ICCM 2007 · Vol. II · 123–130

Riemannian Geometry in Applied Statistics

Yat Sun Poon*

Abstract

In a recent paper with co-author W. Y. Poon and titled "Sectional curvature and joint influence", the author employs sectional curvature as a tool to study joint influence in the realm of local influence analysis. The development of this piece of work is based on Cook's application of normal curvature in assessing local influence through log-likelihood displacement function. The first half of this talk reviews the key concepts in Cook's approach and the concept of sectional curvature in his setting. The second half of this talk presents Ricci curvature and scalar curvature as auxiliary tool to mathematically justify the appropriateness of using sectional curvature. The full justification and argument could be found in our joint paper.

2000 Mathematics Subject Classification: 53B99, 62-07.
Keywords and Phrases: Joint influence, Local influence, Sectional curvature, Normal curvature.

1 Introduction

Once a statistical model is postulated, the parameters in its mathematical formulation is mostly driven by a data set. Yet both the model and the data collected are seldom exact reflection of the entire population. Therefore, one must subject the model formulation to critiques and modifications. From Weisberg's perspective, there are at least two areas for examination [21]. One is a critique of the underlying model. Another is a critique of the observed data. The former is known as diagnosis and the later influence analysis.

In this talk, we concern ourselves with the latter. Within influence analysis and beginning with Cook [5], there is a well developed method in assessing sensitivity of data points in a data set from a local perturbation perspective. The method is to study the infinitesimal changes in the log-likelihood scale at the unperturbed model. The quantifier to extract information from the perturbation scheme is geometric, namely from normal curvature of a graph. In this talk, I

*Department of Mathematics, University of California at Riverside, Riverside, CA 92521, U.S.A. E-mail: ypoon@ucr.edu.

shall present my recent joint work with W. Y. Poon [13]. We use classical Riemannian geometry concepts such as sectional curvature and scalar curvature as tools in local influence analysis, and use Ricci curvature as an auxiliary tool. The statistical issue is concerned with *joint influence* in perturbation.

To address a general audience, including undergraduate students, geometers and statisticians, I use this notes as background information for my presentation. I first use linear regression model to illustrate the ideas of log-likelihood displacement function. After a review of the concept of normal curvature, I review the concept of sectional curvature in our setting. At the end of this notes, I review how normal curvature is used in local influence analysis. It leads to the usage of sectional curvature to be presented in details in my talk.

2 Local influence analysis

A linear regression model with n observed data and p explanatory variables is a system of linear equations:

$$y = X\theta + \epsilon \tag{2.1}$$

where y and ϵ are vectors in R^n, $\theta \in R^p$, and X is a $n \times p$ matrix. ϵ is a vector of errors with mean zero and variance σ^2.

Given the set of observed data (y, X), a set of parameters $\hat{\theta}$ is chosen so that the fitted data $X\hat{\theta}$ minimize the function of θ

$$L(\theta) = -\frac{1}{2\sigma^2}\|y - X\theta\|^2. \tag{2.2}$$

This is the so-called *least square linear regression model.* It is an elementary linear algebra or calculus exercise to find that the parameter set $\hat{\theta}$ for the regression model is

$$\hat{\theta} = (X^T X)^{-1} X^T y. \tag{2.3}$$

Given the value of $\hat{\theta}$, the fitted value of the y-variable with the given explanatory data X is given by

$$\hat{y} := X\hat{\theta} = X(X^T X)^{-1} X^T y. \tag{2.4}$$

It is apparent that the data set (y, X) affects the coefficients $\hat{\theta}$. Therefore, there has been a long standing activity in statistics to examine the nature of activity above. From Weisberg's perspective, when we concern ourselves with the effective of (y, X) on the formulated model, we are engaged in influence analysis [21].

Perturbation. In [5], it is proposed to study influence analysis from a perturbation perspective. For example, instead of considering the least square linear regression line for a given data set, one considers a "weighted" least square linear regression line.

$$L(\theta|\omega) := -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\omega_i(y_i - x_i^T\theta)^2. \tag{2.5}$$

The parameters $\{\omega_1, \ldots, \omega_n\}$ are the weight of each case of observation. The parameter $\hat{\theta}_\omega$ is the maximizer of the above function.

The perturbed likelihood function, with the regression parameter $\hat{\theta}_\omega$, is

$$L(\hat{\theta}_\omega|\omega) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\omega_i(y_i - x_i^T\hat{\theta}_\omega)^2.$$

When $\omega = (1, 1, \ldots, 1)$, we have the unperturbed model $L(\theta)$. However, when we set $\omega = (0, 1, \ldots, 1)$, the above model is to ignore the first case in the entire data set because the weight of the first case is given weight "zero" in the entire computation. This perturbation is the well-known case-weight perturbation of linear regression. In such a perturbation, one needs to understand the effect of the perturbation on the postulated model. Therefore, we consider the difference of the log-likelihood functions.

$$f(\omega) = L(\hat{\theta}_{\omega_0}|\omega_0) - L(\hat{\theta}_\omega|\omega_0). \tag{2.6}$$

For example, if we evaluate f as $\omega = (0, 1, \ldots, 1)$ in the above perturbation of linear regression, we detect a quantified difference in the parameter of the regression model *with and without* the first case in the data set. This quantity is in proportion to a well known statistical tool, namely *Cook's distance* [1] [4] [8].

Likelihood displacement function. In general, given the log-likelihood $L(\theta)$ for a postulated model with θ a vector of p unknown parameters, let $L(\theta|\omega)$ be a perturbed model with parameter $\omega^T = (\omega_1, \ldots, \omega_n)$ defined in an open subset Ω in R^n such that for ω_0, $L(\theta|\omega_0) = L(\theta)$. Let $\hat{\theta}_\omega$ be the maximizer of $L(\theta|\omega)$. The log-likelihood displacement function is a function of perturbation parameters.

$$f(\omega) := 2\left(L(\hat{\theta}|\omega_0) - L(\hat{\theta}_\omega|\omega_0)\right). \tag{2.7}$$

It has the following properties.

$$f(\omega_0) = 0; \quad \text{and} \quad f(\omega) \geqslant 0 \tag{2.8}$$

for all $\omega \in \Omega$.

In his seminal paper [5], Cook proposes to use geometric information of this graph near ω_0 as tools in influence analysis, or more precise local influence analysis. Progress in this theory has developed well beyond this point. See for example [2], [7], [9], [11], [17], [18], [20].

3 Geometry of graphs

Geometry of graphs is a subject with very long history. See for example [19]. In this section, I review some basic ideas and refer all detailed computation to [19] and [10].

Suppose that $f(\omega)$ is an infinitely differentiable function defined in an open subset Ω in R^n. Its graph G is a subset of R^{n+1}. We could consider the graph as the image of a function map from Ω into R^{n+1}:

$$\Phi(\omega) := \begin{pmatrix} \omega \\ f(\omega) \end{pmatrix} \tag{3.1}$$

Riemannian metric. The restriction of inner product at the tangent space of G at $\Phi(\omega)$ apparently varies according to ω. It is called the *first fundamental form,* denoted by g [19]. In modern and general terms in geometry, it is also called the *Riemannian metric* of the space G [10]. Via the map Φ, we consider tangent vectors to the graph at the point $\Phi(\omega)$ as images of tangent vectors to the domain Ω at the point ω, then g could be expressed as a function of ω explicitly. To be precise, if ℓ and ℓ' are tangent to Ω at ω, we denote the image of ℓ and ℓ' by $d\Phi(\ell)$ and $d\Phi(\ell')$ respectively, then by definition

$$g(\ell, \ell') := \langle d\Phi(\ell), d\Phi(\ell')\rangle_{R^{n+1}}. \tag{3.2}$$

The computation through differentiation of Φ shows that

$$g(\ell, \ell') = \langle \ell, \ell'\rangle + \langle \nabla f, \ell\rangle\langle \nabla f, \ell'\rangle \tag{3.3}$$

where ∇f is the gradient of the function f at the point ω and the inner product $\langle -, -\rangle$ is now taken in R^n, i.e. the tangent space of Ω.

Normal curvature. Given a tangent vector $d\Phi(\ell)$ and a unit normal vector N at $\Phi(\omega)$, we denote the plane spanned by them P_ℓ. This plane cuts across the graph G along a curve. One could now calculate the concavity of this curve on the plane P_ℓ at the point $\Phi(\omega)$ in terms of the normal component of the second derivative of the curve. This quantity is known as *normal curvature* of the graph G at the point $\Phi(\omega)$ in the direction ℓ [19, Chapter 12]. The intuition is that the curve is tangent to the given vector $d\Phi(\ell)$. When it is concave up with respect to the normal vector N, its normal curvature is positive. The bigger the normal curvature, the sharper the concavity. If the curve is a straight line, it is equal to zero.

On the other hand, an effective way to make sense of the curvature of a graph is to study how a unit normal vector of the graph varies with respect to the variables ω. Therefore, when one is given a tangent vector ℓ, one considers the directional derivative of the normal vector N for any given direction tangent to the graph, denoted by $D_{d\Phi(\ell)}N$. It turns out that this vector is again a tangent vector of the graph and the normal curvature is equal to $S(\ell) := -D_{d\Phi(\ell)}N$. The map S from tangent space to tangent space is called the *shape operator.*

The perception of concavity of a cross section and that of directional derivatives are the same [19, Chapter 9]. To be precise, the normal curvature in the direction ℓ is equal to

$$\Pi(\ell, \ell) := -\langle D_{d\Phi(\ell)}(N), d\Phi(\ell)\rangle = \langle S\ell, \ell\rangle. \tag{3.4}$$

This quantity could be computed explicitly in terms of the parameters ω [19, Chapter 12], and

$$\Pi(\ell, \ell) = \frac{\ell^T H(f)\ell}{\sqrt{1 + \nabla f^T \nabla f}}, \tag{3.5}$$

where $H(f)$ is the Hessian matrix of the function f at the point ω.

At the end, the normal curvature in a direction ℓ is given by

$$C_\ell = \frac{\Pi(\ell, \ell)}{g(\ell, \ell)}. \tag{3.6}$$

The symmetric matrix of functions of ω

$$\frac{1}{\sqrt{1 + \nabla f^T \nabla f}}\ H(f) \tag{3.7}$$

is called *second fundamental form* in classical differential geometry of the graph G [19].

It is also the matrix of the shape operator S. The trace of the second fundamental form is called the *mean curvature.* Its determinant is the *Gaussian* curvature.

Since the matrix S is symmetric, the mean curvature and the Gaussian curvature are respectively the sum and the trace of all eigenvalues of S.

Finally, if we consider normal curvature as a function of the space of unit tangent vectors and seek its critical points, a vector is a critical point if and only if it is an eigenvector.

Sectional curvature. Let ℓ and ℓ' be any pair of vectors in the standard basis of R^n, we could now consider the three-dimensional space spanned by the normal vector N and the tangent plane $\Lambda(\ell, \ell')$ spanned by $d\Phi(\ell)$ and $d\Phi(\ell')$. The cross section of the graph now is a two-dimensional graph in R^3. If the two vectors are the first two in the standard basis of R^n, the cross section is the graph of the function f with respect to the first two variables ω_1 and ω_2. If we now reduce discussion in the last paragraph to this situation, the second fundamental form is a 2-by-2 matrix. By definition, the *sectional curvature* of the plane $\Lambda(\ell, \ell')$ is the Gaussian curvature of this two-dimensional graph.

This quantity could be interpreted through two different but related formula. In terms of the second fundamental form or the sharp operator S, for any pair of linearly independent tangent vectors ℓ, ℓ',

$$\mathrm{Sec}(\ell, \ell') = \frac{g(S\ell, \ell)g(S\ell', \ell') - g(S\ell, \ell')g(\ell, S\ell')}{g(\ell, \ell)g(\ell', \ell') - g(\ell, \ell')^2} \tag{3.8}$$

As we now rotate the vector from ℓ through ℓ' to $-\ell$, we find all possible one-dimensional cross sections. If all these cross sections concave in the same direction with respect to the normal vector, the sectional curvature is positive. If some are concave up and others are concave down, the sectional curvature is negative. If the maximum or minimum normal curvature within the two-dimensional graph is equal to zero, the sectional curvature is equal to zero. In fact, the shape operator restricted to the two-dimensional space $\Lambda(\ell, \ell')$ has a maximum and a minimum. The sectional curvature of this plane is simply the product of these quantities.

A feature of the sectional curvature is that in spite of its definition, it depends on the choice of vectors ℓ and ℓ' only up to their linear combinations. In other words, if e and e' are two linearly independent vectors such that the plane spanned by e and e' is equal to the plane spanned by ℓ and ℓ', then one may use the formula (3.8) above to prove that

$$\mathrm{Sec}(\ell, \ell') = \mathrm{Sec}(e, e') \tag{3.9}$$

The sectional curvature has a modern interpretation in terms of covariant derivatives. By definition, the covariant derivative of ℓ' in the direction of ℓ is simply the tangential component of the directional derivative of the vector $d\Phi(\ell')$ in the direction of $d\Phi(\ell)$. Let ∇ denote covariant derivative and D denote directional derivative. Then the formula for computation is

$$\nabla_\ell \ell' = D_{d\Phi(\ell)}(d\Phi(\ell')) - \langle D_{d\Phi(\ell)}(d\Phi(\ell')), N\rangle N. \tag{3.10}$$

When e_i and e_j are vectors of the standard basis $\{e_1, \ldots e_n\}$ in R^n, then

$$K_{ij} := \mathrm{Sec}(e_i, e_j) = g(\nabla_{e_i}\nabla_{e_j} e_i, e_j) - g(\nabla_{e_i}\nabla_{e_j} e_i, e_j). \tag{3.11}$$

From formula (3.8), one sees that the quantity $\mathrm{Sec}(e_i, e_j)$ does not depend on the choice of basis in the plane $\Lambda(e_i, e_j)$. In particular,

$$K_{ij} = K_{ji} = \mathrm{Sec}(e_j, e_i) = g(\nabla_{e_j}\nabla_{e_i} e_j, e_i) - g(\nabla_{e_j}\nabla_{e_i} e_j, e_i). \tag{3.12}$$

The formula (4.3) has a generalization to $\mathrm{Sec}(\ell, \ell')$ when ℓ and ℓ'. It involves the concept of Riemannian curvature, a concept we skip here.

At each point of the tangent space of the graph, we may now treat sectional curvature as a function of the set of all two-dimensional subspaces of its tangent space. The critical points of this functions occur exactly at $\Lambda(\ell, \ell')$ where both ℓ and ℓ' are eigenvectors of the shape operator S or the second fundamental form Π.

4 Applications

Now we take the general theory in the last section to study the geometry of the graph of the log-likelihood displacement function at the point ω_0. Given the property of the function f at ω_0 at noted in (2.8), both the value of f and its gradient at the point ω_0 are zero. Therefore, the first fundamental form (3.3) is simply the standard inner product on R^n, the second fundamental form (3.7) is the Hessian of the function f at ω_0. The normal curvature is simplified to

$$C_\ell = \frac{\ell^T H(f)\ell}{|\ell|^2}. \tag{4.1}$$

The sectional curvature is simplified as follows.

$$\mathrm{Sec}(\ell, \ell') = \frac{(\ell^T H(f)\ell)(\ell'^T H(f)\ell') - (\ell'^T H(f)\ell)(\ell^T H(f)\ell')}{|\ell|^2|\ell'|^2 - \langle \ell, \ell' \rangle^2}. \tag{4.2}$$

In particular, the sectional curvature of the plane spanned by e_i and e_j in the standard basis is

$$\mathrm{Sec}(e_i, e_j) = \left(\frac{\partial^2 f}{\partial \omega_i^2}\right)\left(\frac{\partial^2 f}{\partial \omega_j^2}\right) - \left(\frac{\partial^2 f}{\partial \omega_i \partial \omega_j}\right)^2. \tag{4.3}$$

Usage of normal curvature To assess local influence through perturbation, Cook pioneered to use the normal curvature [5]. In particular, he suggested that one studies a unit eigenvector $\ell_{\max}$ of the maximum eigenvalue. In particular if with respect to the standard basis $\{e_1, \ldots, e_n\}$,

$$\ell_{\max} = a_1 e_1 + \cdots + a_n e_n, \tag{4.4}$$

then the data point associated to the perturbation parameter ω_i with a relatively large coefficient a_i in the vector $\ell_{\max}$ justifies further analysis.

In the past twenty years, utilization of normal curvature in local influence analysis has been subjected to further justification [16] and modification [11]. It has plentiful of applications. e.g. [2] [7] [9] [15] [17] [18]. In this article, we content ourselves with a brief review of the key geometric concept behind Cook's original idea.

Usage of sectional curvature The concept of Cook's distance in linear regression has a long standing generalization from the deletion of one data point to a simultaneous deletion of several data points [6]. However, citing the work in [1] [3] [14], in [8] Lawrance argued that there is a need to understand, qualify and quantify the difference in measuring the influence of individual data points in different processes. Lawrance clarifies that when one studies the influence of individual points, at least through deletion in linear regression, one has to employ the concept of *conditional influence.* If one studies the influence of a group of data points simultaneously, such analysis should be considered an analysis of *joint influence.* These two concepts raise different issues.

With my co-author W. Y. Poon, I have extended Lawrance's concept of conditional influence in linear regression to the conditional local influence in Cook's perturbation scheme [12]. In this talk, I shall present our recent work on using the sectional curvature as a tool to identify a group of perturbation parameter in the log-likelihood displacement function for joint influence analysis. In order to justify our argument and execute the analysis, we shall employ other classic concepts in Riemannian geometry such as Ricci curvature and scalar curvature.

References

[1] A. C. Atkinson, *Plots, transformations and regression*, Oxford Statistical Science Series 1, Oxford Science Publications, 1985.

[2] R. J. Beckman, C. J. Nachtsheim & R. D. Cook, Diagnostics for mixed-model analysis of variance, Technometrics, 29, (1987) 413–246.

[3] S. Chatterjee & A. S. Hadi, A. S. Influential observations, high leverage points, and outliers in linear regression (with discussion), *Statist. Sci.*, 1 (1986) 379–416.

[4] R. D. Cook, Detection of influential observations in linear regression, *Technometrics* 19 (1977), 15-18.

[5] R. D. Cook, Assessment of local influence (with discussion), *J. R. Statist. Soc. B*, 48 (1986), 133–169.

[6] R. D. Cook & S. Weisberg, *Residuals and influence regression*, Monogrphas on Statistics and Applied Probability 18, Chapman & Hall, 1995.

[7] A. J. Lawrance, Regression transformation diagnostics using local influence, *J. Amer. Statist. Ass.*, 83, (1988) 1067-1072.

[8] A. J. Lawrance, Deletion influence and masking in regression, *J. R. Statist. Soc. B*, 57 (1995) 181-189.

[9] S. Y. Lee & S. J. Wang, Sensitivity analysis of structural equation models, *Psychometrika*, 61 (1996) 93–108.

[10] B. O'Neill, *Semi-Riemannian geometry with application to relativity*, Pure and Applied Mathematics, Academic Press, 1983.

[11] W. Y. Poon & Y. S. Poon, Conformal normal curvature and assessment of local influence, *J. R. Statist. Soc. B*, 61 (1999), 51–61.

[12] W. Y. Poon & Y. S. Poon, Conditional local influence, *J. Applied Statistics*, 34 (2007) 1–13.

[13] Y. S. Poon & W. Y. Poon, Sectional curvature and joint influence, *Preprint*, (2007).

[14] P. J. Rousseeuw & A. M. Leroy, *Robust regression and outlier detection*. New York: Wiley. 1987

[15] N. Sartori, K. Thomaseth & A. Salvan, Local influence analysis when interfacing toxicokinetic and proportional hazard models, *Stat Med* 23 (2004) 2399-2412.

[16] B. Schwarzmann, A connection between local-influence analysis and residual diagnostics, *Technometrics,* 33, (1991) 103–104

[17] L. Shi, Local influence in principal component analysis, *Biometrika*, 84 (1997) 175–186.

[18] R. Y. St Laurent & R. D. Cook, Leverage, local influence and curvature in nonlinear regression, *Biometrika*, 80 (1993), 99–106.

[19] J. A. Thorpe, *Elementary topics in differential geometry*, Undergraduate Texts in Mathematics, Springer, 1985.

[20] W. Thomas & R. D. Cook, Assessing influence on predictions in generalized linear models, *Technometrics*, 32 (1990) 59–65.

[21] S. Weisberg, Discussion, some principles for regression diagnostics and influence analysis, *Technometrics*, 25 190 (1983) 240–245.