Andrej Pázman

Geometry of Gaussian nonlinear regression. Parallel curves and confidence intervals

Kybernetika, Vol. 18 (1982), No. 5, 376--396

Persistent URL: http://dml.cz/dmlcz/125872

KYBERNETIKA — VOLUME 18 (1982), NUMBER 5

GEOMETRY OF GAUSSIAN NONLINEAR REGRESSION — PARALLEL CURVES AND CONFIDENCE INTERVALS

ANDREJ PÁZMAN

The regression model $y_i = \eta_i(\theta_1, \ldots, \theta_m) + \varepsilon_i$; $(i = 1, \ldots, N, \theta \in \Theta)$ is considered, with η_i nonlinear and continuously twice differentiable and with $(\varepsilon_1, \ldots, \varepsilon_N) \sim N(0, \sum)$, $\sum$ known. The one-dimensional case (i.e. $m = 1$) is investigated mainly. A nonlinear geometry of the sample space R^N is proposed and curves parallel to the set $\{(\eta_1(\theta), \ldots, \eta_N(\theta));\ \theta \in \Theta\}$ are considered. The results are used to construct confidence intervals for θ.

1. INTRODUCTION

Let us consider the regression model

$$y_i = \eta_i(\theta) + \varepsilon_i\,; \quad (i = 1, \ldots, N) \tag{1}$$

where $\mathbf{y} := (y_1, \ldots, y_N)'$ is the vector of observed variables, $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_m)' \in \Theta \subset \mathbb{R}^m$ is the vector of unknown parameters and $\boldsymbol{\varepsilon} := (\varepsilon_1, \ldots, \varepsilon_N)$ is the vector of random errors which are supposed to be distributed normally, $N(0, \Sigma)$, with a known nonsingular covariance matrix Σ. The *model functions* $\eta_1(\cdot), \ldots, \eta_N(\cdot)$ are known and generally they are not linear in $\boldsymbol{\theta}$. Such a model is often met in applications. Usually the parameters $\theta_1, \ldots, \theta_m$ are estimated by the minimization of the residual form. The resulting estimates

$$\hat{\theta}(\mathbf{y}) := \operatorname{Arg}\min_{\theta \in \Theta} [\mathbf{y} - \boldsymbol{\eta}(\theta)]' \Sigma^{-1} [\mathbf{y} - \boldsymbol{\eta}(\theta)] \tag{2}$$

are the l.s. (= least squares) estimates.

Usually the model functions $\eta_1, \ldots, \eta_N$ are smooth enough and the parameter set Θ has a simple form (e.g. it is an m-dimensional interval), eventually with a not quite clearly defined boundary. This allows to specify the assumptions on Θ and on the model functions in a mathematically convenient way. We shall suppose that the set Θ is convex and compact and that it has a nonvoid interior in $\mathbb{R}^m$ (i.e. it is really m-dimensional). Further we shall suppose that there is an open set $U \supset \Theta$ such that the functions $\eta_1, \ldots, \eta_N$ are defined and have continuous second order derivatives

$$\frac{\partial^2 \eta_i(\boldsymbol{\theta})}{\partial\theta_j \partial\theta_k}; \quad (i = 1, \ldots, N\,,\ j, k = 1, \ldots, m)$$

on the set U. Finally we shall suppose that the mapping

$$\boldsymbol{\eta} : \theta \in \Theta \mapsto (\eta_1(\boldsymbol{\theta}), \ldots, \eta_N(\boldsymbol{\theta})) \in \mathbb{R}^N \tag{3}$$

is one-to-one and that for every $\boldsymbol{\theta} \in \Theta$ the vectors $\partial\boldsymbol{\eta}(\boldsymbol{\theta})/\partial\theta_1, \ldots, \partial\boldsymbol{\eta}(\boldsymbol{\theta})/\partial\theta_m$ are linearly independent. This last assumption allows to consider, equivalently, instead of the parameter set Θ the set of potentially possible mean values

$$\mathscr{E} := \{\boldsymbol{\eta}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}\,. \tag{4}$$

It is called the solution locus in some papers (cf. [2, 3]). We shall prefer to call it the *mean-values manifold.*

The advantage of the mean-values manifold when compared with the parameter set Θ is that it does not depend on the particular parametrization used in the model.

Most algorithms for computing the l.s. estimates and most inference methods for nonlinear models including the asymptotic methods are based on a local linear approximation of the model. That is, in a neighbourhood of a fixed value $\boldsymbol{\theta}^0 \in \Theta$ the model functions are approximated by

$$\eta_j(\boldsymbol{\theta}) \doteq \eta_j(\boldsymbol{\theta}^0) + \sum_{i=1}^{m} \left.\frac{\partial \eta_j}{\partial \theta_i}\right|_{\boldsymbol{\theta}^0} (\theta_i - \theta_i^0) \tag{5}$$

That means, the mean-values manifold $\mathscr{E}$ is approximated by its tangent plane at the point $\boldsymbol{\eta}(\boldsymbol{\theta}^0)$. The quality of the approximation is to be measured with the aid of the norm

$$\|\mathbf{v}\|_{\boldsymbol{\Sigma}}^2 := \mathbf{v}'\boldsymbol{\Sigma}^{-1}\mathbf{v}\,; \quad (\mathbf{v} \in \mathbb{R}^N)\,. \tag{6}$$

Statistically this means that the approximation is good if the variances of the observed variables $y_1, \ldots, y_N$, and of their linear combinations, are small, when compared with the curvatures of $\mathscr{E}$. If we need to appreciate the covariance matrix of $\hat{\boldsymbol{\theta}}$ within this approximation, we compute the local information matrix

$$\{\mathrm{M}(\boldsymbol{\theta})\}_{ij} := \frac{\partial \boldsymbol{\eta}'(\boldsymbol{\theta})}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_j}; \quad (i, j = 1, \ldots, m) \tag{7}$$

at the point $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ and take $\mathbf{M}^{-1}(\boldsymbol{\theta}^0)$ for the covariance matrix of $\hat{\boldsymbol{\theta}}$.

In [3] an investigation was performed to clarify how nonlinearities of $\mathscr{E}$ perturb the confidence level of linear confidence intervals and confidence regions. Measures of nonlinearity of $\mathscr{E}$, which indicate the adequacy of a linear approximation, were obtained for such purposes (cf. [2]). Such investigations were based, in principle, on differential geometric analysis of the mean-values manifold $\mathscr{E}$ or, equivalently, of the parameter set Θ. Yet in the paper by Rao (cf. [6]) it is proposed to measure the distance of two nearby points $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}$ in Θ by

$$\mathrm{d}(\boldsymbol{\theta}, \boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}) := [\mathrm{d}\boldsymbol{\theta}' \, \mathbf{M}(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}]^{1/2}$$

and the distance of two arbitrary points $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)} \in \Theta$ by the length of the shortest curve in Θ connecting $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$. (For an interesting development of this Rao's idea cf. [1]). Since, according to (7),

$$\mathrm{d}\boldsymbol{\theta}' \, \mathbf{M}(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} = \left[\sum_{i=1}^{m} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_i} \, \mathrm{d}\theta_i\right]' \boldsymbol{\Sigma}^{-1} \left[\sum_{j=1}^{m} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_j} \, \mathrm{d}\theta_j\right], \tag{8}$$

the corresponding distance in the mean-values manifold is the "usual distance" on a manifold in $\mathbb{R}^N$. but instead of the Euclidean norm and of the Euclidean inner product in $\mathbb{R}^N$, the inner product

$$\langle \boldsymbol{u}, \mathbf{v} \rangle_{\boldsymbol{\Sigma}} = \boldsymbol{u}' \boldsymbol{\Sigma}^{-1} \mathbf{v} \, ; \quad (\boldsymbol{u}, \mathbf{v} \in \mathbb{R}^N) \tag{9}$$

and the norm (6) are used.

In the presented paper the stress is not on the geometry of $\mathscr{E}$ or of Θ, but on the geometry of the sample space. Of course the sample space may be considered as a Hilbert space with the inner product (9). This is adequate for linear regression models. It is motivated also by the expression

$$\frac{\mathrm{d}P_{\boldsymbol{\theta}}(\mathbf{y})}{\mathrm{d}\lambda} = \frac{1}{(2\pi)^{N/2} \, [\det \boldsymbol{\Sigma}]^{1/2}} \exp\{-\tfrac{1}{2} \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}}^2\} \tag{10}$$

for the probability density of the observed vector $\mathbf{y}$ with respect to the Lebesgue measure λ in $\mathbb{R}^N$. However in nonlinear regression models such a linear structure of the sample space is in contradiction with the nonlinear structure of the mean-values manifold. In Section 2 we propose a nonlinear (local) geometry of the sample space, which is induced by the nonlinearity of $\mathscr{E}$. To do this we use that

(i) the l.s. estimate $\hat{\boldsymbol{\theta}}(\mathbf{y})$ given in (2) associate a point $\boldsymbol{\eta}[\hat{\boldsymbol{\theta}}(\mathbf{y})]$ of $\mathscr{E}$ with each point $\mathbf{y}$ of the sample space $\mathbb{R}^N$,

(ii) the set of sample points which give the same l.s. estimate $\hat{\boldsymbol{\theta}}$, in symbols,

$$L_{\hat{\boldsymbol{\theta}}} := \{\mathbf{y} : \mathbf{y} \in \mathbb{R}^N, \|\mathbf{y} - \boldsymbol{\eta}(\hat{\boldsymbol{\theta}})\|_{\boldsymbol{\Sigma}} = \min_{\boldsymbol{\theta} \in \Theta} \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}}\}$$

is "locally orthogonal" to any curve in $\mathbb{R}^N$ which is parallel to $\mathscr{E}$ and which intersects $L_{\hat{\boldsymbol{\theta}}}$ (see Section 2 and Appendix).

Hence curves in the sample space, which are parallel to $\mathscr{E}$, are of statistical interest. They are investigated in Appendix. The conditional probability density under the condition that the sample is on a given parallel curve is obtained in Section 2. This is the base to construct, at least in principle, the confidence intervals or the confidence regions for $\theta_1, \ldots, \theta_m$.

To clarify the ideas, the main attention will be paid to the case of the one-dimensional regression. In that case, under some regularity conditions, explicit formulae for computing confidence intervals are obtained.

2. THE ONE-DIMENSIONAL NONLINEAR REGRESSION

2.1. The Conditional Probability on Parallel Curves

We shal consider in this section the special case when

$$\Theta := \langle \underline{\theta}, \bar{\theta} \rangle$$

is an interval in $\mathbb{R}^1$. The model functions $\eta_1(\cdot), \ldots, \eta_N(\cdot)$ define a mapping

$$\theta \in \langle \underline{\theta}, \bar{\theta} \rangle \mapsto \boldsymbol{\eta}(\theta) \in \mathbb{R}^N \tag{11}$$

which can be interpreted geometrically as a curve in $\mathbb{R}^N$. According to the assumptions presented in the Introduction, $\boldsymbol{\eta}$ has continuous derivatives $\mathrm{d}\boldsymbol{\eta}(\theta)/\mathrm{d}\theta$, $\mathrm{d}^2\boldsymbol{\eta}(\theta)/\mathrm{d}\theta^2$, and $\|\mathrm{d}\boldsymbol{\eta}(\theta)/\mathrm{d}\theta\|_{\boldsymbol{\Sigma}} \neq 0$ for every $\theta \in \langle \underline{\theta}, \bar{\theta} \rangle$. The expression

$$T := \int_{\underline{\theta}}^{\bar{\theta}} \left\| \frac{\mathrm{d}\boldsymbol{\eta}(\theta)}{\mathrm{d}\theta} \right\|_{\boldsymbol{\Sigma}} \mathrm{d}\theta$$

is the *length* of the curve. Since $\|\mathrm{d}\boldsymbol{\eta}(\theta)/\mathrm{d}\theta\|_{\boldsymbol{\Sigma}} \neq 0$, we can introduce a new parameter t by

$$\frac{\mathrm{d}t}{\mathrm{d}\theta} = \left\| \frac{\mathrm{d}\boldsymbol{\eta}(\theta)}{\mathrm{d}\theta} \right\|_{\boldsymbol{\Sigma}} ; \quad (\theta \in \langle \underline{\theta}, \bar{\theta} \rangle) \tag{12}$$

and by the initial condition $t(\bar{\theta}) = 0$. With this "natural" parametrization we obtain a new mapping

$$\boldsymbol{\delta}_\eta : t \in \langle 0, T \rangle \mapsto \boldsymbol{\eta}[\theta(t)] \in \mathbb{R}^N \tag{13}$$

which corresponds to the same geometrical picture in $\mathbb{R}^N$ as the mapping (11).

Generally, *a curve* in $\mathbb{R}^N$ will be defined as a mapping

$$\gamma : u \in \langle 0, U \rangle \mapsto \gamma(u) \in \mathbb{R}^N$$

which is continuous on $\langle 0, U \rangle$ and has a continuous second order derivative $\mathrm{d}^2\gamma/\mathrm{d}u^2$ on the set $\langle 0, U \rangle - C_\gamma$, where C_γ is a given closed set in $\mathbb{R}^1$ which contains no intervals. Moreover we shall suppose, unless otherwise stated, that the parameter u is "natural", i.e.

$$\left\| \frac{\mathrm{d}\gamma}{\mathrm{d}u} \right\|_{\boldsymbol{\Sigma}} = 1 ; \quad (u \in \langle 0, U \rangle - C_\gamma) . \tag{14}$$

Points $\gamma(u)$; $(u \in C_\gamma)$ are called *singular*. The curve γ will be called *regular* iff $C_\gamma = \varnothing$

and, for some $\varepsilon > 0$, γ can be extended to a curve

$$\tilde{\gamma} : \langle 0 - \varepsilon, U + \varepsilon \rangle \mapsto \mathbb{R}^N$$

with a continuous derivative $\mathrm{d}^2\tilde{\gamma}/\mathrm{d}u^2$ also at 0 and U.

A curve $\gamma : \langle 0, U \rangle \mapsto \mathbb{R}^N$ will be called *parallel* to a regular curve

$$\boldsymbol{\delta} : \langle 0, T \rangle \mapsto \mathbb{R}^N$$

iff there is a differentiable function

$$\varphi_\gamma : \langle 0, T \rangle \xrightarrow{\text{onto}} \langle 0, U \rangle$$

such that

$$\text{(15)} \quad \text{(i)} \qquad \frac{\mathrm{d}\varphi_\gamma}{\mathrm{d}t} \geqq 0 \,, \quad \text{and} \quad \frac{\mathrm{d}\varphi_\gamma}{\mathrm{d}t} = 0 \Leftrightarrow \varphi_\gamma(t) \in C_\gamma \,,$$

$$\text{(16)} \quad \text{(ii)} \qquad \left\langle \delta(t) - \gamma[\varphi_\gamma(t)] \,, \frac{\mathrm{d}\delta(t)}{\mathrm{d}t} \right\rangle_{\Sigma} = 0 \,; \quad (t \in \langle 0, T \rangle) \,,$$

$$\text{(17)} \quad \text{(iii)} \qquad \frac{\mathrm{d}\delta(t)}{\mathrm{d}t} = \pm \left.\frac{\mathrm{d}\delta}{\mathrm{d}u}\right|_{u = \varphi_\gamma(t)} ; \; (t \in \langle 0, T \rangle, \; \varphi_\gamma(t) \notin C_\gamma)$$

One-sided derivatives are taken in (15) and (17) if $t = 0$ or $t = T$.

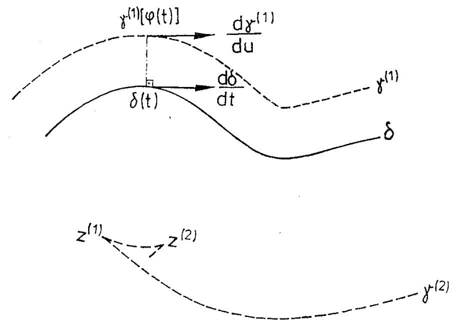

Fig. 1. $\gamma^{(1)}$ and $\gamma^{(2)}$ are curves parallel to δ, $\mathbf{z}^{(1)}$, $\mathbf{z}^{(2)}$ are singular points.

Notation. We shall omit the subscript γ in φ_γ, if there will be no danger of confusion.

Examples. Two concentric circles in $\mathbb{R}^2$ or two parallel straight lines in $\mathbb{R}^N$ are mutually parallel curves. A parallel curve may look quite exotically, as shown in Figure 1. Points where the curve γ changes its direction are points from the set C_γ.

The properties of parallel curves are analysed in the Appendix. We summarize them briefly:

(i) If γ is parallel to a regular curve $\boldsymbol{\delta}$ then

$$\|\gamma[\varphi(t)] - \boldsymbol{\delta}(t)\|_{\boldsymbol{\Sigma}} = \text{const.}\,; \quad (t \in \langle 0, T \rangle)$$

(Proposition A 2). This constant is the distance of γ from $\boldsymbol{\delta}$ and it will be denoted by d_γ.

(ii) A point $\gamma[\varphi(t)]$ is singular if and only if

$$(18) \qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2} \|\gamma(u) - \boldsymbol{\delta}(t)\|^2_{\boldsymbol{\Sigma}}\big|_{u=\varphi(t)} = 0$$

(Corollary to Proposition A 3).

(iii) The expression

$$(19) \qquad \varrho(t) := \left\|\frac{\mathrm{d}^2\boldsymbol{\delta}(t)}{\mathrm{d}t^2}\right\|^{-1}_{\boldsymbol{\Sigma}}$$

is the radius of curvature of the curve $\boldsymbol{\delta}$. As shown in Proposition A 4, if γ is parallel do $\boldsymbol{\delta}$ and if

$$\|\gamma[\varphi(0)] - \boldsymbol{\delta}(0)\|_{\boldsymbol{\Sigma}} < \inf_{t \in \langle 0, T \rangle} \varrho(t)\,,$$

then γ is a regular curve.

(iv) If $\boldsymbol{\delta}$ is a regular curve, if $\mathbf{z}$ is a point in $\mathbb{R}^N$, and if $t_0 \in \langle 0, T \rangle$ is a solution of

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{z} - \boldsymbol{\delta}(t)\|^2_{\boldsymbol{\Sigma}}\big|_{t=t_0} = 0\,,$$

then there is a unique curve γ which is parallel to $\boldsymbol{\delta}$, which contains the point $\mathbf{z}$ (i.e. $\gamma(u) = \mathbf{z}$ for some $u \in \{0, U\rangle$), and which is orthogonal to $\gamma[\varphi(t_0)] - \boldsymbol{\delta}(t_0)$ (i.e. $\langle \gamma[\varphi(t_0)] - \boldsymbol{\delta}(t_0) . \, \mathrm{d}\boldsymbol{\delta}/\mathrm{d}t|_{t_0} \rangle_{\boldsymbol{\Sigma}} = 0$, see Corollary to Proposition A 5). This curve is a solution of the linear differential equations (A 4) and the function φ is given by the integral (A 5).

Let us return to the curve $\boldsymbol{\delta}_{\boldsymbol{\eta}}(t) = \boldsymbol{\eta}[\theta(t)]$ (see Eq. (13)) corresponding to the mean-values manifold $\mathscr{E}$ (see Eq. (4)). Let us consider the following subset of the sample space,

$$(20) \qquad V_{\boldsymbol{\eta}} := \left\{\mathbf{z} : \mathbf{z} \in \mathbb{R}^N\,, \ \exists_{\theta \in (\underline{\theta}, \bar{\theta})} \frac{\mathrm{d}}{\mathrm{d}\theta} \|\mathbf{z} - \boldsymbol{\eta}(\theta)\|^2_{\boldsymbol{\Sigma}} = 0\right\}.$$

The set $V_{\boldsymbol{\eta}}$ contains all samples $\mathbf{y}$ which are such that the l.s. estimate $\hat{\theta}(\mathbf{y})$ is inside $(\underline{\theta}, \bar{\theta})$. Parallel curves are related to local coordinates in $V_{\boldsymbol{\eta}}$ as we shall explain:

First, let us exclude points $\mathbf{z} \in V_{\boldsymbol{\eta}}$ which are singular (i.e. $(\mathrm{d}/\mathrm{d}\theta) \|\mathbf{z} - \boldsymbol{\eta}(\theta)\|^2_{\boldsymbol{\Sigma}} = 0$ and $(\mathrm{d}^2/\mathrm{d}\theta^2) \|\mathbf{z} - \boldsymbol{\eta}(\theta)\|^2_{\boldsymbol{\Sigma}} = 0$), since, as known from differential geometry, the set of such points has a zero Lebesgue measure in $\mathbb{R}^N$, and consequently a zero P_θ measure for every $\theta \in \Theta$ (cf. [5]).

Further let us take a regular point $\mathbf{y}_0 \in V_\eta$ and denote by t_0 a solution of the equation $(d/dt) \|\mathbf{y}_0 - \boldsymbol{\delta}_\eta(t)\|^2_\Sigma = 0$. We can take an open neighbourhood W_{y_0} of $\mathbf{y}_0$ and $\varepsilon > 0$ as explaind in Proposition A 6, and construct local coordinates $u(\mathbf{y}), v_1(\mathbf{y}), \ldots$ $\ldots, v_{N-1}(\mathbf{y})$ for points $\mathbf{y} \in W_{y_0}$. The coordinate $u(\mathbf{y})$ is the value of the natural parameter of the curve γ which is parallel to $\boldsymbol{\delta}_\eta$ and which contains the point $\mathbf{y}$ (i.e. $\gamma[u(\mathbf{y})] = \mathbf{y}$). The coordinates $v_1(\mathbf{y}), \ldots, v_{N-1}(\mathbf{y})$ are defined by

$$v_i(\mathbf{y}) = \langle \mathbf{y} - \boldsymbol{\eta}(\theta_T), \mathbf{b}^{(i)}(t_y) \rangle_\Sigma ; \quad (i = 1, \ldots, N-1),$$

where θ_T is the true value of the vector of parameters, $t_y \in (t_0 - \varepsilon, t_0 + \varepsilon)$ is the solution of the equation $(d/dt) \|\mathbf{y} - \boldsymbol{\delta}_\eta(t)\|^2_\Sigma = 0$, and $\mathbf{b}^{(1)}(t_y), \ldots, \mathbf{b}^{(N-1)}(t_y)$ are orthonormal vectors which are orthogonal to $d\boldsymbol{\delta}(t)/dt|_{t_y}$. More details about the local coordinates are in Appendix before Proposition A 7 where we have to set $\boldsymbol{\eta}(\theta_T)$ instead of $\boldsymbol{\eta}_T$.

Generally, the choice of the local coordinates depends on the choice of the curve γ parallel to $\boldsymbol{\delta}_\eta$ which contains the point $\mathbf{y}$ (it is equivalent to say that it depends on the choice of the solution of the equation $(d/dt) \|\mathbf{y} - \boldsymbol{\delta}_\eta(t)\|^2_\Sigma = 0$) but in the region of regularity which is defined in Section 2.2, this choice is done in a unique way. We can write

$$\mathbf{y} - \boldsymbol{\eta}(\theta_T) = \left\langle \mathbf{y} - \boldsymbol{\eta}(\theta_T), \left.\frac{d\boldsymbol{\delta}_\eta}{dt}\right|_{t_y} \right\rangle_\Sigma \left.\frac{d\boldsymbol{\delta}_\eta}{dt}\right|_{t_y} + \sum_{i=1}^{N-1} \langle \mathbf{y} - \boldsymbol{\eta}(\theta_T), \mathbf{b}^{(i)}(t_y) \rangle_\Sigma \, \mathbf{b}^{(i)}(t_y).$$

Substituting this into the probability density (10) and using that $d\gamma/du|_{u(\mathbf{y})} = \pm d\boldsymbol{\delta}_\eta/dt|_{t_y}$. and the definition of the coordinates $v_1(\mathbf{y}), \ldots, v_{N-1}(\mathbf{y})$, we obtain

$$(21) \quad \frac{dP_{\theta_T}(\mathbf{y})}{d\lambda} = [(2\pi)^N \det \Sigma]^{-1/2} \exp\left\{-\tfrac{1}{2}\left\langle \mathbf{y} - \boldsymbol{\eta}(\theta_T), \left.\frac{d\gamma}{du}\right|_{u(\mathbf{y})}\right\rangle^2_\Sigma + \sum_{i=1}^{N-1} v_i^2(\mathbf{y})\right\}$$
$$= [(2\pi)^N \det \Sigma]^{-1/2} \exp\left\{-\tfrac{1}{2}\left\langle \gamma[u(\mathbf{y})] - \boldsymbol{\eta}(\theta_T), \left.\frac{d\gamma}{du}\right|_{u(\mathbf{y})}\right\rangle^2_\Sigma + \sum_{i=1}^{N-1} v_i^2(\mathbf{y})\right\}.$$

Finally, using the Jacobian (A 8) in Proposition A 7, we obtain from (21)

$$\frac{dP_{\theta_T}(u(\mathbf{y}), v_1(\mathbf{y}), \ldots, v_{N-1}(\mathbf{y}))}{du \, dv_1 \,, \ldots, dv_{N-1}} = [(2\pi)^N \det \Sigma]^{-1/2}$$
$$\exp\left\{-\tfrac{1}{2}\left\langle \gamma[u(\mathbf{y})] - \boldsymbol{\eta}(\theta_T), \left.\frac{d\gamma}{du}\right|_{u(\mathbf{y})}\right\rangle^2_\Sigma + \sum_{i=1}^{N-1} v_i^2(\mathbf{y})\right\}.$$

From this equality follows the theorem:

Theorem 1. The random variables $u(\mathbf{y}), v_1(\mathbf{y}), \ldots, v_{N-1}(\mathbf{y})$ defined on W_{y_0} are independent. The conditional probability density of the coordinate u, under the assumption that the sample point $\mathbf{y}$ is on the curve γ parallel to $\boldsymbol{\delta}_{\boldsymbol{\eta}}$, is given by

$$\text{(22)} \qquad \frac{\mathrm{dP}_{\theta_T}(u \mid \mathbf{y} \in \gamma)}{\mathrm{d}u} \approx \exp\left\{-\tfrac{1}{2}\left\langle \gamma(u) - \boldsymbol{\eta}(\theta_T), \frac{\mathrm{d}\gamma}{\mathrm{d}u}\right\rangle_{\mathbf{\Sigma}}^2\right\}$$

Note. The parameter u used in Eq. (22) is the natural parameter of the curve γ, hence $\|\mathrm{d}\gamma/\mathrm{d}u\|_{\mathbf{\Sigma}} = 1$. The probability density in Eq. (22) is therefore the probability density with respect to the Lebesgue measure on γ in the special case when $\mathbf{\Sigma} = \mathbf{I}$.

A regular curve γ which is parallel to the curve $\boldsymbol{\delta}_{\boldsymbol{\eta}}$ will be called a *regular l.s.* (= *least squares*) *curve* iff for any $\bar{t} \in (0, T)$

$$\bar{t} = \operatorname*{Arg\,min}_{t \in \langle 0, T\rangle} \|\gamma[\varphi(\bar{t})] - \boldsymbol{\delta}_{\boldsymbol{\eta}}(t)\|_{\mathbf{\Sigma}}^2 .$$

In other words a regular l.s. curve is defined by the property: for every sample $\mathbf{y}$ lying on the curve every solution of the equation $(\mathrm{d}/\mathrm{d}\theta)\, \|\mathbf{y} - \boldsymbol{\eta}(\theta)\|_{\mathbf{\Sigma}}^2 = 0$ is a least squares estimate.

Corollary to Theorem 1. If γ is a regular l.s. curve then the probability density of the least squares estimate $\hat{\theta}$, under the condition that the sample lies on the curve γ is given by

$$\text{(23)} \qquad \frac{\mathrm{dP}_{\theta_T}(\hat{\theta} \mid y \in \gamma)}{\mathrm{d}\hat{\theta}} =$$

$$= K_\gamma \exp\left\{-\tfrac{1}{2}\left\langle \gamma(\varphi[t(\hat{\theta})]) - \boldsymbol{\eta}(\theta_T), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \Big/ \left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}}\right\|_{\mathbf{\Sigma}}\right\rangle_{\mathbf{\Sigma}}^2\right\} \left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{t(\hat{\theta})} \left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}}\right\|_{\mathbf{\Sigma}},$$

where K_γ is a constant, the function $t(\theta)$ is defined by (12), $\mathrm{d}\varphi/\mathrm{d}t$ is given by

$$\left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{t(\theta)} = \left|1 + \left\langle \boldsymbol{\eta}(\theta) - \gamma(\varphi[t(\theta)]), \left.\frac{\mathrm{d}^2\boldsymbol{\eta}[\theta(t)]}{\mathrm{d}t^2}\right|_{t(\theta)}\right\rangle_{\mathbf{\Sigma}}^2\right|$$

and where

$$\text{(24)} \qquad \left.\frac{\mathrm{d}^2\boldsymbol{\eta}[\theta(t)]}{\mathrm{d}t^2}\right|_{t(\theta)} = \frac{\mathrm{d}^2\boldsymbol{\eta}}{\mathrm{d}\theta^2} \Big/ \left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\right\|_{\mathbf{\Sigma}}^2 - \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\left\langle\frac{\mathrm{d}^2\boldsymbol{\eta}}{\mathrm{d}\theta^2}, \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\right\rangle_{\mathbf{\Sigma}} \Big/ \left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\right\|_{\mathbf{\Sigma}}^4 .$$

Proof. The function ψ which is defined by

$$\psi(\theta) = \varphi[t(\theta)]$$

maps $\langle \underline{\theta}, \bar{\theta}\rangle$ onto $\langle 0, U\rangle$ and it has a positive derivative

$$\frac{\mathrm{d}\psi}{\mathrm{d}\theta} = \frac{\mathrm{d}\varphi}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\theta} = \frac{\mathrm{d}\varphi}{\mathrm{d}t}\left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\right\|_{\mathbf{\Sigma}} > 0 .$$

This follows from the definition of $t(\theta)$ in Eq. (12), from Eq. (15) and from the regularity of the curve γ. Moreover, from (16) it follows that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left\|\gamma(u)-\boldsymbol{\eta}(\theta)\right\|_{\boldsymbol{\Sigma}}^{2}\Big|_{u=\psi(\theta)} = -2\left\langle \gamma(\varphi[t(\theta)])-\boldsymbol{\eta}(\theta), \frac{\mathrm{d}\boldsymbol{\delta}_{\eta}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\theta}\right\rangle_{\boldsymbol{\Sigma}} = 0\,.$$

Hence $\psi^{-1}(u)$ is the least squares estimate if $\mathbf{y} = \gamma(u)$ is the sample. In the notation introduced in Eq. (2) it means that

$$\hat{\theta}(\mathbf{y}) = \psi^{-1}(u)\big|_{\mathbf{y}=\gamma(u)}\,.$$

Therefore we may write

$$\frac{\mathrm{dP}_{\theta_T}(\hat{\theta}\mid \mathbf{y}\in\gamma)}{\mathrm{d}\hat{\theta}} = \frac{\mathrm{dP}_{\theta_T}(u\mid \mathbf{y}\in\gamma)}{\mathrm{d}u}\frac{\mathrm{d}\psi(\hat{\theta})}{\mathrm{d}\hat{\theta}} =$$

$$= \frac{\mathrm{dP}_{\theta_T}(u\mid \mathbf{y}\in\gamma)}{\mathrm{d}u}\bigg|_{u=\psi(\hat{\theta})}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\bigg|_{t(\hat{\theta})}\left\|\frac{\mathrm{d}\boldsymbol{\eta}(\hat{\theta})}{\mathrm{d}\hat{\theta}}\right\|_{\boldsymbol{\Sigma}}.$$

Comparing this formula with (22) we obtain (23). The derivative $\mathrm{d}\varphi/\mathrm{d}t$ is given by Eq. (A 2) in Appendix. Finally, using Eq. (12) we obtain

$$\frac{\mathrm{d}^2\boldsymbol{\eta}[\theta(t)]}{\mathrm{d}t^2} = \frac{\mathrm{d}^2\boldsymbol{\eta}}{\mathrm{d}\theta^2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 + \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} =$$

$$= \frac{\mathrm{d}^2\boldsymbol{\eta}}{\mathrm{d}\theta^2}\Big/\left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\right\|_{\boldsymbol{\Sigma}}^2 + \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\bigg|_{\theta(t)}, \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\theta}\bigg|_{\theta(t)}\right\rangle_{\boldsymbol{\Sigma}}^{-1/2}$$

which is equal to the right hand side of Eq. (24). □

2.2. Confidence Intervals if Samples are in the Region of Regularity

To find a confidence region for the parameter θ means to construct for every θ a set $W_\theta \subset \mathbb{R}^N$ with the property

$$\mathsf{P}_\theta(W_\theta) = 1 - \alpha$$

where $1 - \alpha$ is the prescribed level of confidence. Then the set

$$\{\theta : \mathbf{y} \in W_\theta\}$$

is the confidence region.

We propose to construct a set $W_{\theta,\gamma} \subset \{\gamma(u) : u \in \langle 0, U\rangle\}$ for each curve γ which is parallel to $\boldsymbol{\delta}_\eta$ in such a way that

$$\mathsf{P}_\theta(\mathbf{y}\in W_{\theta,\gamma}\mid \mathbf{y}\in\gamma) \doteq 1-\alpha\,; \quad (\theta\in\Theta)$$

and to take

$$W_\theta = \bigcup_{\gamma\in\Gamma} W_{\theta,\gamma}$$

where Γ is a "reasonably choosen" set of parallel curves.

This construction can be carried over successfully if the sample lies in the "region of regularity" with probability close to one. and if θ is "distant" from the boundary of Θ, as we shall explain.

We define the *region of regularity* of the curve $\boldsymbol{\eta}$ as the set $R_{\boldsymbol{\eta}}$ of points $\mathbf{z} \in V_{\boldsymbol{\eta}}$ for which

(i) $$\|\mathbf{z} - \boldsymbol{\eta}[\hat{\theta}(z)]\|_{\Sigma} < \inf_{t \in \langle 0, T \rangle} \varrho_{\boldsymbol{\eta}}(t),$$

where $\varrho_{\boldsymbol{\eta}}(t)$ is the radius of curvature of the curve $\boldsymbol{\delta}_{\boldsymbol{\eta}}$ at the point t, (Eq. (19).

(ii) there is a unique solution $(=\hat{\theta}(\mathbf{z}))$ of the equation

$$\frac{\partial}{\partial \theta} \|\mathbf{z} - \boldsymbol{\eta}(\theta)\|_{\Sigma}^2 = 0$$

which is in the set

$$\{\theta : \theta \in \Theta, \|\mathbf{z} - \boldsymbol{\eta}(\theta)\|_{\Sigma} \leqq 2 \inf_{t \in \langle 0, T \rangle} \varrho_{\boldsymbol{\eta}}(t)\}$$

(iii) the symmetric image $\mathbf{z}^*$ of the point $\mathbf{z}$, which is defined by

$$\frac{\mathbf{z}^* + \mathbf{z}}{2} = \boldsymbol{\eta}[\hat{\theta}(\mathbf{z})]$$

also has the properties (i) and (ii).

The number

(25) $$r := \inf_{t \in \langle 0, T \rangle} \varrho_{\boldsymbol{\eta}}(t)$$

will be called *the radius of the region of regularity*.

The property (i) ensures that a parallel curve in $R_{\boldsymbol{\eta}}$ is regular (Proposition A 4). The property (ii) implies that $R_{\boldsymbol{\eta}}$ contains no point of intersection of two parallel curves with distances less than $2r$ and that parallel curves are regular l.s. curves. The property (iii) allows to use symmetric pairs of parallel curves, as we shall explain later.

Take $p_0 \in \langle 0, 1 \}$. We shall say that the regression model is *at least p_0-regular with distant boundaries* iff

$$\mathsf{P}_{\theta}(R_{\boldsymbol{\eta}}) \geqq p_0$$

for every value of the parameter θ which is in a given subset $\Theta_0 \subset \Theta$. Here we suppose that the true value of θ can be only from the set Θ_0 and the values which are in $\Theta - \Theta_0$ are redundant in the model. This is what we mean under the words "distant boundaries". This approach can be justificated by the example of the linear regression, where we take standardly $\Theta = (-\infty, \infty)$, although in applications there are always some constrains on the values of θ.

In an at least p_0-regular model with distant boundaries the formula (23) can be simplified. As a result we have the following theorem:

Theorem 2. In an at least p_0-regular model with distant boundaries a confidence interval for θ is given as the set

$$J_\theta := \{\theta : \hat{\theta}(\mathbf{y}) \in I_\theta\},$$

where the interval I_θ is taken according to

$$\frac{\int_{I_\theta} q_\theta(\hat{\theta})\,\mathrm{d}\hat{\theta}}{\int_{\Theta} q_\theta(\hat{\theta})\,\mathrm{d}\hat{\theta}} = 1 - \alpha$$

and where the function q_θ is given as

$$q_\theta(\hat{\theta}) = \exp\left\{-\left\langle \boldsymbol{\eta}(\hat{\theta}) - \boldsymbol{\eta}(\theta), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}}\right\rangle_{\boldsymbol{\Sigma}}^2 \Big/ \left[2\left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}}\right\|_{\boldsymbol{\Sigma}}^2\right]\right\} \left\|\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}}\right\|_{\boldsymbol{\Sigma}}. \tag{26}$$

The confidence of the interval J_θ is at least $p_0\,(1 - \alpha)$.

Proof. Let γ be a curve parallel to $\boldsymbol{\delta}_\eta$. Let us denote by γ^* the curve which is the symmetric image of γ, that means

$$\gamma^* : t \in \langle 0, T\rangle \mapsto \gamma^*[\varphi_{\gamma^*}(t)], \tag{27}$$

where $\{\gamma^*[\varphi_{\gamma^*}(t)] + \gamma[\varphi_\gamma(t)]\}/2 = \boldsymbol{\delta}_\eta(t)$. (We note that the parameter t in (27) is natural for $\boldsymbol{\delta}_\eta$ but not for γ^*). The difference $\gamma - \boldsymbol{\delta}_\eta$ is a solution of Eq. (A 4) (with $\boldsymbol{\delta}_\eta$ instead of $\boldsymbol{\delta}$). It follows that $\gamma^* - \boldsymbol{\delta}_\eta$ is also a solution of Eq. (A 4), hence γ^* is a curve parallel to $\boldsymbol{\delta}_\eta$ as well.

Let us compare the expressions for $\mathrm{d}\varphi_\gamma/\mathrm{d}t$ and $\mathrm{d}\varphi_{\gamma^*}/\mathrm{d}t$ given by Eq. (A 2). Because $\gamma[\varphi_\gamma(t)] \in R_\eta$, we have

$$\|\gamma[\varphi_\gamma(t)] - \boldsymbol{\delta}_\eta(t)\|_{\boldsymbol{\Sigma}}^2 < \varrho^2(t),$$

hence

$$\left\langle \boldsymbol{\delta}_\eta(t) - \gamma[\varphi_\gamma(t)], \frac{\mathrm{d}^2\,\boldsymbol{\delta}_\eta(t)}{\mathrm{d}t^2}\right\rangle_{\boldsymbol{\Sigma}} < 1.$$

This allows to omit the absolute value in (A2) and to write

$$\frac{\mathrm{d}\varphi_\gamma}{\mathrm{d}t} = 1 + \left\langle \boldsymbol{\delta}_\eta(t) - \gamma[\varphi_\gamma(t)], \frac{\mathrm{d}^2\boldsymbol{\delta}_\eta(t)}{\mathrm{d}t^2}\right\rangle_{\boldsymbol{\Sigma}}.$$

Analogically we obtain

$$\frac{\mathrm{d}\varphi_{\gamma^*}}{\mathrm{d}t} = 1 + \left\langle \boldsymbol{\delta}_\eta(t) - \gamma[\varphi_{\gamma^*}(t)], \frac{\mathrm{d}^2\boldsymbol{\delta}_\eta(t)}{\mathrm{d}t^2}\right\rangle_{\boldsymbol{\Sigma}}.$$

Hence

$$\frac{\mathrm{d}\varphi_\gamma}{\mathrm{d}t} + \frac{\mathrm{d}\varphi_{\gamma^*}}{\mathrm{d}t} = 2\,; \quad (t \in \langle 0, T\rangle). \tag{28}$$

Now let us consider the probability density (23). Since

$$\left\langle \gamma[\varphi_\gamma[t(\hat{\theta})]] - \boldsymbol{\eta}(\hat{\theta}), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\rangle_{\boldsymbol{\Sigma}}^2 = 0$$

for every $\hat{\theta} \in \Theta$, we have

$$\left\langle \gamma[\varphi_\gamma[t(\hat{\theta})]] - \boldsymbol{\eta}(\theta), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\rangle_{\boldsymbol{\Sigma}}^2 = \left\langle \boldsymbol{\eta}(\hat{\theta}) - \boldsymbol{\eta}(\theta), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\rangle_{\boldsymbol{\Sigma}}^2 . \tag{29a}$$

Analogically

$$\left\langle \gamma^*[\varphi_{\gamma^*}[t(\hat{\theta})]] - \boldsymbol{\eta}(\theta), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\rangle_{\boldsymbol{\Sigma}}^2 = \left\langle \boldsymbol{\eta}(\hat{\theta}) - \boldsymbol{\eta}(\theta), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\rangle_{\boldsymbol{\Sigma}}^2 . \tag{29b}$$

Let us set (29a) and (29b) into (23) and into an analogical expression for the probability density $\mathrm{dP}_\theta(\hat{\theta}) \mid \mathbf{y} \in \gamma^*)/\mathrm{d}\hat{\theta}$. We obtain

$$\frac{\mathrm{dP}_\theta(\hat{\theta} \mid \mathbf{y} \in \gamma \text{ or } \mathbf{y} \in \gamma^*)}{\mathrm{d}\hat{\theta}} \approx \exp\left\{ - \left\langle \eta(\hat{\theta}) - \boldsymbol{\eta}(\theta), \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\rangle_{\boldsymbol{\Sigma}}^2 \Big/ \left[2 \left\| \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\|_{\boldsymbol{\Sigma}}^2 \right] \right\} \left\| \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\hat{\theta}} \right\|_{\boldsymbol{\Sigma}} \tag{30}$$

which is the probability density of the l.s. estimate $\hat{\boldsymbol{\theta}}$ under the condition that the sample $\mathbf{y}$ is on the curve γ or on the curve γ^*. The right-hand side of (30) is equal to the function q_θ in (26).

The proposed confidence interval is then obtained as explained in the beginning of Section 2.2, taking for Γ the set of all parallel curves which are in R_η. The level of confidence of this interval is at least equal to $(1 - \alpha)\, p_0$. □

2.3. Examples

1. *The linear model.* Let us take $\boldsymbol{\Sigma} = \mathbf{I}$ and

$$\boldsymbol{\eta}(\theta) = \theta \boldsymbol{f} ; \quad (\theta \in (-\infty, \infty)) ,$$

where $\boldsymbol{f}$ is a given vector. Evidently $R_\eta = \mathbb{R}^N$, $p_0 = 1$. Parallel curves are straight lines: $\gamma(u) = u\boldsymbol{f}/\|\boldsymbol{f}\|_{\boldsymbol{\Sigma}} + \boldsymbol{k}_\gamma$, where $\boldsymbol{k}_\gamma$ is a constant vector. From (26) we obtain

$$q_\theta(\hat{\theta}) = \exp\{-\tfrac{1}{2}(\hat{\theta} - \theta)^2 \boldsymbol{ff}'\} [\boldsymbol{ff}']^{1/2} .$$

Hence the $(1 - \alpha)$ confidence interval is

$$\{\theta : |\theta - \hat{\theta}(\mathbf{y})| [\boldsymbol{ff}']^{1/2} < c_\alpha\} ,$$

where c_α is the $(1 - \alpha/2)$ quantile of $N(0, 1)$. This interval coincides with the confidence interval used in the linear regression, when $\boldsymbol{\Sigma}$ is known.

2. *The exponential model.* Let us take $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, $N = 2$,

$$\eta_1(\theta) = \mathrm{e}^\theta ,$$

$$\eta_2(\theta) = \mathrm{e}^{-\theta} ; \quad (\theta \in (-\infty, \infty)) .$$

Let us denote $\dot{\boldsymbol{\eta}} := \mathrm{d}\boldsymbol{\eta}/\mathrm{d}\theta$ etc. From (24) we obtain

$$\left\|\frac{\mathrm{d}^2\boldsymbol{\eta}}{\mathrm{d}t^2}\right\|_{\Sigma}^2 = \frac{\|\ddot{\boldsymbol{\eta}}\|_{\Sigma}^2}{\|\dot{\boldsymbol{\eta}}\|_{\Sigma}^4} - \frac{\langle\dot{\boldsymbol{\eta}}, \ddot{\boldsymbol{\eta}}\rangle_{\Sigma}^2}{\|\dot{\boldsymbol{\eta}}\|_{\Sigma}^6} = \frac{4\sigma^2}{[\mathrm{e}^{2\theta} + \mathrm{e}^{-2\theta}]^3}.$$

Hence

$$r^2 = \min_{\theta\in(-\infty,\infty)} [\mathrm{e}^{2\theta} + \mathrm{e}^{-2\theta}]^3/4\sigma^2 = 2/\sigma^2 .$$

The set

$$R_{\eta} := \{\mathbf{z} : \mathbf{z} \in \mathbb{R}^2, \min_{\theta} \{[z_1 - \mathrm{e}^{\theta}]^2 + [z_2 - \mathrm{e}^{-\theta}]^2\} < 2\}$$

is the region of regularity. The model is at least p_0-regular with

$$p_0 = \int_0^{2/\sigma^2} h_2(x)\,\mathrm{d}x ,$$

where $h_2(x)$ is the probability density of the χ^2-distribution with 2 degrees of freedom, since every sphere which has its center in the mean-values manifold $\mathscr{E} = \{(e^{\theta}, \mathrm{e}^{-\theta}); \theta \in (-\infty, \infty)\}$, and which has the radius equal to 2, is a subset of R_{η}. (By more subtile considerations we can obtain a better (= larger) value for p_0). Evidently, the model is with distant boundaries for every finite interval $\Theta_0 \subset (-\infty, \infty)$.

The function q_{θ} from Eq. (26) is expressed as

$$q_{\theta}(\hat{\theta}) = (\sqrt{2}/\sigma)\,[\cosh(2\hat{\theta})]^{1/2} \exp\left\{-\frac{1}{2\sigma^2}\frac{[\mathrm{sind}\,(2\hat{\theta}) - \sinh(\hat{\theta} + \theta)]^2}{\cosh(2\hat{\theta})}\right\}.$$

2.4. Curves outside off the Region of Regularity

In the case when, with a large probability, the sample lies outside off the region of regularity it is better to describe the geometry of the sample by curves other than the parallel curves. The reason is that several parallel curves may intersect in points outside the region of regularity.

A curve $\gamma : \langle 0, U\rangle \mapsto \mathbb{R}^N$ will be called a l.s. curve with respect to a regular curve δ iff

(i) for every $u \in \langle 0, U\rangle$ there exists a $t_u \in \langle 0, T\rangle$ such that

$$\min_{t\in\langle 0,T\rangle} \|\gamma(u) - \delta(t)\|_{\Sigma} = \|\gamma(u) - \delta(t_u)\|_{\Sigma}$$

(ii) if $u \notin C_{\gamma}$ then

$$\langle \delta(t_u) - \gamma(u), \mathrm{d}\delta(t)/\mathrm{d}t|_{t_u}\rangle_{\Sigma} = 0 ,$$

and

$$\left.\frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\right|_{t_u} = \pm\frac{\mathrm{d}\gamma(u)}{\mathrm{d}u}.$$

We see that a l.s. curve is composed from parts of parallel curves. which have the same distance from the curve $\boldsymbol{\delta}$.

In the region of regularity a curve is a l.s. curve if and only if it is a curve parallel to $\boldsymbol{\eta}$. Outside the region of regularity the advantage of the l.s. curves, if compared with parallel curves, is that they intersect each other with a negligible probability.

Since l.s. curves coincide with parallel curves in neighbourhoods of regular points, the expression (22) for the conditional probability density remains valid. It follows that we can use the l.s. curves to construct confidence intervals analogically as parallel curves, but in general it is a difficult task from the computational point of view.

3. THE MULTIVARIATE NONLINEAR REGRESSION

Similar ideas as in Section 2 can be used also in the case when set Θ is multidimensional. However computationally the situation is much more complicated than in the case of the onedimensional regression. Here we present only the outlines of a possible approach. We intend to present more developed results for special cases in a future paper.

Let us first consider the case that the regression model is linear,

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = \mathbf{F}\boldsymbol{\theta}\,; \quad (\boldsymbol{\theta} \in \Theta = \mathbb{R}^m)\,,$$

where $\mathbf{F}$ is a known $N \times m$ matrix of rank m. The confidence region for the vector of unknown parameters $\boldsymbol{\theta}$ is well known for this linear model. It is the ellipsoid

$$\mathcal{O} := \{\boldsymbol{\theta} \in \mathbb{R}^m : (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \left(\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F}\right)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leqq c\}\,,$$

where c is a constant. Since $\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F}$ is the information matrix (compare with (7)), the confidence ellipsoid $\mathcal{O}$ is a sphere, if measured in the Rao's distance, that is

$$\mathcal{O} = \{\boldsymbol{\theta} \in \mathbb{R}^m : d(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) \leqq c\}\,. \tag{31}$$

The expression (31) for $\mathcal{O}$ make sense also in the case of a nonlinear regression model, if we suppose that every two points of Θ may be connected by a geodesics. Then the Rao's distance between any point $\boldsymbol{\theta}^*$ on the boundary of $\mathcal{O}$ and the point $\hat{\boldsymbol{\theta}}$ is just the length of the geodetics connecting both points $\boldsymbol{\theta}^*$ and $\hat{\boldsymbol{\theta}}$.

It seems to be resonable to take the set $\mathcal{O}$ as a confidence region for $\boldsymbol{\theta}$ also in the nonlinear regression model, at least for the case of a p_0-regular regression model with distant boundaries, which is to be defined in an analogical way as in Section 2.2. Two difficulties arise if we want to construct such confidence regions:

a) the construction of the set $\mathcal{O}$ means to construct geodesics in Θ,

b) it is necessary to compute the confidence level of such a confidence region.

As to the point a), the geodesics in Θ is a solution $\mathbf{g}(t)$ of the following differential equations:

$$(32) \qquad \left\langle \frac{\mathrm{d}^2\boldsymbol{\eta}[\mathbf{g}(t)]}{\mathrm{d}t^2}, \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_k}\bigg|_{\boldsymbol{\theta}=\mathbf{g}(t)} \right\rangle_{\boldsymbol{\Sigma}} = 0\,; \quad (k = 1, \ldots, m)\,.$$

(The Euler-Lagrange differential equations). To find a solution of these differential equations is a very complicated task. However from (32) it follows that the curve $\boldsymbol{\eta}[\mathbf{g}(t)]$ is a curve in the mean value manifold $\mathscr{E}$ which has the vector of curvature $\mathrm{d}^2\boldsymbol{\eta}[\mathbf{g}(t)]/\mathrm{d}t^2$ orthogonal to the surface $\mathscr{E}$ in every point of $\boldsymbol{\eta}[\mathbf{g}(t)]$. This can be used for an approximative construction of $\mathcal{O}$.

As to the point b), the confidence level can be computed if we know the probability density of the least-squares estimate $\hat{\theta}$. That means we have to find the multidimensional generalization of the expression (26) for $q_\theta(\hat{\theta})$.

APPENDIX

Through the appendix we shall denote by $\boldsymbol{\delta}$ a regular curve $\boldsymbol{\delta} : t \in \langle 0, T \rangle \mapsto \mathbb{R}^N$. The parameters of all considered curves will be natural. By $\gamma \| \boldsymbol{\delta}$ we shall denote that the curve γ is parallel to the curve $\boldsymbol{\delta}$. We shall write $\langle\ ,\ \rangle$ and $\|\ \|$ instead of $\langle\ ,\ \rangle_{\boldsymbol{\Sigma}}$ and $\|\ \|_{\boldsymbol{\Sigma}}$, φ instead of φ_γ, C instead of C_γ.

Proposition A 1. If $\gamma \| \boldsymbol{\delta}$ then the function

$$t \in \langle 0, T \rangle \mapsto \gamma[\varphi(t)]$$

is differentiable, and

$$(\mathrm{A1}) \qquad \frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t} = \pm \frac{\mathrm{d}\boldsymbol{\delta}}{\mathrm{d}t}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\,.$$

Proof. The function φ is nondecreasing. Therefore

$$\frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t_+} = \frac{\mathrm{d}\gamma(u)}{\mathrm{d}u_+}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\bigg|_{\varphi(t)=u}$$

$$\frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t_-} = \frac{\mathrm{d}\gamma(u)}{\mathrm{d}u_-}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\bigg|_{\varphi(t)=u}\,.$$

If $\varphi(t) \in C$ then $\mathrm{d}\varphi/\mathrm{d}t = 0$, otherwise $\mathrm{d}\gamma/\mathrm{d}u = \pm\mathrm{d}\boldsymbol{\delta}/\mathrm{d}t$. □

Proposition A 2. If $\gamma \| \boldsymbol{\delta}$ then the distance $\|\boldsymbol{\delta}(t) - \gamma[\varphi(t)]\|$ does not depend on t.

Proof. From (16), (17) and from Proposition A1 we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{\delta}(t) - \gamma[\varphi(t)]\|^2 = 2\left\langle \boldsymbol{\delta}(t) - \gamma[\varphi(t)], \frac{\mathrm{d}\boldsymbol{\delta}(t)}{\mathrm{d}t} - \frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t} \right\rangle = 0\,. \qquad □$$

Proposition A 3. If $\gamma \parallel \delta$ then

$$\text{(A2)} \qquad \frac{\mathrm{d}\varphi}{\mathrm{d}t} = \tfrac{1}{2}\left|\frac{\mathrm{d}^2}{\mathrm{d}t^2}\,\|\delta(t) - \gamma(u)\|^2_{u=\varphi(t)}\right| = \left|\left\langle \delta(t) - \gamma[\varphi(t)]\,, \frac{\mathrm{d}^2\delta(t)}{\mathrm{d}t^2}\right\rangle + 1\right|$$

Proof. Since $\|\mathrm{d}\delta(t)/\mathrm{d}t\| = 1$ we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \delta(t), \frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\right\rangle = 1 + \left\langle \delta(t), \frac{\mathrm{d}^2\delta(t)}{\mathrm{d}t^2}\right\rangle.$$

Analogically, from Proposition A1

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle \gamma[\varphi(t)], \frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\right\rangle = \pm\frac{\mathrm{d}\varphi}{\mathrm{d}t} + \left\langle \gamma[\varphi(t)], \frac{\mathrm{d}^2\delta(t)}{\mathrm{d}t^2}\right\rangle.$$

Subtracting the terms in the second equality from the terms in the first equality and using Eq. (16) we obtain

$$0 = 1 \pm \frac{\mathrm{d}\varphi}{\mathrm{d}t} + \left\langle \delta(t) - \gamma[\varphi(t)], \frac{\mathrm{d}^2\delta}{\mathrm{d}t^2}\right\rangle = \pm\frac{\mathrm{d}\varphi}{\mathrm{d}t} + \tfrac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\,\|\delta(t) - \gamma(u)\|^2_{u=\varphi(t)}. \quad \square$$

Corollary. A point $\gamma(u)$ is a singular point of the curve γ parallel to δ if and only if $(\mathrm{d}^2/\mathrm{d}t^2)\,\|\delta(t) - \gamma(u)\|^2_{u=\varphi(t)} = 0$. More generally, a point $\mathbf{z} \in \mathbb{R}^N$ is singular at $t \in \langle 0, T\rangle$ (resp. it is regular) iff $(\mathrm{d}/\mathrm{d}t)\,\|\mathbf{z} - \delta(t)\|^2 = 0$ and $(\mathrm{d}^2/\mathrm{d}t^2)\,\|\mathbf{z} - \delta(t)\|^2 = 0$ (resp. iff it is singular at no point of $\langle 0, T\rangle$).

As is known from differential geometry, the expression $\varrho(t) = \|\mathrm{d}^2\delta(t)/\mathrm{d}t^2\|^{-1}$ is the radius of curvature of the curve δ at the point t.

Proposition A 4. If $\gamma \parallel \delta$ and if the distance between the curves satisfies the inequality

$$\text{(A3)} \qquad \|\gamma[\varphi(0)] - \delta(0)\| < \inf_{t \in \langle 0, T\rangle} \varrho(t)\,,$$

then the curve γ has no singular point.

Proof. From Proposition A 2 we obtain, using the Schwarz inequality

$$\langle \delta(t) - \gamma[\varphi(t)], \mathrm{d}^2\delta(t)/\mathrm{d}t^2\rangle \leqq \|\delta(0) - \gamma[\varphi(0)]\|\, \varrho^{-1}(t)\,,$$

hence, according to (A 2), $\mathrm{d}\varphi/\mathrm{d}t \neq 0$ for every $t \in \langle 0, T\rangle$. Thus γ has no singular point. $\square$

Proposition A 5. A curve γ is parallel to δ if and only if the difference

$$\Delta(t) := \gamma[\varphi(t)] - \delta(t)$$

is the solution of the system of linear differential equations

$$\text{(A4)} \qquad \frac{\mathrm{d}\Delta_i(t)}{\mathrm{d}t} = -\frac{\mathrm{d}\delta_i(t)}{\mathrm{d}t}\sum_{j,k=1}^{N} \frac{\mathrm{d}^2\delta_j(t)}{\mathrm{d}t^2}\,\{\Sigma^{-1}\}_{jk}\,\Delta_k(t)\,; \quad (i = 1, \ldots, N)\,,$$

it satisfies the initial condition

$$\left\langle \frac{\mathrm{d}\delta(t)}{\mathrm{d}t}, \Delta(t) \right\rangle \Big|_{t_0} = 0 \quad \text{for some} \quad t_0 \in \langle 0, T \rangle ,$$

and $\varphi(t)$ is defined by

$$\text{(A5)} \qquad \varphi(t) = \int_0^t \left| \left\langle \Delta(v), \frac{\mathrm{d}^2\delta(v)}{\mathrm{d}v^2} \right\rangle - 1 \right| \mathrm{d}v .$$

Proof. Let us suppose first that γ is a curve which is parallel to $\boldsymbol{\delta}$. From Proposition A1 and Eq. (A2) we obtain

$$\frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t} = \pm \left| \left\langle \delta(t) - \gamma[\varphi(t)], \frac{\mathrm{d}^2\delta}{\mathrm{d}t^2} \right\rangle + 1 \right| \frac{\mathrm{d}\delta}{\mathrm{d}t} .$$

Using Eq. (16) we obtain

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} \left\langle \delta(t) - \gamma[\varphi(t)], \frac{\mathrm{d}\delta}{\mathrm{d}t} \right\rangle =$$

$$= 1 + \left\langle \delta(t) - \gamma[\varphi(t)], \frac{\mathrm{d}^2\delta}{\mathrm{d}t^2} \right\rangle - \left\langle \frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t}, \frac{\mathrm{d}\delta}{\mathrm{d}t} \right\rangle ; \quad (t \in \langle 0, T \rangle).$$

Hence

$$\left\langle \frac{\mathrm{d}\gamma}{\mathrm{d}u}, \frac{\mathrm{d}\delta}{\mathrm{d}t} \right\rangle = \operatorname{sign} \left[\left\langle \delta(t) - \gamma[\varphi(t)], \frac{\mathrm{d}^2\delta}{\mathrm{d}t^2} \right\rangle + 1 \right] ; \quad (\gamma[\varphi(t)] \notin C) .$$

It follows that $\mathrm{d}\gamma/\mathrm{d}u = \mathrm{d}\delta/\mathrm{d}t$ (resp. $\mathrm{d}\gamma/\mathrm{d}u = -\mathrm{d}\delta/\mathrm{d}t$) if

$$\left\langle \gamma[\varphi(t)] - \delta(t), \frac{\mathrm{d}^2\delta}{\mathrm{d}t^2} \right\rangle > 1 \quad (\text{resp.} < 1).$$

Hence

$$\frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t} = \left[1 - \left\langle \gamma[\varphi(t)] - \delta(t), \frac{\mathrm{d}^2\delta}{\mathrm{d}t^2} \right\rangle \right] \frac{\mathrm{d}\delta}{\mathrm{d}t} ; \quad (t \in \langle 0, T \rangle)$$

which coincides with Eqs. (A 4). Eq. (A 5) is true due to Proposition A 3.

Conversely, let us suppose that $\gamma[\varphi(t)] = \Delta(t) + \delta(t)$, where $\Delta(t)$ is a solution of (A 4) and let us take φ according to (A 5). Let us define

$$C := \{\varphi(t) : t \in \langle 0, T \rangle, \mathrm{d}\varphi/\mathrm{d}t = 0\} .$$

The set C is compact, since $\varphi(t)$ and $\mathrm{d}\varphi/\mathrm{d}t$ are continuous. If $\mathrm{d}\varphi/\mathrm{d}t|_{t_1} = \mathrm{d}\varphi/\mathrm{d}t|_{t_2} = 0$ then either $\varphi(t_1) = \varphi(t_2)$ or $\mathrm{d}\varphi/\mathrm{d}t > 0$ in a point $t \in (t_1, t_2)$, hence C does not contain intervals.

If $u \notin C$ we obtain from (A 4) and (A5)

$$\frac{\mathrm{d}\gamma}{\mathrm{d}u} = \frac{\mathrm{d}\gamma[\varphi(t)]}{\mathrm{d}t} \left(\frac{\mathrm{d}\varphi}{\mathrm{d}t} \right)^{-1} = \pm \frac{\mathrm{d}\delta}{\mathrm{d}t}$$

and

$$\frac{\mathrm{d}^2\gamma}{\mathrm{d}u^2} = \pm \frac{\mathrm{d}^2\delta}{\mathrm{d}t^2}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^{-1},$$

hence γ has continuous second order derivatives on $\langle 0, \varphi(T)\rangle - C$, Eq. (A 1) is valid and $u = \varphi(t)$ is the natural parameter for γ. Finally we have from (A 4)

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{\left\langle \Delta(t), \frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\right\rangle\right\} = \left\langle \frac{\mathrm{d}\Delta(t)}{\mathrm{d}t}, \frac{\mathrm{d}\delta(t)}{\mathrm{d}t}\right\rangle + \left\langle \Delta(t), \frac{\mathrm{d}^2\delta(t)}{\mathrm{d}t^2}\right\rangle = 0 .$$

Hence $\left\langle \Delta(t), \dfrac{\mathrm{d}\delta(t)}{\mathrm{d}t}\right\rangle = \text{const.} = \left\langle \Delta(t_0), \dfrac{\mathrm{d}\delta(t_0)}{\mathrm{d}t_0}\right\rangle = 0 .$ □

Corollary. Let us take an arbitrary point $\mathbf{z} \in R^N$ and a point $t_0 \in \langle 0, T\rangle$ such that $(\mathrm{d}/\mathrm{d}t)\,\|\mathbf{z} - \boldsymbol{\delta}(t)\|^2 = 0$. Then there is a unique curve γ which contains the point $\mathbf{z}$, which is parallel to $\boldsymbol{\delta}$ and which is such that $\gamma[\varphi(t_0)] - \delta(t_0)$ is orthogonal to $\boldsymbol{\delta}$ at the point t_0.

Proof. As is known (cf. [4] Chapter I, Theorem 5.1) any system of differential equations

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_j a_{ij}(t)\, x_j(t)\,; \quad (i = 1, \ldots, N)$$

with $a_{ij}(t)$ continuous on $\langle a, b\rangle$ has a unique solution $\mathbf{x}(t)$ which satisfies the initial condition $\mathbf{x}(t_0) = \mathbf{z}$ for a given but arbitrary point $\mathbf{z} \in R^N$. □

Proposition A 6. Let be $\mathbf{y}_0 \in R^N$, $t_0 \in (0, T)$, $(\mathrm{d}/\mathrm{d}t)\,\|\mathbf{y}_0 - \delta(t)\|^2|_{t_0} = 0$, $(\mathrm{d}^2/\mathrm{d}t^2)\,.$ $.\,\|\mathbf{y}_0 - \delta(t)\|^2|_{t_0} > 0$ (resp. <0). Then there are a neighbourhood W_{y_0} of $\mathbf{y}_0$ and an $\varepsilon > 0$ such that

a) to every $\mathbf{y} \in W_{y_0}$ there is a unique point in $t_0 - \varepsilon, t_0 + \varepsilon)$ which will be denoted by t_y, such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\|\mathbf{y} - \delta(t)\|^2|_{t_y} = 0\,,$$

b) for every $\mathbf{y} \in W_{y_0}$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\,\|\mathbf{y} - \delta(t)\|^2|_{t_y} > 0$$

(resp. <0).

Proof. Let us denote

$$V := \left\{(\mathbf{y}, t) : \mathbf{y} \in R^N, t \in (0, T), \frac{\mathrm{d}^2}{\mathrm{d}t^2}\,\|\mathbf{y} - \delta(t)\|^2 > 0\right\},$$

$$V_k := \left\{(\mathbf{y}, t) : \frac{\mathrm{d}}{\mathrm{d}t}\,\|\mathbf{y} - \delta(t)\|^2 \in \left(-\frac{1}{k}, \frac{1}{k}\right)\right\}.$$

The sets $V, V_1, V_2, \dots$ are open sets. since the functions

$$(\mathbf{y}, t) \in R^N \times (0, T) \mapsto (\mathrm{d}/\mathrm{d}t) \|\mathbf{y} - \boldsymbol{\delta}(t)\|^2 ,$$
$$(\mathbf{y}, t) \in R^N \times (0, T) \mapsto (\mathrm{d}^2/\mathrm{d}t^2) \|\mathbf{y} - \boldsymbol{\delta}(t)\|^2$$

are continuous. Let us take $(N + 1)$-dimensional open cubes $Q_1, Q_2, \dots$ centered by the point $(\mathbf{y}_0, t_0)$ and such that $Q_k \subset V \cap V_k$; $(k = 1, 2, \dots)$. We have

$$(\mathbf{y}_0, t_0) \in \bigcap_{k=1}^{\infty} Q_k \neq \varnothing . \tag{A6}$$

Because the intersection of a decreasing sequence of open intervals in R^1, $(a_1, b_1) \supset \supset (a_2, b_2) \supset \dots$, is the interval $\langle \lim_i a_i, \lim_i b_i \rangle$, we obtain from (A 6) that there is a k-dimensional open cube W_{y_0} and an open interval $(t_0 - \varepsilon, t_0 + \varepsilon)$ such that

$$(\mathbf{y}_0, t_0) \in W_{y_0} \times (t_0 - \varepsilon, t_0 + \varepsilon) \subset \bigcap_{k=1}^{\infty} Q_k . \qquad \square$$

We shall construct local coordinates in the open set W_{y_0} defined in Proposition A 6. Let us choose $N - 1$ points $\mathbf{z}_1, \dots, \mathbf{z}_{N-1}$ which are in the hyperplane

$$\varkappa_{t_0} = \left\{ \mathbf{z} : \mathbf{z} \in R^N, \left\langle \mathbf{z} - \boldsymbol{\delta}(t_0), \left.\frac{\mathrm{d}\boldsymbol{\delta}(t)}{\mathrm{d}t}\right|_{t_0} \right\rangle = 0 \right\}$$

and are such that the vectors

$$\mathbf{b}^{(i)}(t_0) = \mathbf{z}_i - \boldsymbol{\delta}(t_0) ; \quad (i = 1, \dots, N - 1)$$

are normed and orthogonal vectors. Following Corollary to Proposition A 5, there are $N - 1$ curves, say $\gamma^{(1)}, \dots, \gamma^{(N-1)}$, containing the points $\mathbf{z}_1, \dots, \mathbf{z}_{N-1}$, which are parallel to $\boldsymbol{\delta}$ and are such that

$$\left\langle \gamma^{(i)}[\varphi^{(i)}(t_0)] - \boldsymbol{\delta}(t_0), \left.\frac{\mathrm{d}\boldsymbol{\delta}(t)}{\mathrm{d}t}\right|_{t_0} \right\rangle = 0 ; \quad (i = 1, \dots, N - 1) .$$

For every $t \in (t_0 - \varepsilon, t_0 + \varepsilon)$ we shall denote

$$\mathbf{b}^{(i)}(t) := \gamma^{(i)}[\varphi^{(i)}(t)] - \boldsymbol{\delta}(t) ; \quad (i = 1, \dots, N - 1) .$$

According to Proposition A 2 we have

$$\mathbf{b}^{(i)}(t) = 1 ; \quad (i = 1, \dots, N - 1) .$$

Analogically as in the proof of Proposition A 2 we may write

$$\frac{\mathrm{d}}{\mathrm{d}t} \langle \mathbf{b}^{(i)}(t), \mathbf{b}^{(j)}(t) \rangle = \left\langle \gamma^{(i)}[\varphi^{(i)}(t)] - \boldsymbol{\delta}(t), \frac{\mathrm{d}\boldsymbol{\delta}}{\mathrm{d}t} \right\rangle \left[\pm \frac{\mathrm{d}\varphi^{(j)}(t)}{\mathrm{d}t} - 1 \right] +$$
$$+ \left\langle \gamma^{(j)}[\varphi^{(j)}(t)] - \boldsymbol{\delta}(t), \frac{\mathrm{d}\boldsymbol{\delta}}{\mathrm{d}t} \right\rangle \left[\pm \frac{\mathrm{d}\varphi^{(i)}(t)}{\mathrm{d}t} - 1 \right] = 0 ,$$

where we used the property (16) of parallel curves. Hence

$$\langle \mathbf{b}^{(i)}(t), \mathbf{b}^{(j)}(t) \rangle = \langle \mathbf{b}^{(i)}(t_0), \mathbf{b}^{(j)}(t_0) \rangle = 0 . \tag{A7}$$

We recall that we denote by t_y the unique point which is in the interval $(t_0 - \varepsilon, t_0 + \varepsilon)$ and which is defined by

$$\left\langle \mathbf{y} - \delta(t_y), \left.\frac{\mathrm{d}\delta}{\mathrm{d}t}\right|_{t_y} \right\rangle = 0$$

(Proposition A6). The curve parallel to δ which contains $\mathbf{y}$ will be denoted by γ. We have

$$\left\langle \gamma[\varphi(t_y)] - \delta(t_y), \left.\frac{\mathrm{d}\delta}{\mathrm{d}t}\right|_{t_y} \right\rangle = 0 .$$

As before, we denote by u the natural parameter of the curve γ and by $u(\mathbf{y})$ its value at the point $\mathbf{y}$ (i.e. $\mathbf{y} = \gamma[u(\mathbf{y})]$). Finally. let us use the notation

$$v_i(\mathbf{y}) := \langle \mathbf{y} - \boldsymbol{\eta}_T, \boldsymbol{b}^{(i)}(t_y) \rangle \; ; \quad (i = 1, \ldots, N - 1) ,$$

where $\boldsymbol{\eta}_T$ is an arbitrary but fixed point of $\mathbb{R}^N$.

Proposition A 7. Let $\mathbf{y}_0$ be as in Proposition A 6. The mapping

$$\boldsymbol{\tau} : \mathbf{y} \in W_{y_0} \mapsto (u(\mathbf{y}), v_1(\mathbf{y}), \ldots, v_{N-1}(\mathbf{y})) \in \mathbb{R}^N$$

defines differentiable local coordinates in W_{y_0}. The Jacobian of $\boldsymbol{\tau}$ is

$$\text{(A8)} \qquad \left| \det\left(\left\{ \frac{\partial \tau_i(\mathbf{y})}{\partial y_j} \right\}_{i,j=1}^{N} \right) \right| = 1 .$$

Proof. Let us denote by $\mathbf{e}_1, \ldots, \mathbf{e}_N$ the orthonormal basis in $\mathbb{R}^N$ (i.e. $y_i = \langle \mathbf{e}_i, \mathbf{y} \rangle$; $(i = 1, \ldots, N)$). We have

$$\frac{\partial y_i}{\partial u} = \left.\frac{\mathrm{d}\gamma_i(u)}{\mathrm{d}u}\right|_{u(\mathbf{y})} ; \quad (i = 1, \ldots, N) .$$

Further

$$\frac{\partial y_i}{\partial v_j} = \langle \mathbf{e}_i, \boldsymbol{b}^{(j)}(t_y) \rangle \; ; \quad (i = 1, \ldots, N \, , \; j = 1, \ldots, N - 1) .$$

Hence

$$\begin{bmatrix} \frac{\partial y_1}{\partial u}, \ldots, \frac{\partial y_N}{\partial u} \\ \frac{\partial y_1}{\partial v_1}, \ldots, \frac{\partial y_N}{\partial v_1} \\ \cdots\cdots\cdots \\ \frac{\partial y_N}{\partial v_1}, \ldots, \frac{\partial y_N}{v_N} \end{bmatrix} = \begin{bmatrix} \frac{\mathrm{d}\gamma'}{\mathrm{d}u} \\ \boldsymbol{b}^{(1)}(t_y) \\ \vdots \\ \boldsymbol{b}^{(N-1)}(t_y) \end{bmatrix} ,$$

$$\left[\det\left(\left\{ \frac{\partial \tau_i(\mathbf{y})}{\partial y_j} \right\}_{i,j=1}^{N} \right) \right]^{-2} = \|\mathrm{d}\gamma/\mathrm{d}u\|^2 \, \|\boldsymbol{b}^{(1)}(t_y)\|^2 \ldots \|\boldsymbol{b}^{(N-1)}(t_y)\|^2 = 1 . \qquad \square$$

(Received November 16, 1981.)

REFERENCES

[1] C. Atkinson, A. F. S. Mitchell: Rao's distance measure. Preprint of the Imperial College of Science and Technology, London 1972.

[2] D. M. Bates, D. G. Watts: Relative curvature measures of nonlinearity. J. Roy. Statist. Soc. *B 42* (1980), 1—25.

[3] E. M. L. Beale: Confidence regions in non-linear estimation. J. Roy. Statist. Soc. *B 22* (1960), 41—88.

[4] E. A. Coddington, N. Levinson: Theory of Ordinary Differential Equations. McGraw-Hill Comp., New York 1955.

[5] J. Milnor: Morse Theory. Princeton Univ. Press, Princeton, N. J. 1963.

[6] C. R. Rao: On the distance between two populations. Sankhya *9* (1949), 246—248.

[7] S. Sternberg: Lectures on Differential Geometry. Second printing. Prentice-Hall, Englewood Cliffs 1965.

RNDr. Andrej Pázman, DrSc., Matematický ústav SAV (Mathematical Institute — Slovak Academy of Sciences), ul. Obrancov mieru 49, 814 73 Bratislava. Czechoslovakia.