Robust Frequency Estimation Using Elemental Sets[*]

Gordon K. Smyth
Department of Mathematics,
University of Queensland, Australia 4072

Douglas M. Hawkins
Department of Applied Statistics
University of Minnesota, St Paul, MN 55108

Abstract

The extraction of sinusoidal signals from time-series data is a classic problem of ongoing interest in the statistics and signal processing literatures. Obtaining least squares estimates is difficult because the sum of squares has local minima $O(1/n)$ apart in the frequencies. In practice the frequencies are often estimated using ad hoc and inefficient methods. Problems of data quality have received little attention. An elemental set is a subset of the data containing the minimum number of points such that the unknown parameters in the model can be identified. This paper shows that, using a variant of the classical method of Prony, parameter estimates for a sum of sinusoids can be obtained algebraically from an elemental set. Elemental set methods are used to construct finite algorithm estimators which approximately minimize the least squares, least trimmed sum of squares or least median of squares criteria. The elemental set estimators prove able in simulations to resolve the frequencies to the correct local minima of the objective functions. When used as the first stage of an MM estimator, the constructed estimators based on the trimmed sum of squares and least median of squares criteria produce final estimators which have high breakdown properties and which are simultaneously efficient when no outliers are present. The approach can also be applied to sums of exponentials, and sums of damped sinusoids. The paper includes simulations with one and two sinusoids and two data examples.

Keywords: frequency estimation, sums of exponential functions, elemental sets, least trimmed sum of squares, least median of squares, MM estimators, high breakdown, high efficiency.

1 Introduction

1.1 A Special Nonlinear Regression Problem

The normal-error nonlinear regression model is

$$y_t = \mu(t) + \epsilon_t, \quad \mu(t) = g(\mathbf{x}_t, \boldsymbol{\beta}), \quad \epsilon_t \sim N(0, \sigma^2)$$

$t = 1, \ldots, n$, where the mean function $\mu(t)$ depends nonlinearly on some unknown coefficient vector $\boldsymbol{\beta}$ and some covariate vector $\mathbf{x}_t$, and we assume normal homoscedastic errors ϵ_t. One special but important subclass of nonlinear regression models is that in which $g(\mathbf{x}_t, \boldsymbol{\beta})$ is a sum of sinusoids in t with the frequencies, phases and amplitudes constituting the $\boldsymbol{\beta}$.

Fitting these sinusoidal models can involve daunting computational problems. The maximum likelihood estimators are ordinary least squares estimators, and the least squares function has local minima spaced $O(n^{-1})$ apart, making the gradient-based search methods of general non-linear regression ineffective without excellent starting values.

This problem is further complicated by the possibility of outliers in the data. These can be incorporated in the distributional model by assuming that the majority of the ϵ_t follow the $N(0, \sigma^2)$ distribution but that a minority are "contaminated". This possibility multiplies the difficulties; not only do we have to find elusive estimators, but have to do so in the face of data that disrupt the usual fitting criteria.

In this paper we develop methods which, with high probability, resolve the frequencies to the correct local minima of the objective function, and which have high breakdown points against outliers.

1.2 The Elemental Set Method

The method of elemental sets is widely used in linear regression as a way of getting reasonable parameter estimates in data sets that may have substantial contamination, even by extreme and badly-placed outliers (Marazzi, 1991). It involves performing many fits to a given data set, each fit made to a subsample just large enough to identify the parameters. For example, in a linear regression with p coefficients, the method involves sampling p of the cases and performing an exact fit to get values for the p coefficients. Repeating this with different subset of p cases leads to different coefficient vectors. A key characteristic of these coefficient vectors is that, by definition, each has some support in the data. See Hawkins (1993) for a discussion of the statistical properties of these coefficient vectors, and Rousseeuw (1984), Hawkins, Bradu and Kass (1984), and Rousseeuw and Leroy (1986) for illustrations of the approach for handling linear regression outliers.

The extension of elemental set methods to get high breakdown estimators in nonlinear regression has been considered recently by Stromberg and Ruppert (1992) and by Stromberg (1993, 1995). One difficulty is that the computation of exact fits to elemental sets is far from trivial in general nonlinear regression. This paper shows that this difficulty does not apply to sinusoidal regressions when the observations are at equi-spaced times as the elemental set estimates can be computed algebraically. The approach actually applies to any function $\mu(t)$ which solves a homogeneous differential equation with constant coefficients, notably sums

of exponential functions and damped sinusoids as well as sums of sinusoids, although we concentrate on the sinusoidal problem here. Even when the times are not equi-spaced, it is sometimes possible to interpolate to equi-spaced times for the purposes of the elemental sets as we show in Section 5.2.

The elemental set method is useful for optimizing criteria which are not smooth, have many local minima or are otherwise not amenable to global optimization by refinement in the parameter space. There are other classes of algorithms, notably simulated annealing and genetic algorithms, which can also deal with multiple minima, but these have so far been less used than elemental sets for detecting outliers in linear models. Atkinson and Weisberg (1991) present some performance figures showing the simulated annealing is not competitive with the other approaches. Genetic algorithms are more closely competitive on the linear problem, and are therefore candidates for the nonlinear problem also. However they would be more complex to program than the elemental set method and seem in general more suited to problems where the objective function has more structure around the local minima.

1.3   Criterion Functions

The method of elemental sets involves many repeated fits using different elemental sets, each fit providing an estimate of the parameters and a set of estimated residuals. We can use these different fits in two ways. One is to "average" the parameter estimates from the different fits in some sensible way. The other is to look across the different elemental sets for that one yielding the minimum of some criterion function. This is logically related to the random search for the least squares frequency estimate suggested by Rosenblatt and Rice (1988), with the difference that each estimate considered has some support in the data so that the candidates are relatively dense in the neighborhood of the true values. In either approach, we can then use the estimate produced by the elemental sets as the starting point for some further iterative refinement.

Write $\hat{\mu}(t)$ for some estimate of the regression function of case t. This corresponds to the estimate of the residual

$$e_t = y_t - \hat{\mu}(t).$$

Conventional estimates are found by the LS criterion – minimizing the sum of squares of all n of these residuals. This criterion is motivated by ideas of statistical efficiency but is inappropriate if some of the residuals may be contaminated. The wish to protect the estimate from such contamination leads to the minimization of outlier-insensitive criteria. Two well-known examples are the "least trimmed squares" or LTS criterion and "least median of squares" or LMS criterion. LTS aims to minimize the sum of squares of the h smallest absolute residuals while LMS aims to minimize the hth smallest squared residual. The coverage h is chosen to reflect the amount of protection against outliers that is felt appropriate, with choices slightly above one-half the data set size reflecting the default of seeking the maximum possible protection.

For any criterion, an estimator can be defined with which minimizes the criterion over a large number of elemental sets. In this paper we report on simulations with the LS, LTS and LMS criteria.

1.4 Multistage Estimators

A feature of the elemental sets method is that it is finite. It can therefore be used to generate preliminary estimators with which to initialize iterative estimation schemes such as maximum likelihood or M estimation. In particular, an estimator which combines high breakdown with high efficiency under normal errors can be achieved by using a high breakdown elemental estimator to initialize the MM estimation scheme of Yohai (1987) and Stromberg (1993). We use our elemental set estimators as part of a multistage estimator similar to that of Stromberg (1993). We find that the elemental LTS and LMS estimators are successful in resolving the frequencies to the correct local minima of the objective function even in the presence of outliers. We find that both criterion are successful as stage one estimators for an MM estimation scheme, and that in simulations the final estimators have both high breakdown properties again outliers and high efficiency when there are no outliers.

2 Computation of Elemental Set Estimates

2.1 Constant Coefficient Differential Equations

The regression functions $\mu(t)$ we consider are those which solve constant coefficient differential equations of the form

$$\sum_{k=1}^{p+1} \xi_k D^{k-1} \mu = 0 \tag{1}$$

where D is the differential operator. Solutions to (1) include complex exponentials, damped and undamped sinusoids and real exponentials, depending on the roots of the polynomial with the ξ_k as coefficients (Brockwell and Davis, 1991; Osborne and Smyth, 1995). Let the roots be β_j, $j = 1, \ldots, p$. If the β_j are distinct, then $\mu(t)$ is a sum of exponentials,

$$\mu(t) = \sum_{j=1}^{p} \alpha_j \exp(\beta_j t) \tag{2}$$

where in general the α_j and β_j may be complex. If the α_j and β_j are not all real, then they must occur in complex conjugate pairs in order that $\mu(t)$ be a real signal. If the roots are purely imaginary, then $\mu(t)$ is a sinusoidal signal

$$\mu(t) = \sum_{j=1}^{p/2} \alpha_j \sin(\omega_j t + \phi_j) \tag{3}$$

for real α_j and ϕ_j and $\omega_j \in [0, \pi)$.

The extraction of sinusoidal signals from time-series data and the least squares fitting of exponential signals are highly nonlinear problems which have attracted enormous attention in the statistics and engineering literatures. Algorithms which fit sums of real exponential functions typically require excellent starting values and frequently have difficulty converging because of collinearity between the exponential decay functions. Sinusoidal signals on

4

the other hand are asymptotically uncorrelated, but give rise to highly non-convex sums of squares having local minima $O(n^{-1})$ apart in the frequencies (Rosenblatt and Rice, 1988). This has typically prevented the use of fully efficient global least squares estimators in practice. Even in the 'clean-data' case then, this class of problems presents substantial computational difficulties. For neither type of problem have the even harder problems of data quality have received much attention. Both however turn out to be amenable to elemental set methods.

The basis of our method is that discrete values of $\mu(t)$ at equi-spaced times satisfy exactly a difference equation, which is related in a known way to the differential equation (1). We suppose that observations, $y_i = \mu(t_i) + \epsilon_i$, are made at equi-spaced times t_i, $i = 1, \ldots, n$, where the ϵ_i are independent with mean zero and variance σ^2. Write $\mu_i = \mu(t_i)$. The discrete μ_i satisfy the difference equation

$$c_1\mu_i + \ldots + c_{p+1}\mu_{i+p} = 0 \tag{4}$$

$i = 1, \ldots, n-p$, where the c_j are the coefficients of the polynomial $p_c(z) = c_1 + c_2 z + \cdots + c_{p+1}z^p$ with roots $e^{\beta_j \delta t}$ and δt is the spacing unit $t_{i+1} - t_i$.

2.2   Elemental Set Estimators

An elemental set estimator consists of equating μ_i to y_i for the sequence of observations comprising the elemental set and solving the resulting system of linear equations to identify the c_j. Doing so requires p equations, each involving p equi-spaced observations. The estimator therefore gives closed-form estimates given a subsequence of $2p$ equi-spaced observations. Having found the c_j, the roots of the polynomial are found to identify the β_j. Then another $p \times p$ linear system is solved to find the α_j.

The elemental set computation is in fact an application of the classical interpolation method of Prony (1795). See Osborne and Smyth (1991, 1995), Kahn et al (1992), Mackisack et al (1994) and Smyth (1999) for modern least squares modifications of Prony's method.

Suppose that $\mu(t)$ is a general solution to the differential equation (1). Consider an elemental set of $2p$ observations spaced s apart, $\mathbf{y}_e = (y_i, y_{i+s}, \ldots, y_{i+(2p-1)s})^T$. The elemental estimate of $\mathbf{c}_s$ is the null vector of the $(p+1) \times p$ matrix

$$Y_i = \begin{pmatrix} y_i & \cdots & y_{i+(p-1)s} \\ \vdots & & \vdots \\ y_{i+ps} & \cdots & y_{i+(2p-1)s} \end{pmatrix}$$

Without loss of generality we can put $c_{p+1} = 1$. Also let $\mathbf{d} = (c_1, \ldots, c_p)^T$. Writing B_i^T for the first p rows of Y_i and $\mathbf{b}_i^T$ for the last row, the elemental estimate is given by $\hat{\mathbf{d}}_i = -B_i^{-1}\mathbf{b}_i$.

Solving the polynomial with coefficients given by $\hat{\mathbf{c}} = (\hat{\mathbf{d}}^T, 1)^T$ gives estimates for the roots $e^{s\beta_1}, \ldots, e^{s\beta_p}$. After extracting estimates for the β_j, the coefficients α_j are found by solving the linear system with the elemental set $\mathbf{y}_e$ as the response and the matrix with elements $\exp(t_{i+(j-1)s}\beta_k)$, $j = 1, \ldots, 2p$, $k = 1, \ldots, p$, as the design matrix.

2.3 Frequency Estimation

We now specialize the discussion to that of a periodic signal. The derived signal $\mu(t)$ is a sum of cosines if and only if the roots of the polynomial $p_c()$ lie on the unit circle. In this case p is even and the number of frequencies is $f = p/2$. The roots of $p_c()$ occur in conjugate pairs, $\exp(\pm\sqrt{-1}\omega_j s)$, $j = 1, \ldots, f$, where the ω_j are the frequencies.

The fact that the roots of $p_c()$ lie on the unit circle implies that $\mathbf{c}$ is symmetric, $c_j = c_{2f+2-j}$ for $j = 1, \ldots, f$ (Kay and Marple, 1981; Kundu and Kannan, 1997). If the signal is known *a priori* to be sinusoidal then it is appropriate to constrain $\mathbf{c}$ to be symmetric. This reduces the number of parameters which need to be estimated and reduces the size of an elemental set to $3f$ data values.

Let Q be the $(2f + 1) \times (f + 1)$ matrix

$$Q = \begin{pmatrix} I_f & 0 \\ 0 & 1 \\ J_f & 0 \end{pmatrix}$$

where I_f is the $f \times f$ identity matrix and J_f is the $f \times f$ anti-diagonal matrix

$$J_f = \begin{pmatrix} 0 & & & 1 \\ & & \cdot & \\ & \cdot & & \\ 1 & & & 0 \end{pmatrix}$$

The symmetry constraint can be represented as $\mathbf{c} = Q\mathbf{e}$ where $\mathbf{e}$ is an unrestricted real vector of dimension $f + 1$.

The elemental estimate of $\mathbf{c}$ is the symmetric null vector of the $(2f + 1) \times f$ matrix

$$Y_i = \begin{pmatrix} y_i & \cdots & y_{i+(f-1)s} \\ \vdots & & \vdots \\ y_{i+2fs} & \cdots & y_{i+(3f-1)s} \end{pmatrix}$$

The convention that $c_{p+1} = 1$ here implies that $e_1 = 1$. Let $\mathbf{d} = (e_2, \ldots, e_{f+1})^T$. If we write $\mathbf{b}_i^T$ for the first row of $Q^T Y_i$ and B_i for its remaining rows, then the elemental estimate of $\mathbf{d}$ is $\hat{\mathbf{d}}_i = -B_i^{-1}\mathbf{b}_i$.

To obtain estimates for the frequencies we compute $\mathbf{e} = (1, \mathbf{d})^T$ and $\mathbf{c} = Q\mathbf{e}$ and solve the polynomial with coefficients $\mathbf{c}$ as before to obtain roots on the unit circle. The roots yield frequencies $\omega_1, \ldots, \omega_f$ and the elemental estimate for μ is obtained by solving

$$X_e \boldsymbol{\alpha} = \mathbf{y}_e$$

for $\boldsymbol{\alpha}$ where X_e is the $3f \times 2f$ matrix

$$X_e = \left\{ \cos(\omega_1 t_{i+(j-1)s}), \ldots, \cos(\omega_f t_{i+(j-1)s}), \sin(\omega_1 t_{i+(j-1)s}), \ldots, \sin(\omega_f t_{i+(j-1)s}) \right\}_{j=1}^{3f}$$

6

We can then obtain an elemental estimate for the entire mean vector from $\boldsymbol{\mu} = X\boldsymbol{\alpha}$ with

$$X = \{\cos(\omega_1 t_i), \ldots, \cos(\omega_f t_i), \sin(\omega_1 t_i), \ldots, \sin(\omega_f t_i)\}_{i=1}^n$$

There are a couple of subtleties concerning sinusoid signals which affect the computational burden of the elemental set method. The first is that not every sequence of $3f$ observations can be interpolated by a signal with 3 sinusoids. This is dealt with easily. In that case not all the roots of the polynomial $p_c()$ will lie on the unit circle, and the elemental set can simply be discarded. The second subtlety is that sequences of observations which can be interpolated by sums of sinusoids can be also interpolated by harmonics of the sinusoids.

If the elemental set $\mathbf{y}_e$ is at unit-spacing, $s = 1$, then the returned roots (those lying above the real line) are $\rho_j = \exp(\sqrt{-1}\omega_j)$, $j = 1, \ldots, f$, with $0 \leq \omega_j < \pi$. The ω_j are obtained as the imaginary parts of the logarithms of the roots. If the elemental set $\mathbf{y}_e$ is at spacing $s > 1$, then the returned roots are $\rho_j = \exp(\sqrt{-1}\omega_j s)$, $j = 1, \ldots, f$. For each root we must consider the s possible complex values for $\rho_j^{1/s}$. Write $\omega_j^0 = \Im \log(\rho_j)/s$ for the base frequencies. The elemental set $\mathbf{y}_e$ at spacing s can be interpolated by a signal using any of the harmonics $(\omega_j^0 + 2\pi k/s)$ modulus π, $k = 0, \ldots, s - 1$. In this way, any elemental set at spacing s yields not one elemental set estimator, but s^f, each of which exactly interpolates the $3f$ observations.

Our approach has been to evaluate the criterion (LS, LTS or LMS) for a selection of $1 + f(s - 1)$ of the possible s^f harmonics for each elemental set. This keeps to the computations to lower order in s, allows many more elemental sets to be considered, and has proved satisfactory in practice. The method consists of a one dimensional search over the harmonics of each frequency in turn. This strategy often succeeds in minimizing the criterion over the harmonics for a particular elemental set because of approximate orthogonality between sinusoidal signals at different frequencies. The criterion is first evaluated at the signal obtained from the base frequencies ω_j^0. The criterion is then minimized by trying the other $s - 1$ harmonics of the first frequency, holding all the other frequencies at their base values. The criterion is then minimized by trying the other $s - 1$ harmonics of the second frequency, holding the first frequency at its best value, and any remaining frequencies at their base values. This process is continued until all of the frequencies have been considered. The final set of harmonics is taken to be the elemental estimator of the frequencies.

It it worth mentioning that there is another way to compute nearly $\binom{n}{f}$ algebraic estimators which avoids the need for harmonics. Let $\mathbf{y}_i = (y_i, \ldots, y_{i+2f})^T$ be a vector containing a sequence of $2f + 1$ consecutive observations starting at y_i. A generalized elemental estimate of the Prony vector $\mathbf{c}$ can be obtained as the symmetric null vector of the matrix

$$Y_{i_1,\ldots,i_f} = \left(\mathbf{y}_{i_1}, \ldots, \mathbf{y}_{i_f}\right)$$

where $i_1, \ldots, i_f$ is any f-subset of $1, \ldots, n - 2f$. These estimates however are not true elemental estimates, as they use more than the minimum number of observations, and therefore this approach is not taken further in this paper.

2.4   How Many Elemental Sets?

High breakdown estimators are found by evaluating the criterion over a large number of elemental sets. The elemental estimate with the lowest value for the criterion is the working estimate. Ideally the procedure would be repeated for all possible elemental sets, but this is computationally expensive, and it is desirable to use a smaller number of elemental sets.

The PROGRESS algorithm used by Rousseeuw and Leroy (1987) to approximate the LMS estimator in linear regression uses 1500 elemental sets when there are 3 unknown coefficients and 3000 elemental sets when there are 6 or more unknown coefficients. Stromberg (1993) showed that a smaller number of elemental sets is often sufficient when the elemental sets are used as part of a multistage algorithm. Stromberg (1993) used $\log(0.001)/\log(1-\epsilon^q)$ elemental sets, where q is the number of unknown parameters and ϵ is the desired breakdown point. This ensures that the probability is approximately 0.999 that at least one elemental set will contain none of the outliers. With $\epsilon = 0.5$ this suggests 52 elemental sets for estimating one frequency, 439 for estimating two frequencies and 3534 for estimating three frequencies. For the frequency estimation problem it is necessary to consider $O(n)$ elemental sets so as to resolve the frequencies to $O(1/n)$. We use as default value Stromberg's suggestion or n, whichever is greater.

Preference is given to smaller values of the spacing s to minimize the number of harmonics which need to be considered. However it is necessary to evaluate elemental sets which span a reasonable range of the time points in order to ensure a high chance of detecting slow frequencies. We require the maximum spacing considered to be at least $s_{7\%} = n/14/(3f-1)$, which ensures that the elemental sets at the larger spacings span at least 7% of the range of the data. If the total number of elemental sets at spacings 1 to $s_{7\%}$ is greater than the target number of elemental set estimators, then only a proportion of the sets at each spacing are used. If thinning is necessary, then the probability of choosing an elemental set at spacing s used at each spacing is chosen inversely proportional to $1/(1 + f(s - 1))$, i.e., to the computation burden in evaluating an elemental set. To bound the computations for larger data sets, software referred to at the end of this paper limits the total number of evaluations of the criterion (including evaluations for harmonics) to 6000. This limit was not relevant for the simulations reported in the next section.

Compared with the PROGRESS algorithm for linear regression, our algorithm performs less computation for smaller data problems because of the economy on the number of elemental sets. This is despite the fact that a highly nonlinear regression function is being estimated. For larger programs the computational complexity is about double that of PROGRESS because of the higher limit on the number of evaluations of the criterion.

3   A Multistage Estimator

Our algorithm is a multistage procedure which was inspired by that of Stromberg (1993). First an estimator of the frequencies and coefficients is obtained using an eigenanalysis method described by Smyth (1999). An initial frequency estimate is obtained using Pisarenko's method adapted to make use of the symmetry constraint described above, and this

estimate is refined using the ORA method from Kahn et al (1992), again adapted to make use of the symmetry constraint. The ORA method is very insensitive to starting values, and produces an estimator which is equivalent in efficiency to the least squares estimator, although it is not robust against outliers. The eigenanalysis parameter estimate will be denoted by $\hat{\boldsymbol{\theta}}_{ORA}$.

The criterion function (LTS or LMS) is also evaluated for each of the elemental set estimates. Let $\hat{\boldsymbol{\theta}}$ be the best elemental set estimate. The least squares fit is then found to that half of the data with the smallest squared residuals, using $\hat{\boldsymbol{\theta}}$ as starting value. The least squares program is a Gauss-Newton algorithm with Levenberg damping to ensure convergence. The resulting least squares estimator replaces $\hat{\boldsymbol{\theta}}$.

Fitted values and hence residuals are obtained from $\hat{\boldsymbol{\theta}}$ for the whole data set and these are used to compute a scale M estimate. Using the same scale estimate, an M estimate of the parameters is obtained using (i) $\hat{\boldsymbol{\theta}}$ as starting value and (ii) $\hat{\boldsymbol{\theta}}_{ORA}$ as starting value. The M estimate with the lower value of the objective function is taken to be the final estimate.

The MM estimator is chosen as in Stromberg (1993). Hampel's redescending psi function ρ_H is used (Hampel, Ronchetti, Rousseeuw and Stahel, 1986). The MM estimator was constructed as recommended by of Yohai (1987) with $\rho_0(u) = \rho_H(u/0.212)$ for the scale and $\rho_1(u) = \rho_H(u/0.9014)$ for estimation of the frequencies and coefficients. The M estimator is implemented as suggested by Yohai (1987) but with Levenberg damping to ensure convergence. Combined with a high breakdown starting estimator, the MM estimator produces parameter estimates with a high breakdown point and with theoretical efficiency 95% under normal errors and no outliers.

Stromberg (1993) used a Nelder-Mead minimization step after finding the least squares fit to half the data and before computing the scale M estimator. The purpose this step was to get closer to the global LMS estimator. We have not used a similar step as a default because it is relatively slow compared with the rest of the computation. In the single sinusoid case it can increase the number of criterion evaluations several-fold. In the simulations reported in the next section we experiment with the use of the S-Plus functions ltsreg and lmsreg to get closer to the LTS and LMS estimators respectively before computing the scale M estimate. Our recommendation is that this step can be dispensed with if the LTS criterion is used.

Let $e_{(i)}^2$ be the ith ordered squared residual. Our algorithm defines the trimmed sum of squares to be $\sum_{i=1}^{r} e_{(i)}^2$ where $r = \lfloor n/2 \rfloor + \lfloor 3f/2 \rfloor$, i.e., half the number of observations plus half the number of parameters. Similarly the median squared residual is $e_{(r)}^2$, which agrees with the PROGRESS algorithm of Rousseeuw and Leroy (1987). In the simulations reported in the next section we have actually used $r = \lfloor n/2 \rfloor + \lfloor f \rfloor$ so that the criterion will agree with the ltsreg and lmsreg functions when these are applied to the linear coefficients only. The small difference in the criterion was not material in the simulations.

Table 1: Simulation results for one cosine signal and no outliers. The true frequency is 0.5.

Method	Mean	StDev	RMSE	Efficiency
ORA	0.499956	0.001020	0.001021	0.982
ELS	0.500341	0.003598	0.003614	0.078
ELS-LS	0.499976	0.001011	0.001011	1.000
LTS	0.500431	0.004096	0.004119	0.060
LTS-LS	0.500247	0.002072	0.002086	0.235
LTS-MM	0.499943	0.001013	0.001015	0.993
LTS-LI1-MM	0.499945	0.001013	0.001014	0.994
LTS-LI-MM	0.499945	0.001013	0.001015	0.993
LMS	0.500377	0.004228	0.004245	0.057
LMS-LS	0.500137	0.001675	0.001680	0.362
LMS-MM	0.499966	0.001158	0.001159	0.762
LMS-LI1-MM	0.499946	0.001012	0.001014	0.995

4   The Performance Of The Estimator

4.1   One Sinusoid

We report here simulation results for estimators based on the LS, LTS and LMS criteria. We have also experimented with an estimator based on the idea of "averaging" the different elemental estimates. For this we took the principal axis of the Prony vectors arising from the elemental fits, treating them as axial observations on the unit sphere. See Fisher *et al* (1987, Section 3.2.4) for a discussion of principal axes. This principal axis estimator performed far less well than the criterion-based estimators on the outlier free data and failed completely when outliers were included. In the results reported here therefore we have ignored the averaging based estimator in favor of the criterion minimization approach.

A simulation study was conducted with one frequency and $n = 100$. Independent Gaussian observations were generated with means

$$\mu_i = \cos(0.5t_i + 0.1)$$

$t_i = 1, \ldots, n$, and standard deviation $\sigma = 0.2$. The Cramer-Rao lower bound for the standard deviation of an unbiased estimator of the frequency for these parameter values is 0.000975.

With $n = 100$ and $f = 1$ our algorithm uses 100 elemental sets. The requirement that the elemental sets span 7% of the time range means that spacings up to $s = 4$ must be used. A fraction of the available elemental sets are used at each spacing, these fractions being 0.33, 0.33, 0.22 and 0.165 respectively. This produces 100 elemental sets and a total of 219 evaluations of the criterion functions for the full data set.

Table 1 gives results for various estimators applied to 100 randomly generated data sets with no outliers. For each of the other estimators the mean, standard deviation, root mean square error (RMSE) and relative efficiency are given. The efficiency is computed relative to

an idealized estimator, a least squares program (a Gauss-Newton algorithm with Levenberg damping) which was started from the true value of $\omega = 0.5$. The performance of this estimator is therefore the best possible, and indeed its RMSE at 0.001012 was very close to the Cramér-Rao Lower Bound.

The first row gives results for the ORA frequency estimator, which is nearly fully efficient. The rows labelled ELS, LTS and LMS give results for the best elemental set estimator using the sum of squares, LTS and LMS criteria respectively. The elemental least squares estimator has efficiency 7.8% which increases to 100% when used as an starting value for a least squares algorithm (ELS-LS). This shows that the elemental sets are able to resolve the frequency to the correct local minima of the sum of squares. The elemental trimmed sum of squares estimator has efficiency 6.0%. This rises to 23.5% when the least squares estimator is found for the half of the data with smallest residuals (LTS-LS). This further rises to 99.3% efficiency when used as the stage 1 estimator for MM estimation (LTS-MM). As described in the previous section the scale M estimator was obtained using the LTS-LS estimate, and the final M estimation algorithm for the frequency and coefficients was started from both the LTS-LS and ORA estimates.

The LTS-LS estimator will not be the actual LTS estimator unless the correct subset of observations has been chosen. Two attempts were made to further reduce the trimmed sum of squares before computing the scale estimator. First the S-Plus function ltsreg (a genetic algorithm) was used to minimize the trimmed sum of squares with respect to the linear coefficients keeping the frequency fixed. This produces the LTS-LI1-MM estimator. The use of ltsreg may change the subset of observations with smallest squared residuals, so least squares on the half of the data with the smallest squared residuals and the ltsreg function were iterated until the selected subset of observations did not change further. This produces the LTS-LI-MM estimator. In Table 1 the LTS-MM estimator is already nearly fully efficient, so that these further steps were not necessary.

The elemental LMS estimator has efficiency 5.7%, which increases to 36% when least squares is used on the best half of the data, and further increases to 76% when used as the first stage of the MM estimator. A further attempt was made to reduce the median squared residual by using the S-Plus function lmsreg with the frequency fixed. This produced the LMS-LI1-MM estimator, which is nearly fully efficient on this problem.

Table 2 gives the results of a similar simulation experiment with 30% outliers. Outliers were generated to have standard deviations 100 times that of the good observations and were associated with a random subset of the t_i. Efficiencies were calculated relative to the least squares algorithm applied to the good 70% of the data and started from the true frequency. The performance of this estimator represents an ideal which is not necessarily achievable in practice without knowledge of the generating process. The RMSE of this estimator at 0.001125 is actually slightly below the Cramér-Rao Lower Bound which is 0.00116 based on the 70% of the observations without outliers.

The first three rows of the table give the results for the ORA estimator, for the ELS estimator over the elemental sets, and for the least squares algorithm started from the elemental estimator with lowest sum of squares. All of these methods fail completely because of the presence of outliers.

Table 2: Simulation results for one cosine signal and 30% outliers. The true frequency is 0.5.

Method	Mean	StDev	RMSE	Efficiency
ORA	1.593491	0.992601	1.476813	0.000
ELS	1.332609	0.854173	1.192832	0.000
ELS-LS	1.332352	0.853603	1.192245	0.000
LTS	0.499542	0.008610	0.008622	0.017
LTS-LS	0.499901	0.003418	0.003420	0.108
LTS-MM	0.499917	0.001251	0.001253	0.806
LTS-LI1-MM	0.499911	0.001247	0.001250	0.811
LTS-LI-MM	0.499911	0.001233	0.001236	0.828
LMS	0.498960	0.008063	0.008130	0.019
LMS-LS	0.499582	0.004589	0.004608	0.060
LMS-MM	0.499861	0.001306	0.001313	0.734
LMS-LI1-MM	0.499899	0.001272	0.001276	0.778

The LTS and LMS criteria are now markedly superior to least squares. The elemental LTS estimator has efficiency 1.7% relative to least squares on the good data. This increases to 10.8% when least squares is used to refine the estimator on the half of the data with smallest squared residuals, and rises again to a very respectable 80.6% when used as the first stage of the MM estimator. Small further rises in efficiency, up to 83%, were achieved by using the ltsreg function to further decrease the trimmed sum of squares before computing the scale M estimator.

The LMS criteria performs at least as well as LTS at the elemental set level, but does not respond as well to the refinements. The elemental LTS estimator has efficiency 1.9%, increasing to 6% when refined by least squares on the half of the data with smallest squared residuals, and to 73% when used as the first stage of the MM estimator. A further increase in efficiency to 78% was achieved by using the lmsreg function to decrease the median squared residual with respective to the linear coefficients.

Of course even better results for the LTS and LMS estimators could be obtained if it were known in advance that only 30% of the observations were contaminated (and not the 50% allowed for by the level of trimming used). In that case further efficiency could be recovered by trimming a smaller proportion of the data.

4.2 Two Sinusoids

Another simulation study was conducted with two frequency and $n = 100$. Independent Gaussian observations were generated with means

$$\mu_i = \cos(0.3t_i + 0.2) + \cos(0.7t_i + 0.1)$$

$t_i = 1, \ldots, n$, and standard deviation $\sigma = 0.2$. With $n = 100$ and $f = 2$ the algorithm uses spacings up to $s = 6$, and uses 0.91 of the available elemental sets at spacings 1 to 5 and

Table 3: Simulation results for the lower of two frequencies with no outliers. The true frequency is 0.3.

Method	Mean	StDev	RMSE	Efficiency
ORA	0.299966	0.000965	0.000966	0.996
ELS	0.300387	0.004057	0.004076	0.056
ELS-LS	0.299979	0.000963	0.000963	1.000
LTS	0.299484	0.005076	0.005103	0.036
LTS-LS	0.299861	0.001754	0.001760	0.300
LTS-MM	0.299980	0.001009	0.001009	0.912
LTS-LI1-MM	0.299979	0.001009	0.001009	0.912
LTS-LI-MM	0.299980	0.001014	0.001014	0.903
LMS	0.300027	0.004901	0.004902	0.039
LMS-LS	0.299931	0.001292	0.001294	0.554
LMS-MM	0.299938	0.001071	0.001072	0.807
LMS-LI1-MM	0.299979	0.001005	0.001005	0.918

0.745 of the available elemental sets at spacing 6. This produces 439 elemental sets and a total of 2417 evaluations of the criteria on the complete data.

The results for with two sinusoids are qualitatively the same as for one sinusoid. Table 3 gives results for the lower of the two sinusoids. In this case least squares started from the true value has RMSE 0.000963 which is very close to the Cramér-Rao lower bound of 0.000975. Table 4 gives the results for the higher of the sinusoids. In this case least squares started from the true value with RMSE at 0.000938 outperforms the Cramér-Rao lower bound somewhat, this being 0.001005. ORA is fully efficient for both frequencies. The elemental estimator with minimum sum of squares has efficiency about 6%, but is fully efficient when used to start the Gauss-Newton least squares algorithm. Again this shows that the elemental sets are able to resolve the global minimum of the sum of squares. The elemental estimator with minimum trimmed sum of squares has efficiency a little under 4%, but this rises to 30% or 60% when refined using least squares on the half of the data with smallest absolute residuals, and to 91 or 92% when used as the first stage of a the MM estimator. Further attempts to minimize the trimmed sum of squares before computing the scale M estimate did not increase the efficiency of the final estimator. The LMS is slightly more efficient that the LTS criteria at the elemental set level, but is slightly less efficient (81 to 85%) when used as the first stage of the MM estimator. When the lmsreg function was used to further reduce the median squared residual before computing the scale M estimator, the efficiency rose to 92 or 93%, similar to LTS.

Tables 5 and 6 give results for the two frequencies when 30% outliers were included in the samples. The results are qualitatively the same as for one frequency. Efficiencies are computed relative to least squares applied to the 70% of the data without outliers and started from the true values. The performance of this estimator was similar to the Cramér-Rao lower bounds based on the good data (RMSE of 0.001183 and 0.001338 versus bounds of

Table 4: Simulation results for the higher of two frequencies with no outliers. The true frequency is 0.7.

Method	Mean	StDev	RMSE	Efficiency
ORA	0.699991	0.000931	0.000931	1.014
ELS	0.700287	0.003686	0.003697	0.064
ELS-LS	0.700016	0.000937	0.000938	1.000
LTS	0.700072	0.004809	0.004810	0.038
LTS-LS	0.700003	0.001215	0.001215	0.596
LTS-MM	0.699989	0.000978	0.000978	0.920
LTS-LI1-MM	0.699988	0.000980	0.000980	0.915
LTS-LI-MM	0.699985	0.000982	0.000982	0.911
LMS	0.700530	0.004375	0.004407	0.045
LMS-LS	0.700142	0.001233	0.001242	0.570
LMS-MM	0.700005	0.001019	0.001019	0.846
LMS-LI1-MM	0.699990	0.000973	0.000973	0.929

Table 5: Simulation results for the lower of two frequencies with 30% outliers. The true frequency is 0.3.

Method	Mean	StDev	RMSE	Efficiency
ELS	0.848431	0.573100	0.793235	0.000
ELS-LS	0.850716	0.574819	0.796056	0.000
LTS	0.298395	0.009731	0.009862	0.014
LTS-LS	0.299615	0.003605	0.003626	0.107
LTS-MM	0.300046	0.001275	0.001275	0.863
LTS-LI1-MM	0.300048	0.001266	0.001267	0.874
LTS-LI-MM	0.300044	0.001236	0.001237	0.917
LMS	0.297937	0.009174	0.009403	0.016
LMS-LS	0.299075	0.005922	0.005994	0.039
LMS-MM	0.299989	0.001399	0.001399	0.717
LMS-LI1-MM	0.300019	0.001338	0.001338	0.784

Table 6: Simulation results for the higher of two frequencies with 30% outliers. The true frequency is 0.7.

Method	Mean	StDev	RMSE	Efficiency
ELS	1.985364	0.674255	1.451475	0.000
ELS-LS	1.984822	0.680069	1.453706	0.000
LTS	0.701049	0.011471	0.011519	0.013
LTS-LS	0.700416	0.004177	0.004197	0.098
LTS-MM	0.699981	0.001435	0.001436	0.837
LTS-LI1-MM	0.699981	0.001427	0.001427	0.847
LTS-LI-MM	0.699967	0.001392	0.001392	0.890
LMS	0.700439	0.011533	0.011542	0.013
LMS-LS	0.700399	0.006650	0.006662	0.039
LMS-MM	0.700020	0.001582	0.001582	0.689
LMS-LI1-MM	0.699977	0.001515	0.001515	0.751

0.001165 and 0.001201 respectively for the two frequencies). As in the one sinusoid case, this represents as ideal which is not necessarily achievable without knowledge of the generating process.

Least squares on the complete data fails completely when the outliers are included. This is true also for the ORA algorithm which is therefore not included in the tables. The elemental estimator with lowest trimmed sum of squares has efficiency about 1.4%, which rises to about 10% when refined by least squares on half of the data, and to 84 or 86% when used at the first stage of the MM estimator. Further minimization of the trimmed sum of squares before computing the scale M estimator increased the efficiency to about 90%.

The LMS criterion was at least as efficient as the LTS criterion at the elemental set level, but did not respond quite as well to refinement by least squares on half of the data or to MM estimation. Using the LMS criterion as the first stage for MM estimation produced about 70% efficiency, which was increased further to 75% or 78% using the lmsreg function before computing the scale M estimator.

In summary we find here that the elemental set estimators which minimize the LTS or LMS criteria can produce nearly fully efficient estimators when using as part of the multistage algorithm, the LTS based estimator when first used with the MM estimator, and the LMS estimator after further minimization with respect to the linear coefficients. The greater ease with which the LTS criterion can be refined may be due to the closer relationship between the criterion and the refinement methods. Least squares on half the data does in fact produce LTS estimator if the correct half of the data is chosen.
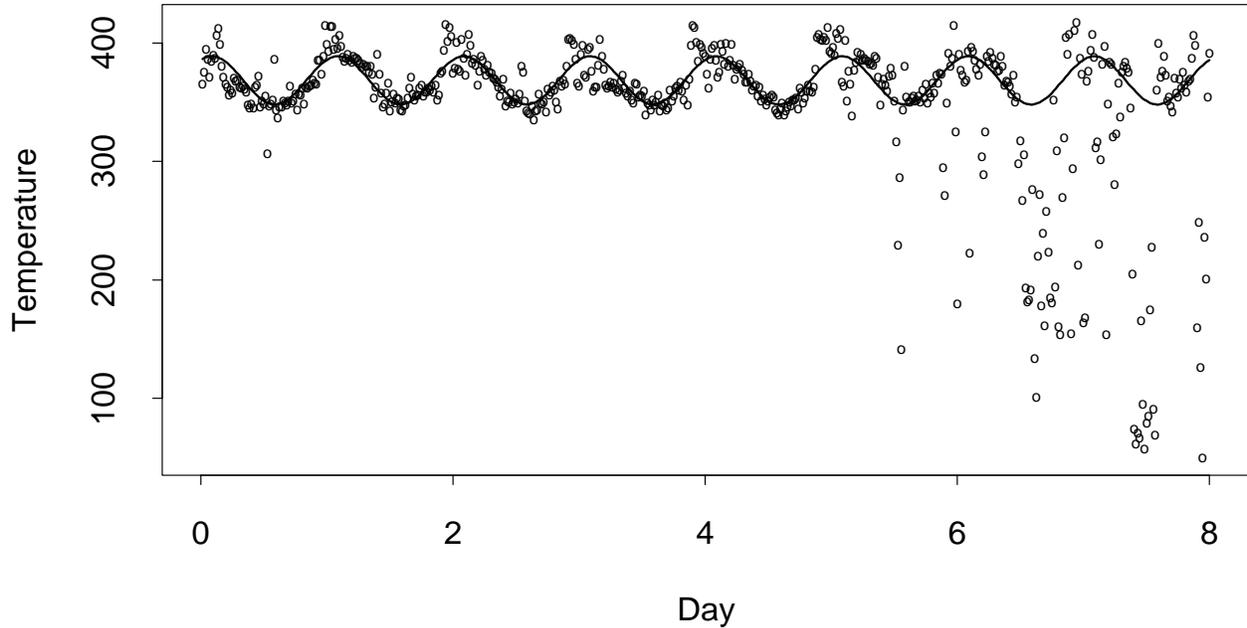
Figure 1: Circadian rhythms for the long-tail pocket mouse: 20-minute temperature averages.

5   Data Examples

5.1   Circadian Rhythms

The data are temperature recordings made at 2 minute intervals over 3 months on a nocturnal mammal, Perognathus formosus (long-tail pocket mouse – common in the Nevada desert). The animal was given 8 days of 12 hours light and 12 hours dark as an adjustment period, followed by about 73 days of constant darkness (Andrews and Herzberg, 1985, Data Set 48.3). The data was collected at Princeton University in the late 1960's under the direction of Dr C. S. Pittendrich. The interest of the researchers was to observe periodicities in the behavior of the animals. Problems occurred during the experiment associated with transient failures of the monitoring equipment and with imperfections in the data logging process, which was based on a conversion to mechanical keypunch operation. Obvious errors, such as card punch jams or misfeeds, were corrected during the experiment, but a proportion of outliers remain. The data was obtained electronically from the Statlib database (lib.stat.cmu.edu). It has been reformatted by one of the authors of this paper for easy input to statistical programs, and is available from `http://www.statsci.org/data/general/pformosu.html`.

Here we analyze 20-minute averages of the temperatures. Figure 1 displays the temperatures for the first 8 days, together with the best fitting cosine curve. The frequency was found by using the LTS elemental-set estimator as starting value for the MM estimator. The constant term was initially set equal to the median of the data during the elemental-set stage, and then included as an estimated coefficient during the MM estimation step. A large number of outliers apparently due to equipment failures are obvious after about 5.5 days,
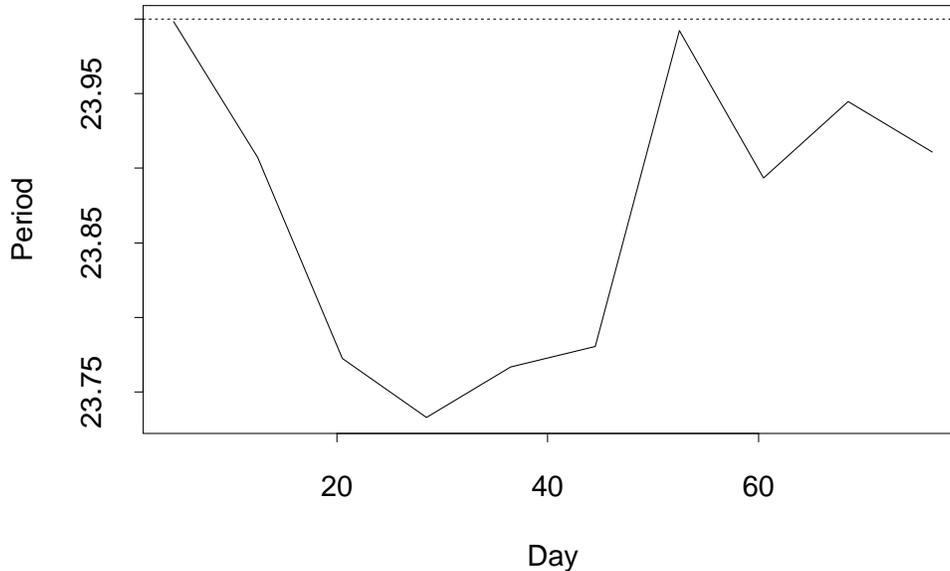
16

Figure 2: Circadian rythms for the long-tail pocket mouse: variation in the period of temperature cycles over time. The dashed line corresponds to a 24 hour cycle.

but the fitted curve successfully ignores them and follows the daily periodic trend. There are also a number of less obvious outliers closer to the fitted curve.

We use only one frequency is this example because we want to focus on the period of the dominant cycle with and without the stimulus of external lighting. Adding two more frequencies does not in any case dramatically improve the visual fit. The estimated frequency for the first 8 days is 0.087273 corresponding to a period of 23.998 hours, almost precisely equal to the known 24-hour period imposed by the lighting regime during the first 8 days. Least squares based methods fail on this data. The ORA algorithms fails to converge and terminates with a frequency of 0.164, a little less than twice as fast as the known daily cycle. Similar cosine curves were fitted to successive series of 8 days of observations during which the cage was in total darkness. Figure 2 plots the estimated periods for 10 successive sequences of 8 days. It can be seen that the animal maintained, in the absence of external stimuli, a period very close to but slightly shorter than its 24-hour period in the wild. During the first few weeks the period fell from 24 hours to 23.75 hours, then recovered towards the end of the experiment.

5.2 Variable Star

The determination of the periodicities of a variable star and the shape of its light curve is important in studies of stellar structure and evolution. The relationship between period and magnitude is used to determine distances on a cosmic scale, for example. The data in this example gives observations on the magnitude of a variable star made from the Mount Stromlo Observatory near Canberra in Australia over a period of about 250 days (Reimann,
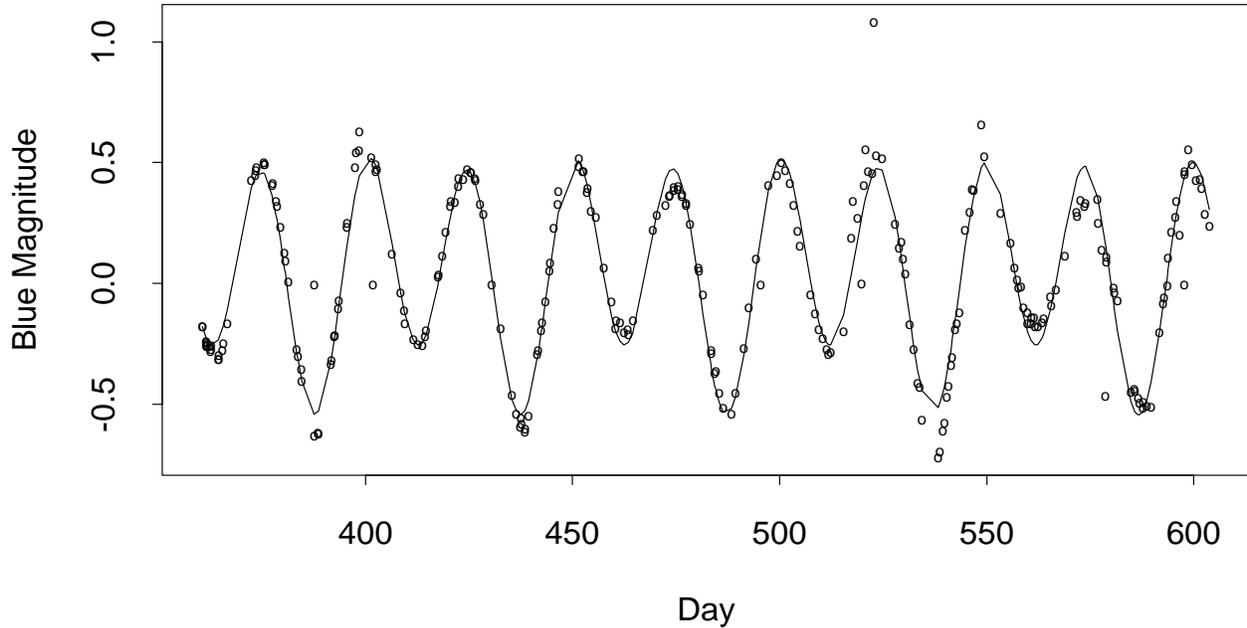
Figure 3: Blue band magnitudes of a variable star.

1994). The data was supplied by Professor John Rice of the University of California, Berkeley (www.stat.berkeley.edu/users/rice/UBCWorkshop/). It has been reformatted for easy input to statistical programs, and is available from `http://www.statsci.org/data/oz/ceph2.html`. Magnitudes were recorded separately for the blue and red bands; the blue band measurements are considered here. Some of the observations were considered to be unreliable due to observation conditions.

Observation times were irregularly spaced depending on sky conditions and the observation schedule. For the purposes of using elemental sets, the data was interpolated linearly onto an equally spaced grid of time points of the same length. The LTS elemental-set estimator was used to obtain initial frequency estimates for the MM estimator. The MM estimator used the exact time points and data rather than the interpolated data. Interpolation of the data was used therefore only as a computational device to obtain starting values, and does not directly affect the final estimates.

Figure 3 plots the data with the best fitting curve consisting of a constant plus two sinusoids. The data includes one obvious outlier, and a number of other probable outliers with smaller deviations from the curve. The estimated frequencies are 0.126 and 0.253, corresponding to periods of 50.0 and 24.9 days. The star is therefore determined to be periodic with period about 50 days. The light curve is well described by a cosine curve and one harmonic. In this case, increasing the number of frequencies to four does not dramatically improve the visual fit of the curve. The period and the relative brightness of this star help to classify it as a star of the Cepheid class.

18

6 Summary

This paper has considered a special but important class of nonlinear regression problems for which elemental set estimators can be computed in closed form using Prony's method. We concentrated on estimating periodic signals, but the methods considered here can also be applied to damped sinusoids and to transient exponential signals. We considered criterion-based elemental estimators, which minimize a criterion over the elemental sets and also an estimator which directly averages the elemental estimators by computing the principal axis of the Prony coefficient vectors. The principal axis estimator requires less computation than the criterion based estimators, but proves to be inefficient and to have poor breakdown properties.

Our suggested algorithm uses an economical number of elemental sets as part of a multistage process. This produces an estimator with relatively small computational demands. The LTS and LMS criterion based elemental estimators have high breakdown properties and are able to resolve the frequencies to the correct local minima of the objective function. The elemental based estimators can therefore be refined further using various local methods including least squares for data points with small residuals and MM estimation. When used as the first stage of MM estimation, and combined with a high efficiency least squares type estimator which becomes relevant with good data, the elemental set based estimators produce an overall estimator which simultaneously has high breakdown and high efficiency in the absence of outliers.

Using the LTS, LMS or LS criteria, elemental sets provide a solution to the problem of obtaining fully efficient frequency estimators discussed by Rosenblatt and Rice (1988).

S-Plus functions for robust frequency estimation have been developed and are available from the `http://www.statsci.org/s/`.

References

Andrews, D. F., and Herzberg, A. M. (1985). *Data. A collection of problems from many fields for the student and research worker.* Springer, New York.

Atkinson, A. C., and Weisberg, S. (1991). Simulated annealing for the detection of multiple outliers using least squares and least median of squares fitting. In *Directions on Robust Statistics: Part I*, eds. W. Stahel and S. Weisberg, Springer, pp. 7–20.

Bloomfield, P. (1976). *Fourier Analysis of Time Series.* Wiley, New York.

Brockwell, P. J., and Davis, R. A. (1991). *Time series. Theory and methods (2nd edition).* Springer, New York.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data*, Cambridge UP.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics. The Approach Based on Influence Functions.* Wiley, New York.

Hawkins, D. M. (1993). The accuracy of elemental set approximations for regression. *J. Amer. Statist. Assoc.* **88**, 580–589.

Hawkins, D. M., Bradu, D., and Kass, G. V. (1984). Location of several outliers in multiple regression data using elemental sets. *Technometrics* **26**, 197–208.

Kahn, M., Mackisack, M.S., Osborne, M.R. and Smyth, G.K. (1992). On the consistency of Prony's method and related algorithms. *J. Comput. Graph. Statist.* **1**, 329–349.

Kundu, D., and Kannan, N. (1997). Constrained maximum likelihood estimators for super-imposed exponential signals. *Comm. Statist. Simulation Comput.* **26**, 733–764.

Kay, S.M., and Marple, S.L. (1981). Spectrum analysis—a modern perspective. *Proc. IEEE* **69**, 1380–1419.

Mackisack, M. S., Osborne, M. R., and Smyth, G. K. (1994). A modified Prony algorithm for estimating sinusoidal frequencies. *J. Statist. Comput. Simul.* **49**, 111–124.

Marazzi, M. (1991). Algorithms and programs for robust linear regression. In W. Stahel and S. Weisberg (eds), *Directions in Robust Statistics and Diagnostics, Part I*, Springer-Verlag, New York, pp. 183–200.

Osborne, M. R. and Smyth, G. K. (1991). A modified Prony algorithm for fitting functions defined by difference equations, *SIAM J. Sci. Stat. Comput.* **12**, 362–382.

Osborne, M. R., and Smyth, G. K. (1995). A modified Prony algorithm for exponential function fitting. *SIAM J. Sci. Statist. Comput.* **16**, 119-138.

Prony, R. (1795). Essai éxperimental et analytique. *J. de L'École Polytechnique* **2**, 24–76.

Reimann, J. D. (1994). *Frequency estimation using unequally-spaced astronomical data.* PhD Thesis, University of California, Berkeley.

Rice, J. A., and Rosenblatt, M. (1988) On frequency estimation. *Biometrika* **75**, 477–484.

Rousseeuw, P. J. (1984). Least-median of squares regression. *J. Amer. Statist. Assoc.* **79**, 871–880.

Roussseeuw, P. J. and Leroy, A. M. (1986). *Robust Regression and Outlier Detection.*, Wiley, New York.

Smyth, G. K. (1999). Employing symmetry constraints for improved frequency estimation by eigenanalysis Methods. *Technometrics.* To appear.

Stromberg, A. J. (1993). Computation of high breakdown nonlinear regression parameters. *J. Amer. Statist. Assoc.* **88**, 237–244.

Stromberg, A. J. (1995). Consistency of the least median of squares estimator in nonlinear least squares. *Commun. Statist.—Theory Meth.* **24(8)**, 1971–1984.

Stromberg, A. J., and Ruppert, D. (1992). Breakdown in nonlinear regression. *J. Amer. Statist. Assoc.* **87**, 991–997.

Yohai, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15**, 642–656.