Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Journal of Computational and Applied Mathematics 166 (2004) 535–551

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Global dynamics of a system governing an algorithm for regression with censored and non-censored data under general errors

Carlos Rivero[a], Ángela Castillo[b], Pedro J. Zufiria[b,*], Teófilo Valdés[a]

[a] *Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas, UCM, Madrid 28040, Spain*

[b] *Departamento de Matemática Aplicada a las Tecnologías de la Información, ETSI de Telecomunicación, UPM, Madrid 28040, Spain*

Received 14 February 2002; received in revised form 14 July 2003

Abstract

We present an investigation into the dynamics of a system, which underlies a new estimating algorithm for regression with grouped and nongrouped data. The algorithm springs from a simplification of the well-known EM algorithm, in which the expectation step of the EM is substituted by a modal step. This avoids awkward integrations when the error distribution is assumed to be general. The sequences generated by the estimating procedure proposed here define our objective system, which is piecewise linear. The study tackles the system's asymptotic stability as well as its speed of convergence to the equilibrium point. In this sense, to reduce the speed of convergence, we propose an alternative estimating procedure. Numerical examples illustrate the theoretical results, compare the proposed procedures and analyze the precision of the estimate.

Keywords: Global dynamics; Iterative estimation; Censored data; Regression; Imputation

1. Introduction

This paper focuses on the global dynamics of a system that derives from an estimating algorithm proposed here, which serves to fit a linear model when some of the dependent values are interval-censored and the error distribution is general.

* Corresponding author. Tel.: +34-9133-67284; fax: +34-9133-67289.
E-mail address: pzz@mat.upm.es (P.J. Zufiria).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2003.09.048

The algorithm mentioned above implies a modification of the well-known EM algorithm [2,3,5–8]. This algorithm starts from an arbitrary initial point and the iteration consists of two steps. The E-step computes the expectation of the complete log-likelihood conditioned on the observed data and the current value of the parameter. The M-step substitutes this current value of the parameter by the argument which maximizes the former expectation.

In the case of exponential families, it is known that only the sufficient statistics need to be taken into consideration. In the case of a linear model with censored data, the sufficient statistic is the sample mean and the E-step is equivalent to filling each censored data by its conditional expectation. Finally, in the exponential family case the M-step agrees with the calculation of the maximum likelihood estimate using the sufficient statistic obtained from the observed data and the above imputations for the censored data. When the errors distribute normally this estimate can be obtained by ordinary least squares.

We propose an algorithm which, first, imputes the censored data by a conditional modal step and, second, the estimate is updated by simple least squares. As we will see, the modal step takes a very simple form and avoids the awkward integrations which appear in the E-step when the distribution of the errors is arbitrary. The simplicity of the algorithm lies also on the least-squares step whose implementation may be carried out using standard computer statistical packages.

The study of the convergence of the algorithm proposed here is the main goal of this paper. The main contribution of the paper characterises the algorithm as a multidimensional piecewise linear system, whose in-depth study allows us to conclude that it is globally convergent to a single equilibrium point. This will define our proposed estimate of the regression coefficients. Additionally, the piecewise characterisation suggests an alternative algorithm, which, on some occasions, may allow us to speed up the search of the equilibrium point. Also, we comment on some issues concerning the selection of convenient initial points, in order to diminish the number of iterations to convergence.

The paper is organised as follows. Section 2 presents the regression model and the iterative procedure for estimating the model parameters. In Section 3 the evolution equations are formulated for the different regions of the state space, given by the censored data. The stability analysis of the global system is addressed in Section 4, together with its piecewise characterisation. An alternative algorithm to seek the unique equilibrium point is presented in Section 5, whereas numerical results are illustrated in Section 6. These results refer to convergence rates of the algorithm and also to the precision of the final estimates. Some concluding remarks are stated in Section 7.

2. Original model and the dynamic system

Let us consider the standard linear regression model

$$z_i = a^{\mathrm{t}} x_i + v_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where a is a fixed m-dimensional parameter to be estimated, x_i is an m-dimensional vector of independent variables, z_i is a real dependent variable and the errors v_i are independent identically distributed random variables following a known density function f, which may be general with the only restriction of being symmetrical and unimodal about zero. It will also be assumed that the dependent data has been collected from different sources and may be either grouped or nongrouped. Then, the index set $I = \{1, 2, \ldots, n\}$ may be partitioned into I^g and I^{ng} containing, respectively, the

indices i depending on whether or not their corresponding values z_i have been grouped. For $i \in I^g$, only an interval $(l_i, u_i]$ which contains z_i is known, whereas, on the contrary, if $i \in I^{ng}$ the exact value z_i is observed.

The ordinary least-squares (OLS) estimate for the regression parameter a, $\hat{a} = (X^t X)^{-1} X^t z$, where $X^t = (x_1, \ldots, x_n)$ and $z = (z_1, \ldots, z_n)^t$ cannot be directly applied if there is grouped data, since the interval-censored z_i's are unknown. In this case, the simple substitution of each grouped z_i by an arbitrary value within its grouping interval yields, in general, undesirable bias. Assuming normal errors and taking into consideration the sufficient statistic, it can be seen that the EM algorithm runs as follows:

Initialisation: Let a_0 be an arbitrary initial value.
Iteration: Assume that the current estimate a_k is known.

- (E-step) Compute the vector $z^*(a_k) = (z_1^*(a_k), \ldots, z_n^*(a_k))^t$, where $z_i^*(a_k) = z_i$ if $i \in I^{ng}$, and

$$z_i^*(a_k) = a_k^t x_i + E(v_i | -a_k^t x_i + l_i < v_i \leqslant -a_k^t x_i + u_i),$$

  if $i \in I^g$ and $z_i \in (l_i, u_i]$.
- (M-step) Evaluate the maximum likelihood estimate, $a_{k+1} = (X^t X)^{-1} X^t z^*(a_k)$.

The conditional expectation is given by the integral equality

$$E(v_i | -a_k^t x_i + l_i < v_i \leqslant -a_k^t x_i + u_i) = \frac{\int_{-a_k^t x_i + l_i}^{-a_k^t x_i + u_i} t f(t)\,\mathrm{d}t}{\int_{-a_k^t x_i + l_i}^{-a_k^t x_i + u_i} f(t)\,\mathrm{d}t}, \tag{2}$$

where f is the density function of the errors. It is known that the EM algorithm converges to the maximum likelihood estimate for the censored data. Thus its asymptotic distribution, as $n \to \infty$, is normal, centred on the true parameter a. We propose to analyse the behaviour of the following algorithm, which is a modification of the former, and to use it for the general class of error distributions commented on above.

Initialisation: Let a_0 be an arbitrary initial value.
Iteration: Assume that the current estimate a_k is known.

- (Modal step) Compute the components of the vector $z(a_k) = (z_1(a_k), \ldots, z_n(a_k))^t$, which are given by $z_i(a_k) = z_i$, if this value is not grouped, and

$$\begin{aligned} z_i(a_k) &= a_k^t x_i + \mathrm{Mode}(v_i | -a_k^t x_i + l_i < v_i \leqslant -a_k^t x_i + u_i) \\ &= \begin{cases} 0 & \text{if } l_i \leqslant a_k^t x_i < u_i, \\ -a_k^t x_i + u_i & \text{if } u_i \leqslant a_k^t x_i, \\ -a_k^t x_i + l_i & \text{if } a_k^t x_i < l_i, \end{cases} \end{aligned} \tag{3}$$

  otherwise.

• (OLS step) Update the current a_k through the equation

$$a_{k+1} = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}z(a_k). \tag{4}$$

The use of conditional modes instead of the conditional expectations of the EM avoids the integration shown in (2), which is substituted by the three alternatives given on the right of (3). These alternatives are much easier to compute than the integrals of (2) with general errors. After the modification, it is not obvious whether or not the transformation $T(\beta) = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}z(\beta)$ has any fixed point β^* and, if it has, if this point is asymptotically stable, globally or locally.

3. Dynamic system explicit formulation

In this section, we prove that the state space $\mathbb{R}^m$, where dynamic system (4) evolves, can be partitioned in different regions. In each one of these regions the behaviour of such a dynamic system is ruled by a time invariant linear difference equation. When the state vector a_k jumps from one region to another, the coefficients of the dynamic system change accordingly, delineating a piecewise characterization.

Let us consider an index $i \in I^g \subset \{1,\dots,n\}$, corresponding to a grouped datum $z_i \in (l_i, u_i]$, where only the extreme values $-\infty \leqslant l_i < u_i \leqslant \infty$ are known. This index $i \in I^g$, determines three sets denoted by

$$R_i^l = \{a \in \mathbb{R}^m | a^{\mathrm{t}}x_i \leqslant l_i\},$$

$$R_i^o = \{a \in \mathbb{R}^m | l_i < a^{\mathrm{t}}x_i \leqslant u_i\},$$

$$R_i^u = \{a \in \mathbb{R}^m | a^{\mathrm{t}}x_i > u_i\}.$$

If $l_i = -\infty$, then $R_i^l = \emptyset$, whereas, if $u_i = \infty$, then $R_i^u = \emptyset$.

Obviously, the sets R_i^l, R_i^u and R_i^o define a partition of the state space $\mathbb{R}^m$ associated with the ith grouped datum. When i varies over I^g, the intersection of these sets provides a new partition

$$\mathscr{P} = \left\{ \bigcap_{i \in I^g} R_i^{h_i} \right\}_{h_i \in \{l,o,u\}}$$

containing, at most, 3^w different nonempty disjoint convex subsets, where w is the cardinal of I^g. For simplicity, we will denote this partition by $\mathscr{P} = \{S_j\}_{j=1,\dots,r}$, where $r \leqslant 3^w$. The following theorem proves that the dynamic system (4) is linear time invariant within every set S_j.

Theorem 1. *If a_k is in the set S_j, then the transition to a_{k+1} is ruled by a linear time invariant dynamic system*

$$a_{k+1} = c(j) + B(j)a_k, \tag{5}$$

where $c(j) \in \mathbb{R}^m$ and $B(j)$ is a square matrix of order m.

Proof. Following a constructive procedure, we will explicitly evaluate the coefficients $c(j)$ and $B(j)$ related to the set S_j. We begin by writing

$$S_j = \bigcap_{i\in I_{L_j}} R_i^l \cap \bigcap_{i\in I_{U_j}} R_i^u \cap \bigcap_{i\in I_{O_j}} R_i^o,$$

where $I_{L_j} = \{i \in I^g \mid h_i = l\}$, $I_{U_j} = \{i \in I^g | h_i = u\}$ and $I_{O_j} = \{i \in I^g | h_i = o\}$ define a partition of I^g. Now, when $a_k \in S_j$, it holds that

$$z_i(a_k) = \begin{cases} z_i & \text{if } i \in I^{ng}, \\ l_i & \text{if } i \in I_{L_j}, \\ x_i^{\mathrm{t}} a_k & \text{if } i \in I_{O_j}, \\ u_i & \text{if } i \in I_{U_j}. \end{cases}$$

Additionally,

$$\begin{aligned} a_{k+1} &= (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}z(a_k) \\ &= (X^{\mathrm{t}}X)^{-1}(X_{ng}^{\mathrm{t}}, X_{L_j}^{\mathrm{t}}, X_{O_j}^{\mathrm{t}}, X_{U_j}^{\mathrm{t}}) \begin{pmatrix} z_{ng} \\ L_j \\ X_{Oj}a_k \\ U_j \end{pmatrix}. \end{aligned} \tag{6}$$

In this notation, $X_{ng}^{\mathrm{t}} = (x_i)_{i\in I^{ng}} = (\ldots x_i \ldots)$, $X_{L_j}^{\mathrm{t}} = (x_i)_{i\in I_{L_j}}$, $X_{O_j}^{\mathrm{t}} = (x_i)_{i\in I_{O_j}}$, $X_{U_j}^{\mathrm{t}} = (x_i)_{i\in I_{U_j}}$, $z_{ng} = (z_i)_{i\in I^{ng}} = (\ldots z_i \ldots)^{\mathrm{t}}$, $L_j = (l_i)_{i\in I_{L_j}}$ and $U_j = (u_i)_{i\in I_{U_j}}$.

Eq. (6) can be rewritten as

$$\begin{aligned} a_{k+1} &= (X^{\mathrm{t}}X)^{-1}(X_{ng}^{\mathrm{t}}z_{ng} + X_{L_j}^{\mathrm{t}}L_j + X_{U_j}^{\mathrm{t}}U_j) + (X^{\mathrm{t}}X)^{-1}X_{O_j}^{\mathrm{t}}X_{Oj}a_k \\ &= c(j) + B(j)a_k, \end{aligned}$$

where $c(j) = (X^{\mathrm{t}}X)^{-1}(X_{ng}^{\mathrm{t}}z_{ng} + X_{L_j}^{\mathrm{t}}L_j + X_{U_j}^{\mathrm{t}}U_j)$ and $B(j) = (X^{\mathrm{t}}X)^{-1}X_{O_j}^{\mathrm{t}}X_{Oj}$, thus completing the proof. □

4. Global stability analysis of the system

In this section we prove, under weak conditions, the global stability of the dynamic system (4), that is, its convergence to a single equilibrium point, regardless the selected initial point. This result is based on a general theorem on the stability of piecewise linear contractive dynamic systems.

First, we show that all the linear dynamic systems (5) are globally asymptotically stable.

Theorem 2. *Let us assume that the matrix $(X^{\mathrm{t}}X)^{ng}$ is positive definite. For all indices $j \in \{1,\ldots,r\}$, the linear dynamic system*

$$a_{n+1} = c(j) + B(j)a_n, \quad n = 1, 2, \ldots,$$

converges to a unique point $a(j) \in \mathbb{R}^m$, whichever initial point a_0 is selected.

Proof. The dynamic system will globally converge to a unique equilibrium point if, and only if, the spectral radius of the matrix $B(j)$ is less than 1. The matrix $B(j)$ is not symmetric but it has real eigenvalues, since it is similar to the symmetric and positive definite matrix $(X^tX)^{-1/2}X_{O_j}^tX_{Oj}(X^tX)^{-1/2}$. Denoting by $\rho(A)$ the spectral radius of matrix A, it holds that

$$\begin{aligned}\rho(B(j)) &= \rho((X^tX)^{1/2}B(j)(X^tX)^{-1/2})\\ &= \rho((X^tX)^{-1/2}X_{O_j}^tX_{Oj}(X^tX)^{-1/2})\\ &= \max_{|u|=1}|u^t(X^tX)^{-1/2}X_{O_j}^tX_{Oj}(X^tX)^{-1/2}u|\\ &= \max_{|u|=1}\frac{u^tX_{O_j}^tX_{Oj}u}{u^tX^tXu} \leqslant \left(1+\frac{\min_{|u|=1}u^t(X^tX)^{ng}u}{\max_{|u|=1}u^tX_{O_j}^tX_{Oj}u}\right)^{-1} = \tau_j < 1,\end{aligned}$$

having assumed that $I_{O_j} \neq \emptyset$. Otherwise, $B(j) = 0$ and $\rho(B(j)) = 0 < 1$. Therefore, in both cases, the theorem holds. □

Corollary 1. *The linear function*

$$f_j(a) = (X^tX)^{1/2}(c(j) + B(j)(X^tX)^{-1/2}a)$$

is strictly contractive in $a \in \mathbb{R}^m$.

Proof. Take arbitrary $x, y \in \mathbb{R}^m$ and observe that, since $(X^tX)^{1/2}B(j)(X^tX)^{-1/2}$ is a symmetric matrix, we can write

$$\begin{aligned}\|f_j(x) - f_j(y)\| &= \|(X^tX)^{1/2}B(j)(X^tX)^{-1/2}(x-y)\|\\ &\leqslant \tau_j\|x-y\|,\end{aligned}$$

where $\tau_j < 1$. □

The following lemma will lead to prove the global asymptotic stability of piecewise contractive functions.

Lemma 1. *Let* $f : \mathbb{R}^m \to \mathbb{R}^m$ *be defined by* $f(x) = \sum_{j=1}^r I_{S_j}(x)f_j(x)$, *where the functions* f_j *are strictly contractive and* I_{S_j} *is the indicator function of* S_j, *which are convex sets forming a partition of* $\mathbb{R}^m$. *If* f *is a continuous function, then* f *is strictly contractive.*

Proof. For any $x, y \in S_j$

$$\|f(x) - f(y)\| = \|f_j(x) - f_j(y)\| \leqslant \tau_j\|x-y\| \leqslant \tau\|x-y\|,$$

where $\tau = \max_{i=1}^r \tau_i < 1$. On the contrary, if $x \in S_i$ and $y \in S_j$ $(i \neq j)$, and denoting $[x, y]$ the linear segment of extreme points x and y, there exist points $a_k \in [x, y]$, $k \in \{1, \ldots, s\} \subseteq \{1, \ldots, r\}$,

such that

$$[x, y] = \bigcup_{k=1}^{s-1} [a_k, a_{k+1}],$$

$\{(a_k, a_{k+1})\}_{k=1}^{s-1}$ are disjoint intervals, $x = a_1$, $y = a_s$, and $a_k, a_{k+1} \in \mathrm{cl}(S_{h_k})$. From the continuity of f, it follows, for all $k = 1, \ldots, s-1$, that

$$f(a_k) = f_{h_k}(a_k) \quad \text{and} \quad f(a_{k+1}) = f_{h_k}(a_{k+1}) = f_{h_{k+1}}(a_{k+1}).$$

Hence, we can write

$$\begin{aligned}
\|f(x) - f(y)\| = \|f_i(x) - f_j(y)\| &\leqslant \sum_{k=1}^{s-1} \|f(a_k) - f(a_{k+1})\| \\
&= \sum_{k=1}^{s-1} \|f_{h_k}(a_k) - f_{h_k}(a_{k+1})\| \\
&\leqslant \sum_{k=1}^{s-1} \tau_{h_k} \|a_k - a_{k+1}\| \leqslant \tau \sum_{k=1}^{s-1} \|a_k - a_{k+1}\| \\
&= \tau \|x - y\|. \qquad \square
\end{aligned}$$

Corollary 2. *The piecewise linear function*

$$f(a) = (X^{\mathrm{t}}X)^{-1/2} X^{\mathrm{t}} z((X^{\mathrm{t}}X)^{-1/2} a)$$

is contractive in $a \in \mathbb{R}^m$.

Proof. Since $z(a)$ is a continuous function f is also continuous. It suffices to consider the contractive functions defined in Corollary 1, $f_j(a) = (X^{\mathrm{t}}X)^{1/2}(c(j) + B(j)(X^{\mathrm{t}}X)^{-1/2}a)$, when $(X^{\mathrm{t}}X)^{-1/2}a \in S_j$, and apply Lemma 1. $\square$

Theorem 3. *The piecewise linear dynamic system*

$$a_{n+1} = (X^{\mathrm{t}}X)^{-1} X^{\mathrm{t}} y(a_n), \quad a_0 \in \mathbb{R}^m$$

converges to a unique point, whichever initial point is taken.

Proof. Define $b_n = (X^{\mathrm{t}}X)^{1/2} a_n$, and observe that

$$b_{n+1} = (X^{\mathrm{t}}X)^{1/2} a_{n+1} = f((X^{\mathrm{t}}X)^{1/2} a_n) = f(b_n),$$

where f agrees with that defined in Corollary 2. Using this, we guarantee that b_n converges to the unique fixed point, say $\hat{b}$, of the strictly contractive function f. As the matrix $(X^{\mathrm{t}}X)^{1/2}$ is definite positive, we can guarantee that a_n also converges, for any initial point a_0, to $\hat{a} = (X^{\mathrm{t}}X)^{-1/2}\hat{b}$, the

unique solution of the equation

$$\hat{a} = (X^{t}X)^{-1}X^{t}z(\hat{a}),$$

thus completing the proof. □

Remarks. (i) Subsystem (5) associated with the region S_j is globally convergent to the point $a(j) = (I - B(j))^{-1}c(j)$, which can be computed via standard routines. Since

$$(I - B(j))^{-1} = (X^{t}X)^{-1}_{I-O_j}(X^{t}X),$$

where $X^{t}_{I-O_j} = (x_i)_{i\in I-I_{O_j}}$ and $(X^{t}X)_{I-O_j} = X^{t}_{I-O_j}X_{I-O_j}$, it holds that

$$a(j) = (X^{t}X)^{-1}_{I-O_j}X^{t}{}_{I-O_j}\begin{pmatrix} z_{ng} \\ L_j \\ U_j \end{pmatrix}.$$

Thus, $a(j)$ agrees with the OLS estimate considering only the censored data which are imputed as l_i if $i \in I_{L_j}$ and u_i if $i \in I_{U_j}$.

(ii) In each region S_j there exists a fixed number of constraints of the type $l_i < a^{t}x_i \leqslant u_i$, which are held by all $a \in S_j$. This number s $(0 \leqslant s \leqslant w)$ determines the complexity of the accompanying matrix of the subsystem associated with S_j. If $s = 0$, then $I_{O_j} = \emptyset$, which implies that $B(j) = 0$ and subsystem (5) is constant, $a_{k+1} = c(j)$. Thus, if the objective system (4) reaches the state $a_k \in S_j$ at step k, then $a_{k+1} = c(j)$. Therefore, $a_{k+1} = a_{k+2} = \cdots = c(j)$ if $c(j) \in S_j$, and $a_{k+2} = c(j^*) + B(j^*)c(j)$, if $c(j) \in S_{j^*}$. If $s = 1$, then I_{O_j} is unitary. Let i be the only point of I_{O_j}. Then $X^{t}_{O_j} = x_i$, and subsystem (5) associated to S_j becomes

$$a_{k+1} = c(j) + (X^{t}X)^{-1}x_i x_i^{t}a_k.$$

When a_k, $a_{k+1} \in S_j$ it follows that the differences $a_{k+2} - a_{k+1}$ are proportional to the constant vector $(X^{t}X)^{-1}x_i$. In general, for $s > 1$, the rank of the matrix $B(j) = (X^{t}X)^{-1}X^{t}_{O_j}X_{O_j}$ is at most $\min\{m, s\}$.

(iii) The most likely situation is that the limit point $\widehat{a} \in \mathrm{Int}(S_j)$, for some region S_j. In this case, objective system (4) maintains in the region S_j from a certain step onward. This means that the jth subsystem governs the asymptotic behaviour of (4), which tends to be within the vector space generated by the eigenvectors associated with the dominant eigenvalue. If algebraic multiplicity of the latter is 1, then the system approximates the limit point in the direction given by the unique eigenvector. If $\hat{a} \notin \mathrm{Int}(S_j)$, for some j, there may be several regions S_j for which $\hat{a} \in Fr(S_j)$. In this case, the asymptotic behaviour of objective system (4) may be ruled by one or several subsystems.

5. An alternative algorithm

It is known that system (4) converges to a unique equilibrium point, $\hat{a}$, which satisfies the implicit equation

$$\hat{a} = (X^{t}X)^{-1}X^{t}z(\hat{a}).$$

If we assume that $\hat{a} \in S_j$, then $\hat{a} = c(j) + B(j)\hat{a}$ and $\hat{a} = a(j)$ is the unique limit point of the jth linear subsystem. Additionally, it is clear that there exists a unique index $j \in \{1, \ldots, r\}$ such that $a(j) \in S_j$. Thus, the search of the point $\hat{a}$, can be restricted over the finite set $\{a(1), \ldots, a(r)\}$.

This suggests the following new algorithm for determining the point $\hat{a}$ and also provides an optimum stop criterion.

1. Take an arbitrary value $a \in \mathbb{R}^m$ in a region which is not labelled.
2. Determine the index j such that $a \in S_j$. If region S_j is labelled, return to 1. Otherwise, evaluate $a(j) = (I - B(j))^{-1}c(j)$, the fixed point of the linear subsystem (5), and put a label on the region S_j.
3. If $a(j) \in S_j$, the algorithm stops (since $\hat{a} = a(j)$). Otherwise, take $a = a(j)$ and return to 2.

Since the regions S_j are polytopes, the selection of a point in such regions can be performed via specific procedures (see [4]). The new algorithm determines the solution in a number of iterations less than r. Since each iteration requires solving a linear algebraic system, if direct methods are employed, the number of operations (sums and products) needed in each iteration is $O(m^3)$, unlike the algorithm proposed in (4) which requires $O(m^2)$ operations. If we denote the number of iterations of both the original algorithm (4) and the alternative by N_1 and N_2, respectively, then the latter algorithm is more efficient than the former if $mN_1 > N_2$, as it occurs in the example shown in Section 6.

6. Some numerical results

The following data was presented in [1] and, later, analysed in [9,10]. The data is the result of a temperature accelerated life test on electrical insulation in 40 motorettes. Ten motorettes were tested at each one of the four following temperatures, 150°, 170°, 190° and 220° (Table 1). Testing was terminated at different times at each temperature, resulting in a total of 17 failed units and 23 unfailed ones. The exact results of the failure times are given below, where l_i^+ indicates that the correspondent datum is still on test without failure at l_i hours.

Table 1
Results of failure times for the numerical example

150°	170°	190°	220°
8064^+	1764	408	408
8064^+	2772	408	408
8064^+	3444	1344	504
8064^+	3542	1344	504
8064^+	3780	1440	504
8064^+	4860	1680^+	528^+
8064^+	5196	1680^+	528^+
8064^+	5448^+	1680^+	528^+
8064^+	5448^+	1680^+	528^+
8064^+	5448^+	1680^+	528^+

As in [10], we consider the linear model was fitted

$$z_i = a_1 + a_2 t_i + v_i, \quad i = 1, \ldots, 40,$$

where $z_i = \log(i\text{th failture time})$, $t_i = 1000(\text{temperature} + 273.2)^{-1}$ and v_i is an arbitrary unimodal distribution of error (in [10] normal distributions were assumed). Observe that the values with sign $+$ are censored in the interval (l_i^+, ∞) and therefore we are ready to apply algorithm (4) presented in Section 2. The 23 censored values partition the parameter space $\mathbb{R}^2$ in $r = 11$ different regions, $\{S_j\}_{j=0,1\ldots,10}$. Observe that there are four different values for t_i and l_i^+:

$$t_1 = 2.3629, \quad l_1^+ = \log(8064),$$

$$t_2 = 2.2563, \quad l_2^+ = \log(5448),$$

$$t_3 = 2.1588, \quad l_3^+ = \log(1680),$$

$$t_4 = 2.0275, \quad l_4^+ = \log(528).$$

The regions S_j are built from the four hyperplanes

$$H_i\colon\ a_1 + t_i a_2 = \log(l_i^+), \quad i = 1, 2, 3, 4$$

which are plotted in Fig. 1. Each hyperplane H_i determines two halfspaces, $R_i^l = \{a_1 + \tau_i a_2 \leqslant l_i^+\}$ and $R_i^o = \{a_1 + \tau_i a_2 > l_i^+\}$, and the intersection of all of these halfspaces determines the aforementioned regions $\{S_j\}_{j=0,\ldots,10}$.

Let us pay attention to the following three important regions:

$$S_0 = R_1^l \cap R_2^l \cap R_3^l \cap R_4^l,$$

$$S_1 = R_1^o \cap R_2^l \cap R_3^l \cap R_4^l,$$

$$S_2 = R_1^o \cap R_2^o \cap R_3^o \cap R_4^o,$$

which are showed in Fig. 1. No constraint of the type $l_i < a^{\mathrm{t}} x_i \leqslant u_i$ $(i \in I^g)$ is fulfilled in S_0, exactly one (for $i = 1$) is fulfilled in S_1 and, finally, all the constraints are fulfilled in S_2. Therefore,

$$B(0) = 0,$$

$$\operatorname{rank} B(1) = \operatorname{rank}(X_{O_1}) = \operatorname{rank}\begin{pmatrix} 1 \\ t_1 \end{pmatrix} = 1 \text{ and}$$

$$\operatorname{rank} B(2) = \operatorname{rank}(X_{O_2}) = \operatorname{rank}\begin{pmatrix} 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 \end{pmatrix} = 2.$$

Next, we illustrate the dynamics of the algorithm for different initial points. First, we have verified that, independently on the initial point $a_0 \in \mathbb{R}^2$, algorithm (4) converges to a unique limit point $\hat{a} = (-5.2250, 3.8868)^{\mathrm{t}}$, as Theorem 3 states. Taking an arbitrary initial point $a_0 \in S_0$ the system firstly jumps to the state value $a_1 = c(0) = (-4.9326, 3.7480)^{\mathrm{t}}$ that belongs to S_1 and then remains indefinitely in S_1 converging to $\hat{a} \in S_1$. Since $\operatorname{rank} B(1) = 1$, the points a_k lie in a line. This fact

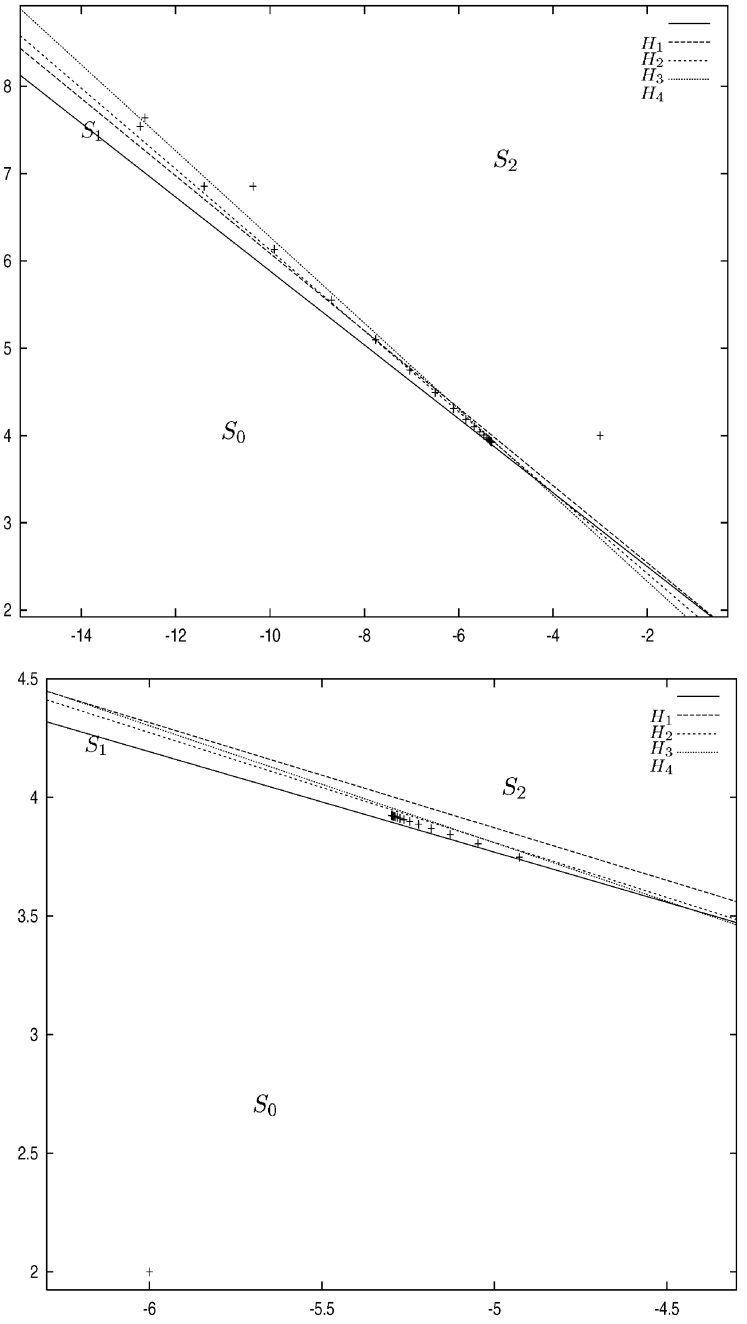

Fig. 1. Two algorithm trajectories from different initial conditions in the numerical example. The top figure corresponds to $a_0 = (-3, 4) \in S_2$, whereas the bottom figure corresponds to $a_0 = (-6, 2) \in S_0$.

Table 2
Characterisation of a trajectory with initial point $(a_{0,1}, a_{0,2})$ in S_0 for the numerical example (the line indicates change of region)

Iteration	$a_{k,1}$	$a_{k,2}$	$v_{k,1}$	$v_{k,2}$	$\|d_k\|$
0	$a_{0,1}$	$a_{0,2}$			
1	−4.9326	3.7480	0.8790	0.4767	5.76452
2	−5.0271	3.7929	−0.9034	0.4287	0.10464
3	−5.0911	3.8232	−0.9034	0.4287	0.07081
4	−5.1344	3.8438	−0.9034	0.4287	0.04791
5	−5.1637	3.8577	−0.9034	0.4287	0.03242
6	−5.1835	3.8671	−0.9034	0.4287	0.02194
7	−5.1969	3.8734	−0.9034	0.4287	0.01484
8	−5.2060	3.8777	−0.9034	0.4287	0.01004
9	−5.2121	3.8807	−0.9034	0.4287	0.00679
10	−5.2163	3.8826	−0.9034	0.4287	0.00460
11	−5.2191	3.8840	−0.9034	0.4287	0.00311
12	−5.2210	3.8849	−0.9034	0.4287	0.00210
13	−5.2223	3.8855	−0.9034	0.4287	0.00142
14	−5.2232	3.8859	−0.9034	0.4287	0.00096
15	−5.2237	3.8862	−0.9034	0.4287	0.00065
20	−5.2248	3.8867	−0.9034	0.4287	0.00009
25	−5.2250	3.8868	−0.9034	0.4287	0.00001
26	−5.2250	3.8868	−0.9034	0.4287	0.00000

can be recognised from Table 2, which shows: (i) the points $a_k = (a_{k,1}a_{k,2})^t$, (ii) the normalized difference vectors $n_k = d_k/\|d_k\|$ (where $d_k = a_k - a_{k-1}$), and (iii) the norm $\|d_k\|$, from iteration $k = 1$–26. The line in the table indicates that the sequence changes from one region to another. The sequence with starting point $a_0 = (-6, 2)^t$ is shown in Fig. 1.

Now, for the initial point $a_0 = (-3, 4)^t \in S_2$ we observe that a_1 and a_2 stay in S_2 where the system is governed by

$$a_{k+1} = B(2)a_k + c(2),$$

where the matrix

$$B(2) = \begin{pmatrix} -1.9381 & -5.7979 \\ 1.1416 & 3.2166 \end{pmatrix}$$

has rank two, and $c(2) = (7.0128, -2.5868)^t$.

The sequence generated by the algorithm is shown in Table 3 (with the same structure as Table 2). In the iteration $k = 8$ the sequence enters in the region S_1, keeping the vector n_{k+1} invariant from this iteration on. This is a consequence of the fact that the dynamic system associated with region S_1 is determined by the rank one matrix

$$B(1) = \begin{pmatrix} -5.5670 & -13.1544 \\ 2.6424 & 6.2437 \end{pmatrix}.$$

Table 3
Characterization of a trajectory with initial point in S_2 for the numerical example (the line indicates change of region)

Iteration	$a_{k,1}$	$a_{k,2}$	$v_{k,1}$	$v_{k,2}$	$\|d_k\|$
0	−3.0000	4.0000			
1	−10.3645	6.8549	−0.9323	0.3614	7.89851
2	−12.6435	7.6307	−0.9466	0.3222	2.40743
3	−12.7248	7.5247	−0.6084	−0.7936	0.13364
4	−11.3325	6.8200	0.8922	−0.4515	1.56041
5	−9.8261	6.0906	0.9000	−0.4357	1.67368
6	−8.6069	5.5036	0.9009	−0.4338	1.35315
7	−7.6601	5.0480	0.9011	−0.4335	1.05074
8	−6.9331	4.6985	0.9012	−0.4333	0.80663
9	−6.3936	4.4414	0.9027	−0.4301	0.59765
10	−6.0157	4.2621	0.9034	−0.4287	0.41822
11	−5.7601	4.1407	0.9034	−0.4287	0.28300
12	−5.5871	4.0586	0.9034	−0.4287	0.19150
13	−5.4700	4.0031	0.9034	−0.4287	0.12958
14	−5.3908	3.9655	0.9034	−0.4287	0.08769
15	−5.3372	3.9400	0.9034	−0.4287	0.05933
20	−5.2409	3.8943	0.9034	−0.4287	0.00841
25	−5.2272	3.8878	0.9034	−0.4287	0.00119
30	−5.2253	3.8869	0.9034	−0.4287	0.00016
35	−5.2250	3.8868	0.9034	−0.4287	0.00002
36	−5.2250	3.8868	0.9034	−0.4287	0.00001
37	−5.2250	3.8868	0.9034	−0.4287	0.00001
38	−5.2250	3.8868	0.9034	−0.4287	0.00000

The sequences converge to the limit point $\hat{a}=(-5.2250, 3.8868)^{\mathrm{t}}$ following the eigenvector associated with the dominant eigenvalue of $B(1)$, $(0.9034, -0.4288)^{\mathrm{t}}$, which appears in Tables 2 and 3.

In order to analyse the algorithm speed of convergence, a stop condition must be fixed. In our case, this condition was $\|d_k\|_1 = \|a_k - a_{k-1}\|_1 < 10^{-4}$. Fig. 2 depicts the number of iterations required to achieve this stop condition, for different initial points a_0. A grid of initial conditions was employed to generate this figure, both on a large rectangle (top part) and on a small one (bottom part) containing the limit point $\hat{a}$. It can be seen that the number of iterations remains constant for all $a_0 \in S_0$. When the starting point is chosen in S_2, a type of staircase evolution occurs, and the number of iterations increases as the initial point moves away from $\hat{a}$, accordingly with matrix $B(2)$. The bottom part of Fig. 2 shows a sink around the limit point $\hat{a}$, where the number of iterations reaches the value zero.

Note that, although there exist initial points for which the required number of iterations is small (e.g., less than 20), they are within a very narrow region. From a practical point of view, it is unlikely that one may choose one of these such points as an initial value (if we think in terms of volume of the regions). It is more likely to select an arbitrary initial point in the unbounded region S_0. In this case, 20 iterations are required to reach the limit point within the fixed accuracy.

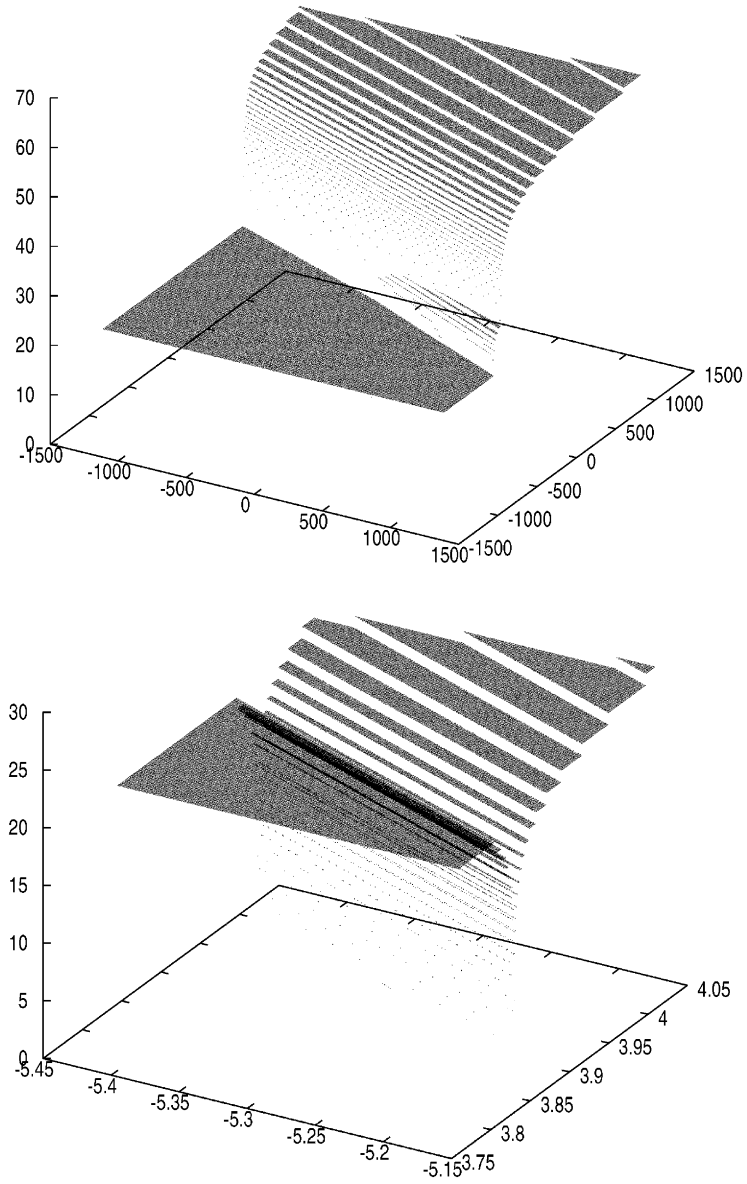

Fig. 2. Number of algorithm iterations required for convergence in the numerical example. The bottom figure is focused in a small neighborhood of the limit point.

If we start from region S_2 the required number of iterations may be too large. One can see that this number may have a significant dependency on the selected initial value. This suggests the need for developing some appropriate criteria for choosing the initial iterate.

The behaviour of the alternative algorithm presented in Section 5 is also illustrated. The selected initial iteration points are the same as in the previous computations. Table 4 indicates the behaviour

Table 4
Alternative algorithm iteration results for any initial point in S_0

Iteration	$a_{k,1}$	$a_{k,2}$
0		
1	−4.932643	3.748032
2	−5.225032	−3.886814

Table 5
Alternative algorithm iteration results for any initial point in S_2

Iteration	$a_{k,1}$	$a_{k,2}$
0		
1	−5.149491	3.819083
2	−4.932643	3.748032
3	−5.225032	−3.886814

for any initial point belonging to region S_0, whereas Table 5 displays such behaviour for any initial point belonging to region S_1. Note that the number of iterations is considerably smaller, $N_2 = 2$ or 3. Since $m = 2$, the linear systems can be solved by direct methods and, in general, $N_2 \ll mN_1$, which shows the high efficiency of the alternative procedure.

Finally, practical statisticians are interested in measuring the precision of the estimates. To this end, we have used the jackknife method to evaluate the covariance matrix, Σ, of the point estimate $\hat{a} = (-5.2250, 3.8868)^{\mathrm{t}}$. As it is known (see [11, pp. 154–155]), once the sample data has been randomly partitioned in g groups of m observations the jackknife estimate of Σ is given by

$$\hat{\Sigma}_{g,m} = \frac{1}{g(g-1)} \sum_{\alpha=1}^{g} (J_\alpha - \hat{a})(J_\alpha - \hat{a})^{\mathrm{t}},$$

where the pseudovalue J_α is defined as

$$J_\alpha = g\hat{a} - (g-1)\hat{a}_\alpha$$

and $\hat{a}_\alpha$ is the estimate of a computed from the reduced sample of $m(g-1)$ observations which results after omitting the observations of the αth group. For values given in Table 1 and taking $m = 1$ and $g = 40$ we have obtained

$$\hat{\Sigma}_{40,1} = \begin{pmatrix} 0.2687 & -0.1285 \\ -0.1285 & 0.0616 \end{pmatrix},$$

which leads to the following confidence intervals for a_1 and a_2 at level 0.95:

$$I(a_1) = (4.6820, 5.7680) \text{ and } I(a_2) = (3.7623, 4.0113).$$

Equally, for $m = 2$ and $g = 20$ the jackknife estimate of Σ and the corresponding intervals for a_1 and a_2 at level 0.95 have been

$$\hat{\Sigma}_{20,2} = \begin{pmatrix} 0.2460 & -0.1170 \\ -0.1170 & 0.0559 \end{pmatrix}$$

and

$$I(a_1) = (4.7101, 5.7399) \text{ and } I(a_2) = (3.7698, 4.0038).$$

7. Concluding remarks

In this paper an analysis of the global dynamics of an iteration scheme for performing regression with grouped data has been presented. From a practical point of view, the estimation algorithm proposed here is an alternative to the EM algorithm which avoids the integration involved in the E-step as it is substituted by a much simpler step. This implies the reduction of both the programming and the computation time.

The dynamical system of the iteration procedure has been shown to be piecewise linear time invariant, having a unique equilibrium point, which is asymptotically stable. In addition, a partition has been defined in the whole state space, so that in each composing region the dynamics evolve in some affine subspaces. The nature of the problem has led us also to define an alternative algorithm for computing the equilibrium point of the objective system in a finite number of iterations. Finally, on the one hand, a numerical example shows the performance of system (5), whose speed of convergence may be reduced with the alternative procedure; on the other hand, the covariance matrix of the estimate that we have obtained has been evaluated using the jackknife method.

Acknowledgements

This work has been partially supported by Projects PB97-0566-C02, BFM2000-1475 and SEC990402 of the MEC and MCyT of Spain.

References

[1] D.E. Crawford, Analysis of incomplete life test data on motorettes, Insulation/Circuits 16 (10) (1970) 43–48.
[2] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. B 39 (1977) 1–38.
[3] K. Lange, A gradient algorithm locally equivalent to de EM algorithm, J. Roy. Statist. Soc. B 57 (1995) 425–437.
[4] D.M.W. Leenaerts, W.M.G. van Bokhoven, Piecewise Linear Modeling and Analysis, Kluwer, Boston, MA, 1998.
[5] R.J.A. Little, D.B. Rubin, Statistical Analysis With Missing Data, Wiley, New York, 1987.
[6] T.A. Louis, Finding the observed information matrix when using the EM algorithm, J. Roy. Statist. Soc. Ser. B 44 (1982) 226–233.

[7] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, Wiley, New York, 1997.
[8] I. Meilijson, A fast improvement to the EM algorithm on its own terms, J. Roy. Statist. Soc. Ser. B 51 (1989) 127–138.
[9] J. Schmee, G.J. Hahn, A simple method for regression analysis with censored data, Technometrics 21 (1979) 417–432.
[10] M.A. Tanner, Tools for Statistical Inference. Observed Data and Data Augmentation Methods, Springer, Berlin, 1993.
[11] K.M. Wolter, Introduction to Variance Estimation, Springer, Berlin, 1985.