RECONSTRUCTION OF UNEQUALLY–SAMPLED RANDOM PROCESSES

David J. Thomson

Queen's University, Kingston Ontario, djt@mast.queensu.ca (613) 533-2426
KEYWORDS: Interpolation, missing data, global warming, space physics

ABSTRACT

This paper describes a method to reconstruct a random process from data that has significant numbers of missing samples. The method uses an inverse–theory projection operator based on Slepian sequences, with an elliptical likelihood constraint. A variant can also be used to combine various proxy series and data with unequal observational accuracy.

1. INTRODUCTION

It is perhaps more common than otherwise for time–series data in the physical sciences to have missing observations, outliers, or, in some important cases, to consist primarily of proxies. Because much of this data is historical, irreplacible, and/or cost billions of dollars to collect, the problem cannot be ignored. Typical examples include ground–based optical astronomy where bad weather, etc. cause gaps, see [1]. Data from spacecraft similarly has gaps in coverage caused by everything from bad data packets to lack of capacity in the deep space network, see [2]. An exceptionally important example is provided by the ongoing controversy surrounding reconstruction of the climate of last few thousand years, see [3, 4], the news article [5]. Instrumental temperature data only exists for a short time with the longest such record, that from "Central England" beginning in 1670 as monthly averages, and the daily series from Uppsala beginning in 1722, and even these have been seriously misinterpreted, see [6].

Here I propose another method, a hybrid of the the local regression methods now common in statistics[7, 8] with multitaper projection–filter techniques[9, 10, 11]. The major considerations in this process are:

1. Gaps and missing data often occur in "clumps" interspersed with relatively dense strings of good observations. The distribution of missing data is rarely uniform.
2. As in the climate data problem mentioned above, one needs to be able to combine data sets with different characteristics. Thus one must be able to segregate, and weight, data by frequency. Standard transverse filters effectively lose half their impulse duration at each end of the series, unacceptable in such problems.
3. Functions equivalent to Slepian sequences defined only on an irregular set of observations usually have poorer concentration properties than the regular Slepian sequences. Because they are optimized for energy concentration, they are poorly matched to the interpolation problem.
4. The interpolation process should preserve level shifts, trends, and similar low–degree polynomial components.
5. Apart from trends and a few strong periodic terms, many of the series appear to be approximately stationary. Asymptotically, a discrete-time process has a one-point interpolation error of
$$\sigma_I^2 = \left[\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{S(f)} df\right]^{-1} \quad (1)$$
so it is better to interpolate, then use standard Slepian sequences than to use "gappy" basis functions.
6. The interpolation process should be insensitive to frequency shifts so that translating the data in frequency should result in the same interpolation shifted in frequency.

Various forms of the method sketched below have been tested on astronomical, [1]; space physics, [12, 2]; and global temperature data, [6, 13, 14, 15]. Before turning to details, note that many of the missing data problems one encounters can be solved with "simple" interpolators. The Appendix of [2] shows a test of an iterative Wiener interpolator applied to Ulysses magnetic field data where about 10% of the data was randomly marked missing and the power spectra estimated from the original and interpolated data compared. In this paper I begin the process of unifying various methods that have proved useful in "difficult" problems.

2. BACKGROUND

Interpolations of the type considered here frequently depend on filtering data to extract bands of particular interest. In many problems of interest, the data consists of a superposition of many discrete modes. Individually, these modes may remain stable for weeks and so in principle should be easy to predict or interpolate. Collectively, however, effects of measurement noise, poor frequency resolution from data gaps, etc. limits predictability. Thus we begin with a discussion of filtering in relatively short data segments.

The advantages of transverse filters for data analysis are well-known; stability, simple gain characterization, zero-phase, easily adapted to unusual conditions and, most important, easily explained to non–specialists. The advantages of these filters, however, are accompanied by two major disadvantages; first, poor performance on gappy data, and second, the filter impulse response must be quite long to obtain good performance so, if the filter length is $M = 2J + 1$ samples, J samples are lost at each end of the record. Specifically, if we have N observations, $x(t)$, for $t = 0, 1, ..., N-1$ and the filter is specified by a non-causal "impulse response" $h(j), j = -J, J$ (usually with $h(j) = h(-j)$) the output at time t is $x_{LP}(j) = \sum_{j=-J}^{+J} h(j)x(t+j)$ and so is only valid for $J \leq t \leq N - J - 1$. Under the usual constraint that the weights h sum to one, (so the transfer function has unit gain

Thanks to the Canada Research Chair program for funding.

at zero frequency) offsetting the input of such a filter offsets the output by the same quantity.

2.1. Slepian Sequence Expansions

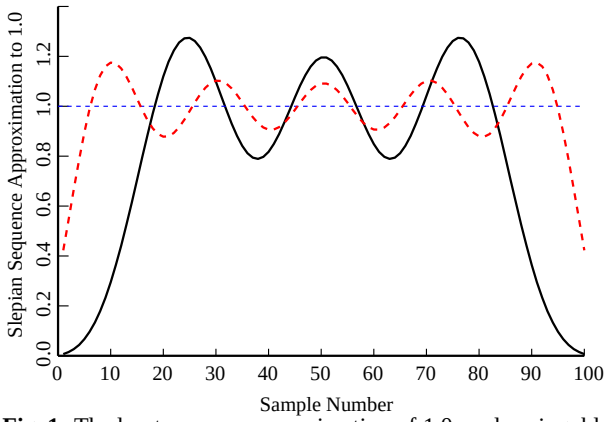

Fig. 1. The least–squares approximation of 1.0 made using, black line, the first $K = 6$ Slepian sequences with $N = 100$ and $NW = 5$. The large ripple is the motivation for the level–invariant inversions and disappears completely when the polynomial inversion is used. The red dashed line is the equivalent with $K = 10$.

An alternative to using conventional low-pass filters for such applications is to expand the entire data set in Slepian, or discrete prolate spheroidal, sequences of the desired bandwidth. Given a desired bandwidth of W, $0 < W < \frac{1}{2}$ the expansion coefficients of the Slepian sequences $v_n^{(k)}(N, W)$, see *e.g.* [16, 17], are

$$y_k = \sum_{n=0}^{N-1} v_n^{(k)}(N, W) x(n) \tag{2}$$

so, for $K = \lfloor 2NW \rfloor$, the standard-pass low-pass projection is

$$y(t) = \sum_{k=0}^{K-1} y_k \, v_t^{(k)}(N, W) \tag{3}$$

here for $t = 0, 1, ..., N-1$. Unlike the filter approach, one does not loose the ends of the data. Note that the pair of operations (2) and (3) is a projection operator, that is, if the process is iterated, the result is identical.

Here, as before, there are drawbacks. First, as with other orthogonal expansions, truncation and Gibb's ripples cause even the expansion of a constant to have amplitude ripple. This is shown in Figure 1 where the ripple is about $\pm 20\%$ across the center of the band and the response is zero at the ends. Here 6 of the $2NW = 10$ sequences were used, good for eliminating leakage, but clearly a poor choice for a fit. The red dashed line in this figure shows that more, $K = 10$, terms gives a better approximation but this is paid for with sideband leakage. Second, denoting this filter by $\mathcal{H}$ it is clear that

$$\mathcal{H}\{x(t) + a\} \neq \mathcal{H}\{x(t)\} + a \tag{4}$$

and, defining the Slepian function at zero frequency as

$$U_k = U_k(0) = \sum_{n=0}^{N-1} v_n^{(k)} \ , \tag{5}$$

the usual Bessel inequality shows that the squared error in expanding a constant is proportional to $|a|$ and, on average, the expansion is biased towards zero by a factor $(\sum U_k^2)/N < 1$. Since a constant might be regarded as the ultimate low-pass signal, this is unacceptable.

2.2. Level-Invariant Slepian Sequence Inversions

While (3) is the simplest inverse, it is not the only valid one and there are a set of inversions that are also projection operators. For motivation, consider the simplest of these, and write

$$z(t) = x(t) + a \tag{6}$$

so the eigencoefficients become $z_k = y_k + aU_k$ where U_k was defined in (5). If one estimates a using linear regression on the z_k's by minimizing

$$\sum_{k=0}^{K-1} |z_k - \hat{a} U_k|^2 \tag{7}$$

the least-squares estimate is

$$\hat{a} = \sum_{k=0}^{K-1} U_{0,k} z_k / \sum_{k=0}^{K-1} U_{0,k}^2 \tag{8}$$

with residuals $r_k = z_k - \hat{a} U_{0,k}$ We now write the inverse

$$\hat{z}(t) = \hat{a} + \sum_{k=0}^{K-1} r_k \, v_n^{(k)} \tag{9}$$

Re-expanding $\hat{z}(t)$ shows that this sequence of operations is also a projection operator. It is also obvious that if the mean value, a is large compared to the rest of the process, the inverse (9) is much "smoother" than (3). More important, the low-pass filter implicit in inverse (9) preserves constant offsets. It should be noted that this process is the same as that used to estimate the amplitudes of periodic components in the "harmonic-F" test, [18, 17]. Thus, $\hat{a}$ considered as an estimate of the mean, has variance close to the Cramer-Rao bound. We now generalize this procedure so the projection operator preserves polynomials.

2.3. Associated Polynomials

This generalization uses a special set of polynomials associated with the Slepian sequences. Let $R_j(N, W, K; n)$ or, more concisely, $R_j(n)$ be a polynomial of degree j, and $U_{j,k}$ be the expansion coefficients with respect to the k^{th} Slepian sequence

$$U_{j,k} = \sum_{n=0}^{N-1} R_j(N, W, K; n) \, v_n^{(k)}(N, W) \tag{10}$$

with the polynomials defined by the orthogonality condition

$$\sum_{k=0}^{K-1} U_{i,k} U_{j,k} = \delta_{j,k} \tag{11}$$

These polynomials are computed by a Gram-Schmidt like process. Plotted, they resemble the Gegenbauer polynomials $C_n^{(1)}(x)$ with $[-\frac{1}{2}, N - \frac{1}{2}]$ is mapped onto $[0, 1]$.

Denote by $\mathbb{V}$ the $N \times K$ matrix of the K lowest order Slepian sequences, $v_n^{(k)}(N,W)$, with $n = 0, 1, \ldots N-1$, where $K = \lfloor 2NW \rfloor$.

$$\mathbb{V} = \begin{bmatrix} v_0^{(0)} & v_0^{(1)} & \cdots & v_0^{(K-1)} \\ v_1^{(0)} & v_1^{(1)} & \cdots & v_1^{(K-1)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{N-1}^{(0)} & v_{N-1}^{(1)} & \cdots & v_{N-1}^{(K-1)} \end{bmatrix} \tag{12}$$

so, by the orthonormality of the Slepian sequences

$$\mathbb{V}'\mathbb{V} = \mathbf{I}_K \tag{13}$$

where $'$ denotes transpose and $\mathbf{I}_K$ the $K \times K$identity matrix.

Similarly, let $\mathbf{R}$ be the $N \times P$ matrix of polynomials $R_j(t)$ evaluated at $t = 0, 1, \ldots, N-1$ of degrees $0, \ldots, P-1$. Denote the $K \times P$ matrix of inner products by

$$\mathbf{U} = \mathbb{V}'\mathbf{R} \tag{14}$$

and define the polynomials by requiring

$$\mathbf{U}'\mathbf{U} = \mathbf{I}_P \ . \tag{15}$$

Note that $\mathbf{R}'\mathbf{R} \neq \mathbf{I}_P$ Denote a given data block (possibly translated in frequency) of N samples by the vector $\mathbf{X}$ and the corresponding eigencoefficients by

$$\mathbf{Y} = \mathbb{V}'\mathbf{X} \tag{16}$$

These are separated into polynomial coefficients,

$$\mathbf{a} = \mathbf{U}'\mathbf{Y} \tag{17}$$

and residuals,

$$\mathbf{r} = \mathbf{Y} - \mathbf{U}\mathbf{a} \ . \tag{18}$$

The degree $P-1$ reconstruction is

$$\widehat{\mathbf{X}} = \mathbb{V}\mathbf{r} + \mathbf{R}\mathbf{a} \ . \tag{19}$$

It is easily shown that the eigencoefficients corresponding to $\widehat{\mathbf{X}}$ are identical to those of $\mathbf{X}$ so the projection property is preserved even though the sequences, $\widehat{\mathbf{X}}$, can look quite different for different P. We note, however, that effects such as Gibb's and truncation ripples are proportional to the amplitude of the signal and, by effectively removing local polynomial trends, the ripples are greatly attenuated.

3. MAXIMUM–CONCENTRATION WITH IRREGULAR SAMPLING

It has been suggested that one approach to this problem is to use a basis set equivalent to Slepian sequences but with the concentration problem on the existing samples. Construction of these sequences was done in [19]. Denote the set of observation times in $[0, N-1]$ by $\mathbb{O}$ and optimize the energy concentration in the same way as done in the equally–spaced case. This gives

$$\beta_k z_k(t) = \sum_{u \in \mathbb{O}} \frac{\sin 2\pi W(t-u)}{\pi(t-u)} z_k(u) \tag{20}$$

where $t \in \mathbb{O}$ in the eigenvalue problem and unrestricted when interpolating the eigenvectors. If there are $\mathbb{P}$ observations in $[0, N-1]$, the eigenvalue problem (20) is $\mathbb{P} \times \mathbb{P}$ with, obviously, $\mathbb{P} \leq N$. Figure 2 compares the eigenvalues λ_k and β_k for the case $N = 100$, $W = 0.05$ for the relatively benign situation where the 25 missing values are contiguous and near the center of the data block. Clearly, the available concentration of the gappy problem is much poorer than that of the Slepians; $1 - \beta_0 \approx 3.3 \times 10^{-5}$ compared to $1 - \lambda_0 \approx 4.1 \times 10^{-13}$. The eigenvalues, however, do not tell the whole story. Figure 3 shows the first two eigenvectors for the two cases, with Figure 4 the $k = 10$ eigenvectors. The low–order eigenvectors have, as they were designed to, concentrated most of thir energy at existing data samples. One expects about $2NW = 10$ eigenvectors with good energy concentration plus a few transition cases.

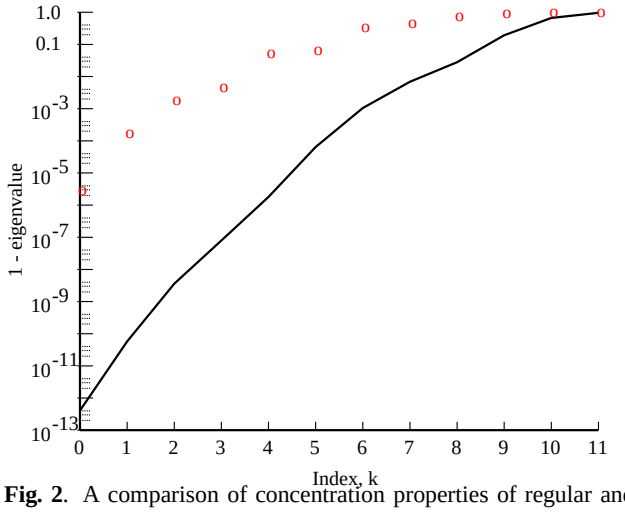

Fig. 2. A comparison of concentration properties of regular and irregular sampling. The solid curve shows $1 - \lambda_k$ *vs* k for Slepian sequences with $N = 100$ and $W = 0.05$. The red circles are $1 - \beta_k$ for the case where samples 37 to 61, inclusive are missing.

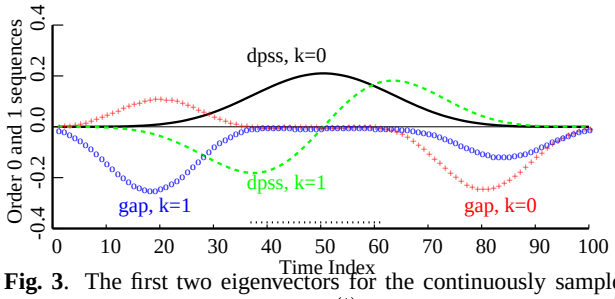

Fig. 3. The first two eigenvectors for the continuously sampled problem, the Slepian sequences $v_0^{(t)}(100, 0.05)$ (black) and $v_1^{(t)}$ (green). For comparison, the red curve is $z_0(t)$ and the blue curve is $z_1(t)$ with the locations of the missing values shown by the "rug plot" at the bottom of the grid. Note that the z_k's are almost zero in this region.

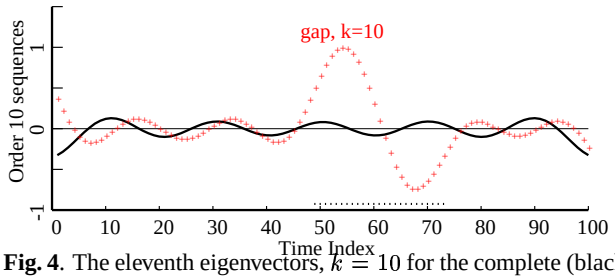

Fig. 4. The eleventh eigenvectors, $k = 10$ for the complete (black) and gappy (red) situations. Note that $z_{10}(t)$ is large in the data gap.

4. INTERPOLATION STRATEGY

The basic idea is to use an iterative scheme where the bandwidth is effectively doubled at each iteration. Starting with raw data, at the beginning of each iteration find the longest remaining gap in the data including the results of previous iterations as "data" Counting the longest gap as two Nyquist samples, $2\Delta^{(i)}$, the bandwidth at the i^{th} iteration is $W^{(i)} = 1/2\Delta^{(i)}$. Using Slepian sequences as a basis, compute a series of time–offset, band–limited approximations to the observations. For simplicity, I have written the following section at baseband, but the same ideas have been used where, as in the SOHO data shown later, the peak in the spectrum is not at the origin and one simply frequency translates the approximations.

The basic approximation is of the form

$$x(t) \approx \sum_{k=0}^{K-1} a_k \, v_t^{(k)}(N, W) \tag{21}$$

for $t \in [0, N-1]$. While this equation is similar to (3), the difference is that with incomplete data, the a_k's are best considered as an estimate of the y_k's. Beginning with a base time t_b, define the approximation

$$x(t|t_b) \approx \sum_{k=0}^{K-1} a_k(t_b) \, v_t^{(k)}(N, W) \tag{22}$$

The expansion coefficients are obtained by minimizing

$$e^2(t_b) = \sum_{t \in \mathbb{O}(t_b)} \left[y(t) - \sum_{k=0}^{K-1} a_k(t_b) \, v_{t-t_b}^{(k)}(N, W) \right]^2 \tag{23}$$

where $\mathbb{O}(t_b)$ denotes the set of observation times in $[t_b, t_b + N - 1]$. Because the Slepian sequences, the $v_k^{(t)}(N, W)$'s are orthogonal on $[0, N-1]$ and not on $\mathbb{O}(t_b)$ the expansion coefficients are found by direct minimization of (23) using a QR algorithm. Denoting the lower edge of a gap by t_L and the upper edge by t_H with $t_H - t_L \approx 2\Delta^{(i)}$ estimate the center point

$$t_c = \frac{t_L + t_H}{2}$$

by

$$\hat{x}(t_c) = \underset{t_b}{\text{ave}} \; x(t_c|t_b) \tag{24}$$

Because interpolators can "overshoot" in long gaps, a useful check is to compare the mean-square data in each block

$$E_d(t_b) = \frac{1}{\mathbb{P}(t_b)} \sum_{t \in \mathbb{O}(t_b)} y(t)^2 \tag{25}$$

where $\mathbb{P}(t_b)$ is the number of observations in $\mathbb{O}(t_b)$, with the mean square value of the interpolation for the block

$$E_i(t_b) = \frac{1}{N} \sum_{t=0}^{N-1} x(t|t_b)^2 \frac{1}{N} \sum_{k=0}^{K-1} a_k(t_b)^2 \tag{26}$$

and examine any block where they differ by much more than the residual $e^2(t_b)$ carefully. Note that Bessel's inequality is not satisfied here, so one can have

$$E_i(t_b) > E_d(t_b) \; .$$

This is because $E_d(t_b)$ represents the energy in the data while $E_i(t_b)$ is that in the interpolations. Among other reasons it appears that the irregular sampling can result in very badly conditioned normal equations where some observation times exery undue influence on the solution. We have recently [20] found an iterative weighting scheme where samples corresponding to excessive values of the "hat" matrix are downweighted. While this, or similar robust regression schemes could be applied here, they essentially discard observations and, as many of the problems occur where observations are sparse, the constraint method of the following section appears to be better suited for this problem.

4.1. Constrained Normal Equations

It was observed that the interpolation variance exceeded the data variance a distressing fraction of the time, so various constrained solutions were attempted. Constraints on integrated squared derivatives attenuate high frequencies excessively, but the following likelihood constraint appears to give satisfactory results. Defining the coefficient vector of the interpolator as

$$\mathbf{A} = [a_0, a_1, \ldots, a_{K-1}]'$$

and impose the likelihood constraint

$$L = e^2(t_b) + \mu \mathbf{A}^T \mathbf{\Sigma}^{-1} \mathbf{A} \tag{27}$$

that is the squared misfit (normally standardized by the observation variance) plus a penalty term. Here $\mathbf{\Sigma}$ is the eigencoefficient covariance matrix,

$$\mathbf{\Sigma} = \mathbf{E}\{\mathbf{A}\mathbf{A}^T\} \; , \quad \mathbf{B} = \mathbf{\Sigma}^{-1} \tag{28}$$

that is typically estimated from stretches of good data and μ a parameter to be determined. It should be noted that this constraint is matched to the data and unlike arbitrary "smoothness" constraints allows rapid oscillations if this is what the data contains.

The expected values of the terms in (27) are

$$\mathbf{E}\{e^2(t_b)\} = \mathbb{P}(t_b) - K \tag{29}$$

$$\mathbf{E}\{\mathbf{A}^T \mathbf{B} \mathbf{A}\} = K \tag{30}$$

Temporarily dropping the base–time dependence of the $\{a_k\}$'s and writing L as

$$L = \sum_{t \in \mathbb{O}(t_b)} \frac{1}{\sigma^2(t)} \left[y(t) - \sum_{k=0}^{K-1} a_k \, v_{t-t_b}^{(k)} \right]^2 + \mu \sum_{j,k=0}^{K-1} a_j (\mathbf{B})_{j,k} a_k \tag{31}$$

where $\sigma(t)$ is the estimated error of the data at time t, and differentiating with respect to the coefficients gives

$$\frac{\partial L}{\partial a_j} = 0 = \sum_{t\in\mathbb{O}(t_b)} \frac{1}{\sigma^2(t)}\, v^{(j)}_{t-t_b} \left[y(t) - \sum_{k=0}^{K-1} a_k\, v^{(k)}_{t-t_b} \right] + \mu \sum_{k=0}^{K-1} (\mathbf{B})_{j,k} a_k \tag{32}$$

Denoting

$$c_j = \sum_{t\in\mathbb{O}(t_b)} \frac{1}{\sigma^2(t)} y(t)\, v^{(j)}_{t-t_b} \,, \tag{33}$$

and

$$\Psi_{j,k} = \sum_{t\in\mathbb{O}(t_b)} \frac{1}{\sigma^2(t)} v^{(j)}_{t-t_b}\, v^{(k)}_{t-t_b} \,, \tag{34}$$

so (32) becomes

$$\mathbf{C} = \Psi\mathbf{A} + \mu\mathbf{B}\mathbf{A} \,. \tag{35}$$

Defining the eigenvectors $\mathbf{X}_j$ and eigenvalues θ_j of the generalized eigenvalue problem

$$\Psi\mathbf{X}_j = \theta_j \mathbf{B}\mathbf{X}_j \tag{36}$$

with the standardization

$$\mathbf{X}_j^\dagger \Psi \mathbf{X}_k = \delta_{j,k} \tag{37}$$

$$\mathbf{X}_j^\dagger \mathbf{B} \mathbf{X}_k = \theta_j \delta_{j,k} \tag{38}$$

and the coefficients

$$c_q = \mathbf{X}_q^\dagger \mathbf{C} \tag{39}$$

and expanding $\mathbf{A}$ as

$$\mathbf{A} = \sum_p \alpha_p \mathbf{X}_p \tag{40}$$

(35) becomes

$$c_q = \sum_p \alpha_p (1 + \mu\theta_p)\delta_{p,q} \tag{41}$$

so

$$\alpha_p = \frac{c_p}{1+\mu\theta_p} \tag{42}$$

Expanding the likelihood constraint with the preceding definitions (30) becomes

$$K = \mathbf{A}^T\mathbf{B}\mathbf{A} \tag{43}$$

$$= \sum_{p,q} \alpha_p \alpha_q \mathbf{X}_p^\dagger \mathbf{B} \mathbf{X}_q \tag{44}$$

$$= \sum_p \alpha_p^2 \theta_p = \sum_p \frac{c_p^2 \theta_p}{(1+\mu\theta_p)^2} \tag{45}$$

Thus one can solve for μ. Much of the motivation for this approach comes from [21] and, in common with his work, the normal equations formulation can be replaced with a generalized SVD to improve accuracy. Here, however, the constraint (30) can either be satisfied exactly, or, because $\mathbf{A}^\mathbf{T}\mathbf{B}\mathbf{A}$ is distributed as χ^2_K, in distribution.

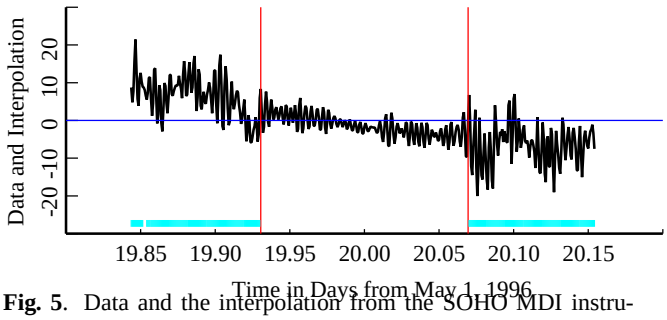

Fig. 5. Data and the interpolation from the SOHO MDI instrument, the $l = 1$, $m = 0$ spherical harmonic coefficient. The data gap, marked by the vertical red lines, is about 220 samples long.

5. DISCUSSION

Some additional aspects of this process are:

- The procedure has much in common with EM algorithms and the EM formalism, see *e.g.* [22], can be invoked to justify the iterative aspect of the process and show convergence.
- If, instead of having a "red" spectrum, the process has its maximum power away from the origin, one should use Slepian sequences centered around the peak frequency or frequencies.
- The reconstruction shown schematically in (22) is, in practice, replaced with (19).
- The simple average (24) is weighted by the estimated reliability of the different blocks.
- $\mathbf{\Sigma}$ is estimated by averaging over good data blocks. This imposes constraints as, typically, one needs many more "good" blocks than basis sequences.
- Clearly, the length of the longest gap will tend to halve at each step, so the process converges in roughly $\log_2(\text{longest gap})$ iterations.
- Although not mentioned explicitly, in practice one includes a stage where short gaps, where "short" depends on the data, are simply filled with a Wiener interpolator as in [2]. In doing this I have found that the best results are obtained if one estimates multitaper spectra on the longest data sections available, averages them, and takes the Fourier transforms to generate autocorrelations. In most physical applications at least 64–bit floating point arithmetic is necessary when working with autocorrelations. Autocorrelation estimates derived using lagged products tend to result in unstable interpolators.

Figure 5 is an example of a longish fill ($\sim$ 220 samples) in the $l = 1$, $m = 0$ spherical coefficient series from the SOHO MDI helioseismology instrument. In addition to the major gap marked by the vertical red lines, there are several small gaps in the plotted data. (The cyan "rug" at the bottom of the frame marks original samples, and one of the short gaps is visible near day 19.85) Overall, the data file consisted of approximately 1.1 million samples at a sampling rate of 1 minute interrupted by 2083 gaps of lengths ranging from single samples to four gaps exceeding 1000 samples. To compensate, there were 14 sections with over 8000 contiguous

samples in each. The longest good section was 15512 samples. The longest gap was 2288 samples and is only marginally recoverable without resorting to interpolating individual modes, a scheme suggested by John Liebacker. This clearly works for p-modes that are well–mapped, but its performance for g–modes is less certain because their frequencies are largely unknown. When the markers are removed the interpolations are essentially indistinguishable from the original data in all but the longest gaps.

6. REFERENCES

[1] R. Schild and D. J. Thomson, "The twin QSO Q0957-561 time delay data set," *Astronomical Journal*, vol. 109, pp. 1970–, 1995.

[2] D. J. Thomson, L. J. Lanzerotti, and C. G. Maclennan, "The interplanetary magnetic field: Statistical properties and discrete modes," *J. Geophysical Res.*, vol. 106, pp. 15,941–15,962, 2001.

[3] P. D. Jones and M. E. Mann, "Climate over past millenia," *Reviews of Geophysics*, vol. 42, pp. 1–42, 2004, doi:10.1029/2003RG000143.

[4] A. Moberg, D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Kartén, "Highly variable Northern Hemisphere temperatures reconstructed from low– and high–resolution proxy data," *Nature*, vol. 433, pp. 613–617, 2005.

[5] R. A. Kerr, "Millennium's hottest decade retains its title, for now," *Science*, vol. 307, pp. 828–829, 2005, News of the Week.

[6] D. J. Thomson, "The seasons, global temperature, and precession," *Science*, vol. 268, pp. 59–68, 1995.

[7] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London, 1996.

[8] W. S. Cleveland and C. Loader, "Smoothing by local regression: Princilples and methods," in *Statistical Theory and Computational Aspects of Smoothing*, W. Hardle and M. G. Schimek, Eds., pp. 10–49. Springer-Verlag, New York, 1996.

[9] D. J. Thomson, "Projection filters for data analysis," in *Proc. Seventh IEEE SP Workshop on Stat. Sig. and Array Proc.*, Quebec, 1994, pp. 39–42.

[10] D. J. Thomson, "Inverse-constrained projection filters," *Proceedings of the SPIE*, vol. 4478, pp. 172–183, 2001.

[11] David J. Thomson, "Multitaper analysis of nonstationary and nonlinear time series data," in *Nonlinear and Nonstationary Signal Processing*, W. Fitzgerald, R. Smith, A. Walden, and P. Young, Eds., pp. 317–394. Cambridge Univ. Press, 2001.

[12] David J. Thomson, Carol G. Maclennan, and Louis J. Lanzerotti, "Propagation of solar oscillations through the interplanetary medium," *Nature*, vol. 376, pp. 139–144, 1995.

[13] D. J. Thomson, "Dependence of global temperatures on atmospheric CO_2 and solar irradiance," *Proc. Natl. Acad. Sci. USA*, vol. 94, pp. 8370–8377, 1997.

[14] C. Kuo, C. Lindberg, and D. J. Thomson, "Coherence established between atmospheric carbon dioxide and global temperature," *Nature*, vol. 343, pp. 709–714, 1990, (Reprinted in pp 395-400 of *Coherence and Time Delay Estimation*, G. C. Carter, Ed., IEEE Press, 1993.).

[15] D. J. Thomson, "Time series analysis of holocene climate data," *Phil. Trans. R. Soc. Lond. A*, vol. 330, pp. 601–616, 1990.

[16] David Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: the discrete case," *Bell System Tech. J.*, vol. 57, pp. 1371–1429, 1978.

[17] D. J. Thomson, "Quadratic-inverse spectrum estimates: applications to paleoclimatology," *Phil. Trans. R. Soc. Lond. A*, vol. 332, pp. 539–597, 1990.

[18] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, pp. 1055–1096, 1982.

[19] T. P. Bronez, "Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences," *IEEE Trans. on Signal Processing*, vol. 36, pp. 862–873, 1988.

[20] A. D. Chave and D. J. Thomson, "A bounded influence regression estimator based on the statistics of the hat matrix," *J. Roy. Stat. Soc., Ser. C (Applied Statistics)*, vol. 52, pp. 307–322, 2003.

[21] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Review*, vol. 34, pp. 561–580, 1992.

[22] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley and Sons, New York, 1997.