J. R. Statist. Soc. B (1999)
61, *Part* 1, *pp.* 51–61

Conformal normal curvature and assessment of local influence

Wai-Yin Poon

Chinese University of Hong Kong, Hong Kong

and Yat Sun Poon

University of California, Riverside, USA

[Received January 1997. Final revision March 1998]

Summary. In 1986, R. D. Cook proposed differential geometry to assess local influence of minor perturbations of statistical models. We construct a conformally invariant curvature, the conformal normal curvature, for the same purpose. This curvature provides a measure of local influence ranging from 0 to 1, with objective bench-marks to judge largeness. We study various approaches to using the conformal normal curvature and the relationships between these approaches.

Keywords: Aggregate contribution; Conformal normal curvature; Influence graph; Linear regression; Normal curvature

1. Introduction

A general method was introduced by Cook (1986) for assessing the influence of local departures from the assumptions underlying a statistical model. Using different perturbation schemes, this local influence approach has been applied successfully in various analyses. Typical examples are the applications of the approach to the diagnostics and influence analyses in mixed model analysis of variance (Beckman *et al.*, 1987), in regression transformations (Lawrance, 1988), in generalized linear modelling (Thomas and Cook, 1990), in non-linear regression (St Laurent and Cook, 1993), in structural equation models (Lee and Wang, 1996) and in principal components analysis (Shi, 1997). Some effort has also been devoted to relating Cook's approach to other work in the literature. For example, Farebrother (1992) pointed out that there is a close relationship between the condition number that is traditionally used by numerical analysts and the measure of influence in the local influence approach.

The method is powerful because it is relatively simple to use. It utilizes certain ideas from differential geometry to assess the behaviour of the likelihood displacement function. More specifically, the normal curvature along a direction l at the optimal point of the function, C_l, is computed. Large values of C_l indicate strong local influence. The directions which give large normal curvatures carry important information about how to perturb the postulated model. In particular, the direction corresponding to the maximum of C_l is of special interest. More details are summarized in Section 2.1.

Address for correspondence: Wai-Yin Poon, Department of Statistics, Chinese University of Hong Kong, Shatin, Hong Kong.
E-mail: wypoon@hp735.sta.cuhk.edu.hk

 1369–7412/99/61051

Although the method has been demonstrated to be very useful, several issues were raised in the discussion of Cook (1986) (see, for example, Beckman (1986), Lawrance (1986), Loynes (1986) and Prescott (1986)). The normal curvature may take any value and is not invariant under a uniform change of scale. As a result, there is no objective criterion to judge the largeness of normal curvatures and the relative size of the components of the directions corresponding to large normal curvatures. Although in some situations the lack of objective criteria does not cause significant difficulties, an objective and yet flexible criterion is surely welcome and could enhance the applicability of the method. Furthermore, the method requires the computation of eigenvalues and eigenvectors. This hinders its popularity in problems with large dimensions and leads to intractability in theoretical developments.

To facilitate the application and interpretation of the local influence approach, it is necessary to address these issues. In this respect, there are the contributions by Schall and Dunne (1992). They demonstrated the close relationship between the concepts of parameter collinearity and local influence in regression diagnostics, and they introduced a modification of Cook's normal curvature, named the scaled curvature. This scaled curvature is invariant over reparameterizations of the perturbation scheme and has an upper bound of 1. However, the other issues remain unresolved. In fact, unlike Cook's normal curvature, the scaled curvature does not have a clear geometric basis, and only in special circumstances are the two curvatures equivalent diagnostics.

The purpose of this paper is to develop methods which address these difficulties. With a clear geometrical basis, a measure, named the conformal normal curvature, which is a one-to-one function of the normal curvature, and assumes values in the interval [0, 1], is constructed. One of the many properties of the conformal normal curvature is its conformal invariance (Kobayashi, 1972). On the basis of this and its other properties, objective bench-marks to judge largeness are provided. Using the conformal normal curvature, an aggregate measure for each basic perturbation vector of the perturbation space is constructed. Different numbers of eigenvectors can be used in the aggregation. When all eigenvectors are considered, the aggregate contribution for each basic perturbation vector is equivalent to its conformal normal curvature. A perturbation vector is basic if it is in the direction of one and only one perturbation parameter. It turns out that the conformal normal curvature of a basic perturbation vector is an effective measure of local influence, and the computation of eigenvectors is no longer necessary.

In the next section, we construct the conformal normal curvature and examine its basic properties. In Section 3, we discuss various approaches and the advantages of using the conformal normal curvature to assess local influence. In Section 4, an example in linear regression is presented. The paper is concluded with a discussion in Section 5.

2. Conformal normal curvature

2.1. Influence graph and normal curvature

Let $L(\theta)$ denote the log-likelihood for a postulated model, where θ is a $p \times 1$ vector of unknown parameters. Let $L(\theta|\omega)$ be the log-likelihood corresponding to the perturbed model for a given ω, where $\omega^{\mathrm{T}} = (\omega_1, \ldots, \omega_n)$ is an $n \times 1$ vector in Ω of R^n, and Ω represents the set of relevant perturbations (n does not necessarily represent the sample size). It is assumed that there is an ω_0 such that $L(\theta|\omega_0) = L(\theta)$ for all θ. Let $\hat{\theta}$ and $\hat{\theta}_\omega$ be the maximum likelihood estimators of θ under $L(\theta|\omega_0)$ and $L(\theta|\omega)$ respectively. The likelihood displacement function (Cook, 1986) is given by

$$f(\omega) = 2\{L(\hat{\theta}|\omega_0) - L(\hat{\theta}_\omega|\omega_0)\} \tag{2.1}$$

and can be used to assess the local influence of a minor perturbation. Let

$$\alpha(\omega) = (\omega^{\mathrm{T}}, f(\omega))^{\mathrm{T}} \tag{2.2}$$

be the influence graph formed by the values of the $(n+1) \times 1$ vector. Cook (1986) proposed the normal curvature to assess the local influence of the perturbation.

Geometrically, the normal curvature is a function of the first fundamental form $\mathbf{I}$ (Thorpe, 1979) and the second fundamental form Π (Thorpe, 1979) of the graph of the likelihood displacement function f. They are symmetric matrices represented by

$$\begin{aligned} \mathbf{I}_{ij} &= \delta_{ij} + \frac{\partial f}{\partial \omega_i}\frac{\partial f}{\partial \omega_j}, \\ \Pi_{ij} &= \frac{1}{(1+|\nabla_f|^2)^{1/2}}\frac{\partial^2 f}{\partial \omega_i \partial \omega_j}, \end{aligned} \tag{2.3}$$

where δ_{ij} equals 1 when $i = j$ and is 0 otherwise, and $|\nabla_f|$ represents the norm of the gradient vector of f. These two forms are evaluated on vectors v and w by $\mathbf{I}(v, w) = v^{\mathrm{T}}\mathbf{I}w$ and $\Pi(v, w) = v^{\mathrm{T}}\Pi w$. A straight line in Ω passing through ω_0 is defined by $\omega(a) = \omega_0 + al$, where a is in R^1 and ω_0 and l are fixed column vectors in R^n. Then the normal curvature of the graph α in a direction l at the point ω_0 is

$$C_l = C(l, l) = \frac{\Pi(l, l)}{\mathbf{I}(l, l)} = \frac{l^{\mathrm{T}} H_f l}{l^{\mathrm{T}}(I_n + \nabla_f \nabla_f^{\mathrm{T}})l(1+|\nabla_f|^2)^{1/2}}\Big|_{\omega=\omega_0}, \tag{2.4}$$

where I_n is the $n \times n$ identity matrix and

$$H_f = \left(\frac{\partial^2 f}{\partial \omega_i \partial \omega_j}\right)$$

is the Hessian matrix.

We are primarily interested in the curvature at a critical point on the graph. When ω_0 is a critical point, then $\nabla_f(\omega_0) = 0$. If we choose l to be such that $l^{\mathrm{T}}l = 1$, then equation (2.4) is reduced to

$$C_l = l^{\mathrm{T}} H_f l|_{\omega=\omega_0}. \tag{2.5}$$

Cook (1986) proposed to use the normal curvature to study characteristics of influence graphs. According to equation (2.5), he further deduced that

$$C_l = -2(l^{\mathrm{T}}\ddot{F}l)|_{\omega=\omega_0}, \tag{2.6}$$

where $\ddot{F}$ is the $n \times n$ matrix with elements $\partial^2 L(\hat{\theta}_\omega)/\partial\omega_i\partial\omega_j$. Let Δ be the $p \times n$ matrix with elements $\Delta_{ij} = \partial^2 L(\theta|\omega)/\partial\theta_i\partial\omega_j$ and $\ddot{L}$ be the $p \times p$ matrix with elements

$$\ddot{L}_{ij} = \frac{\partial^2 L(\theta|\omega)}{\partial\theta_i\partial\theta_j} \tag{2.7}$$

evaluated at $\theta = \hat{\theta}$ and $\omega = \omega_0$. Then equation (2.6) can be written as (Cook, 1986)

$$C_l = -2\{l^{\mathrm{T}}\Delta^{\mathrm{T}}(\ddot{L})^{-1}\Delta l\}|_{\theta=\hat{\theta},\,\omega=\omega_0}. \tag{2.8}$$

The matrix $-\ddot{F}$ is positive semidefinite at ω_0 because the likelihood displacement function achieves its minimum at this point. Let $C_{\max} = \max_l(C_l)$ which corresponds to the maximum eigenvalue of $-\ddot{F}$, and let $\mathbf{e}_{\max}$ be the eigenvector for $C_{\max}$. Cook (1986) suggested that a large value of $C_{\max}$ is an indication of a serious local problem, and if the ith element in $\mathbf{e}_{\max}$ is relatively large special attention should be paid to the element being perturbed by ω_i. Although this approach has been demonstrated to be very useful, several issues that were raised in the discussion of Cook (1986) and mentioned in Section 1 remain unresolved. To address these issues, we introduce the conformal normal curvature.

2.2. Conformal normal curvature and its properties

Definition 1. The conformal normal curvature at a point ω_0 of a graph α in the direction l is

$$B_l = \frac{\Pi(l, l)}{\mathbf{I}(l, l)\{\mathrm{tr}(\Pi^2)\}^{1/2}}\Big|_{\omega=\omega_0}. \tag{2.9}$$

Since the matrix representation of Π is symmetric, when the eigenvalues of the second fundamental form are λ_i, $1 \leqslant i \leqslant n$, then $\mathrm{tr}(\Pi^2) = \Sigma_{i=1}^n \lambda_i^2$. Except when the second fundamental form is identically 0, the conformal normal curvature is well defined. Define $\|H_f\| = \sqrt{\mathrm{tr}(H_f^2)}$. Then, by equations (2.4) and (2.9),

$$B_l = \frac{1}{l^{\mathrm{T}}(I_n + \nabla_f \nabla_f^{\mathrm{T}})l} \frac{l^{\mathrm{T}} H_f l}{\|H_f\|}\Big|_{\omega=\omega_0}. \tag{2.10}$$

From equations (2.5)–(2.8), it can be further deduced that the conformal normal curvature of the graph in a direction l at a critical point ω_0 is

$$B_l = \frac{l^{\mathrm{T}} H_f l}{\|H_f\|}\Big|_{\omega=\omega_0} = -\frac{l^{\mathrm{T}} \ddot{F} l}{\sqrt{\mathrm{tr}(\ddot{F}^2)}}\Big|_{\omega=\omega_0} = -\frac{l^{\mathrm{T}} \Delta^{\mathrm{T}} (\dot{L})^{-1} \Delta l}{\sqrt{\mathrm{tr}\{\Delta^{\mathrm{T}} (\dot{L})^{-1} \Delta\}^2}}\Big|_{\theta=\hat{\theta}, \omega=\omega_0}. \tag{2.11}$$

It is clear from equation (2.11) that the computation of B_l requires almost no more effort than does the computation of C_l. Moreover, the conformal normal curvature enjoys numerous properties. They are summarized in the next few theorems.

First, we concern ourselves with *conformal reparameterizations*. When Ω represents the set of perturbations, a reparameterization is a smooth map ϕ: $\Omega \to \Theta$ from the domain Ω to a new domain Θ with the same dimension such that the Jacobian matrix of ϕ is non-singular throughout Ω. In particular, it is a one-to-one map and hence has an inverse map. We consider reparameterizations as 'modifications of the perturbation scheme' (Cook (1986), rejoinder). Furthermore, an $n \times n$ matrix M is a conformal matrix if there is a positive number τ such that $MM^{\mathrm{T}} = \tau I_n$. A reparameterization is conformal at ω_0 if its Jacobian matrix at ω_0 is a conformal matrix.

Theorem 1. If a reparameterization of Ω is conformal at a critical point ω_0 on the graph of f over Ω, then the conformal normal curvature in any direction at ω_0 is invariant under reparameterization.

Proof. Let $\phi(\omega)$ and $\psi(x)$ be the reparameterization and its inverse respectively. Let g be the composition $g(x) = f \circ \psi(x)$. When ω_0 is a critical point for f, $\phi(\omega_0)$ is a critical point for g. Therefore, when we apply the chain rule to calculate the Hessian for g in terms of derivatives of f, we have $H_g = \Psi^{\mathrm{T}} H_f \Psi$, where Ψ is the Jacobian matrix of ψ. Since ϕ is conformal at ω_0,

so is its inverse at $\phi(\omega_0)$. Therefore, at $\omega = \omega_0$, $\Psi^{\mathrm{T}}\Psi = \tau I_n$ for a positive number τ, and hence $\|H_g\| = \tau\|H_f\|$. Let Φ be the Jacobian matrix of ϕ. The reparameterization takes the direction l at ω_0 to $\Phi(l)$ at $\phi(\omega_0)$. At $\omega = \omega_0$, $\Phi = (1/\tau)\Psi^{\mathrm{T}}$; then, by equation (2.4),

$$\hat{C}_{\Phi(l)} = \frac{(\Phi l)^{\mathrm{T}} H_g (\Phi l)}{(\Phi l)^{\mathrm{T}}(\Phi l)} = \frac{(\Psi^{\mathrm{T}} l)^{\mathrm{T}}(\Psi^{\mathrm{T}} H_f \Psi)(\Psi^{\mathrm{T}} l)}{(\Psi^{\mathrm{T}} l)^{\mathrm{T}}(\Psi l)} = \frac{l^{\mathrm{T}} \tau H_f \tau l}{l^{\mathrm{T}} \tau l} = \tau C_l,$$

where the derivatives are evaluated at ω_0. By equation (2.11), $\hat{B}_{\Phi(l)} = \hat{C}_{\Phi(l)}/\|H_g\| = \tau C_l/\tau\|H_f\| = B_l$ as claimed. □

An example of a conformal reparameterization is $\phi(\omega) = M\omega + c$ such that M is a conformal matrix. For example, in linear regression analysis, when $\omega^{\mathrm{T}} = (\omega_1, \ldots, \omega_n)$ is the vector of case weights for a sample of size $N = n$ (Cook (1986), section 4), the conformal normal curvature is invariant with respect to the reparameterization $(1 + \omega_i)/2$ that was studied by Loynes (1986). A class of reparameterizations studied by Cook (1986) (rejoinder) is when $\phi(\omega) = (k(\omega_1), \ldots, k(\omega_n))^{\mathrm{T}}$ where $k(t)$ is a smooth function. Such a reparameterization is conformal at $\omega_0 = (a_1, \ldots, a_n)^{\mathrm{T}}$ if there is a non-zero constant a such that derivatives $\dot{k}(a_i)$ are all equal to a. The conformal normal curvature is also invariant when a single explanatory variable in a linear regression model is perturbed with different units or when a set of explanatory variables is perturbed accordingly (see Section 4.3).

Theorem 2. For any direction l, B_l satisfies the condition that $0 \leqslant |B_l| \leqslant 1$.

Proof. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis of eigenvectors of Π with eigenvalues λ_i, $i = 1, \ldots, n$, then, for any $l = \Sigma_i a_i \mathbf{e}_i$, $\Pi(l, l) = \Sigma_i \lambda_i a_i^2$. Then

$$|B_l| = \left|\sum_i \lambda_i a_i^2\right| \Big/ \left(\sum_i a_i^2\right)\left(\sum_i \lambda_i^2\right)^{1/2}.$$

By the Schwarz inequality and $(\Sigma_i a_i^2)^2 - \Sigma_i a_i^4 = 2\,\Sigma_{i<j} a_i^2 a_j^2 \geqslant 0$, $|B_l| \leqslant 1$. □

Theorem 2 tells us that B_l is a normalized measure, and thus it becomes easier to interpret its magnitude. Furthermore, $B_{\mathbf{e}_i}$ is equal to the normalized eigenvalue $\hat{\lambda}_i$ which is

$$\hat{\lambda}_i = \lambda_i \Big/ \left(\sum_{k=1}^{n} \lambda_k^2\right)^{1/2}. \tag{2.12}$$

It is an eigenvalue of the matrix of the bilinear form of B_l in equation (2.10). Therefore, we have the following simple observation.

Theorem 3. If $\{\mathbf{e}_i : 1 \leqslant i \leqslant n\}$ is a collection of orthonormal eigenvectors of Π, then $\Sigma_i B_{\mathbf{e}_i}^2 = 1$.

This theorem has useful consequences. The curvature of the graph at a point is uniform in all directions if the second fundamental form is diagonalizable to a constant multiple of the identity matrix. Such a point is called an umbilic point in differential geometry (Do Carmo, 1976). To study the uniformity of the curvature of an influence graph at an extremal point, we compare the curvature with a totally umbilic space whose curvature is the mean curvature of the graph at the given point, i.e. we measure the departure of the graph from being spherical (Do Carmo, 1976). Since $\Sigma_i B_{\mathbf{e}_i}^2 = 1$, if the conformal normal curvatures for all eigenvectors are identical, then they are all equal to $1/\sqrt{n}$. With reference to this, we can assess the local influence more systematically and objectively.

3. Assessment of local influence

Normal curvature C_l and the associated direction l are used to assess local influence. In particular, Cook (1986) suggested inspecting the eigenvector $\mathbf{e}_{\max}$ with maximum normal curvature $C_{\max}$ regardless of its size. Since normal curvature and its conformal counterpart differ only by an overall positive factor, the two curvatures are equivalent diagnostic measures, and the eigenvector $\mathbf{e}_{\max}$ gives the maximum conformal normal curvature. However, the conformal normal curvature is preferred because its invariance property and normalized nature facilitate its interpretation. Moreover, the discussion in the last section suggests that an objective bench-mark to judge the effect of $B_{\mathbf{e}_i}$ and B_l at various levels can be determined with reference to the mean curvature by using geometric concepts. Therefore, the following definition is established.

Definition 2. An eigenvector $\mathbf{e}$ is q influential if $|B_{\mathbf{e}}| \geqslant q/\sqrt{n}$.

3.1. Influence of individual eigenvector

An influential eigenvector can be further examined. Let E_i be the column vector in R^n whose ith entry is 1 and all other entries are 0. We call E_i the ith basic perturbation vector of the perturbation space. To analyse the contribution of basic perturbation vectors to the influence of an influential eigenvector $\mathbf{e}$, we can find basic perturbation vectors which are close to $\mathbf{e}$. When $\hat{\lambda}_i$ is the normalized eigenvalue given in equation (2.12), let $\{\mathbf{e}_i: 1 \leqslant i \leqslant n\}$ be a collection of orthonormal eigenvectors, with corresponding normalized eigenvalues $\hat{\lambda}_i$. The collection of basic perturbation vectors $\{E_i: 1 \leqslant i \leqslant n\}$ at an extremal point is also orthonormal with respect to the first fundamental form and the usual dot product. When $\mathbf{e}_i = \Sigma_{j=1}^n a_{ij}E_j$, then $\Sigma_{j=1}^n a_{ij}^2 = 1$. It means that for any fixed i, if the contribution of all the a_{ij} is uniform, then $|a_{ij}| = 1/\sqrt{n}$. It could be used to construct bench-marks to judge largeness. Furthermore, the line spanned by $\mathbf{e}_i$ is close to the line spanned by E_j if $|a_{ij}|$ is close to 1. This method can be applied to study $\mathbf{e}_{\max}$ or any individual influential eigenvectors.

3.2. Aggregate contribution of basic perturbation vectors

More generally, we can analyse the influence of basic perturbation vectors to all influential eigenvectors. Define $\mu_i = |\hat{\lambda}_i|$. We arrange the absolute values of the normalized eigenvalues by

$$\mu_{\max} = \mu_1 \geqslant \ldots \geqslant \mu_k \geqslant q/\sqrt{n} > \mu_{k+1} \ldots \mu_n \geqslant 0$$

and use a_{ij} to denote the jth element of the normalized eigenvector corresponding to μ_i.

Definition 3. The aggregate contribution of the jth basic perturbation vector to all q-influential eigenvectors is $m(q)_j = \sqrt{(\Sigma_{i=1}^k \mu_i a_{ij}^2)}$.

As $\Sigma_{j=1}^n m(q)_j^2 = \Sigma_j (\Sigma_{i=1}^k \mu_i a_{ij}^2) = \Sigma_{i=1}^k \mu_i (\Sigma_{j=1}^n a_{ij}^2) = \Sigma_{i=1}^k \mu_i$, if the contribution of all basic perturbation vectors is uniform then each is equal to

$$\bar{m}(q) = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{k}\mu_i\right)}. \tag{3.1}$$

Therefore, we refer to $\bar{m}(q)$ when determining the significance of the contribution of individual basic perturbation vectors.

There are two extremes in this method. One is to allow q to be sufficiently big that we consider the contribution of individual basic perturbation vectors to $\mathbf{e}_{\max}$ only. Then $m(q)_j$

$= \sqrt{\mu_{\max}}|a_{1j}|$, and this method is equivalent to comparing a_{1j} with $1/\sqrt{n}$ as suggested in Section 3.1. Another extreme is to allow $q = 0$ so that all eigenvalues are included in our consideration. This is the total contribution $m_j = m(0)_j = \sqrt{(\Sigma_{i=1}^n \mu_i a_{ij}^2)}$. If the contribution of all basic perturbation parameters is uniform, then each is equal to

$$\bar{m} = \bar{m}(0) = \sqrt{\left(\sum_{i=1}^n \mu_i \Big/ n\right)} = \sqrt{\left\{\sum_{i=1}^n |\lambda_i| \Big/ n \sqrt{\left(\sum_{i=1}^n \lambda_i^2\right)}\right\}}. \tag{3.2}$$

The total contribution m_j and the conformal normal curvature B_{E_j} of the basic perturbation vector E_j are deeply related. With the notation of Section 3.1, $E_l = \Sigma_i a_{il}\mathbf{e}_i$. Let B be the matrix of the bilinear form of B_l. Since $B_\upsilon = \upsilon^{\mathrm{T}} B \upsilon$, and $B\mathbf{e}_i = \hat{\lambda}_i \mathbf{e}_i$,

$$B_{E_j} = \sum_{i,k} a_{ij} a_{kj} \mathbf{e}_i^{\mathrm{T}} \hat{\lambda}_k \mathbf{e}_k = \sum_{i=1}^n \hat{\lambda}_i a_{ij}^2.$$

Therefore, if all eigenvalues are non-negative, B_{E_j} is equal to the square of the total contribution of the jth basic perturbation vector. To summarize, we have the following theorem.

Theorem 4. If the Hessian H_f is semipositive definite and all eigenvalues are non-negative, then $m_j^2 = B_{E_j}$, for all j.

If the contribution of all B_{E_j}s is uniform, then each is equal to

$$b = \mathrm{tr}(\Pi)/n\sqrt{\mathrm{tr}(\Pi^2)}. \tag{3.3}$$

Under the condition of theorem 4, $\bar{m}^2 = b$. We shall use b to set up objective bench-marks for the curvatures of the basic perturbation vectors.

Theorem 4 represents an important development in the local influence approach. Since the Hessian matrix of $f(\omega)$ in equation (2.1) always satisfies the assumption of theorem 4, the B_{E_j}s are useful to determine local influence. Note from equation (2.11) that B_{E_j} can be obtained easily once $\ddot{F}$ is available, and the computation of eigenvectors is no longer required. In many practical applications, it is even possible to obtain an explicit expression for B_{E_j}. This not only enhances the practical applicability of the local influence approach but also removes a real hindrance for its further development.

We can modify the definition of the aggregate contribution by changing the weight of the eigenvalues. For instance, we can consider the second-order aggregate contribution $M(q)_j = \sqrt{(\Sigma_{i=1}^k \mu_i^2 a_{ij}^2)}$. In this case, $\bar{M}(q) = \sqrt{\{(1/n)\,\Sigma_{i=1}^k \mu_i^2\}}$ is used to construct bench-marks. We can also consider aggregate contributions of various orders.

In the discussion of Cook (1986), Lawrance raised the question of the relevance of the individual elements of $\mathbf{e}_{\max}$, and some comments were given by Cook (1986) in his rejoinder. The developments in this section provide more insight into this problem.

4. Application in linear regression

4.1. Case weights in normal linear regression

Consider the linear regression model

$$Y = X\beta + \epsilon, \tag{4.1}$$

where X is an $N \times p$ matrix and ϵ is an $N \times 1$ normal random vector with $E(\epsilon) = 0$ and $\mathrm{var}(\epsilon) = \sigma^2 I_N$. Assuming that σ^2 is known, let $n = N$ and ω be the $n \times 1$ vector of case weights, so that the log-likelihood for the perturbed model is

$$L(\beta|\omega) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\omega_i(y_i - x_i^{\mathrm{T}}\beta)^2, \tag{4.2}$$

where ω_i and y_i are the ith elements of ω and Y respectively and x_i^{T} is the ith row of X. It can be shown that the normal curvature in a direction l at $\omega_0 = (1, \ldots, 1)^{\mathrm{T}}$ is given by (Cook (1986), equation (29))

$$C_l = 2l^{\mathrm{T}}DPDl/\sigma^2, \tag{4.3}$$

where D is a diagonal matrix whose entries are the residuals $(r_1, \ldots, r_n)$ and $P = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ is the projection matrix. Let E_i, $i = 1, \ldots, n$, be basic perturbation vectors of dimension n. Then it can be shown that (Cook (1986), equation (32))

$$C_{E_i} = 2r_i^2h_{ii}/\sigma^2, \tag{4.4}$$

where h_{ii} is the (i, i)th element of P. The conformal normal curvatures at ω_0 along the directions l and E_i are respectively given by

$$B_l = l^{\mathrm{T}}DPDl/\{\mathrm{tr}(DPD)^2\}^{1/2}, \tag{4.5}$$

$$B_{E_i} = r_i^2h_{ii}/\{\mathrm{tr}(DPD)^2\}^{1/2}. \tag{4.6}$$

From equation (3.15), if the contribution of all B_{E_i}s is uniform, then each is equal to

$$b = \mathrm{tr}(DPD)/n\{\mathrm{tr}(DPD)^2\}^{1/2}. \tag{4.7}$$

If σ^2 is unknown and only β is of interest, Cook calculated the normal curvature of the influence graph of the function $LD_s(\omega)$ (Cook (1986), equation (7)). He found that the normal curvature is given by equation (4.3) with σ^2 replaced by $\hat{\sigma}^2$, the usual maximum likelihood estimate. This is because the log-likelihood in equation (4.2) is a product of a function of σ^2 and a function of β, and the matrix of observed information in equation (2.8) of β and σ^2 is a diagonal block matrix with blocks corresponding to β and σ^2 respectively. Since the conformal normal curvature is independent of σ as demonstrated in equations (4.5) and (4.6), if we are only interested in β, an analysis using conformal normal curvature would not be affected by the estimate of σ.

In general, if $\theta^{\mathrm{T}} = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}})$, and if only θ_1 is of interest, then the conformal normal curvature would not depend on θ_2 if the likelihood function and the partition of θ follow a structure similar to that of β and σ^2 as above.

4.2. An example

As an illustration, consider the rat data reported in Weisberg (1985), p. 121. The data set consists of 19 cases, and there are four explanatory variables in the model. They are $x_0 =$ constant, $x_1 =$ body weight, $x_2 =$ liver weight and $x_3 =$ relative dose. The dependent variable is the percentage of the dose in the liver. Using standard diagnostic techniques in linear regression, Weisberg (1985) found that the leverage and Cook's distance are relatively large for case 3. Moreover, Cook (1986) gave a scatterplot of the absolute values of the elements of $\mathbf{e}_{\max}$ *versus* x_2 and found that the cases which have relatively small values of x_2 are more influential locally.

The data set was analysed by the methods suggested in the previous sections. The non-zero normalized eigenvalues are 0.816, 0.462, 0.338 and 0.076. For each case, the conformal normal curvature of the basic perturbation vector B_{E_j} and the aggregate contributions $m_j(q)$,

Table 1. Influential measures of influential cases: case weight perturbation

	Number of influential eigenvectors	*Mean*	*Bench-mark*	*Values for the following cases:*			
				Case 1	*Case 5*	*Case 13*	*Case 19*
$m_j(3)$	1	0.207	0.293	0.389	0.127	0.536	0.474
$m_j(2)$	2	0.259	0.367	0.511	0.412	0.543	0.523
$m_j(1)$	3	0.292	0.413	0.511	0.417	0.543	0.559
B_{E_j}	19	0.089	0.178	0.264	0.174	0.295	0.312

with q varying from 0 to 3, were computed. Results for selected cases are presented in Table 1. When $q = 3$, only one eigenvector, i.e. $\mathbf{e}_{\max}$, is considered as influential, and the $m_j(3)$s are in one-to-one correspondence with the entries in $\mathbf{e}_{\max}$. $m_j(0)$ is not presented because it is equivalent to B_{E_j}. Following the common practice of considering greater than two times the average as influential, we use $2b$ and $\bar{m}(q)\sqrt{2}$ as the bench-marks for B_{E_j} and $m_j(q)$ respectively. We use $\sqrt{2}$ instead of 2 in the bench-mark of $m_j(q)$ because the relationship between m_j and B_{E_j} given in theorem 4 is quadratic. With the chosen bench-marks, cases 13, 19 and 1 have been identified as influential by all measures. When $q = 1$ and $q = 2$, case 5 is included as influential. When $q = 0$, case 5 is marginally influential. In an analysis of all eigenvectors that is not presented here, we found case 5 to be the most influential element in the eigenvector of the second-largest eigenvalue.

4.3. Perturb explanatory variables in normal linear regression

For the standard linear regression model (4.1), a minor perturbation of the explanatory variables can be given in the form (Cook (1986), equation (36))

$$X_\omega = X + WS, \tag{4.8}$$

where $W = (\omega_{ij})$ is an $N \times p$ matrix of perturbations, $S = \mathrm{diag}(s_1, \ldots, s_p)$ and s_j converts the perturbation ω_{ij} to the appropriate size and units so that $\omega_{ij}s_j$ is compatible with the (i, j)th element of X. Under this perturbation scheme, and assuming that σ^2 is known, Cook (1986) showed that the non-zero eigenvalues of the matrix $-\ddot{F}$ in equation (2.6) are given by

$$\lambda_i = r^{\mathrm{T}} r \delta_i / \sigma^2 + \sum_j \hat{\beta}_j^2 s_j^2 / \sigma^2, \tag{4.9}$$

where r is the vector of residuals and δ_i, $i = 1, \ldots, p$, is the ith eigenvalue of $S(X^{\mathrm{T}}X)^{-1}S$. It is clear from equation (4.9) that the normal curvature depends on the choice of s_j. The normalized non-zero eigenvalues of the matrix $-\ddot{F}$ are given by

$$\hat{\lambda}_i = \left(r^{\mathrm{T}} r \delta_i + \sum_j \hat{\beta}_j^2 s_j^2 \right) \Big/ \left\{ \sum_i \left(r^{\mathrm{T}} r \delta_i + \sum_j \hat{\beta}_j^2 s_j^2 \right)^2 \right\}^{1/2}. \tag{4.10}$$

Again $\hat{\lambda}_i$ is independent of the value of σ^2. In addition, by theorem 1 or directly from equation (4.10), $\hat{\lambda}_i$ is invariant for the transformation $S \to aS$, i.e., if the ratios of the scales of perturbation for different explanatory variables are invariant, the conformal normal curvatures are invariant.

By setting some of the s_j to 0, the procedure can be applied to situations in which only some of the explanatory variables are perturbed. In practical applications, perturbing a single explanatory variable, say the kth variable, is of special interest because it can isolate a

Table 2. Influential measures of influential cases: perturbation of x_2

	Number of influential eigenvectors	*Mean*	*Bench-mark*	*Values for the following cases:*			
				Case 1	*Case 2*	*Case 4*	*Case 19*
$m_j(1)$	1	0.299	0.324	0.473	0.361	0.370	0.466
B_{E_j}	19	0.059	0.119	0.227	0.133	0.140	0.224

Table 3. Influential measures of influential cases: perturbation of x_1 and x_3

	Number of influential eigenvectors	*Mean*	*Bench-mark*	*Values for the following cases:*			
				Case 1, x_1	*Case 3, x_1*	*Case 1, x_3*	*Case 3, x_3*
$m_j(2)$	1	0.151	0.214	0.279	0.271	0.268	0.268
$m_j(1)$	4	0.213	0.302	0.323	0.313	0.303	0.307
B_{E_j}	38	0.045	0.091	0.104	0.098	0.092	0.094

particular component of a case. The conformal normal curvature in this situation is invariant to the choice of s_k.

The rat data set was analysed again. When only x_2 is perturbed, $n = 19$ and the normalized non-zero eigenvalues are 0.997 and 0.044 with multiplicity 3. Cases 1, 19, 4 and 2 in this order are identified as influential by each measure. The values of B_{E_j} and $m_j(1)$ for these cases are presented in Table 2. Since the maximum normalized eigenvalue is close to 1, the aggregate contributions are dominated by $\mathbf{e}_{\max}$.

When x_1 and x_3 are perturbed simultaneously, the dimension of the perturbation space is $n = 38$. Since the standard deviations of x_1 and x_3 are 16.49 and 0.085 respectively, we choose s_1 and s_3 such that the ratio s_1/s_3 is 200. For a chosen ratio, the conformal normal curvature is invariant with respect to the values of s_1 and s_3. (case 1, x_1), (case 3, x_1), (case 3, x_3) and (case 1, x_3) in this order are identified by each measure as influential. Their values are given in Table 3. We can see that the conformal normal curvature is in a restricted form of invariance; it is invariant under conformal reparameterizations, but it may not be invariant under a general reparameterization. This is nice because in some situations the invariance property is definitely not appropriate. For example, when different s_1 and s_3 are used so that the ratio s_1/s_3 is changed, invariance of the conformal normal curvature is not wanted. For the rat data set, if the ratio is changed to 100, the conformal normal curvature is changed and x_1 becomes not influential at all.

5. Discussion

Although the total contribution formally involves all eigenvectors, only those corresponding to non-zero eigenvalues are considered because the aggregation is weighted by the eigenvalues. In some situations, such as linear regression, the Hessian matrix H_f is not of full rank, and hence the number of non-zero eigenvalues is not equal to the dimension of ω. Therefore we can determine the bench-marks with reference to the rank of the Hessian rather than the dimension of the matrix. We prefer to use the dimension of the Hessian because of its generality and simplicity of presentation, and because of the geometric meaning of the mean curvature. In practical situations, e.g. in linear regression, the number of cases is always much

greater than the number of explanatory variables. In such cases, one should carefully decide whether to use the rank or the dimension of the Hessian matrix to determine the bench-marks. No matter which bench-marks are used, one must pay attention to the non-linear relationship between m_j and B_{E_j}.

Although the displacement function (2.1) has been emphasized, the methods can be generalized to other objective functions. However, if a chosen objective function leads to a Hessian matrix which does not satisfy the condition of theorem 4, the conformal normal curvature of the basic perturbation vector is not sufficient to determine influence.

Nevertheless, different measures and their objective bench-marks are constructed not only systematically but also flexibly. It is therefore easy to implement them in statistical software, and to provide a set of measures which can address the need of researchers who have different requirements for the sensitivity level. For further analysis, these measures can be used to construct plots. Finally, the results of the numerical examples support the claim that B_{E_j} is a convenient and effective measure for assessing local influence.

Acknowledgements

This research was partially supported by a Chinese University of Hong Kong Direct Grant for Research and a University of California at Riverside field research travel grant. Y. S. Poon thanks the Department of Statistics at the Chinese University of Hong Kong and the Institut des Hautes Etudes Scientifiques for hospitality. Both authors thank the referees, the Associate Editor and the Joint Editor for useful comments.

References

Beckman, R. (1986) Discussion on Assessment of local influence (by R. D. Cook). *J. R. Statist. Soc.* B, **48**, 161–162.
Beckman, R. J., Nachtsheim, C. J. and Cook, R. D. (1987) Diagnostics for mixed-model analysis of variance. *Technometrics*, **29**, 413–426.
Cook, R. D. (1986) Assessment of local influence (with discussion). *J. R. Statist. Soc.* B, **48**, 133–169.
Do Carmo, M. P. (1976) *Differential Geometry of Curves and Surfaces*. Englewood Cliffs: Prentice Hall.
Farebrother, R. W. (1992) Relative local influence and the condition number. *Communs Statist. Simuln Computn*, **21**, 707–710.
Kobayashi, S. (1972) *Transformation Groups in Differential Geometry*. New York: Springer.
Lawrance, A. J. (1986) Discussion on Assessment of local influence (by R. D. Cook). *J. R. Statist. Soc.* B, **48**, 157–159.
———(1988) Regression transformation diagnostics using local influence. *J. Am. Statist. Ass.*, **83**, 1067–1072.
Lee, S. Y. and Wang, S. J. (1996) Sensitivity analysis of structural equation models. *Psychometrika*, **61**, 93–108.
Loynes, R. M. (1986) Discussion on Assessment of local influence (by R. D. Cook). *J. R. Statist. Soc.* B, **48**, 156–157.
Prescott, P. (1986) Discussion on Assessment of local influence (by R. D. Cook). *J. R. Statist. Soc.* B, **48**, 161.
Schall, R. and Dunne, T. T. (1992) A note on the relationship between parameter collinearity and local influence. *Biometrika*, **79**, 399–404.
Shi, L. (1997) Local influence in principal component analysis. *Biometrika*, **84**, 175–186.
St Laurent, R. Y. and Cook, R. D. (1993) Leverage, local influence and curvature in nonlinear regression. *Biometrika*, **80**, 99–106.
Thomas, W. and Cook, R. D. (1990) Assessing influence on predictions in generalized linear models. *Technometrics*, **32**, 59–65.
Thorpe, J. A. (1979) *Elementary Topics in Differential Geometry*. New York: Springer.
Weisberg, S. (1985) *Applied Linear Regression*, 2nd edn. New York: Wiley.