjes transform of a probability measure as a continued fraction [Studden (1980, 1982) and Lau and Studden (1988)], and some key properties of the transform are briefly reviewed in the Appendix. Of course, an advantage of an analytical description of the optimal design is that properties of the design can be examined more fully.

In the next section we provide the setup for our design problem for estimating parameters in a heteroscedastic polynomial model and discuss the optimality criterion and the scope of the problem we seek to solve. Section 3 contains our main results and gives the optimal Bayesian designs for the polynomial regression model under various assumptions on the heteroscedastic structure. The methodology is illustrated in Section 4 with two examples, followed by, in Section 5, a revisit to the design problem for the generalized exponential growth model.

2. Optimality criterion. Let χ be a given design space, $f(\cdot)$ be a known $k \times 1$ vector of regression functions defined on χ and $\lambda(x, \theta)$ be known except for the values of θ. Let $\tilde{\pi}(\theta)$ be a prior distribution on θ with support on a given set Θ. Here and throughout, only approximate designs in Kiefer's sense are considered [Kiefer (1960)]. This means an arbitrary design ξ on χ is treated as a probability measure on χ. If we assume the number of observations is fixed at the onset of the experiment, then these observations are taken at the support points of the probability measure and the number of observations at each of these points is proportional to the mass of the probability measure at each of its support points.

Given a design ξ and a known value of the parameter θ, the information matrix for the model in Section 1.2 is given by

$$M(\xi, \theta) = \int_\chi \lambda(x, \theta) f(x) f^T(x)\, d\xi(x).$$

This measures the information contained in ξ and is frequently used in the literature; see Fedorov (1972) and Pazman (1986). When θ is unknown, which is the case here, this information matrix is still meaningful. The integrand of this matrix corresponds to the upper left $k \times k$ submatrix in the block diagonal Fisher information matrix for (κ, θ) at the point x derived in Atkinson and Cook (1995), equation 6:

$$\begin{pmatrix} \lambda(x, \theta) f(x) f^T(x) & 0 \\ 0 & \dfrac{\lambda(x, \theta)^2}{2} \dfrac{\partial \lambda^{-1}(x, \theta)}{\partial \theta} \left(\dfrac{\partial \lambda^{-1}(x, \theta)}{\partial \theta} \right)^T \end{pmatrix},$$

where they considered a more general heteroscedastic model. The lower right block submatrix in their information matrix is irrelevant here because we are only interested in estimating κ and θ is treated as a nuisance parameter. Hence, in designing an experiment to gain maximal information on κ, it is reasonable to choose a design which maximizes the determinant of $M(\xi, \theta)$ or some function thereof, after averaging out the plausible values of θ by a prior.

The optimality criterion of interest here is, for a given prior $\tilde{\pi}$,

$$\text{(2.0)} \qquad \Phi_p(\xi) = \left\{ \int_\Theta \left\{ \frac{|M(\xi, \theta)|}{|M(\xi_\theta, \theta)|} \right\}^{p/k} d\tilde{\pi}(\theta) \right\}^{1/p}, \qquad -\infty < p \leq 1,$$

with the interpretation that the case $p = 0$ corresponds to

$$\Phi_0(\xi) = \exp\left\{ \frac{1}{k} \int_\Theta \log\left\{ \frac{|M(\xi, \theta)|}{|M(\xi_\theta, \theta)|} \right\} d\tilde{\pi}(\theta) \right\}.$$

Here ξ_θ is the locally D-optimal design [Chernoff (1953)] for the problem when θ is the true parameter. The optimal design $\xi_{p, \tilde{\pi}}$ is the one which maximizes $\Phi_p(\xi)$ over the set of all designs on χ and is called Bayesian

Φ_p-optimal with respect to the prior $\tilde{\pi}$. The optimal design within the class of all k-point designs is called a Bayesian Φ_p-optimal k-point design with respect to the prior $\tilde{\pi}$. When the prior $\tilde{\pi}(\theta)$ is a degenerate distribution and $p = 1$, $\Phi_1(\xi)$ gives the D-efficiency of ξ.

The criterion (2.0) is motivated largely as a robust criterion in the sense that we are averaging the usual D-optimality criterion with respect to some prior distribution. This seems to be a common practice for studying nonlinear models; see Ford, Titterington and Kitsos (1989) and Pronzato and Walter (1985, 1988a). The case $p = 0$ gives the criterion considered in Läuter (1974) and may be the most interesting case from the Bayesian point of view, in the sense that the optimal design maximizes the expected increase in Shannon information provided by the experiment; see Chaloner and Larntz (1989) and Chaloner (1993). The limiting case $p = -\infty$ gives the maximum criterion studied in Pronzato and Walter (1988b). See also Walter and Pronzato (1990) for a more detailed discussion of this type of criteria, including the cases $p = 1$ and $p = -1$. The other values of p are less interpretable and they are included to resemble Kiefer's L_p-class of criteria [Pazman (1986), page 94].

In this work we focus on the polynomial regression function $f^T(x) = (1, x, x^2, \dots, x^n)$ and choose $\lambda(x, \theta)$ to be one of several classes of efficiency functions commonly used for modeling heteroscedasticity; see, for example, Fedorov (1972), page 88, and Pazman (1986), page 179. They are:

$$
\begin{aligned}
&\text{(i)} \quad \lambda(x, \theta) = \exp(-\theta x), && x \ge 0,\ \theta > 0;\\
&\text{(ii)} \quad \lambda(x, \theta) = x^{\alpha}\exp(-\beta x), && x > 0,\ \theta^T = (\alpha, \beta),\ \alpha > 0,\ \beta > 0;\\
&\text{(iii)} \quad \lambda(x, \theta) = (1 - x)^{\alpha}(1 + x)^{\beta}, && -1 < x < 1,\ \theta^T = (\alpha, \beta),\\
& && \alpha > 0,\ \beta > 0;\\
&\text{(iv)} \quad \lambda(x, \theta) = \exp(-\theta x^2), && -\infty < x < \infty,\ \theta > 0.
\end{aligned}
$$

These efficiency functions have different shapes that are sufficiently versatile to accommodate many efficiency functions that may be of practical interest. For example, monotone decreasing efficiency functions are represented in (i), while unimodal efficiency functions are richly embedded in (ii) and (iii). The ubiquitous bell-shaped curves are represented by (iv). For these efficiency functions, we determine the Bayesian Φ_p-optimal $(n + 1)$-point designs for all prior distributions. The question whether these $(n + 1)$-point designs are also optimal within the class of all designs depends on the specific prior. We will demonstrate this in Section 4. As will be seen, the Bayesian Φ_p-optimal $(n + 1)$-point designs depend on the prior $\tilde{\pi}$ with varying degree; the case with $p = 0$ depends only on the expectation of $\tilde{\pi}$, whereas other cases depend on the derivative of the logarithm of the Laplace transform of a measure induced by the prior. In all cases analytic solutions are found although they rarely exist in a compact form. The support points of these optimal designs are roots of certain orthogonal polynomials whose coefficients depend on the prior distribution.

3. An equivalence theorem and main results. In order to simplify our statements of the main results, it is helpful to identify each prior distribution $\tilde{\pi}$ with an associated prior π defined by

$$d\pi(\theta) = |M(\xi_\theta, \theta)|^{-q}\, d\tilde{\pi}(\theta), \tag{3.0}$$

where ξ_θ is the locally D-optimal design for θ in a polynomial model of degree n and $q = p/k$. Note that if $p = 0$, $\pi = \tilde{\pi}$ and π is still a proper prior. The effect of (3.0) is to transform (2.0) into an equivalent formulation, where the task now is to find a design which maximizes

$$\tilde{\Phi}_q(\xi) = \left\{\int_\Theta \{|M(\xi, \theta)|\}^q\, d\pi(\theta)\right\}^{1/q}, \qquad -\infty < q \le 1/k,$$

over the set of all designs on χ. To compute $d\pi(\theta)$ for a given prior $d\tilde{\pi}(\theta)$, it is necessary to calculate $|M(\xi_\theta, \theta)|$. For the efficiency functions in Section 2, closed-form formulas exist for the locally D-optimal designs and their determinants. They are found using arguments similar to Karlin and Studden (1966), page 330.

LEMMA 3.1. *Suppose $\lambda(x, \theta)$ is one of the efficiency functions in Section 2, and the regression function is $f^T(x) = (1, x, \dots, x^n)$ defined on χ. If ξ_θ denotes the locally D-optimal design for the problem, the determinant of the information matrix $M(\xi_\theta, \theta)$ for fixed θ is*:

(i) $\theta^{-n(n+1)} \prod_{j=1}^{n} j^{2j} \exp[-n(n+1)]$ *if* $\lambda(x, \theta) = \exp(-\theta x)$, $\theta > 0$;

(ii) $\beta^{-(n+1)(n+\alpha)}(n+\alpha)^{n+\alpha} \prod_{j=1}^{n} j^j (j+\alpha-1)^{j+\alpha-1} \exp[-(n+1)(n+\alpha)]$
if $\lambda(x, \theta) = x^\alpha \exp(-\beta x)$ *and* $\theta^T = (\alpha, \beta)$, $\alpha > 0$, $\beta > 0$;

(iii) $$2^{(n+1)(n+\alpha+\beta)} \prod_{j=1}^{n} j^j \prod_{j=1}^{n+1} (j+\alpha-1)^{j+\alpha-1} (j+\beta-1)^{j+\beta-1} \times \left\{\prod_{j=1}^{n+1} (j+n-1+\alpha+\beta)^{j+n-1+\alpha+\beta}\right\}^{-1}$$
if $\lambda(x, \theta) = (1-x)^\alpha (1+x)^\beta$, $\theta^T = (\alpha, \beta)$, $\alpha > 0$, $\beta > 0$;

(iv) $(2\theta \exp(1))^{-(n+1)n/2} \prod_{j=1}^{n} j^j$ *if* $\lambda(x, \theta) = \exp(-\theta x^2)$.

We are now ready to state our main results. Let $\tilde{\pi}$ be a given prior with finite first moment, π be its associated prior defined in (3.0) and $q = p/k$.

The following result gives us a practical way of checking if a design is Φ_p-optimal with respect to the prior $\tilde{\pi}$. The proof relies on standard arguments used in optimal design theory [Pazman (1986) and Pukelsheim (1993)] and is thus omitted.

THEOREM 3.1 (Equivalence theorem). *Let* $1 \geq p > -\infty$, $f(x)$ *be a* $k \times 1$ *vector of known regression functions defined on a given design space* χ, $\lambda(x, \theta)$ *be a known efficiency function apart from the values of* θ *and* $q = p/k$. *If* $\tilde{\pi}$ *is the prior for* θ *on a given set* Θ *and* π *is its associated prior defined in* (3.0), *then the design* $\xi_{p,\tilde{\pi}}$ *is* Φ_p-*optimal with respect to the prior* $\tilde{\pi}$ *if and only if*

$$g(x, \xi_{p,\tilde{\pi}}) = \int_{\Theta} \{|M(\xi_{p,\tilde{\pi}}, \theta)|^q \lambda(x, \theta) f^T(x) M^{-1}(\xi_{p,\tilde{\pi}}, \theta) f(x) \\ - k|M(\xi_{p,\tilde{\pi}}, \theta)|^q\} \, d\pi(\theta) \leq 0 \quad \text{for } x \in \chi. \tag{3.1}$$

Note that the special case with $p = 0$ yields $q = 0$, $\pi = \tilde{\pi}$ and $\xi_{0,\tilde{\pi}}$ is Φ_0-optimal with respect to the prior $\tilde{\pi}$ if and only if, for all $x \in \chi$,

$$\int_{\Theta} \lambda(x, \theta) f^T(x) M^{-1}(\xi_{0,\tilde{\pi}}, \theta) f(x) \, d\tilde{\pi}(\theta) - k \leq 0,$$

a result due to Läuter (1974).

In the rest of the paper, our interest will be in polynomial models of degree n, and so we have $k = n + 1$. Theorems 3.2–3.5 give the Φ_p-optimal $(n + 1)$-point designs for each of the efficiency functions described earlier. The proofs are deferred to the Appendix.

THEOREM 3.2. *Let* $\chi = [0, \infty)$, $f_n^T(x) = (1, x, x^2, \dots, x^n)$, $\lambda(x, \theta) = \exp(-\theta x)$ *and* $\Theta = (0, \infty)$. *Assume that* $\tilde{\pi}$ *is a prior on* Θ *and* π *is its associated prior defined in* (3.0) *such that*, *for* $q = p/(n + 1)$,

$$\text{(i)} \qquad \int_0^{\infty} \exp(-\theta q x) \, d\pi(\theta) < \infty \quad \text{for all } x > 0,$$

and, *if* $q > 0$,

$$\text{(ii)} \qquad \lim_{x \to \infty} x^{n(n+1)} \left\{ \int_0^{\infty} \exp(-\theta q x) \, d\pi(\theta) \right\}^{1/q} = 0.$$

Then the Φ_p-*optimal* $(n + 1)$-*point design with respect to the prior* $\tilde{\pi}$ *has equal mass at the* 0's *of the polynomial*

$$x L_n^{(1)}\left(\frac{n(n + 1)x}{z} \right).$$

Here $L_n^{(1)}(x)$ *is the nth generalized Laguerre polynomial of degree n orthogonal with respect to the measure* $x \exp(-x)\, dx$, *and* z *is a positive root of the equation*

$$zF(qz) = -n(n + 1), \tag{3.2}$$

where

$$F(x) = -\int_0^\infty \theta \exp(-\theta x)\, d\pi(\theta) \bigg/ \int_0^\infty \exp(-\theta x)\, d\pi(\theta).$$

If $q \le 0$, *the solution of* (3.2) *is unique.*

REMARK 3.0. The function $F(x)$ is the derivative of the logarithm of the Laplace transform of the measure $d\pi(\theta)$, and plays a crucial role here and throughout.

REMARK 3.1. It is straightforward to show that assumption (ii) holds whenever the infimum of the set of support points of the associated prior π is positive independently of $q \le 1/(n+1)$.

THEOREM 3.3. *Let* $\chi = (0, \infty)$, $f_n^T(x) = (1, x, x^2, \dots, x^n)$, $\lambda(x, \theta) = x^\alpha \exp(-\beta x)$, $\theta^T = (\alpha, \beta)$ *and* $\Theta = (0, \infty) \times (0, \infty)$. *Assume that* $\tilde{\pi}$ *is a prior on* Θ *and* π *is its associated prior defined in* (3.0) *such that, for* $q = p/(n+1)$,

$$\text{(i)} \qquad \int_\Theta x^{\alpha q} \exp(-\beta q x)\, d\pi(\alpha, \beta) < \infty \quad \textit{for all } x > 0,$$

and, if $q > 0$,

$$\text{(ii)} \qquad \lim_{x \to \infty} x^{n(n+1)} \left\{ \int_\Theta x^{\alpha q(n+1)} \exp(-q\beta x)\, d\pi(\alpha, \beta)) \right\}^{1/q} = 0.$$

Define, for $i = 1, 2$,

$$\begin{aligned} F_i(x_1, x_2) &= -\int_\Theta \theta_i \exp(-\theta_1 x_1 - \theta_2 x_2)\, d\pi(\theta_1, \theta_2) \\ &\qquad \times \left\{ \int_\Theta \exp(-\theta_1 x_1 - \theta_2 x_2)\, d\pi(\theta_1, \theta_2) \right\}^{-1} \end{aligned} \tag{3.3}$$

and let $L_n^{(\alpha)}(x)$ *be the nth generalized Laguerre polynomial orthogonal with respect to the measure* $x^\alpha \exp(-x)\, dx$. *Then the* Φ_p-*optimal* $(n+1)$-*point design with respect to the prior* $\tilde{\pi}$ *has equal mass at the* 0's *of the polynomial*

$$L_{n+1}^{(-F_1(-qy, qz) - 1)}(-F_2(-qy, qz)x),$$

where (y, z) *is a solution of the simultaneous equations*

$$\begin{aligned} zF_2(-qy, qz) &= (n+1)[F_1(-qy, qz) - n], \\ y &= \sum_{i=0}^{n} \log\left\{ \frac{z}{n+1} + \frac{i}{F_2(-qy, qz)} \right\}. \end{aligned} \tag{3.4}$$

If $q \le 0$, *the solution in* (3.4) *is unique.*

Unlike previous results, the next theorem does not require side conditions.

THEOREM 3.4. *Let* $\chi = (-1,1)$, $f_n^T(x) = (1, x, x^2, \dots, x^n)$, $\lambda(x,\theta) = (1-x)^{\alpha}(1+x)^{\beta}$, $\theta^T = (\alpha, \beta)$ *and* $(\alpha, \beta) \in \Theta = (0,\infty)\times(0,\infty)$. *If* $\tilde{\pi}$ *is a prior for* $\theta^T = (\alpha,\beta)$ *on* Θ *and* π *is its associated prior defined in* (3.0), *then the* Φ_p*-optimal* $(n+1)$*-point design with respect to* $\tilde{\pi}$ *has equal mass at the* 0*'s of the Jacobi polynomial* $P_{n+1}^{(\mu,\nu)}(x)$ *orthogonal with respect to the measure* $(1-x)^{\mu}(1+x)^{\nu}\,dx$, *where* $\mu = -F_1(-qz, -qy) - 1$, $\nu = -F_2(-qz, -qy) - 1$, $F_i(x_1, x_2)$ *is as defined in* (3.3), $i = 1,2$, *and* (z, y) *solves the simultaneous equations*

$$z = \sum_{j=1}^{n} \log\left\{\frac{n-j+\mu+\nu+2}{2n-2j+3+\mu+\nu}\right\} + \sum_{j=1}^{n+1} \log\left\{\frac{n-j+2+\mu}{2n-2j+4+\mu+\nu}\right\}, \tag{3.5}$$
$$z = y + \sum_{j=1}^{n+1} \log\left\{\frac{n-j+2+\mu}{n-j+2+\nu}\right\}.$$

Moreover, the solution of (3.5) *is unique if* $q \le 0$.

REMARK 3.2. This result simplifies, if we have an associated prior π for α and β such that $F_1(x_1,x_2) = F_2(x_1,x_2)$. In this case, the support of the optimal $(n+1)$-point design is given by the 0's of $P_{n+1}^{(\mu,\mu)}(x)$, where $\mu = -F_1(-qz,-qz) - 1$ and z solves the equation

$$z = \sum_{j=1}^{n} \log\left\{\frac{n-j+2+2\mu}{2n-2j+3+2\mu}\right\} - (n+1)\log 2.$$

THEOREM 3.5. *Let* $\chi = (-\infty,\infty)$, $f_n^T(x) = (1, x, x^2, \dots, x^n)$, $\lambda(x,\theta) = \exp(-\theta x^2)$ *and* $\Theta = (0,\infty)$. *Assume that* $\tilde{\pi}$ *is a prior on* Θ *and* π *is its associated prior defined in* (3.0) *such that*:

(i) $$\int_0^{\infty} \exp(-\theta q x)\, d\pi(\theta) < \infty \quad \text{for all } x > 0,$$

and, if $q > 0$,

(ii) $$\lim_{x\to\infty} x^{n(n+1)}\left\{\int_0^{\infty} \exp(-\theta q x)\, d\pi(\theta)\right\}^{1/q} = 0.$$

Let

$$F(x) = -\int_0^{\infty} \theta \exp(-\theta x)\, d\pi(\theta) \Big/ \int_0^{\infty} \exp(-\theta x)\, d\pi(\theta)$$

and let $H_n(x)$ *be the Hermite polynomial of degree n orthogonal with respect to the measure* $\exp(-x^2)\,dx$. *Then the* Φ_p*-optimal* $(n+1)$*-point design with* *respect to* $\tilde{\pi}$ *has equal mass at the* 0*'s of the polynomial*

$$H_{n+1}\left(\sqrt{\frac{n(n+1)}{4z}}\,x\right),$$

where z *is a positive solution of*

$$4zF(2qz) = -n(n+1). \tag{3.6}$$

Moreover, if $q \le 0$, *then the solution of* (3.6) *is unique.*

COROLLARY 3.1. *If* $\tilde{\pi}_1$ *and* $\tilde{\pi}_2$ *are two prior distributions for* θ *such that*

$$d\tilde{\pi}_2(\theta) = d\tilde{\pi}_1(c\theta) \quad \text{for some } c > 0,$$

then the support points of the Bayesian Φ_p*-optimal* $(n+1)$*-point design with respect to the prior* $\tilde{\pi}_2$ *can be obtained by multiplying the support points of the Bayesian* Φ_p*-optimal* $(n+1)$*-point design with respect to the prior* $\tilde{\pi}_1$ *by*:

(i) c *in Theorem* 3.2;
(ii) $\sqrt{c}$ *in Theorem* 3.5.

Corollary 3.1 is deduced by observing how the z_i's in the theorem change from one prior to the other. For example, in Theorem 3.2, it is readily shown from the relationship between $\tilde{\pi}_2$ and $\tilde{\pi}_1$, we have $F^2(x) = F^1(x/c)/c$, where

$$F^i(x) = -\int_0^\infty \theta \exp(-\theta x)\, d\pi_i(\theta) \Big/ \int_0^\infty \exp(-\theta x)\, d\pi_i(\theta), \qquad i = 1, 2.$$

Consequently, the solutions z_1 and z_2 of (3.2) corresponding to $F_1(x)$ and $F_2(x)$ satisfy $z_2 = cz_1$ and Corollary 3.1(i) holds. Likewise, Corollary 3.1(ii) follows by observing that $z_2 = \sqrt{c}\,z_1$. Note that Corollary 3.1(i) includes an analogous result for Theorem 3.3 if the parameter α is assumed to be known. These corollaries are quite intuitive since multiplying the parameter θ with a constant c may be interpreted as corresponding to a rescaling of the design spaces in Theorems 3.1 and 3.5 with $1/c$ and $1/\sqrt{c}$, respectively. Therefore, the designs change in a predictable way.

Theorems 3.2–3.5 provide a necessary condition for an $(n+1)$-point design to be Bayesian Φ_p-optimal within the class of all $(n+1)$-point designs. These conditions are also sufficient if the solutions of the equations in the corresponding theorem are unique. This can be proved if $q \le 0$. However, if $q > 0$, there could exist more than one solution in (3.2), (3.4), (3.5) and (3.6), in which case the optimal $(n+1)$-point design has to be found among these candidates. In the examples of Section 4, the solution of the nonlinear equations in Theorems 3.2–3.5 are also unique when $q > 0$.

Before we present examples, several comments are in order. First, Theorems 3.2–3.5 reduce to the results in Fedorov [(1972), page 88], when the prior distribution on Θ is degenerate, corresponding to the case when the efficiency function is completely specified. For instance, if a degenerate prior at $\theta_0 = 1$ is used in Theorem 3.2, then (3.2) gives $z = n(n+1)$, which coincides with case (iii) in Fedorov (1972), page 89. Second, Φ_p-optimal

$(n+1)$-point designs always have equal mass at each of the support points. This fact can easily be shown using the same argument in Silvey [(1980), page 42]. Third, when $p = 0$, the Φ_0-optimal $(n+1)$-designs depend only on the mean of the prior distribution. This can be seen, for example, in Theorem 3.3, where the Bayesian Φ_0-optimal $(n+1)$-point design has equal mass at the 0's of $L_{n+1}^{(E_1(\alpha)-1)}(E_2(\beta)x)$. Likewise, in Theorem 3.4, the support of the Bayesian Φ_0-optimal $(n+1)$-point design is given by the 0's of $P_{n+1}^{(E_1(\alpha)-1,\, E_2(\beta)-1)}(x)$, where E_i denotes the expectations of the marginal distributions of $\tilde{\pi}$, $i = 1, 2$. Finally, we emphasize that Theorems 3.2–3.5 produce designs which are optimal within the class of all designs supported on a minimal set of support points. To check if this optimal design is also optimal within the class of all designs, condition (3.1) must be invoked and checked. For example, an optimal design found from Theorem 3.5 is Bayesian Φ_p-optimal with respect to the prior $\tilde{\pi}$ if

$$g(x, \xi_{p,\tilde{\pi}}) \le 0 \qquad \text{for all } x \in R.$$

Here

$$g(x, \xi_{p,\tilde{\pi}}) = \int_0^\infty \{|M(\xi_{p,\tilde{\pi}}, \theta)|^q \exp(-\theta x^2) f_n^T(x) M^{-1}(\xi_{p,\tilde{\pi}}, \theta) f_n(x) \\ - (n+1)|M(\xi_{p,\tilde{\pi}}, \theta)|^q\}\, d\pi(\theta),$$

and π is the associated prior of $\tilde{\pi}$ defined in (3.0).

In general, analytic verification of (3.1) is rather difficult especially for continuous priors; see Dette and Neugebauer (1996a) and Dette and Sperlich (1994) where they proved some two-point Bayesian designs are optimal within the class of all designs. In practice, (3.1) is checked graphically by plotting the function $g(x, \xi_{p,\tilde{\pi}})$ over the design space.

4. Examples. In this section we illustrate our method with two examples where the efficiency function is (i) $\exp(-\theta x)$, $\theta > 0$, and (ii) $\exp(-\theta x^2)$, $\theta > 0$. In the first example, when the regression function is $f^T(x) = (1, x, \dots, x^n)$, we show that a more detailed description of the Bayesian Φ_p-optimal $(n+1)$-point design is possible with a gamma prior density on θ. In the second example, numerically Bayesian Φ_p-optimal $(n+1)$-point designs are computed using Mathematica™ [Wolfram (1988)] for the quadratic regression model. Our numerical results provide some insight into the robustness properties of these designs with respect to the choice of the prior distribution. The prior distribution on θ may have several (known) parameters, but sometimes it is not necessary to vary all their values to study how these optimal designs behave when the values of the parameters are changed. This is demonstrated in Example 4.1.

EXAMPLE 4.1. Let $\chi = [0, \infty)$, $f^T(x) = (1, x, \dots, x^n)$, $\lambda(x, \theta) = \exp(-\theta x)$ and $\theta > 0$. Suppose a gamma prior $\tilde{\pi}(\theta)$ with parameters (r, s) is used, that is,

$$\tilde{d}\pi(\theta) = \theta^{r-1} s^r \exp(-s\theta)\, d\theta / \Gamma(r), \qquad \theta > 0.$$

To find the optimal design, first note $q = p/(n+1)$ and use Lemma 3.1 to verify that the associated prior π induced by the gamma prior is also a gamma distribution but with parameters $(pn + r, s)$. In order to ensure that conditions (i) and (ii) of Theorem 3.2 apply, we require $q \geq 0$. Further algebra shows $F(x) = -(r + pn)/(s + x)$ and the (unique) solution of (3.2) is given by

$$z = \frac{sn(n+1)}{pn + r - qn(n+1)} = \frac{sn(n+1)}{r}.$$

This shows that the Bayesian Φ_p-optimal $(n+1)$-point design with respect to the gamma prior with parameters r and s depends on the prior mean (r/s) and does not depend on p. By Theorem 3.2, the Bayesian Φ_p-optimal $(n+1)$-point design has equal masses at the 0's of the polynomial $xL_n^{(1)}(rx/s)$, where $L_n^{(1)}(x)$ is the nth Laguerre polynomial defined by

$$L_n^{(1)}(x) = \frac{1}{n!}\frac{\exp(x)}{x}\left(\frac{d}{dx}\right)^n \{x^{n+1}\exp(-x)\}.$$

The first three polynomials are $L_1^{(1)}(x) = 2 - x$, $L_2^{(1)}(x) = (x^2 - 6x + 6)/2$ and $L_3^{(1)}(x) = (-x^3 + 12x^2 - 36x + 24)/6$; see Szegö (1959). Alternatively, they can be computer-generated from Mathematica$^{\text{TM}}$ (1988).

Note that, in this example, we can study changes in the Bayesian Φ_p-optimal $(n+1)$-point design with respect to the parameters of the prior very easily. We calculate the optimal design for the gamma prior with parameters $(1, 1)$. By Corollary 3.1, if a different scaling parameter s is used, the support points of the resulting optimal design have to be multiplied by s. More generally, if the mean of the gamma prior (r/s) is changed to $\delta r/s$ for some positive constant δ, then the support points of the resulting optimal design have to be divided by δ.

EXAMPLE 4.2. Let $\chi = (-\infty, \infty)$, $f^T(x) = (1, x, x^2)$, $\lambda(x, \theta) = \exp(-\theta x^2)$ and $\theta > 0$. Suppose the prior on θ is $\tilde{\pi}(\theta) = U[1\colon t]$, a discrete uniform probability distribution on the integer points $1, 2, \dots, t$. Table 2 shows the characteristics of the Bayesian Φ_p-optimal three-point designs, including their support points and the criterion value $\Phi_p(\xi_{p,\tilde{\pi}})$ for $p = 1$, 0 and -1, and $t = 2$ and 10. An asterisk on the criterion value column indicates the Bayesian Φ_p-optimal three-point design is not universally optimal.

This example demonstrates whether the Bayesian Φ_p-optimal design found from our procedure is optimal within the class of all designs [i.e., satisfies (3.1)] depends on the prior distribution and the value of the parameter p. When $p = -1$, 0 and 1 and the prior is uniformly supported on the points 1 and 2, the Bayesian Φ_p-optimal designs are all universally optimal. However, when the prior is uniformly supported on the points $1, 2, \dots, 10$, only the Bayesian Φ_1-optimal design is universally optimal. A more in-depth study of

TABLE 2
Numerically Bayesian Φ_p-optimal three-point designs for quadratic regression model using various uniform priors on θ and $\lambda(x,\theta) = \exp(-\theta x^2)$, $\theta > 0$

	z	Support points	Criterion value $\Phi_p(\xi_{p,\tilde{\pi}})$
$\tilde{\pi}(\theta) = \boldsymbol{U}[1\text{: }2]$ (uniform two-point prior at 1 and 2)			
$p = 1$	0.995063	$\pm 0.99753, 0$	0.94290
$p = 0$	1.00000	$\pm 1.00000, 0$	0.94281
$p = -1$	1.00399	$\pm 1.00199, 0$	0.94274
$\tilde{\pi}(\theta) = \boldsymbol{U}[1\text{: }10]$ (uniform 10-point prior at 1–10)			
$p = 1$	0.254875	$\pm 0.50485, 0$	0.774595
$p = 0$	0.272727	$\pm 0.52223, 0$	0.705004*
$p = -1$	0.293427	$\pm 0.54169, 0$	0.795368*

*These cases do not satisfy the equivalence condition of Theorem 3.1, and so the optimal three-point designs are not optimal within the class of all designs.

the robustness properties of these Bayesian Φ_p-optimal designs with respect to choice of criterion and model and prior assumptions is currently underway and the results will be reported in another paper.

5. An application to exponential growth models. We now show that the results in Section 3 can provide an optimal strategy for taking observations to estimate the parameters in a generalized exponential growth model described at the beginning of this paper.

Recall that the generalized exponential growth model is given by

$$y = f(x, \theta, a) + \varepsilon, \tag{5.0}$$

where

$$f(x, \theta, a) = x^v \exp(-\theta x) \sum_{j=0}^{n-1} a_j x^j$$

and ε is normally distributed with mean 0 and constant variance. For the generalized exponential growth model (5.0), observe that, at the point $x \geq 0$,

$$\left(\frac{\partial}{\partial a^T} f(x, \theta, a), \frac{\partial}{\partial \theta} f(x, \theta, a) \right)^T = x^v \exp(-\theta x) \begin{pmatrix} 1 \\ x \\ \vdots \\ x^{n-1} \\ -\sum_{j=0}^{n-1} a_j x^{j+1} \end{pmatrix} = Ag(x, \theta),$$

where $g^T(x,\theta) = x^v \exp(-\theta x)(1, x, \dots, x^n)$ and

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{pmatrix}.$$

The Fisher information matrix of a design ξ is given by

$$M(\xi,\theta,a) = A\left\{\int g(x,\theta) g^T(x,\theta)\, d\xi(x)\right\} A^T$$

and its determinant is

$$|M(\xi,\theta,a)| = a_{n-1}^2 |N(\xi,\theta)|, \tag{5.1}$$

where $N(\xi,\theta) = \int g(x,\theta)g^T(x,\theta)\,d\xi(x)$. Let $\tilde{\pi}$ denote a prior for the parameters (a^T,θ) with marginal distribution $\tilde{\pi}_2$ for θ and let ξ_θ denote the locally D-optimal design for the model (5.0). By (5.1), this design does not depend on the linear parameter a. The Bayesian Φ_p-optimal design with respect to the prior $\tilde{\pi}$ for the model (5.0) maximizes

$$\begin{aligned}&\left\{\int\left\{\frac{|M(\xi,\theta,a)|}{|M(\xi_\theta,\theta,a)|}\right\}^{p/(n+1)} d\tilde{\pi}(\theta,a)\right\}^{1/p}\\ &\qquad = \left\{\int_0^\infty\left\{\frac{|N(\xi,\theta)|}{|N(\xi_\theta,\theta)|}\right\}^{p/(n+1)} d\tilde{\pi}_2(\theta)\right\}^{1/p}\end{aligned}$$

and depends only on the marginal distribution $\tilde{\pi}_2$ for the parameter θ.

THEOREM 5.1. *Assume $\tilde{\pi}_2$ is a prior on θ which satisfies assumptions* (i) *and* (ii) *of Theorem* 3.3. *The Bayesian Φ_p-optimal $(n+1)$-point design with respect to $\tilde{\pi}_2$ in the model* (5.0) *has equal mass at the* 0's *of the polynomial*

$$L_{n+1}^{(2v-1)}\left(\frac{2(n+2v)(n+1)}{z}x\right) \quad \text{if } v > 0$$

and

$$xL_n^{(1)}\left(\frac{2n(n+1)}{z}x\right) \quad \text{if } v = 0.$$

Here $L_n^{(\alpha)}(x)$ denotes the nth generalized Laguerre polynomial orthogonal with respect to the measure $x^\alpha \exp(-x)\,dx$ and z is a solution of

$$zF_2(qz) = -(n+1)(2v+n), \tag{5.2}$$

where $q = p/(n+1)$,

$$F_2(x) = -\int_0^\infty \theta \exp(-\theta x)\, d\pi_2(\theta) \Big/ \int_0^\infty \exp(-\theta x)\, d\pi_2(\theta)$$

and π_2 *is the associated prior of* $\tilde{\pi}_2$ *defined in* (3.0), *that is*, $d\pi_2(\theta) = \theta^{p(n+2v)}\, d\tilde{\pi}_2(\theta)$. *In addition, if* $q \le 0$, *the solution of* (5.2) *is unique.*

PROOF. Consider the case $v > 0$ and the situation of Theorem 3.3 for the prior $\delta_{2v} \otimes \tilde{\pi}_2$, where δ_{2v} is the Dirac measure at the point $2v$ (i.e., $\alpha = 2v$ is assumed to be known) and $\tilde{\pi}_2$ is a prior for the parameter β. By Theorem 3.3, the optimal $(n+1)$-point design has equal masses at the 0's of the polynomial

$$L_{n+1}^{(2v-1)}\left(\frac{(n+v)(n+1)}{z} x\right),$$

where z is a solution of (5.2) and π_2 is the associated prior of $\tilde{\pi}_2$. Because the optimal design problem for the model (5.0) with respect to the prior $\tilde{\pi}_2$ coincides with the problem in Theorem 3.3 with respect to the prior

$$d\hat{\pi}_2(\theta) = d\tilde{\pi}_2(\theta/2),$$

the assertion of the theorem follows. The case $v = 0$ is treated in the same way using Theorem 3.2. □

COROLLARY 5.1. *The Bayesian D-optimal* $(n+1)$*-point design for the model* (5.0) *with respect to the prior* $\tilde{\pi}_2$ *puts equal mass at the* 0*'s of the polynomial* $L_{n+1}^{(2v-1)}(2E_2(\theta)x)$ *if* $v > 0$ *and* $xL_n^{(1)}(2E_2(\theta)x)$ *if* $v = 0$. *Here* $E_2(\theta)$ *denotes the expectation with respect to the prior distribution* $\tilde{\pi}_2$ *for the parameter* θ.

EXAMPLE 5.1. Assume that $n = 1$, $v \ge 0$. From Szegö (1959), we obtain $L_2^{(2v-1)}(t) = \{t^2 - 2(2v+1)t + 2v(2v+1)\}/2$, $L_1^{(1)}(t) = 2 - t$ and Corollary 5.1 shows that the best two-point design with respect to the Bayesian D-optimality criterion has equal mass at the points

$$\frac{2v + 1 - \sqrt{2v+1}}{2E_2(\theta)} \quad \text{and} \quad \frac{2v + 1 + \sqrt{2v+1}}{2E_2(\theta)}.$$

This provides an alternative proof of a result in Mukhopadhyay and Haines (1995), Dette and Neugebauer (1996a) ($v = 0$) and Dette and Sperlich (1994) ($v \ge 0$).

APPENDIX

Since the proofs of Theorems 3.2–3.5 are similar, we supply the proof of Theorem 3.2 only. Before this can be done, we first briefly review some basic results from the theory of continued fraction expansion and orthogonal polynomials. Some useful references for background reading are Shohat and Tarmarkin (1943) and Wall (1948). For proving Theorems 3.3–3.5, the work of Lau and Studden (1988) is helpful.

It is known that the Stieltjes transform of any probability measure ξ on $[0,\infty)$ with existing moments can be written as a continued fraction expansion, commonly denoted by

$$\Psi(z) = \int_0^\infty \frac{d\xi(x)}{z-x} = \frac{1|}{|z} - \frac{d_1|}{|1} - \frac{d_2|}{|z} - \cdots - \frac{d_{2n}|}{|z} - \frac{d_{2n+1}|}{|1} - \cdots$$

where $d_j \geq 0$ for all j. The Stieltjes transform is defined for all complex z outside of $(0,\infty)$ and is an analytic function in this area. Moreover, the continued fraction on the right-hand side converges uniformly on every compact (complex) subset with positive distance from the nonnegative line [Perron (1954)].

If ξ is supported at $n+1$ support points, this expansion terminates at the $(2n+1)$ term with $d_j > 0$ for $j \leq 2n$ and $d_{2n+1} \geq 0$, and the above expansion simplifies to a rational polynomial $P_n(z)/Q_{n+1}(z)$. The degrees of these polynomials are n and $n+1$, respectively, and their coefficients can be expressed in terms of the d_i's. They can also be found using recursive relationships [Lau and Studden (1988)]. The support points of the design ξ are given by the roots of $Q_{n+1}(x) = 0$ and the mass of ξ at the support point x_j is given by

$$(z - x_j)\Psi(z)|_{z=x_j} = \frac{P_n(x_j)}{(d/dz)Q_{n+1}(z)|_{z=x_j}}.$$

We are now ready to prove Theorem 3.2.

PROOF OF THEOREM 3.2. Since we are interested in seeking an optimal design ξ with $n+1$ support points, the Stieltjes transform of ξ has a continued fraction expansion of the form

$$\int_0^\infty \frac{d\xi(x)}{z-x} = \frac{1|}{|z} - \frac{d_1|}{|1} - \frac{d_2|}{|z} - \cdots - \frac{d_{2n}|}{|z} - \frac{d_{2n+1}|}{|1}, \tag{A.1}$$

with $d_i > 0$ for $i \leq 2n$ and $d_{2n+1} \geq 0$. From the results of Lau and Studden (1988), the determinants $M(\xi, \theta)$ can be expressed in terms of the d_i's if ξ is an $(n+1)$-point design, that is,

$$|M(\xi,\theta)| = \prod_{i=1}^{n} (d_{2i-1}d_{2i})^{n-i+1} \exp\left[-\theta \sum_{i=1}^{2n+1} d_i\right].$$

Thus, if optimization in (2.0) [or, equivalently, (3.1)] is restricted to $(n+1)$-

point designs, the optimality criterion in (3.1) can be rewritten as

$$\text{(A.2)}\quad \tilde{\phi}_q(\xi) = \prod_{i=1}^{n} (d_{2i-1}d_{2i})^{n-i+1}\left[\int_0^\infty \exp\left[-\theta q \sum_{i=1}^{2n+1} d_i\right] d\pi(\theta)\right]^{1/q},$$

where the integral exists by assumption (i) of Theorem 3.2. If $d_j \to 0$, we have $\tilde{\phi}_q(\xi) \to 0$, $j = 1, \dots, 2n$, and, in the case $d_j \to \infty$, assumption (ii) implies that $\tilde{\phi}_q(\xi) \to 0$, $j = 1, \dots, 2n$. Consequently, $\tilde{\phi}_q(\xi)$ is maximized for some point $(d_1, d_2, \dots, d_{2n}) \in (0, \infty)^{2n}$ and $d_{2n+1} = 0$. From the theory of exponential families [Lehmann (1986), page 59, for example] and assumption (i), it follows that $\log[\tilde{\phi}_q(\xi)]$ is differentiable with respect to $d_1, d_2, \dots, d_{2n}$ with partial derivatives

$$\frac{\partial}{\partial d_{2i}} \log \tilde{\phi}_q(\xi) = \frac{n-i+1}{d_{2i}} + F\left(q \sum_{i=1}^{2n+1} d_i\right), \qquad i = 1, 2, \dots, n,$$

$$\frac{\partial}{\partial d_{2i-1}} \log \tilde{\phi}_q(\xi) = \frac{n-i+1}{d_{2i-1}} + F\left(q \sum_{i=1}^{2n+1} d_i\right), \qquad i = 1, 2, \dots, n,$$

where

$$F(x) = -\int_0^\infty \theta \exp(-\theta x)\, d\pi(\theta) \Big/ \int_0^\infty \exp(-\theta x)\, d\pi(\theta).$$

Putting $z = \sum_{i=1}^{2n+1} d_i$, the optimal $d_1, d_2, \dots, d_{2n+1}$ have to satisfy

$$\text{(A.3)}\quad d_{2i-1} = d_{2i} = -\frac{n-i+1}{F(qz)}, \qquad i = 1, 2, \dots, n \text{ and } d_{2n+1} = 0.$$

From the definition of z, we obtain that z is a root of the equation

$$z = \sum_{i=1}^{n} (d_{2i-1} + d_{2i}) = -\frac{n(n+1)}{F(qz)},$$

which gives (3.2). Furthermore, by (A.3),

$$\text{(A.4)}\quad d_{2i-1} = d_{2i} = \frac{(n-i+1)z}{n(n+1)}, \qquad i = 1, 2, \dots, n \text{ and } d_{2n+1} = 0.$$

If ξ has support points $x_0, x_1, \dots, x_n$ with masses $\xi_0, \xi_1, \dots, \xi_n$ and corresponding continued fraction expansion (A.1), then it is easy to see that, for $\gamma > 0$, the measure $\tilde{\xi}$ with support points $\gamma x_0, \gamma x_1, \dots, \gamma x_n$, and masses $\xi_0, \xi_1, \dots, \xi_n$ has the continued fraction expansion (A.1), where the d_i's in (A.1) have to be replaced by $\tilde{d}_i = \gamma d_i$, $i = 1, 2, \dots, 2n+1$. The measure $\tilde{\xi}$ corresponding to the sequence

$$\text{(A.5)}\qquad \tilde{d}_{2i-1} = \tilde{d}_{2i} = n - i + 1, \qquad i = 1, 2, \dots, n \text{ and } \tilde{d}_{2n+1} = 0$$

can be obtained as follows. For the Stieltjes transform of $\tilde{\xi}$, we obtain by an odd contraction [Perron (1954)] from (A.1) that

$$\begin{aligned}\Psi(z) &= \int_0^\infty \frac{d\tilde{\xi}(x)}{z-x}\\ &= \frac{1}{z} + \frac{\tilde{d}_1}{z}\left\{\frac{1|}{|z-\tilde{d}_1-\tilde{d}_2} - \frac{\tilde{d}_2\tilde{d}_3|}{|z-\tilde{d}_3-\tilde{d}_4} - \cdots - \frac{\tilde{d}_{2n-2}\tilde{d}_{2n-1}|}{|z-\tilde{d}_{2n-1}-\tilde{d}_{2n}}\right\}\\ &= \frac{G_n(z)+nG_{n-1}(z)}{zG_n(z)},\end{aligned}$$

where the polynomial $G_n(x)$ is defined recursively by [see Perron (1954)]

$$\begin{aligned}G_{k+1}(x) &= \left(x-\tilde{d}_{2n-2k-1}-\tilde{d}_{2n-2k}\right)G_k(x) - \tilde{d}_{2n-2k}\tilde{d}_{2n-2k+1}G_{k-1}(x)\\ &= (x-2k-2)G_k(x) - k(k+1)G_{k-1}(x), \qquad k = 1,2,\dots,n-1,\end{aligned}$$

with $G_0(x) = 1$ and $G_1(x) = x - 2$. Observing the recurrence relation for the Laguerre polynomials $L_n^{(1)}(x)$ [see, e.g., Abramowitz and Stegun (1964), Section 22.7.12], it follows that $G_n(x) = (-1)^n n!\, L_n^{(1)}(x)$ and for the Stieltjes transform of $\tilde{\xi}$,

$$\Psi(x) = \frac{L_n^{(1)}(x) - L_{n-1}^{(1)}(x)}{xL_n^{(1)}(x)}.$$

Consequently, the support points of $\tilde{\xi}$ are given by the 0's of $xL_n^{(1)}(x)$ and the mass at the support point $\tilde{x}_j$, $j = 0,1,\dots,n$, is

$$\begin{aligned}\tilde{\xi}(\tilde{x}_j) = (z-\tilde{x}_j)\Psi(z)|_{z=\tilde{x}_j} &= \frac{L_n^{(1)}(\tilde{x}_j) - L_{n-1}^{(1)}(\tilde{x}_j)}{(d/dz)zL_n^{(1)}(z)|_{z=\tilde{x}_j}}\\ &= \frac{L_n^{(1)}(\tilde{x}_j) - L_{n-1}^{(1)}(\tilde{x}_j)}{L_n^{(1)}(\tilde{x}_j)(n+1) - L_{n-1}^{(1)}(\tilde{x}_j)(n+1)} = \frac{1}{n+1};\end{aligned}$$

see Szegö (1959), formula 5.1.14. Therefore, $\tilde{\xi}$ is a uniform distribution with support at the 0's of the polynomials $xL_n^{(1)}(x)$. The first assertion in Theorem 3.2 now follows from the discussion after (A.4) which relates the designs ξ and $\tilde{\xi}$ corresponding to the sequences in (A.4) and (A.5). [Note that the derivation of $\tilde{\xi}$ and the corresponding statement in Lau and Studden (1988) is incorrect.] Finally, we remark that it is straightforward to show that the function

$$H_q(z) = -\frac{n(n+1)}{F(qz)}$$

is strictly decreasing in z if $q < 0$ and that $H_q(0) = n(n+1)/E_\pi(\theta)$. Therefore, the equation $H_q(z) = z$ has exactly one positive root which is given by (3.2). When $q > 0$, a positive solution always exists by standard optimal design theory arguments, and the right "z" will have to be selected by trial and error if there is more than one positive solution. □

Acknowledgments. We would like to thank the referees and the editors for helpful comments. Parts of the paper were written while H. Dette was visiting the University of Göttingen and the Department of Biostatistics at UCLA. This author would like to thank the Institut für Mathematische Stochastik and the Department of Biostatistics for their hospitality.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1966). *Handbook of Mathematical Functions*. Dover, New York.

ATKINSON, A. C. and COOK, R. D. (1995). *D*-optimum designs for heteroscedastic linear models. *J. Amer. Statist. Assoc.* **90** 204–212.

BOX, G. E. P. and LUCAS, H. (1959). Designs of experiments in non-linear situations. *Biometrika* **46** 77–90.

CHALONER, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* **12** 771–781.

CHALONER, K. (1993). A note on optimal Bayesian design for nonlinear problems. *J. Statist. Plann. Inference* **37** 229–235.

CHALONER, K. and LARNTZ, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inference* **21** 191–208.

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24** 586–602.

DETTE, H. and NEUGEBAUER, H. M. (1996a). Bayesian *D*-optimal designs for exponential regression models. *J. Statist. Plann. Inference*. To appear.

DETTE, H. and NEUGEBAUER, H. M. (1996b). Bayesian optimal one point designs for one parameter nonlinear models. *J. Statist. Plann. Inference*. To appear.

DETTE, H. and SPERLICH, S. (1994). A note on Bayesian *D*-optimal designs for a generalization of the exponential growth model. *South African Statist. J.* **28** 103–117.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments* (translated by W. J. Studden and E. M. Klimko). Academic Press, New York.

FORD, I., TITTERINGTON, D. M. and KITSOS, C. P. (1989). Recent advances in nonlinear experimental design. *Technometrics* **31** 49–60.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience, New York.

KIEFER, J. (1960). Optimum experimental designs V, with applications to systematic and rotatable designs. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 381–405. Univ. California Press, Berkeley.

LAU, T. S. and STUDDEN, W. J. (1988). On an extremal problem of Fejer. *J. Approx. Theory* **53** 184–194.

LÄUTER, E. (1974). Experimental planning in a class of models. *Math. Operationsforsch. Statist. Ser. Statist.* **36** 1627–1655.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*. Wiley, New York.

MUKHOPADHYAY, S. and HAINES, L. (1995). Bayesian *D*-optimal design for the exponential growth model. *J. Statist. Plann. Inference* **44** 385–398.

PAZMAN, A. (1986). *Foundations of Optimum Experimental Design*. Reidel, Dordrecht.

PERRON, O. (1954). *Die Lehre von den Kettenbrüchen*. Teubner, Stuttgart.

PILZ, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models*. Wiley, New York.

PRONZATO, L. and WALTER, E. (1985). Robust experimental design via stochastic approximation. *Math. Biosci.* **75** 103–120.

PRONZATO, L. and WALTER, E. (1988a). Robust experimental design for nonlinear regression. In *Model Oriented Data Analysis* (V. V. Fedorov and E. Läuter, eds.) 77–86. Springer, Berlin.

PRONZATO, L. and WALTER, E. (1988b). Robust experimental design via maximin optimality. *Math. Biosci.* **89** 161–176.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

RATKOWSKY, D. A. (1983). *Nonlinear Regression Modelling*. Dekker, New York.

SHOHAT, J. A. and TAMARKIN, J. D. (1943). The problem of moments. *Math. Surveys* **1**.

SILVEY, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

STUDDEN, W. J. (1980). D_s-optimal designs for polynomial regression. *Ann. Statist.* **8** 1132–1141.

STUDDEN, W. J. (1982). Optimal designs for weighted polynomial regression using canonical moments. *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) 335–350. Academic Press, New York.

SZEGÖ, G. (1959). *Orthogonal Polynomials*, 4th ed. Amer. Math. Soc. Colloq. Publ. **23**. Amer. Math. Soc., Providence, RI.

WALL, H. S. (1948). *Analytic Theory of Continued Fractions*. Van Nostrand, New York.

WALTER, E. and PRONZATO, L. (1990). Qualitative and quantitative experiment design for phenomenological models—a survey. *Automatica* **26** 195–213.

WOLFRAM, S. (1988). *Mathematica*™. *A System for Doing Mathematics by Computer*. Addison-Wesley, Reading, MA.

FAKULTÄT UND INSTITUT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: holger.dette@rz.ruhr-uni-bochum.de

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CALIFORNIA 90095
E-MAIL: wkwong@sunlab.ph.ucla.edu