The Annals of Statistics
1986, Vol. 14, No. 4, 1399–1418

CURVATURES FOR PARAMETER SUBSETS IN NONLINEAR REGRESSION[1]

BY R. DENNIS COOK AND MIRIAM L. GOLDBERG

University of Minnesota and U.S. Environmental Protection Agency

The relative curvature measures of nonlinearity proposed by Bates and Watts (1980) are extended to an arbitrary subset of the parameters in a normal, nonlinear regression model. In particular, the subset curvatures proposed indicate the validity of linearization-based approximate confidence intervals for single parameters. The derivation produces the original Bates–Watts measures directly from the likelihood function. When the intrinsic curvature is negligible, the Bates–Watts parameter-effects curvature array contains all information necessary to construct curvature measures for parameter subsets.

1. Introduction. Confidence regions for parameters of a normal nonlinear regression model are commonly constructed by using linear regression methods, replacing the solution locus with the tangent plane at the maximum likelihood estimate. Such linear regions are generally easier to construct and comprehend than corresponding likelihood regions. Likelihood regions, on the other hand, are not influenced by parameter-effects nonlinearity and generally have true coverage closer to the nominal level than do linear regions. Under suitable regularity conditions and with a sufficiently large sample size, linear and likelihood regions will be in good agreement, but in any particular problem the strength of this agreement is uncertain.

Bates and Watts (1980) propose measures of intrinsic and parameter-effects curvature for assessing the adequacy of the linear approximation: Relatively small values for both the maximum intrinsic curvature Γ^{η} and the maximum parameter-effects curvature Γ^{τ} indicate that the linear approximation is reasonable, while relatively large values for either Γ^{η} or Γ^{τ} indicate that this approximation is questionable. These ideas are extended and refined by Bates and Watts (1981), and Hamilton, Bates and Watts (1982). For a review of related literature, see Bates and Watts (1980) and Ratkowsky (1983). Programs for calculating Γ^{η} and Γ^{τ} are given by Bates, Hamilton, and Watts (1983).

The material in Bates and Watts (1980) represents an important step forward, but their method for assessing the adequacy of the linear approximation applies only to the full parameter vector, as indicated by Cook and Witmer (1985) and Linssen (1980). It is fairly easy to construct examples where Γ^{τ} is relatively large and yet there is good agreement between the linear and likelihood regions for a

Received April 1985; revised December 1985.

[1]This work was undertaken while the first author was a visiting professor at the Mathematics Research Center, University of Wisconsin-Madison, and was sponsored in part by the United States Army under Contract No. DDAAG29-80-0041 and the Monsanto Company.

AMS 1980 *subject classifications.* Primary 62J02; secondary 62F25.

Key words and phrases. Confidence regions, curvature measures, Fieller–Creasy problem, least squares, likelihood.

subset of the parameters. One such example is given in Section 2, which also contains a brief review of the linear approximation and the Bates–Watts methodology. This inability to assess the adequacy of subset regions reflects an important gap in our understanding and ability to deal with nonlinear models.

In Section 3 we develop measures for assessing the agreement between linear and likelihood regions for an arbitrary subset of parameters from a nonlinear regression model. The measures require the same building blocks as needed for the construction of Γ^τ, and reduce to Γ^τ when the full parameter vector is considered. Computationally, these measures require little more effort than Γ^τ itself. Section 4 contains several examples and our concluding comments are given in Section 5.

Subsequent developments are based on the nonlinear regression model

$$y_i = f(x_i, \theta) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1}$$

where y_i is the ith response, x_i is a vector of known variables, θ is a $p \times 1$ vector of unknown parameters, the response function f is a known, scalar-valued function that is twice continuously differentiable in θ, and the errors ε_i are independent and identically distributed normal random variables with mean 0 and variance σ^2. For notational convenience, let $f_i(\theta) = f(x_i, \theta)$ and let V denote the $n \times p$ matrix with elements $f_i^r = \partial f_i / \partial \theta_r$, $i = 1, \dots, n$, $r = 1, \dots, p$. Here and in what follows all derivatives are evaluated at $\hat{\theta}$, the maximum likelihood estimate of θ, unless explicitly indicated otherwise.

Various quadratic approximations to be used in the following sections involve the $p \times p$ matrices W_i, $i = 1, \dots, n$, with elements $f_i^{rs} = \partial^2 f_i / \partial \theta_r \, \partial \theta_s$, $r, s = 1, \dots, p$. These matrices can be written conveniently in an $n \times p \times p$ array W (Bates and Watts, 1980). The abth "column" of W is the abth second derivative vector W_{ab} with elements f_i^{ab}, $i = 1, \dots, n$, while the ith face W_i of W is the $p \times p$ matrix consisting of the ith elements of the second derivative vectors W_{ab}.

2. Curvatures and the linear approximation. Let $F(\theta)$ denote the $n \times 1$ vector with elements $f_i(\theta)$. The standard elliptical confidence region for θ based on a linear approximation of $F(\theta)$ about $\hat{\theta}$ can be written as

$$\{\theta \colon \phi^T V^T V \phi \le s^2 G\}, \tag{2}$$

where $\phi = (\phi_a) = \theta - \hat{\theta}$, $s^2 = \Sigma(y_i - f_i(\hat{\theta}))^2/(n-p)$, $G = pF_\alpha(p, n-p)$, and $F_\alpha(v_1, v_2)$ is the upper α probability point of an F distribution with v_1 and v_2 degrees of freedom.

To assess the adequacy of the region in (2), we need the standard quadratic expansion of F about $\hat{\theta}$:

$$F(\theta) \simeq F(\hat{\theta}) + V\phi + \tfrac{1}{2}\phi^T W \phi. \tag{3}$$

Multiplication involving three-dimensional arrays is defined as in Bates and Watts (1980) so that $\phi^T W \phi$ is an $n \times 1$ vector with elements $\phi^T W_i \phi$, $i = 1, \dots, n$. Generally, if F is quadratic over a sufficiently large neighborhood of $\hat{\theta}$ and the quadratic term of (3) is sufficiently small relative to the linear term, the linear region (2) should be reasonable; otherwise, this approximation may be in doubt.

Bates and Watts (1980, 1981) implement this idea by first decomposing each column of W as $W_{ab} = P_V W_{ab} + (I - P_V)W_{ab} = W^{\tau}_{ab} + W^{\eta}_{ab}$, where P_V is the orthogonal projection operator for the column space of V. With this decomposition, the quadratic expansion (3) becomes

$$F(\theta) \simeq F(\hat{\theta}) + V\phi + \tfrac{1}{2}\phi^T W^{\tau}\phi + \tfrac{1}{2}\phi^T W^{\eta}\phi, \tag{4}$$

where W^{τ} and W^{η} are the $n \times p \times p$ arrays with columns W^{τ}_{ab} and W^{η}_{ab}, respectively.

Next, the adequacy of the linear region is assessed by using the maximum parameter-effects curvature

$$\Gamma^{\tau} = \max \frac{\|\phi^T W^{\tau}\phi\|}{\|V\phi\|^2}\sqrt{p}\, s \tag{5}$$

and the maximum intrinsic curvature

$$\Gamma^{\eta} = \max \frac{\|\phi^T W^{\eta}\phi\|}{\|V\phi\|^2}\sqrt{p}\, s, \tag{6}$$

where the maximum is taken over all ϕ in R^p. Bates and Watts (1980) suggest that (2) should be adequate if Γ^{η} and Γ^{τ} are both small compared to the guide $c = (F_{\alpha}(p, n-p))^{-1/2}$. When Γ^{η} or Γ^{τ} is greater than c, the linear approximation and the circular approximation that is the basis of the curvature measures both break down within the linear region. Thus, Ratkowsky (1983) proposes that $c/2$ be used as a cutoff level, beyond which the linear region is presumed inadequate.

This method, which was designed specifically for normal nonlinear regression, fits within the larger context of curved exponential families and Amari's (1982) α connections. The relationship between the work of Bates and Watts (1980, 1981), Amari (1982), and others is discussed in detail by Kass (1984).

To demonstrate the importance of extending the Bates–Watts methodology to subsets of θ, we consider the Fieller–Creasy problem in which the ratio of the means of two normal populations is of interest:

$$f(x_i, \theta) = \theta_1 x_i + \theta_1\theta_2(1 - x_i), \tag{7}$$

where x_i is an indicator variable that takes the values 1 and 0 for populations 1 and 2, respectively. For convenience we assume equal sample sizes for the two populations $n_1 = n_2 = n/2$ and, without loss of generality, we assume that σ^2 is known.

The model given in (7) is intrinsically linear so that $\Gamma^{\eta} = 0$. Further, Cook and Witmer (1985) show that

$$\Gamma^{\tau} = \frac{\sqrt{2}\,\sigma\left\{\left(\hat{\theta}_2^2 + 1\right)^{1/2} + |\hat{\theta}_2|\right\}}{|\hat{\theta}_1|\sqrt{n}}, \tag{8}$$

which may be arbitrarily large depending on θ_1.

In this case the guide for judging the adequacy of the linear approximation is $c = (\chi(\alpha; 2))^{-1/2}$ where $\chi(\alpha; v)$ is the upper α probability point of the χ^2

distribution with v degrees of freedom. However, it is clear that standard methods can be used to form exact confidence intervals for θ_1, the mean of the first population, regardless of the value of Γ^τ. In other words, the linear and likelihood regions for θ_1 are identical for all Γ^τ.

A similar phenomenon occurs in connection with θ_2. Let $r = 2\sigma^2\chi(\alpha; 1)/n\hat{\theta}_1^2$. Assuming that $r < 1$, the $1 - \alpha$ likelihood region for θ_2 can be written as

$$\left[\hat{\theta}_2 \pm \left\{r(1-r) + r\hat{\theta}_2^2\right\}^{1/2}\right]\Big/(1-r). \tag{9}$$

The level associated with this region is exact. The corresponding linear region is

$$\hat{\theta}_2 \pm \left(r + r\hat{\theta}_2^2\right)^{1/2}. \tag{10}$$

Clearly, (9) and (10) will be close only if r is sufficiently small. For any fixed value of r, however, Γ^τ may be large or small depending on the value of $\hat{\theta}_2$. We will return to this example at the end of the next section.

3. Subsets. The Bates–Watts curvatures measure relevant local properties of the manifold defined by $F(\theta)$. To develop similar curvatures for a selected subset of θ, we require a submanifold that captures relevant statistical information for inference about the parameter subset. Such a submanifold is not uniquely defined and hence an additional statistical criterion is needed to guide the selection. Once the submanifold has been selected, curvatures for the submanifold can be constructed in a manner similar to that used by Bates and Watts (1980) for the ambient manifold. The following development of these ideas makes implicit use of the inheritance relationship between an affine connection on a submanifold with that on the ambient manifold. This relationship is well known in differential geometry (Eisenhart (1964)) and also appears in the recent statistical literature (Amari (1982), Section 4; Kass (1984)).

Let $L(\theta, \sigma^2)$ denote the log likelihood for model (1), partition $\theta^T = (\theta_1^T, \theta_2^T)$, where θ_i is a $p_i \times 1$ vector, $i = 1, 2$, and assume that θ_2 is the parameter subset of interest. The submanifold that we use is obtained from the likelihood region for θ_2 formed by inverting the corresponding likelihood ratio test. To construct this region, let $(g^T(\theta_2), \tilde{\sigma}^2(\theta_2))$ denote the vector-valued function that maximizes $L(\theta_1, \theta_2, \sigma^2)$ over θ_1 and σ^2 for given θ_2. Then after a little simplification the likelihood region for θ_2 can be written in the form (cf. Cox and Hinkley (1974), page 343)

$$\left\{\theta_2: \Sigma\left(y_i - f_i(g(\theta_2), \theta_2)\right)^2 < \rho\right\} \tag{11}$$

where ρ, a selected positive constant, is used to set the nominal level. Clearly, the form of this region is governed by the vector-valued function $h(\theta_2) = F(g(\theta_2), \theta_2)$. If h is essentially linear over a sufficiently large neighborhood of $\hat{\theta}_2$, the contours defined by (11) will be elliptical and we can expect (11) and the corresponding linear region to agree; otherwise these regions will tend to be dissimilar. To determine when these regions are in substantial agreement, we investigate the behavior of h by using the method described in Section 2, except that F is replaced by h which, in combination with $Y = (y_i)$, contains essential information on θ_2. Thus, in exact analogy with the Bates–Watts development, we will produce expressions for the curvature of the submanifold defined by h. Where necessary for clarity, we refer to this as "subset curvature." Similarly, "subset parameter effects," and "subset intrinsic" refer to the decomposition of the subset curvature into components in the submanifold tangent plane and its orthogonal complement.

Let $\alpha^T(\theta_2) = (\alpha_i(\theta_2)) = (g^T(\theta_2), \theta_2^T)$, let Δ_1 denote the $p \times p_2$ matrix with elements $\partial\alpha_i/\partial\theta_{2j}$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p_2$, and let Δ_2 denote the $p \times p_2 \times p_2$ array with ith face Δ_{2i}, $i = 1, 2, \ldots, p$; the elements of Δ_{2i} are $\partial^2\alpha_i/\partial\theta_{2j}\,\partial\theta_{2k}$, $j, k = 1, \ldots, p_2$. We assume that g is a twice continuously differentiable function of θ_2. With these definitions the straightforward quadratic approximation of $h(\theta_2)$ about $\hat{\theta}_2$ can be written as

(12a) $$h(\theta_2) \simeq F(\hat{\theta}) + V\Delta_1\phi_2$$

(12b) $$+ \tfrac{1}{2}\phi_2^T\Delta_1^T W \Delta_1\phi_2$$

(12c) $$+ \tfrac{1}{2}V\big(\phi_2^T\Delta_2\phi_2\big),$$

where $\phi_2 = \theta_2 - \hat{\theta}_2$.

3.1. *Refining equation* (12). For the quadratic expansion in (12) to be useful, we need to develop explicit forms for Δ_1 and Δ_2 to produce a reexpression that displays the (subset) parameter-effects and intrinsic components of h at $\hat{\theta}_2$. To avoid interruption, the details of this development have been relegated to the Appendix. Here we discuss the final form.

The final form of (12) is based on the assumption that the intrinsic curvature of F at $\hat{\theta}$ is negligible. That assumption is somewhat restrictive but it is valid in the important class of problems where the parameters of interest are nonlinear functions of the location parameters in a linear model. In any event, we judge the practical advantages of allowing for substantial intrinsic curvatures to be minimal since experience has shown (see Bates and Watts (1980) and Ratkowsky (1983)) that they are typically small. Of course, Γ^{η} can and should be evaluated in practice so that this assumption can be checked.

In the remainder of this paper we use $C(M)$ and $C'(M)$ to indicate the column and orthogonal spaces, respectively, of the matrix M; the corresponding orthogonal projection operators will be denoted by P_M and P_M', respectively.

In their development of the intrinsic and parameter-effects curvatures for the full parameter vector, Bates and Watts (1980) found it convenient and revealing to work in transformed coordinates. Similarly, the quadratic expansion (12) is most easily understood in terms of these same transformed coordinates: Let $V = UR$ denote the unique QR factorization of V where R is upper triangular and the columns of the $n \times p$ matrix U form an orthonormal basis for $C(V)$. Next, partition R as

(13) $$R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix},$$

where R_{ii} is $p_i \times p_i$, $i = 1, 2$. Transformed coordinates $\tilde{\phi}$ can now be defined as

$\tilde{\phi}^T = (\tilde{\phi}_1^T, \tilde{\phi}_2^T) = \phi^T R^T$ so that

$$\tilde{\phi}_1 = R_{11}\phi_1 + R_{12}\phi_2 \tag{14}$$

and

$$\tilde{\phi}_2 = R_{22}\phi_2. \tag{15}$$

In the following, any quantity with a tilde added above indicates evaluation in the $\tilde{\phi}$ coordinates. Thus, for example, $\tilde{V} = U$ and $\tilde{W} = R^{-T}WR^{-1}$. Partition the ith face $\tilde{W}_i$ of $\tilde{W}$ as

$$\tilde{W}_i = \begin{pmatrix} \tilde{W}_{i11} & \tilde{W}_{i12} \\ \tilde{W}_{i21} & \tilde{W}_{i22} \end{pmatrix}, \qquad i = 1, \dots, n, \tag{16}$$

where the dimension of $\tilde{W}_{ijj}$ is $p_j \times p_j$, $j = 1, 2$. Next, define $\tilde{W}_{22}$ to be the $n \times p_2 \times p_2$ subarray of $\tilde{W}$ with ith face $\tilde{W}_{i22}$ and similarly define $\tilde{W}_{12}$ to be the $n \times p_1 \times p_2$ subarray of $\tilde{W}$ with ith face $\tilde{W}_{i12}$, $i = 1, \dots, n$. Finally, partition $V = (V_1, V_2)$ and $U = (U_1, U_2)$, where U_i and V_i are $n \times p_i$ matrices.

With this structure, the quadratic expansion of h can be reexpressed informatively as

$$h(\theta_2) \simeq F(\hat{\theta}) + U_2\tilde{\phi}_2 \tag{17a}$$

$$+ \tfrac{1}{2}\tilde{\phi}_2^T [P_{U_2}][\tilde{W}_{22}]\tilde{\phi}_2 \tag{17b}$$

$$- U_1[\tilde{\phi}_2^T U_2^T][\tilde{W}_{12}]\tilde{\phi}_2, \tag{17c}$$

where the brackets $[\cdot][\cdot]$ indicate column (sample space) multiplication as defined in Bates and Watts (1980), and discussed briefly in the Appendix. Term (17a) describes the plane tangent to h at $\hat{\theta}_2$. Since $C(U_2) = C(P'_{V_1}V_2)$, this plane is simply the affine subspace $F(\hat{\theta}) + C(P'_{V_1}V_2)$. This is the same as the subspace obtained when using the linear approximation to form a confidence region for θ_2. In other words, the confidence contours based on (2) will coincide with those based on substituting the linear approximation of h into (11), as expected.

Term (17b) contains the projections of the columns of $\tilde{W}_{22}$ onto the plane tangent to h at $\hat{\theta}_2$. Thus, this term reflects the (subset) parameter-effects curvature of h in the direction $\tilde{\phi}_2$. The maximum parameter-effects curvature Γ_s^τ for the subset θ_2 can now be defined as

$$\Gamma_s^\tau(\theta_2) = \max\|d^T[P_{U_2}][\tilde{W}_{22}]d\|\sqrt{p_2}\,s, \tag{18}$$

where the maximum is taken over all d in $D = \{d\colon d \in R^{p_2}, \|d\| = 1\}$. Since $\tilde{\phi}_2$ is a linear transformation of ϕ_2 as described in (15), $\Gamma_s^\tau(\theta_2)$ will be the same in both coordinate systems.

To further understand (18), partition the ith face A_i of the $p \times p \times p$ unscaled parameter-effects curvature array $A = [U^T][\tilde{W}]$ as

$$A_i = \begin{pmatrix} A_{i11} & A_{i12} \\ A_{i21} & A_{i22} \end{pmatrix}, \tag{19}$$

where the dimension of A_{ijj} is $p_j \times p_j$, $j = 1, 2$, $i = 1, \dots, p$. Next, let A_{22}

denote the $p_2 \times p_2 \times p_2$ subarray of A with faces A_{i22}, $i = p_1 + 1, \ldots, p$. Then

$$[P_{U_2}][\tilde{W}_{22}] = [U_2][A_{22}]$$

and

$$\Gamma_s^{\tau}(\theta_2) = \max_D \|d^T A_{22} d\| \sqrt{p_2}\, s. \tag{20}$$

In this form it is clear that the maximum parameter-effects curvature for the subset problem depends only on the behavior of the $\tilde{\phi}_2$ parameter curves. The elements of A_{22} can be used to understand the behavior of these parameter curves in terms of arcing, compansion, fanning, and torsion, as described in Bates and Watts (1981).

Term (17c) is clearly in $C(V_1)$ and is thus orthogonal to the submanifold tangent plane. This term then reflects the intrinsic curvature of h at $\hat{\theta}_2$ so that the maximum intrinsic curvature can be defined as

$$\Gamma_s^{\eta}(\theta_2) = \max_D \|[d^T U_2^T][\tilde{W}_{12}] d\| 2\sqrt{p_2}\, s. \tag{21}$$

Note that (21) contains the extra factor 2, corresponding to the absence of the factor $\frac{1}{2}$ in (17c).

This curvature can also be expressed in terms of a subarray of A. Let A_{12} denote the $p_2 \times p_1 \times p_2$ subarray of A that has faces A_{i12}, $i = p_1 + 1, \ldots, p$. Then $A_{12} = [U_2^T][\tilde{W}_{12}]$ and

$$\begin{aligned} \Gamma_s^{\eta}(\theta_2) &= \max_D \|[d^T][A_{12}] d\| 2\sqrt{p_2}\, s \\ &= \max_D \Big\| \sum_{j=p_1+1}^{p} d_j A_{j12} d \Big\| 2\sqrt{p_2}\, s, \end{aligned} \tag{22}$$

where d_j is the $(j - p_1)$th element of d. Interestingly, the intrinsic curvature for the subset problem depends only on fanning and torsion components of A; compansion and arcing play no role in the determination of Γ_s^{η}. The fanning and torsion terms of A depend in part on how the columns of V are ordered. Since we have assumed that the last p_2 columns of V correspond to θ_2, it is the fanning and torsion with respect to this ordering that are important.

If both Γ_s^{η} and Γ_s^{τ} are sufficiently small, the likelihood and linear confidence regions for θ_2 will be similar; otherwise we can expect these regions to be dissimilar. Following Bates and Watts (1980), $c = (F_{\alpha}(p_2, n - p))^{-1/2}$ can be used as a rough guide for judging the size of these curvatures. As noted earlier, our experience indicates that curvatures must be substantially less than c to insure close agreement between linear and likelihood regions. This will be illustrated in Sections 3.3 and 4.

Finally, we combine the intrinsic and parameter-effects components of (17) to define the total curvature $\Gamma_s(\theta_2)$ of h at $\hat{\theta}_2$ as

$$\Gamma_s(\theta_2) = \sqrt{p_2}\, s \max_D \left\{ \|d^T A_{22} d\|^2 + 4\|[d^T][A_{12}] d\|^2 \right\}^{1/2}. \tag{23}$$

As will be demonstrated in the next subsection, the total subset curvature Γ_s

may be more relevant than both Γ_s^η and Γ_s^τ. For example, it is possible to have $\Gamma_s^\eta < c$ and $\Gamma_s^\tau < c$ while $\Gamma_s > c$. In such situations Γ_s^τ and Γ_s^η may incorrectly indicate that the tangent plane approximation is adequate, while Γ_s correctly indicates otherwise. Since Γ_s lies in the interval $[\max(\Gamma_s^\eta, \Gamma_s^\tau), \Gamma_s^\eta + \Gamma_s^\tau]$, its exact computation may be unnecessary once Γ_s^η and Γ_s^η are known.

When the full parameter θ is of interest, we have $\theta_2 = \theta$ and $p_2 = p$. In this case, the subset intrinsic curvature (22) is zero, A_{22} is the Bates–Watts parameter-effects array, and both (20) and (23) represent the maximum parameter-effects curvature for θ.

The main conclusion of this section is that the unscaled parameter-effects curvature array A for the full parameter contains all necessary information for evaluating the adequacy of linear confidence regions for certain subsets of θ. For example, if the last parameter θ_p is of interest, then $\Gamma_s^\tau(\theta_p)$ is simply $s|a_{ppp}|$ where a_{ijk} is the (j, k)th element of the ith face of A. Similarly,

$$\Gamma_s^\eta(\theta_p) = 2s\left(\sum_{i=1}^{p-1} a_{pip}^2\right)^{1/2}. \tag{24}$$

Thus, for a single parameter, compansion and fanning are the only effects that are relevant to an assessment of the agreement between likelihood and linear confidence regions.

3.2. *Computation.* Recall that the developments of this section are based on the assumption that the last p_2 columns of V correspond to the parameters of interest. This assumption is necessary to maintain the collective identity of θ_2 as indicated in (15). This implies that the ordering of the columns of V is critical and consequently θ_p is the only single parameter for which curvatures can be constructed from a given parameter-effects array A. The A array for other orderings can be constructed by permuting the columns of V and beginning again, of course.

Alternatively, a computationally more efficient method for obtaining the A array in a rotated coordinate system can be constructed as follows. Let $\phi_z = Z\phi$ where Z is a selected $p \times p$ permutation matrix. In what follows, the subscript z added to any quantity indicates evaluation in the coordinates ϕ_z. Clearly, $V_z = VZ^T = URZ^T$. Let U^{*T} be an orthogonal matrix such that $R^* = U^*RZ^T$ is upper triangular. Since the QR factorization of V_z is unique, it follows that $V_z = U_zR_z$, where $U_z = UU^{*T}$ and $R_z = R^*$. Using this structure it is not difficult to verify that

$$A_z = [U^*][U^*AU^*T]. \tag{25}$$

Thus, to find A_z, the parameter-effects curvature array for the rotated coordinates ϕ_z, we need only the $p \times p$ matrix U^* to diagonalize RZ^T. A single call to LINPACK (Dongarra et al. (1979)) routine SCHEX produces R^*, $[U^{*T}][A]$, and the information necessary to construct U^*.

3.3. *Fieller–Creasy again.* To apply Γ_s^η and Γ_s^τ in the Fieller–Creasy problem when θ_2 is the subset of interest, we require only the $2 \times 2 \times 2$ parameter-effects curvature array A for

$$V = (x + \hat{\theta}_2(b - x), \hat{\theta}_1(b - x)),$$

where x is the $n \times 1$ vector with elements x_i as defined following (7) and b is an $n \times 1$ vector of ones. The faces A_i of A are (Cook and Witmer (1985))

$$A_1 = \frac{\hat{\theta}_2\sqrt{2}}{\hat{\theta}_1\{n(1 + \hat{\theta}_2)\}^{1/2}} \begin{pmatrix} 0 & 1 \\ 1 & -2\hat{\theta}_2 \end{pmatrix} \tag{26}$$

and

$$A_2 = \frac{A_1}{\hat{\theta}_2}. \tag{27}$$

Reading directly from this array we have

$$\Gamma_s^{\tau}(\theta_2) = \sigma|a_{222}| = \frac{2^{3/2}\sigma}{\sqrt{n}\,|\hat{\theta}_1|} \frac{|\hat{\theta}_2|}{(1 + \hat{\theta}_2^2)^{1/2}} \tag{28}$$

and

$$\Gamma_s^{\eta}(\theta_2) = 2\sigma|a_{212}| = \frac{2^{2/3}\sigma}{\sqrt{n}\,|\hat{\theta}_1|} \frac{1}{(1 + \hat{\theta}_2^2)^{1/2}}. \tag{29}$$

Recall that we are assuming σ to be known in this example, so that the guide for assessing the magnitudes of Γ_s^{η} and Γ_s^{τ} is $c = (\chi(\alpha; 1))^{-1/2}$.

From (28) we see that $\Gamma_s^{\tau}(\theta_2)$ will be zero only if $\hat{\theta}_2 = 0$; in this case $\Gamma_s^{\eta}(\theta_2) = 2^{3/2}\sigma/\sqrt{n}\,|\hat{\theta}_1| < c$ or, equivalently, $r = 2\sigma^2\chi(\alpha; 1)/n\hat{\theta}_1^2 < \frac{1}{4}$ is necessary for the subset intrinsic curvature to be less than the guide. Further $r < \frac{1}{4}$ is a sufficient—although not necessary—condition for both $\Gamma_s^{\eta}(\theta_2)$ and $\Gamma_s^{\tau}(\theta_2)$ to be less than c when $\hat{\theta}_2$ is arbitrary.

Next, using (23) it follows that the total subset curvature is simply

$$\Gamma_s(\theta_2) = 2^{3/2}\sigma\big/\sqrt{n}\,|\hat{\theta}_1|, \tag{30}$$

and thus $\Gamma_s(\theta_2) < c$ if and only if $r < \frac{1}{4}$. When $r > 1$, the likelihood region for θ_2 will be either the complement of an interval or else the entire real line; otherwise, this region will be the interval given in (9). In this example, the total subset curvature recovers the critical quantity r as introduced in Section 2, and the condition $\Gamma_s < c$ insures that (10) will in fact be approximating a likelihood interval rather than some dissimilar region. This condition also provides for an added measure of agreement between these intervals since it is equivalent to $r < \frac{1}{4}$ rather than simply $r < 1$.

Applying (20) and (22) when θ_1 is the subset of interest gives $\Gamma_s^{\eta}(\theta_1) = \Gamma_s^{\tau}(\theta_1) = 0$, as expected. Notice that this conclusion cannot be obtained by inspecting the A array given in (26) and (27). As mentioned previously, different subsets in general require different orderings for the columns of V and thus different coordinates. This is the case here.

Finally, we consider the special case characterized by $(\theta_1, \theta_2) = (3, 0)$ and $r = 0.428$. These conditions correspond to $n = 2\sigma^2$. From (8), $\Gamma^{\tau} = 0.33 < 0.41 =$

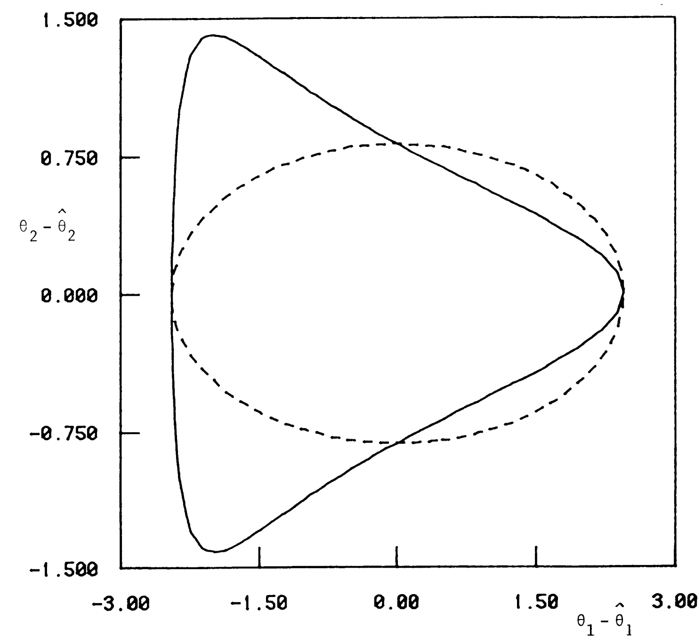

FIG. 1. *95% confidence regions for* (θ_1, θ_2) *from the Fieller–Creasy model* (7): $(\hat{\theta}_1, \hat{\theta}_2) = (3, 0)$. *Likelihood region* ——; *linear region* ----.

$\chi^{-1/2}(0.05; 2) = c$. From Figure 1 (Cook and Witmer (1985)), we see that the likelihood region, whose level is exact in this case, does not seem to be adequately approximated by the linear region for small values of θ_1.

Further insight into this problem can be gained by inspecting marginal likelihood regions for θ_1 and θ_2. Generally, marginal regions for subsets can be obtained by projecting all points in the joint region onto the appropriate subspaces. The levels of these marginal regions will be somewhat larger than that for the joint region. In Figure 1, projecting all points onto the θ_i axis will yield a $\Pr(\chi(1) < \chi(0.05; 2)) = 98.6\%$ interval for θ_i, $i = 1, 2$. Further, projecting the regions in Figure 1 onto the θ_1 axis shows that the likelihood and linear intervals for θ_1 will be identical, as expected. By contrast, projecting onto the θ_2 axis shows that the resulting 98.6% likelihood interval will be about 60% longer than the corresponding tangent plane interval. This dissimilarity is clearly indicated by $\Gamma_s^{\eta}(\theta_2) = 0.67 > 0.41 = \chi^{-1/2}(0.014; 1)$.

Our experience leads to the following heuristic characterization of the problem described in the previous paragraph. Consider a p_2-dimensional subset θ_2 with guide $c_2 = (F_\alpha(p_2, n - p))^{-1/2}$ and partition $\theta_2^T = (\theta_{21}^T, \theta_{22}^T)$, where θ_{2i} is $p_{2i} \times 1$, $i = 1, 2$. The guide corresponding to the confidence region for θ_{2i} obtained by projecting the selected $1 - \alpha$ region for θ_2 is simply $c_{2i} = c_2(p_{2i}/p_2)^{1/2}$, $i = 1, 2$. When the subset curvatures for θ_{21} are large relative to c_{21} and the subset curvatures for θ_{22} are near zero, it can happen that the curvatures for θ_2

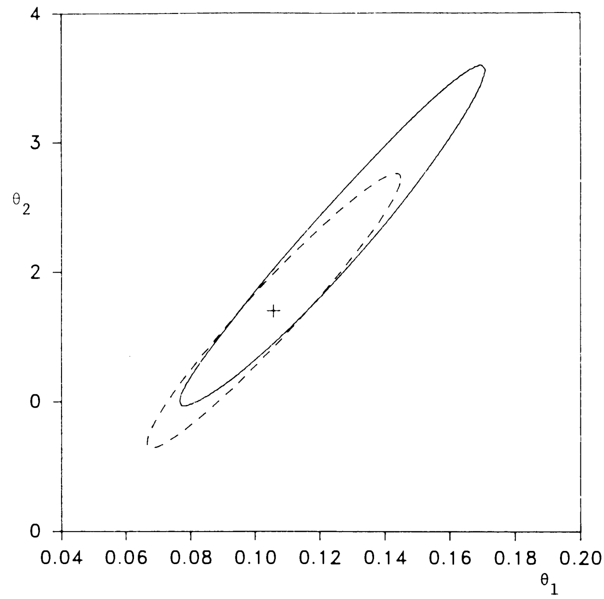

FIG. 2. *Nominal* 87% *bivariate confidence regions with* 95% *marginal regions for* (θ_1, θ_2) *from model* (31) *and the Bates–Watts* (1980) *data. Likelihood* ——; *linear* ----.

are moderate. In such cases the curvatures for θ_2 can provide a misleading indication that the tangent plane and likelihood regions for θ_2 are in acceptable agreement. As hinted above, this problem might be overcome by requiring that all subsets θ_{2i} of θ_2 have curvatures less than the respective guides c_{2i}. When $\theta_2 = \theta$ this added requirement seems to represent a useful fine tuning of the basic Bates–Watts methodology.

4. Illustrations. In this section we present several numerical examples to illustrate selected results of the previous sections.

For the first example we use the Michaelis–Menton model,

$$f_i = \theta_1 x_i/(\theta_2 + x_i), \tag{31}$$

in combination with the 12 observations reported in Bates and Watts (1980). Figure 2 gives 87% linear (broken contour) and likelihood (solid contour) confidence regions for (θ_1, θ_2). Here and in the following examples the levels of displayed bivariate confidence regions are chosen so that the corresponding univariate marginal regions have a nominal 95% coverage rate. It seems clear from Figure 2 that the linear region for (θ_1, θ_2) is not an adequate approximation of the likelihood region, although directly interpreting the Bates–Watts guide as the cutoff value would lead to the opposite conclusion, since $\Gamma^\tau = 0.598 < c = 0.635$. (The value $\Gamma^\tau = 0.771$ reported by Bates and Watts is based on using replicate error with 6 d.f. to estimate σ^2; the value $\Gamma^\tau = 0.598$ reported here is based on using s^2 with 10 d.f.) The subset curvatures for θ_1 and θ_2 are listed in

TABLE 1
Subset curvatures

Model / data	Parameter subset	Γ_s^τ	Γ_s^η	Γ_s	C
(31)	θ_1	0.330	0.183	0.377	0.449
Bates and Watts	θ_2	0.393	0.089	0.403	0.449
(31)	θ_1	0.014	0.025	0.029	0.389
Michaelis and Menten	θ_2	0.050	0.019	0.053	0.389
(32)	θ_1	0.165	0.180	0.244	0.484
Ratkowsky	θ_2	0.003	0.059	0.059	0.484
	θ_3	0.153	0.132	0.203	0.484
	(θ_1, θ_3)	1.07	0.008	1.07	0.542
	(θ_2, θ_3)	0.518	0	0.518	0.542
(32)	θ_1	1.75	0.190	1.76	0.408
Hunt	θ_2	1.80	0.256	1.82	0.408
	θ_3	0.018	0.094	0.095	0.408
	(θ_2, θ_3)	36.4	0	36.4	0.441

Table 1; the corresponding guide is $c = [F_{0.05}(1, 10)]^{-1/2} = 0.449$. Again, the curvatures are less than the guide while the marginal likelihood regions do not seem to be well represented by the corresponding linear regions. This reinforces our previous remark that curvatures must be substantially less than c to insure close agreement. With this interpretation we see that all curvatures successfully indicate the dissimilarity between the various likelihood and linear regions in Figure 2.

Figure 3 gives 88% likelihood and linear regions for (θ_1, θ_2) obtained by using model (31) and the seven observations reported by Michaelis and Menten (1913). For these data $\Gamma^\tau = 0.079$. This value and the subset curvatures reported in Table 1 are relatively small, indicating reasonable agreement between the regions displayed in Figure 3.

For the three-parameter asymptotic regression model

$$f_i = \theta_1 + \theta_2 \exp(\theta_3 x_i) \tag{32}$$

and the 27 observations reported in Ratkowsky (1983, page 101, data set 1), we obtain $\Gamma^\tau = 1.53$. The corresponding guide is $c = [F_{0.05}(3, 24)]^{-1/2} = 0.58$. This suggests that the 95% likelihood region for $\theta^T = (\theta_1, \theta_2, \theta_3)$ cannot be adequately approximated by the corresponding linear region. The subset curvatures for selected subsets of θ are listed in Table 1. From these curvatures alone we would reach the following conclusions:

(1) The likelihood and linear regions for θ_2 are in very close agreement.
(2) The marginal regions for θ_1 and θ_3 will be noticeably different, but the agreement is probably adequate for most purposes.
(3) The usual 95% linear regions for (θ_1, θ_3) and (θ_2, θ_3) should be used for only very rough analyses, although lower level regions may be acceptable replacements for the corresponding likelihood regions.

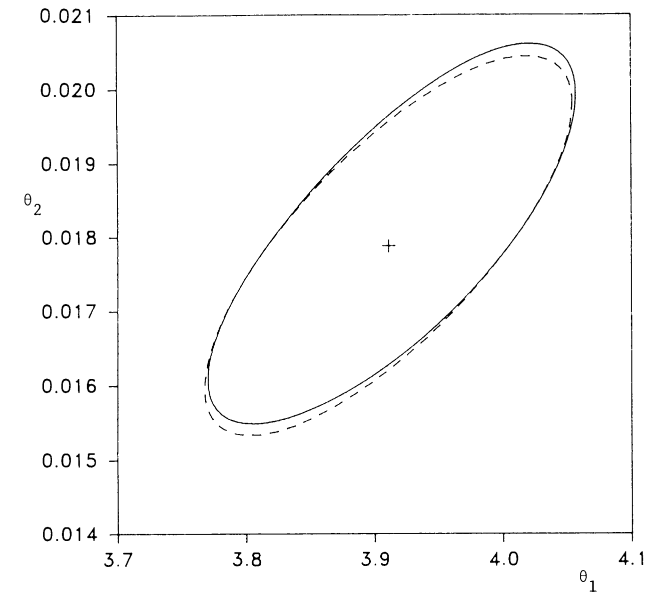

FIG. 3. *Nominal* 88% *bivariate regions with* 95% *marginal regions for* (θ_1, θ_2) *from model* (31) *and the Michaelis–Menten* (1913) *data. Likelihood* ——; *linear* ----.

These conclusions are supported by the 86% regions for (θ_2, θ_3) and (θ_1, θ_3) shown in Figures 4 and 5, respectively. The value $\Gamma_s^{\eta}(\theta_2, \theta_3) = 0$ reported in Table 1 for model (32) will always occur since $\tilde{W}_{12} = 0$ for this model.

For our final example we again use the asymptotic regression model (32), this time in combination with the nine observations reported by Hunt (1970). Subset curvatures for four parameter subsets are listed in Table 1. The subset curvature for θ_3 is small, indicating good agreement between the corresponding likelihood and linear regions. The subset curvatures for the remaining subsets, particularly (θ_2, θ_3), are large.

The 87% likelihood and linear confidence regions for (θ_2, θ_3) are given in Figure 6. The large total curvature, $\Gamma_s(\theta_2, \theta_3) = 36.4$, correctly indicates that use of the linear region as an approximation of the likelihood region would be a disaster for this pair of parameters. In fairness, however, it should be recalled that the approximations used to derive the subset curvatures are local so that $\Gamma_s(\theta_1, \theta_2)$ is responding primarily to the disagreement between the linear region and the portion of the likelihood region that contains $\hat{\theta}$. Similar comments apply when only θ_2 is of interest.

From Figure 6, there is reasonable agreement between the linear and likelihood regions for θ_3, as indicated by the small curvature $\Gamma_s(\theta_3) = 0.095$. It can be argued justifiably, however, that this correct indication from the curvature is largely fortuitous since the curvatures do not recognize the contribution of the smaller piece of the likelihood region for (θ_2, θ_3) to the likelihood region for θ_3. Under this argument, the subset curvature measure for θ_3 has failed to indicate

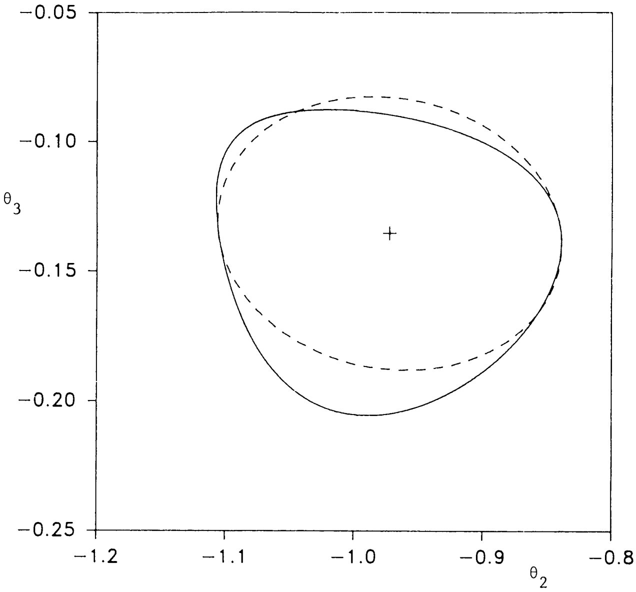

FIG. 4. *Nominal* 86% *bivariate regions with* 95% *marginal regions for* (θ_2, θ_3) *from model* (32) *and the Ratkowsky* (1983) *data. Likelihood* ——; *linear* ----.

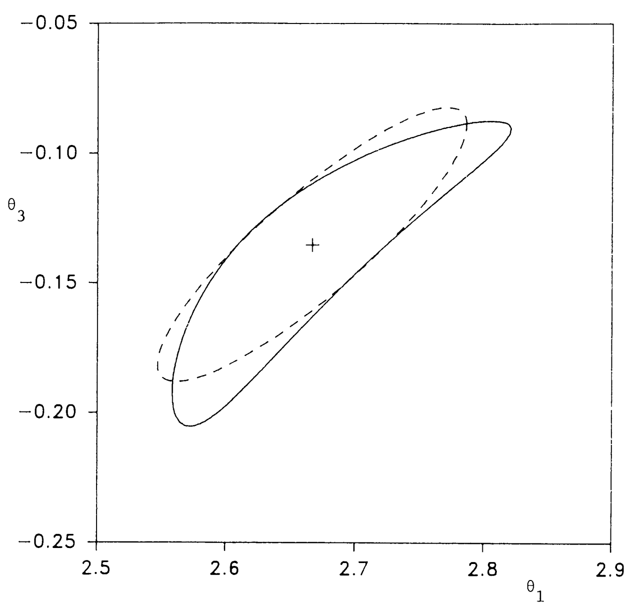

FIG. 5. *Nominal* 86% *bivariate regions with* 95% *marginal regions for* (θ_1, θ_3) *from model* (32) *and the Ratkowsky* (1983) *data. Likelihood* ——; *linear* ----.

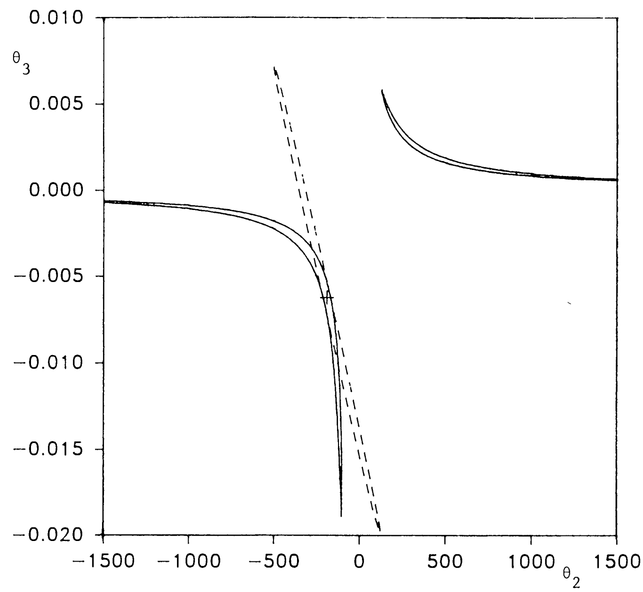

FIG. 6. *Nominal* 87% *bivariate regions with* 95% *marginal regions for* (θ_2, θ_3) *from model* (32) *and the Hunt* (1970) *data. Likelihood* ——; *linear* ----.

the dissimilarity between the tangent plane region for θ_3 and the likelihood region $(-0.0191, 0)$ obtained by using only the larger subregion that contains $\hat{\theta}$.

The reason that the curvatures give some inappropriate indications in this final example is that both the linear and quadratic approximations to the model function fail. This failure is evident from a very low R^2 from the regression used by Goldberg, Bates and Watts (1983) to obtain numerical curvatures, and from related measures of "lack of quadraticity" explored by the present authors. In cases where the quadratic approximation to the model function is poor, curvature measures based on that approximation may not be meaningful.

Nevertheless, these subset curvature measures represent an important advance in our understanding of nonlinear models, and provide useful information about the adequacy of the linear approximation when the quadratic approximation is appropriate. Further work is needed on methods of identifying cases where the quadratic approximation may fail.

5. Conclusions. The subset curvatures developed in this paper appear to be reliable indicators of the adequacy of linear confidence regions for most nonlinear models. In particular, the curvature for a single parameter is a useful tool for assessing the agreement between standard large sample confidence intervals and corresponding marginal likelihood regions. This ability to deal with subsets greatly extends the usefulness of the Bates–Watts methodology.

Because the original Bates–Watts framework applies only to the complete parameter vector, guidelines developed in that framework can be misleading when the adequacy of the linear approximation is very different for different subsets. To ensure good agreement between the tangent plane and likelihood regions, the maximum curvature must be considerably smaller than the Bates–Watts guide. However, this criterion can be too stringent for certain parameter subsets if the whole-parameter curvature Γ^{τ} is used. By contrast, the subset curvature describes the shape of the likelihood region in the parameter subspace of interest. Thus, the subset curvature is more directly relevant to the linearization adequacy question and, based on the examples described above, is evidently more accurate.

The practical usefulness of the methods described here depends, in part, on their ease of implementation. The subset curvatures for any selected subset can be computed directly from the Bates–Watts parameter-effects curvature array. This array can be obtained either analytically (Bates and Watts (1980)) or numerically by using the procedure given in Goldberg, Bates and Watts (1983).

The usefulness of the subset curvatures depends also on the restriction that the intrinsic curvature of F at $\hat{\theta}$ is small. This restriction is not of great practical importance since it has been found to hold in most cases. Nevertheless, a unified approach that incorporates the intrinsic curvature component might offer further insight in some situations.

Another area for further research is the development of measures that indicate when the subset curvatures themselves may be unreliable due to the failure of the second-order approximation to the model function. While the possibility of such failure is of concern, the class of models adequately described by a quadratic function is considerably larger than the class for which the linear approximation alone is adequate.

APPENDIX

Derivation of equation (17). To develop equation (17) from equation (12), we first require explicit expressions for Δ_1 and Δ_2.

A.1. Δ_1 *and* Δ_2. Let $\ddot{L}$ and $\dddot{L}$ denote the $p \times p$ matrix and $p \times p \times p$ array of second and third partial derivatives of the log likelihood L with respect to the elements of θ, respectively. Let g_a denote the ath component of g as defined near (11) and partition $\ddot{L}$ as

$$\ddot{L} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix},$$

where L_{jj} is $p_j \times p_j$, $j = 1, 2$.

Since g maximizes $L(\theta_1, \theta_2)$ for each fixed value of θ_2,

$$\frac{\partial L(g(\theta_2), \theta_2)}{\partial g_a}\Big|_{g=g(\theta_2)} = 0 \tag{A.1}$$

for $a = 1, 2, \dots, p_1$ and all θ_2. This identity will be used as the basis for obtaining Δ_1 and Δ_2.

Differentiating both sides of (A.1) with respect to θ_2 and evaluating at $\hat{\theta}_2$ gives

$$(L_{11}, L_{12})\Delta_1 = 0.$$

Since the submatrix consisting of the last p_2 rows of Δ_1 is an identity matrix it follows that

$$\Delta_1 = \begin{pmatrix} -L_{11}^{-1}L_{12} \\ I \end{pmatrix}. \tag{A.2}$$

Let $e_i = y_i - f_i(\hat{\theta})$. The first term of

$$\ddot{L} = \left(\sum_{i=1}^{n} e_i W_i - V^T V \right) \Big/ \sigma^2$$

represents effective residual intrinsic curvature of F at $\hat{\theta}$ (Bates and Watts (1982)). Since this curvature is assumed to be negligible, $\ddot{L} = -V^T V/\sigma^2$ and, therefore,

$$\Delta_1 = \begin{pmatrix} -(V_1^T V_1)^{-1} V_1^T V_2 \\ I \end{pmatrix} = \begin{pmatrix} -R_{11}^{-1}R_{12} \\ I \end{pmatrix}, \tag{A.3}$$

where $V = (V_1, V_2)$ and R_{ij} is defined in (13).

An expression for Δ_2 can be obtained similarly by taking second partial derivatives of (A.1) with respect to θ_{2r} and θ_{2s}, $r, s = 1, 2, \ldots, p_2$. This yields

$$\sum_{b=1}^{p_1} L_{ab} \frac{\partial^2 \alpha_b}{\partial \theta_{2r}\, \partial \theta_{2s}} = -\sum_{b=1}^{p} \sum_{c=1}^{p} L_{abc} \frac{\partial \alpha_c}{\partial \theta_{2r}} \frac{\partial \alpha_b}{\partial \theta_{2s}}, \tag{A.4}$$

where L_{ab}, L_{abc}, and α_b denote the indicated elements of $\ddot{L}$, $\dddot{L}$, and $\alpha^T = (g^T(\theta_2), \theta_2^T)$, respectively, and $a = 1, 2, \ldots, p_1$. The component $\partial^2\alpha_b/\partial\theta_{2r}\,\partial\theta_{2s}$ is the (r, s)th element of the bth face Δ_{2b} of Δ_2. Since $\Delta_{2b} = 0$ for $b = p_1 + 1, \ldots, p$, the summation on the left of (A.4) need only range from 1 to p_1. Notice also that $\partial\alpha_c/\partial\theta_{2r}$ is simply the (c, r)th element of Δ_1. Expressing (A.4) in matrix notion and solving for Δ_2 gives

$$\Delta_2 = -\begin{bmatrix} L_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} [\Delta_1^T \dddot{L} \Delta_1]. \tag{A.5}$$

As indicated following (A.2), we will take $L_{11} = -V_1^T V_1/\sigma^2$.

In (A.5) and the following, brackets [ ][ ] indicate column multiplication as defined in Bates and Watts (1980). Generally, if A is an $a \times b$ matrix and B is a $b \times c \times d$ array, then the elements of the ith face C_i, $i = 1, \ldots, a$, of the $a \times c \times d$ array $C = [A][B]$ are $A_i^T B_{jk}$, $j = 1, 2, \ldots, c$, $k = 1, 2, \ldots, d$, where A_i^T is the ith row of A and B_{jk} is the jkth column of B. Further, if D and E are $n \times c$ and $d \times m$ matrices, respectively, then $DCE = D[A][B]E = [A][DBE]$. This property is used frequently in the following development.

To further evaluate Δ_2, we require the $p \times p \times p$ array $\dddot{L}$. Straightforward algebra will verify that

$$L_{abc} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(e_i f_i^{abc} - f_i^a f_i^{bc} - f_i^b f_i^{ac} - f_i^c f_i^{ab} \right).$$

Neglecting the first term of $\ddot{L}_{abc}$, which corresponds to effective residual intrinsic curvature, the ath face $\ddot{L}_a$ of $\ddot{L}$ is

$$\ddot{L}_a = -\frac{1}{\sigma^2}\{[b_a^T V^T][W] + V^T K_a + K_a^T V\}, \tag{A.6}$$

where b_a is the ath standard basis vector for R^p and $K_a = b_a^T W$ is the $n \times p$ matrix with W_{ac} as the cth column. Finally, it follows from (A.6) that for an arbitrary $p \times 1$ vector Z,

$$Z^T \ddot{L} Z = -\frac{1}{\sigma^2}\{Z^T[V^T][W]Z + 2[Z^T V^T][W]Z\}. \tag{A.7}$$

This form will be useful in later developments.

A.2. *Tangent plane, term* (17a). It follows immediately from (A.3) that

$$V\Delta_1 = P'_{V_1} V_2 = U_2 R_{22}, \tag{A.8}$$

where U_2 is defined following (16). Thus, the relevant tangent plane is the affine subspace $F(\hat{\theta}) + C(P'_{V_1} V_2)$. Transforming term (12a) according to (14) and (15) immediately gives term (17a).

A.3. *Parameter effects, term* (17b). From the form Δ_2 given by (A.5), it is clear that term (12c) is in $C(V_1)$ and is thus orthogonal to the θ_2-subspace tangent plane. The parameter-effects component of (12) must therefore come from term (12b).

The three-dimensional array W in (12b) can be decomposed into the sum of three arrays with orthogonal columns,

$$W = [P_V - P_{V_1}][W] + [P_{V_1}][W] + [P'_V][W]. \tag{A.9}$$

The first term in this decomposition contains the projections of columns of W onto $C(P'_{V_1} V_2)$ and thus it represents parameter-effects curvature for the subset problem. The second and third terms are intrinsic components for h and F, respectively. Since the intrinsic curvature of F at $\hat{\theta}$ is assumed to be negligible, the third term of (A.9) is set to zero. Addend (12b) can now be reexpressed as

$$\tfrac{1}{2}\phi_2^T \Delta_1^T W \Delta_1 \phi_2 = \tfrac{1}{2}\phi_2^T \Delta_1^T [P_V - P_{V_1}][W]\Delta_1 \phi_2 \tag{A.10a}$$

$$+ \tfrac{1}{2}\phi_2^T \Delta_1^T [P_{V_1}][W]\Delta_1 \phi_2. \tag{A.10b}$$

From (16) and (A.3) it follows that

$$\tilde{W}_{22} = R_{22}^{-T} \Delta_1^T W \Delta_1 R_{22}^{-1}.$$

Using this in combination with (15) and (A.8) to transform the coordinates in term (A.10a) gives term (17b).

A.4. *Intrinsic curvature, term* (17c). In the expansion of h given in (12), we still have the sum of terms (12c) and (A.10b) to deal with. We first consider (12c).

Using (A.5) and (A.7) with $Z = \Delta_1\phi_2$ we have

$$\begin{aligned}\tfrac{1}{2}V(\phi_2^T\Delta_2\phi_2) &= \tfrac{1}{2}\sigma^2 M(\phi_1^T\Delta_1^T\ddot{L}\Delta_1\phi_2)\\ &= -\tfrac{1}{2}M\{\phi_2^T\Delta_1^T[V^T][W]\Delta_1\phi_2\}\\ &\quad -M[\phi_2^T\Delta_1^T V^T][W]\Delta_1\phi_2,\end{aligned} \tag{A.11}$$

where $M = (V_1(V_1^TV_1)^{-1}, 0)$. The first term of (A.11) is exactly the negative of term (A.10b) so that in an obvious notation

$$\begin{aligned}(12c) + (A.10b) &= -M[\phi_2^T\Delta_1^TV^T][W]\Delta_1\phi_2\\ &= -[\phi_2^T\Delta_1^TV^T][MW\Delta_1]\phi_2.\end{aligned} \tag{A.12}$$

From (16) and the definition of $\tilde{W}$, it can be shown that

$$MW\Delta_1 = U_1\tilde{W}_{12}R_{22}.$$

Finally, using this relationship with (A.8) and (15) to transform the coordinates in (A.12) we obtain term (17c).

Acknowledgments. Data computations for this work were performed on the research computer at the University of Wisconsin Statistics Department. The authors thank Douglas Bates for access to software and data libraries used for some examples presented, and for suggestions on computational methods. Philip Hougaard, C. L. Tsai, and the referees gave many useful comments on an earlier version of this manuscript.

REFERENCES

AMARI, S.-I. (1982). Differential geometry of curved exponential families—curvatures and information loss. *Ann. Statist.* **10** 357–385.

BATES, D. M., HAMILTON, D. C. and WATTS, D. G. (1983). Calculation of intrinsic and parameter-effects curvatures for nonlinear regression models. *Comm. Statist. B—Simulation Comput.* **12** 469–477.

BATES, D. M. and WATTS, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *J. Roy. Statist. Soc. Ser. B* **42** 1–25.

BATES, D. M. and WATTS, D. G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Ann. Statist.* **9** 1152–1167.

COOK, R. D. and WITMER, J. A. (1985). A note on parameter-effects curvature. *J. Amer. Statist. Assoc.* **80** 872–878.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

DONGARRA, J. J., MOLER, C. B., BUNCH, J. R. and STEWART, G. W. (1979). *LINPACK Users' Guide*. SIAM, Philadelphia.

DRAPER, N. R. and SMITH, H. (1981). *Applied Regression Analysis*. Wiley, New York.

EISENHART, L. P. (1964). *Riemannian Geometry*. Princeton Univ. Press.

GOLDBERG, M. L., BATES, D. M. and WATTS, D. G. (1984). Curvature measures avoiding second derivatives. Technical Report No. 739, Department of Statistics, Univ. of Wisconsin.

HAMILTON, D. C., WATTS, D. G. and BATES, D. M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *Ann. Statist.* **10** 386–393.

HUNT, W. F. (1970). The influence of leaf death on the rate of accumulation of green herbage during pasture. *J. Appl. Ecology* **7** 41–50.

KASS, R. E. (1984). Canonical parameterizations and zero parameter-effects curvature. *J. Roy. Statist. Soc. Ser. B* **46** 86–92.

LINSSEN, H. N. (1980). Discussion of the paper by Dr. Bates and Professor Watts. *J. Roy. Statist. Soc. Ser. B* **42** 21.

MICHAELIS, L. and MENTEN, M. L. (1913). Kinetik der Invertinwirkung. *Biochemische Z.* **49** 333.

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

DEPARTMENT OF APPLIED STATISTICS
UNIVERSITY OF MINNESOTA
1994 BUFORD AVENUE
ST. PAUL, MINNESOTA 55108

EPA OFFICE OF WATER REGULATION
AND STANDARDS, WH586
401 M STREET, S.W.
WASHINGTON, D.C. 20460