

SIZING UP THE REGIONS OF UNIQUE MINIMA IN THE LEAST SQUARES NONLINEAR REGRESSION

LEONID KHINKIS, MILBURN CROTZER AND ADINA OPRISAN

Abstract. In nonlinear regression analysis, the residual sum of squares may possess multiple local minima. This complicates finding the global minimum and adversely affects reliability of the relevant statistical methods. Identifying and sizing up the regions of a *readily identifiable global minimum* (RIGM) is therefore of both theoretical and practical interest. This paper addresses the issue by using equidistant function previously introduced by the first two co-authors of this paper.

1. INTRODUCTION

Nonlinear regression analysis is a powerful predictive tool used in many scientific and business fields. Similar to linear regression, the estimated parameters of a nonlinear model are those for which the residual sum of squares attains a minimum. Unlike linear regression however, analysis of nonlinear models can lead to multiple minima depending on the distribution of the predictors and the data uncertainties. The analysis is further complicated by the linearized iterative method of analysis and the choice of starting values. These problems are illustrated in Figures 1 and 2. Geometrically, for a given predictor design, a nonlinear model will have a curved expectation surface. Each figure shows the expectation surface for a 2-parameter Hill model, discussed later, with 3 design points. In Figure 1, starting from a global minimum on the surface and proceeding outward in a normal direction (asterisks), a point is reached where the minimum sum of squares leads to a different position on the surface (squares), i.e., a local minimum with a different solution. Moreover, the position of the local minimum can change, depending on the starting point for the analysis, as shown in Figure 2. In practical terms, a highly variable response, i.e., large residuals, can lead to an inaccurate estimate of the model parameters and the associated confidence intervals. Haines illustrated the multiple local minima phenomenon in nonlinear models in [5], in connection with some concepts from differential geometry. The problem was extensively studied by Demidenko in [3] and [4] where the levels of local convexity and local unimodality of the sum of squares were introduced and a relationship between these concepts and the curvature measures of the model were established. Specifically, it was shown in [4] that the level of the local unimodality of the sum of squares equals to the minimum squared radius of the intrinsic curvature of the nonlinear regression model. Demidenko's contributions ([3] and [4]) have further

MSC (2010): primary 62J02, 62K05, 62P10.

Keywords: nonlinear regression, parameter estimation, residual sum of squares.

advanced the utility of the curvature measures in nonlinear regression. These measures were originally introduced by Beale ([2]) and Bates and Watts ([1]).

In a series of papers summarized in [7], Pazman developed “an almost exact” theory of nonlinear regression inference. Under some essential assumptions, it is shown that there exists a region of a sufficiently large size (as measured by its probability being close to one) such that any data point belonging to that region results in a global minimum of the sum of squares function. The goal of this paper is to develop a computational method for finding such a region, that we call a *readily identifiable global minimum* (RIGM) region. We obtain a formula for computing the maximum radius of any RIGM region based on the properties of an *equidistant function* that the first two authors introduced in [6]. This function also plays a prominent role in the definition of extended intrinsic curvature of a nonlinear model presented in [8]. While closely related to the intrinsic curvature locally, the equidistant function reflects global rather than local properties of a nonlinear model.

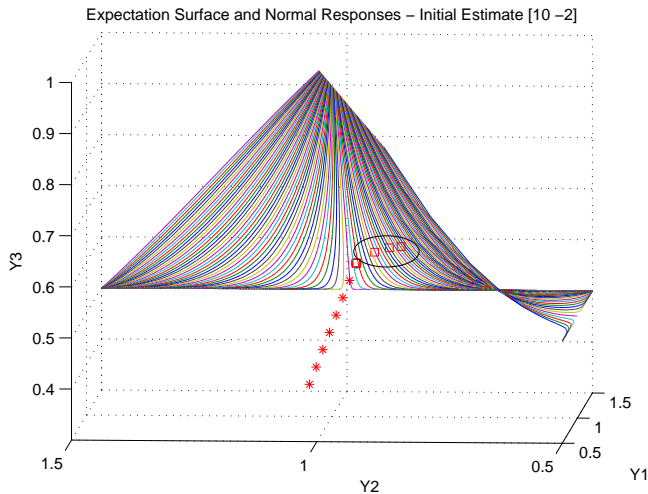


Figure 1. Expectation surface and normal responses.

2. THE EQUIDISTANT FUNCTION AND THE RADIUS OF INTRINSIC CURVATURE

Let's consider a nonlinear regression model given by

$$y = \eta(\theta) + \varepsilon, \quad \theta \in \Theta, \quad E(\varepsilon) = 0, \quad \text{var}(\varepsilon) = \sigma^2 W, \quad (2.1)$$

where $\theta = (\theta_1, \dots, \theta_m)^t$ is a vector of unknown parameters. Assume that $\theta \in \Theta$, and that the (known) parameter space Θ is a subset of \mathbb{R}^m such that $\Theta \subseteq \overline{\text{int } \Theta}$. Thus $y \in \mathbb{R}^N$ is the vector of observed data, $\varepsilon \in \mathbb{R}^N$ is the error vector, σ is the parameter of the variance component which may, but needs not be known, W is a known positive semi-definite matrix. The parameter σ equals to the standard deviation of the error of an individual observation under the assumption of a

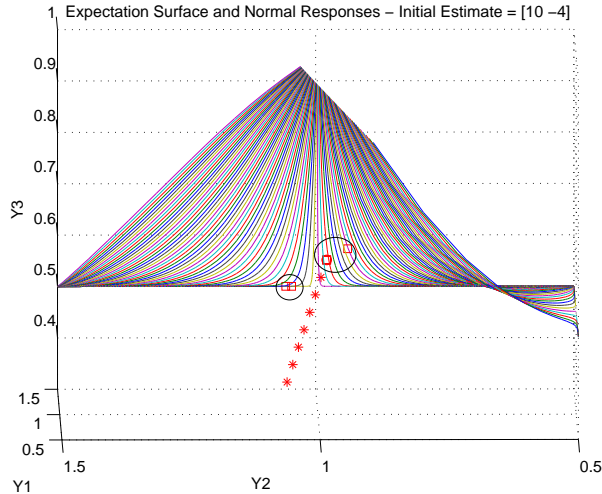


Figure 2. Expectation surface and normal responses.

constant variance in which case $W = I$, the identity matrix. We assume that the errors are normally distributed. The mapping $\eta : \Theta \rightarrow \mathbb{R}^N$ is a known, twice continuously differentiable mapping on $\text{int } \Theta$.

The expectation surface is $E = \{\eta(\theta) : \theta \in \Theta\}$. The least squares estimate (LSE) of θ is

$$\hat{\theta} = \hat{\theta}(y) = \arg \min_{\theta \in \Theta} \|y - \eta(\theta)\|_W^2,$$

where the definition of the squared norm $\|a\|_W^2 = a^t W^{-1} a$ is used. This norm corresponds to the inner product $\langle a, b \rangle_W = a^t W^{-1} b$.

Any LSE $\hat{\theta}(y) \in \text{int } \Theta$ satisfies the system of m normal equations (the stationary conditions)

$$\frac{\partial}{\partial \theta} \|y - \eta(\theta)\|_W^2 = 0. \quad (2.2)$$

The linear span of the vectors $\frac{\partial \eta(\theta)}{\partial \theta_i}$, ($i = 1, \dots, m$), forms a plane, $T(\theta)$, known as the tangent to the expectation surface E at $\theta \in \Theta$. Define the normal plane $NO(\theta)$ as the hyperplane orthogonal to the expectation surface at the point $\eta(\theta)$,

$$NO(\theta) := \left\{ n \in \mathbb{R}^N : \left\langle n, \frac{\partial \eta(\theta)}{\partial \theta_i} \right\rangle_W = 0, (i = 1, \dots, m) \right\}.$$

Let $P(\theta)$ be an orthogonal projector onto the tangent plane $T(\theta)$. Then any vector $h \in \mathbb{R}^N$ can be represented as a sum of its mutually orthogonal components

$$h = P(\theta)h + (I - P(\theta))h, \quad (2.3)$$

so that $P(\theta)h \in T(\theta)$ while $(I - P(\theta))h \in NO(\theta)$. Let $NO_1(\theta)$ be the set of all unit vectors in $NO(\theta)$, $NO_1(\theta) := \{n \in NO(\theta) : \|n\|_W = 1\}$ and \mathbb{R}_+ be the set of all positive real numbers.

A *directional* equidistant function, $t(\vartheta_1, \vartheta_2, n)$ was defined in [6, Eq. 3] as

$$t(\vartheta_1, \vartheta_2, n) = \frac{\|\eta(\vartheta_2) - \eta(\vartheta_1)\|_W^2}{2\langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W}. \quad (2.4)$$

Here $\vartheta_1 \in \text{int } \Theta$, $\vartheta_2 \in \Theta$ and $n \in NO_1(\vartheta_1)$. Clearly, $t(\vartheta_1, \vartheta_2, n)$ is defined by (2.4) only if $\langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W \neq 0$. As pointed out in [6], $y = \eta(\vartheta_1) + tn$ is equidistant from two different points, $\eta(\vartheta_1)$ and $\eta(\vartheta_2)$ if and only if $\langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W > 0$ and $t = t(\vartheta_1, \vartheta_2, n)$. Here $t \in \mathbb{R}_+$, $\vartheta_1 \in \text{int } \Theta$, $\vartheta_2 \in \Theta$ and $n \in NO_1(\vartheta_1)$.

The equidistance property is

$$\|y - \eta(\vartheta_1)\|_W = \|y - \eta(\vartheta_2)\|_W.$$

This property will not hold for any $y \in \mathbb{R}^N$, $\vartheta_1 \in \text{int } \Theta$, $\vartheta_2 \in \Theta$, $n \in NO_1(\vartheta_1)$ and $t \in \mathbb{R}_+$, such that $y = \eta(\vartheta_1) + tn$ and $\langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W \leq 0$. By Theorem 1 from [6], for a fixed $\vartheta_2 \in \text{int } \Theta$, $t(\vartheta_1, \vartheta_2, n)$ is the supremum of the values $d \in \mathbb{R}_+$ such that $y = \eta(\vartheta_1) + dn$ results in the unique least square estimate $\hat{\theta}(y) = \vartheta_1$. Let's define $N_1(\vartheta_1, \vartheta_2)$ as $N_1(\vartheta_1, \vartheta_2) = \{n \in NO_1(\vartheta_1) : \langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W > 0\}$.

Proposition 2.1. *Let $t(\vartheta_1, \vartheta_2)$ be the equidistant function defined as*

$$t(\vartheta_1, \vartheta_2) = \min_{n \in N_1(\vartheta_1, \vartheta_2)} t(\vartheta_1, \vartheta_2, n).$$

Then

$$t(\vartheta_1, \vartheta_2) = \frac{\|\eta(\vartheta_2) - \eta(\vartheta_1)\|_W^2}{2\|(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1))\|_W}, \quad \vartheta_1 \in \text{int } \Theta, \vartheta_2 \in \Theta. \quad (2.5)$$

Proof. Due to orthogonal decomposition (2.3),

$$\langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W = \langle n, (I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1)) \rangle_W.$$

On the other hand, by Cauchy inequality we get:

$$\begin{aligned} \langle n, (I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1)) \rangle_W &\leq \|n\|_W \|(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1))\|_W \\ &= \|(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1))\|_W. \end{aligned}$$

Therefore

$$n_{max} := \frac{(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1))}{\|(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1))\|_W}$$

gives the largest value of $\langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W$ over $N_1(\vartheta_1, \vartheta_2)$, thus

$$\max_{n \in N_1(\vartheta_1, \vartheta_2)} \langle n, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W = \|(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1))\|_W.$$

A simple calculation shows that $n_{max} \in N_1(\vartheta_1, \vartheta_2)$ when $(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1)) \neq 0$.

Thus, the equidistant function (2.5) follows:

$$\begin{aligned} t(\vartheta_1, \vartheta_2) &= \min_{n \in N_1(\vartheta_1, \vartheta_2)} t(\vartheta_1, \vartheta_2, n) = \frac{\|\eta(\vartheta_2) - \eta(\vartheta_1)\|_W^2}{2\langle n_{max}, \eta(\vartheta_2) - \eta(\vartheta_1) \rangle_W} \\ &= \frac{\|\eta(\vartheta_2) - \eta(\vartheta_1)\|_W^2}{2\|(I - P(\vartheta_1))(\eta(\vartheta_2) - \eta(\vartheta_1))\|_W}. \end{aligned}$$

□

In [7], the intrinsic curvature of a nonlinear regression model was defined as

$$K_{int}(\vartheta) = \sup_{v \in \mathbb{R}^m \setminus \{0\}} \frac{\|(I - P(\vartheta))v^t H(\vartheta)v\|}{v^t M_W(\vartheta)v},$$

where

$$\begin{aligned} P(\vartheta) &= J(\vartheta)M_W^{-1}(\vartheta)J^t(\vartheta)W^{-1}, \\ M_W(\vartheta) &= J^t(\vartheta)W^{-1}J(\vartheta), \\ J(\vartheta) &= \frac{\partial \eta(\vartheta)}{\partial \vartheta^t}, \quad H(\vartheta) = \frac{\partial^2 \eta(\vartheta)}{\partial \vartheta \partial \vartheta^t}. \end{aligned}$$

Remark 2.2. Geometrically, $K_{int}(\vartheta)$ equals to the maximal curvature of the geodesic curves on E passing through $\vartheta \in \text{int } \Theta$. The radius of intrinsic curvature $R_{int}(\vartheta) := \frac{1}{K_{int}(\vartheta)}$ represents the infimum of $t(\vartheta, \vartheta_2)$ when ϑ_2 approaches ϑ along all possible directions.

Also, in [4], Demidenko defined the upper local unimodality level for the sum of squares of a nonlinear regression model as

$$\bar{S}_{LU} = \min_{\theta \in \Theta} R_{int}^2(\theta)$$

and a local unimodality level, S_{LU} , as any number less or equal to \bar{S}_{LU} .

3. THE REGION OF A READILY IDENTIFIABLE GLOBAL MINIMUM

Definition 3.1. A region $D \subseteq \mathbb{R}^N$ is called RIGM (readily identifiable global minimum) region if for any $y \in D$ there exists at most one $\vartheta \in \text{int } \Theta$ satisfying both the stationary conditions (2.2) and

$$\|y - \eta(\vartheta)\| < h(\vartheta)$$

for some function $h(\vartheta)$.

Consequently, $\vartheta = \hat{\vartheta}(y)$.

Define a tube $T_1(r)$ about expectation surface E as

$$T_1(r) := \{y = \eta(\vartheta) + sn, n \in NO_1(\vartheta), \vartheta \in \text{int } \Theta, |s| < r\}.$$

As follows from Theorem 1 in [6], the maximal radius of a RIGM tube $T_1(r)$ equals to

$$r = \inf_{\vartheta \in \text{int } \Theta, \vartheta_a \in \Theta, \vartheta \neq \vartheta_a} t(\vartheta, \vartheta_a). \quad (3.1)$$

Let $r_0 > 0$ and a parameter value $\bar{\vartheta} \in \Theta$. Define a subset of the parameter space, $\Theta(2r_0)$, called the restricted parameter space, as

$$\Theta(2r_0) := \{\vartheta \in \text{int } \Theta : \|P(\vartheta)(\eta(\vartheta) - \eta(\bar{\vartheta}))\|_W \leq r_0 \text{ and } \|\eta(\vartheta) - \eta(\bar{\vartheta})\|_W \leq 2r_0\}.$$

Additionally, define a subset of the sample space, $T^*(2r_0)$, called the restricted sample space, as

$$T^*(2r_0) := \cup_{\vartheta \in \Theta(2r_0)} L(\vartheta, r_0),$$

where

$$L(\vartheta, r_0) = \{y : \langle y - \eta(\vartheta), J(\vartheta) \rangle_W = 0 \text{ and } \|y - \Psi_{\bar{\vartheta}}(\vartheta)\|_W \leq r_0\}$$

and

$$\Psi_{\bar{\vartheta}}(\vartheta) := P(\vartheta)(\eta(\vartheta) - \eta(\bar{\vartheta})) + \eta(\bar{\vartheta}).$$

The set $T^*(2r_0)$ constructed this way is a tube of the radius r_0 around the surface $\{\Psi_{\bar{\vartheta}}(\vartheta) : \vartheta \in \Theta(2r_0)\}$. A ball of the radius r centered at $\eta(\bar{\vartheta})$ will be denoted by $G(r)$:

$$G(r) := \{y \in \mathbb{R}^N : \|y - \eta(\bar{\vartheta})\|_W < r\}.$$

In [7] it is proved that there exists a radius r_0 such that

$$G(r_0) \subseteq T^*(2r_0) \subseteq G(2r_0), \quad (3.2)$$

under two assumptions:

A1: *The assumption of the bounded curvature:* The inequality

$$[K_{int}(\vartheta)]^{-1} > 2r_0,$$

holds for every $\vartheta \in \Theta(2r_0)$.

A2: *The assumption of non-overlapping:* There is no $y \in T^*(2r_0)$ such that the normal equation(s) (2.2) have two solutions $\vartheta^1, \vartheta^2 \in \text{int } \Theta$, and

$$\|y - \eta(\vartheta^i)\|_W < 2r_0, \quad (i = 1, 2).$$

Thus, $G(r_0)$ is a region of unique minima of the residual sum of squares function, meaning that for every $y \in G(r_0)$, the function $\|y - \eta(\vartheta)\|_W^2$ cannot have more than one local minimum attainable at $\hat{\vartheta} = \hat{\vartheta}(y)$, where a global minimum is also reached.

Given that $\|y - \eta(\bar{\vartheta})\|_W^2/\sigma^2$ follows the χ^2 distribution with $N - 1$ degrees of freedom under assumptions made in (2.1), the probability statement consistent with (3.2) is

$$\mathbb{P} \left\{ \chi_{N-1}^2 \leq \left(\frac{r_0}{\sigma} \right)^2 \right\} \leq \mathbb{P} \{y \in T^*(2r_0)\} \leq \mathbb{P} \left\{ \chi_{N-1}^2 \leq \left(\frac{2r_0}{\sigma} \right)^2 \right\}.$$

Remark 3.2. If r_0 is too large, then the two assumptions could be violated, while if it is too small, then the accuracy of the approximations will be unsatisfactory. Therefore, r_0 need to be large enough so that $pr \left\{ \chi_{N-1}^2 \leq \left(\frac{r_0}{\sigma} \right)^2 \right\} = 1 - \alpha$, or equivalently, $\sigma(r_0, \alpha) = \frac{r_0}{\sqrt{\chi_{N-1, \alpha}^2}}$.

In what follows we assume the following condition that ensures that $T^*(2r_0)$ is a RIGM region:

(H): Let $r_0 > 0$ and $\vartheta \in \Theta(2r_0)$. If $y \in T^*(2r_0)$ satisfies the stationary conditions (2.2) then $\vartheta = \hat{\vartheta}(y)$.

Theorem 3.3. *A necessary and sufficient condition for a radius r_0 to satisfy condition (H) is*

$$r_0 \leq \inf_{\vartheta \in \Theta(2r_0), \vartheta_a \in \Theta, \vartheta_a \neq \vartheta} r(\vartheta, \vartheta_a) \quad (3.3)$$

where

$$r(\vartheta, \vartheta_a) = \frac{\|\eta(\vartheta_a) - \eta(\vartheta)\|_W^2 + 2\langle (I - P(\vartheta))(\eta(\vartheta) - \eta(\bar{\vartheta})), \eta(\vartheta_a) - \eta(\vartheta) \rangle_W}{2\|(I - P(\vartheta))(\eta(\vartheta_a) - \eta(\vartheta))\|_W}$$

Proof. In order to simplify the expressions, the following abbreviations are used below:

$$\begin{aligned} c &= c(\vartheta_a, \vartheta) = \eta(\vartheta_a) - \eta(\vartheta), \\ f &= f(\vartheta, \bar{\vartheta}) = \eta(\bar{\vartheta}) - \eta(\vartheta), \\ k &= k(\vartheta, \bar{\vartheta}) = \|\eta(\vartheta) - \eta(\bar{\vartheta})\|_W^2, \quad e = e(\vartheta_a, \vartheta) = \|\eta(\vartheta_a) - \eta(\vartheta)\|_W^2, \\ l &= l(y, \vartheta, \bar{\vartheta}) = y - \psi_{\bar{\vartheta}}(\vartheta), \quad g = g(y, \vartheta) = y - \eta(\vartheta), \\ q &= (I - P(\vartheta))f(\vartheta, \bar{\vartheta}), \quad u = (I - P(\vartheta))c(\vartheta_a, \vartheta), \quad t = t(\vartheta, \vartheta_a, n). \end{aligned}$$

Let $y \in T^*(2r_0)$. The assumption (H) will be met if and only if

$$\|l + q\|_W \leq t(\vartheta, \vartheta_a, n) \quad (3.4)$$

for any $\vartheta \in \Theta(2r_0)$, $\vartheta_a \in \Theta$, $n = \frac{l+q}{\|l+q\|_W} \in NO_1(\vartheta)$ such that $\langle n, c \rangle_W > 0$. Note that both $l, q \in NO(\vartheta)$.

Given (2.4), the inequality (3.4) is equivalent to

$$2\langle l + q, c \rangle_W \leq e.$$

This inequality is satisfied automatically if $\langle n, c \rangle_W \leq 0$, therefore one can neglect the sign of $\langle n, c \rangle_W$ in connection with (3.4) and rearrange it as

$$2\langle l, c \rangle_W \leq e - 2\langle q, c \rangle_W. \quad (3.5)$$

The maximum of the left-hand side of (3.5) over all $l \in NO(\vartheta)$ with a fixed $\|l\|_W$ equals $2\|l\|_W\|u\|_W$ and is attained at $l = u \frac{\|l\|_W}{\|u\|_W}$. Combined with the inequality $\|l\|_W \leq r_0$, this leads to the inequality

$$2r_0\|u\|_W \leq e - 2\|q, c\|_W.$$

Thus, the assumption (H) holds true if and only if

$$2r_0\|(I - P(\vartheta))(\eta(\vartheta_a) - \eta(\vartheta))\|_W \leq \|\eta(\vartheta_a) - \eta(\vartheta)\|_W^2 - 2\langle (I - P(\vartheta))(\eta(\vartheta) - \eta(\bar{\vartheta})), \eta(\vartheta_a) - \eta(\vartheta) \rangle_W.$$

We obtained that a necessary and sufficient condition in order for r_0 to satisfy assumption (H) that ensures $T^*(2r_0)$ a RIGM region is

$$r_0 \leq \frac{\|\eta(\vartheta_a) - \eta(\vartheta)\|_W^2 + 2\langle (I - P(\vartheta))(\eta(\vartheta) - \eta(\bar{\vartheta})), \eta(\vartheta_a) - \eta(\vartheta) \rangle_W}{2\|(I - P(\vartheta))(\eta(\vartheta_a) - \eta(\vartheta))\|_W}, \quad (3.6)$$

for any $\vartheta \in \Theta(2r_0)$ and $\vartheta_a \in \Theta$ such that $\vartheta_a \neq \vartheta$. Denote the right-hand side of equation (3.6) with $r(\vartheta, \vartheta_a)$. The largest value of r_0 that makes the tube $T^*(2r_0)$ a RIGM region is

$$r_0 = \inf_{\vartheta \in \Theta(2r_0), \vartheta_a \in \Theta, \vartheta_a \neq \vartheta} r(\vartheta, \vartheta_a), \quad (3.7)$$

which represents a global infimum of

$$\frac{\|\eta(\vartheta_a) - \eta(\vartheta)\|_W^2 + 2\langle (I - P(\vartheta))(\eta(\vartheta) - \eta(\bar{\vartheta})), \eta(\vartheta_a) - \eta(\vartheta) \rangle_W}{2\|(I - P(\vartheta))(\eta(\vartheta_a) - \eta(\vartheta))\|_W}$$

over $\vartheta \in \text{int } \Theta$, $\vartheta_a \in \Theta$, $\vartheta_a \neq \vartheta$, subject to the constraints:

$$\|P(\vartheta)(\eta(\vartheta) - \eta(\bar{\vartheta}))\|_W \leq r_0, \quad \text{and} \quad \|\eta(\vartheta) - \eta(\bar{\vartheta})\|_W \leq 2r_0.$$

□

Next, let's consider the tube around the expectation surface defined in [8]. Let $G(r)$ denotes a ball of the radius r centered at $\eta(\bar{\vartheta})$.

Definition 3.4. A parameter value, ϑ , is called an r -projection of y if

$$\|y - \eta(\vartheta)\|_W < r$$

and either one of the following holds:

- (i) stationary conditions (2.2) are satisfied;
- (ii) $\vartheta = \hat{\vartheta}(y)$ lies on the boundary of Θ .

The set of all r -projections of the points belonging to $G(r)$ is denoted as $B(r)$. The tube $T(r)$ around the expectation surface E is

$$T(r) = \{y \in \mathbb{R}^N : \exists \vartheta \in B(r) \text{ such that } \vartheta \text{ is an } r\text{-projection of } y\}.$$

In [8], Pronzato and Pazman made the following assumption:

(H_S): there exists $r > 0$ such that:

- (a) the probability $P_{\bar{\vartheta}}[G(r)] = \mathbb{P}(\|y - \eta(\bar{\vartheta})\|_W < r)$ is close to one;
- (b) every $y \in T(r)$ has one r -projection only.

Our assumption (H_{SM}) is a variant of (H_S) which replaces (b) with

- (bm) If $y \in T(r)$ satisfies the stationary conditions (2.2) with some $\vartheta \in \text{int } \Theta$ then $\vartheta = \hat{\vartheta}(y)$.

The next theorem provides a computational method for obtaining a maximal radius for the tube $T(r)$ to be a RIGM region.

Theorem 3.5. *The maximal value of the radius r satisfying the assumption (H_{SM}) is obtained as the solution to a constrained optimization problem:*

$$r_{max} = \inf_{\vartheta \in \text{int } \Theta, \vartheta_a \in \Theta, \vartheta \neq \vartheta_a} t(\vartheta, \vartheta_a),$$

subject to the constraint $\|P(\vartheta)f\|_W < r_{max}$.

Proof. Let r be the maximal radius of the ball $G(r)$ described in Theorem 3.3. Using the abbreviations introduced in the proof of Theorem 3.3, we write the Law of Cosines for a triangle formed by $y - \eta(\bar{\vartheta})$, f , and g as:

$$\|y - \eta(\bar{\vartheta})\|^2 = k + s^2 - 2s\langle f, n \rangle < r^2.$$

The expression $k + s^2 - 2s\langle f, n \rangle$ is a quadratic function of s that reaches its minimum of $k - \langle f, n \rangle^2$ at $s = \langle f, n \rangle$. Thus, the inequality $k - \langle f, n \rangle^2 < r^2$ is a necessary and sufficient condition that there exists $s \in \mathbb{R}$ such that $y = \eta(\vartheta) + sn$ belongs to the ball $G(r)$. Additionally,

$$\min_{n \in NO_1(\vartheta)} (k - \langle f, n \rangle^2) = k - \|(I - P(\vartheta))f\|_W^2 = \|P(\vartheta)f\|_W^2,$$

meaning that for a given $\vartheta \in \text{int } \Theta$, there exists $s \in \mathbb{R}$ and $n \in NO_1(\vartheta)$ such that $y = \eta(\vartheta) + sn$ belongs to $G(r)$ if and only if $\|P(\vartheta)f\|_W < r$. In terms of r -projections, this means that every $\vartheta \in \text{int } \Theta \cap B(r)$ satisfies the inequality $\|P(\vartheta)f\|_W < r$, and vice versa. Since a continuous function $\|y - \eta(\vartheta)\|_W^2$ reaches

its absolute (global) minimum on a compact set Θ , such minimum will be reached either at $\vartheta \in \text{int } \Theta$ or at $\vartheta \in \partial\Theta$. The condition (b) of (H_S) requires that

$$\inf_{\vartheta \in \text{int } \Theta \cap B(r), \vartheta_a \in \theta, \vartheta_a \neq \vartheta} t(\vartheta, \vartheta_a) \geq r.$$

Indeed, assume there are $\vartheta \in \text{int } \Theta \cap B(r)$, $\vartheta_a \in \theta$, $\vartheta_a \neq \vartheta$ such that $t(\vartheta, \vartheta_a) < r$. Then $y = \eta(\vartheta) + t(\vartheta, \vartheta_a)(I - P(\vartheta))c$ is equidistant from $\eta(\vartheta)$ and $\eta(\vartheta_a)$:

$$\|y - \eta(\vartheta)\|_W = \|y - \eta(\vartheta_a)\|_W = t(\vartheta, \vartheta_a).$$

As shown in [6], $y_1 = \eta(\vartheta) + (t(\vartheta, \vartheta_a) + \varepsilon)(I - P(\vartheta))c$ will satisfy an inequality

$$\|y_1 - \eta(\vartheta_a)\|_W < \|y_1 - \eta(\vartheta)\|_W$$

meaning that ϑ is not a global minimum of $\|y_1 - \eta(\vartheta)\|_W^2$ on Θ . This gives the maximal value of r as the solution to a constrained optimization problem:

$$r_{max} = \inf_{\vartheta \in \text{int } \Theta \cap B(r_{max}), \vartheta_a \in \Theta, \vartheta \neq \vartheta_a} t(\vartheta, \vartheta_a) = \inf_{\vartheta \in \text{int } \Theta, \vartheta_a \in \Theta, \vartheta \neq \vartheta_a} t(\vartheta, \vartheta_a), \quad (3.8)$$

subject to the constraint $\|P(\vartheta)f\|_W < r_{max}$. \square

It should be noted that, the importance of constructing the regions of unique minima and then estimating their size is not limited to applications to flat models as in Pazman's work.

Application of the formula (3.1) developed in this paper requires global optimization while equations (3.7) and (3.8) require constrained global optimization along with numerically solving these equations. The left-hand sides of these equations are monotonically increasing functions of r while the right-hand sides are monotonically non-increasing functions of r . Hence, each of these equations has a single solution which can be found numerically by zeroing in on it in a systematic fashion using a variable increment. One can use $\sqrt{\bar{S}_{LU}}$ as the appropriate initial value for this purpose. Until more advanced methods for solving unconstrained and constrained optimization problems are developed, it is advisable to try and mitigate potential non-optimal solutions by using a number of different initial values when computing right-hand sides of equations (3.1), (3.7) and (3.8). This is precisely how the numerical calculations were performed in the two examples presented below.

Note also that in the presence of a linear parameter (partially linear models), such as Michaelis–Menten model illustrated below, the calculations are much simplified. When $\eta(\vartheta) = a\varphi(\theta)$ (here $\vartheta = (a, \theta)$), the expression (2.5) for the function $t(\vartheta, \vartheta_a)$ becomes

$$t(a, \theta, a_a, \theta_a) = \frac{a_a^2 \|\varphi(\theta_a)\|_W^2 - 2a_a a \langle \varphi(\theta_a), \varphi(\theta) \rangle_W + a^2 \|\varphi(\theta)\|_W^2}{2a_a \|(I - P(\vartheta))\varphi(\theta_a)\|_W}.$$

Here $a_a > 0$ and also assume $a > 0$. A simple calculation of the first and second derivatives of $t(a, \theta, a_a, \theta_a)$ with respect to a_a shows that the absolute minimum of $t(a, \theta, a_a, \theta_a)$ over a_a is reached at $a_a = \frac{a \|\varphi(\theta)\|_W}{\|\varphi(\theta_a)\|_W}$ and equals to

$$\frac{a(\|\varphi(\theta)\|_W \|\varphi(\theta_a)\|_W - \langle \varphi(\theta_a), \varphi(\theta) \rangle_W)}{\|(I - P(\vartheta))\varphi(\theta_a)\|_W}.$$

4. APPLICATION TO TWO POPULAR MODELS

The following models were used for an illustration purpose in this paper:

- (1) Michaelis–Menten model

$$\eta(\theta) = (f(x_1, \theta), \dots, f(x_N, \theta)); \quad \theta = (\theta_1, \theta_2);$$

$$f(x, \theta) = \frac{\theta_1 x}{\theta_2 + x}; \quad x_i = 0.2, 0.222, 0.286, 0.4, 0.667, 2; \quad (i = 1, \dots, 6); \quad W = I.$$

The model was used with three different parameter spaces:

- (A) $\Theta = [0, 50] \times [0, 20]$;
 (B) $\Theta = [2, 50] \times [0, 20]$;
 (C) $\Theta = [2, 50] \times [0.2, 2]$.

The design used in this model originally appeared in Watts [10] where it was duplicated. It has been widely referenced in the literature [9]. The design points are not replicated in this paper.

- (2) Two-parameter Hill model

$$\eta(\theta) = (f(x_1, \theta), \dots, f(x_N, \theta)); \quad \theta = (\theta_1, \theta_2);$$

$$f(x, \theta) = \frac{\left(\frac{x}{\theta_1}\right)^{\theta_2}}{1 + \left(\frac{x}{\theta_1}\right)^{\theta_2}}; \quad x_i = 0.316, 1, 3.16, 10, 31.6, 100; \quad (i = 1, \dots, 6); \quad W = I.$$

The model was used with two different parameter spaces:

- (A) $\Theta = [0.316, 30] \times [-20, 0]$;
 (B) $\Theta = [0.316, 30] \times [-6, 0]$.

The design used here is a geometric design with a dilution factor of $\sqrt{10}$. Such design is evenly spaced on a logarithmic scale and is often used in biopharmaceutical applications of this model. Although a heterogeneous power model is commonly used in these applications, a simpler homogeneous model is used in this paper.

Both models are widely applied in the fields such as pharmacokinetics and biochemistry.

The numerical results illustrating the methodology developed in this paper are presented in Tables 1 and 2. For each of the two models, the minimal radius of the intrinsic curvature, $r_1 = \min_{\theta \in \Theta} R_{int}(\theta)$, its global analogue r_2 given by formula (3.1), the maximal r_0 given by (3.7), and r_{max} , the solution to the constrained optimization problem (3.8), were calculated. In Table 2 it is assumed that the true parameter vector of Michaelis–Menten model, $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2)$, has $\bar{\theta}_1 = 1$ and that $\bar{\theta}_2$ equals to one of the values, 0.5, 1, 10. In Table 1 it is assumed that the true parameter vector of the Hill model, $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2)$, has $\bar{\theta}_1 = 1$ and that $\bar{\theta}_2$ equals to one of the values, -0.5 , -1.5 , -3 . The scenarios (A)–(C) and (A)–(B) considering different parameter spaces are described above. Notably, in the Michaelis–Menten model both r_0 and r_{max} decrease monotonically when $\bar{\theta}_2$ increases from 0.5 to 10. Since θ_2 is the x -value resulting in $\theta_1/2$, one would expect that r_0 and r_{max} become small when $\bar{\theta}_2$ lies outside $[0.2, 2]$, the interval containing all design points. This is apparent from Table 1, also showing that r_0 and r_{max} are close to one another.

Table 1 also illustrates that, for a fixed design, the minimum of the local curvature radius is highly dependent on the assumed parameter space. Interestingly, under each of three scenarios, (A)–(C), the global measure, r_2 , coincided with the local one given by r_1 . Evidently, this is not true in the case (A) of Table 2, which illustrates the value of global measures related to the nonlinearity of the model.

The values of $\sigma(r_{max}, \alpha)$ for which $P(G(r_{max})) = 1 - \alpha$ are given in both tables. Similarly, one can find σ_0 such that $P(G(r_{r_0})) = 1 - \alpha$.

Table 1. 2-parameter Hill model.

Scenario	$\min R_{int}(\theta)$	r (formula (3.1))
(A)	0.235	0.046
(B)	0.232	0.006

θ_2	r_{max} (formula (3.8))	r_0 (formula (3.7))	$\sigma(r_{max}, 0.01)$
-0.5	0.312	0.268	0.080
-1.5	0.250	0.204	0.064
-3	0.243	0.235	0.063

Table 2. Michaelis–Menten model.

Scenario	$\min R_{int}(\theta)$	r (formula (3.1))
(A)	0	0
(B)	0.263	0.263
(C)	1.24	1.24

θ_2	r_{max} (formula (3.8))	r_0 (formula (3.7))	$\sigma(r_{max}, 0.01)$
0.5	0.483	0.683	0.124
1	0.365	0.516	0.094
10	0.083	0.110	0.021

5. CONCLUSION

The residual sum of squares attains a single local minimum in linear regression that is also a global minimum. While this generally is not true in nonlinear regression problems, there exists regions possessing similar property. Tubes around either the expectation surface or another closely related surface represent an important class of such regions and were studied in [7] and [8]. However, a maximal radius of any RIGM region (and, specifically, a tube), although highly desirable, is not currently available in statistical literature. This paper shows how such measures can be computed as solutions of global optimization problems or as solutions to related equations. The method relies heavily on properties of equidistant function originally introduced in [6]. This function plays a prominent role in the definition

of extended intrinsic curvature of a nonlinear model presented in [8]. The measures introduced in this paper are illustrated using two models, Michaelis–Menten and Hill. We anticipate that as computational experience using these measures with different models is accumulated, a relationship between them and different global aspects of nonlinear models will become apparent.

REFERENCES

- [1] D. M. Bates and D. G. Watts, *Relative curvature measures of nonlinearity*, Journal of the Royal Statistical Society, Series B **42** (1980), 1–25.
- [2] E. M. L. Beale, *Confidence regions in nonlinear estimation*, Journal of the Royal Statistical Society, Series B **22** (1960), 41–76.
- [3] E. Demidenko, *Is this the least squares estimate?*, Biometrika **87** (2000), 437–452.
- [4] E. Demidenko, *Criteria for global minimum of sum of squares in nonlinear regression*, Computational Statistics and Data Analysis **51** (2006), 1739–1753.
- [5] L. Haines, *A note on the differential geometry of least squares estimation for nonlinear regression models*, South African Statistical Journal **28** (1994), 73–91.
- [6] L. Khinkis and M. Crotzer, *A new approach for finding global minima in nonlinear least squares regression*, in: Proceedings of the American Statistical Association, Biopharmaceutical Section, Joint Statistical Meeting, 2008, 226–2271.
- [7] A. Pázman, *Nonlinear Statistical Models*, Mathematics and its Applications **254**, Kluwer Academic Publishers, Dordrecht, 1993.
- [8] L. Pronzato and A. Pázman, *Design of Experiments in Nonlinear Models*, Lecture Notes in Statistics **212**, Springer-Verlag, New York, 2013.
- [9] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, 1989.
- [10] D. G. Watts, *An introduction to nonlinear least squares*, in: L. Endrenyi (ed.), Kinetic Data Analysis: Design and Analysis of Enzyme and Pharmacokinetic Experiments, Springer US, 1981, 1–24.

Leonid Khinkis, Department of Mathematics and Statistics, Canisius College, Buffalo, USA
e-mail: khinkis@canisius.edu

Milburn Crotzer, Department of Mathematics and Statistics, Canisius College, Buffalo, USA
e-mail: crotzerm@canisius.edu

Adina Oprisan, Department of Mathematics and Statistics, Canisius College, Buffalo, USA
e-mail: oprisana@canisius.edu