Astronomical Data Analysis Software and Systems IV
ASP Conference Series, Vol. 77, 1995
R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.

Interpolation of Irregularly Sampled Data Series—A Survey

H.-M. Adorf

Space Telescope–European Coordinating Facility, European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching b. München, Germany

Abstract. Many astronomical observations, including spectra and time series, consist of irregularly sampled data series, the analysis of which is more complicated than that of regularly spaced data sets. Therefore a viable strategy consists of resampling a given irregularly sampled data series onto a regular grid, in order to use conventional tools for further analysis. Resampling always requires some form of interpolation, which permits the construction of an underlying continuous function representing the discrete data. This contribution surveys the methods used in astronomy for the interpolation of irregularly sampled one-dimensional data series.

There is virtually no literature
on which method does best
with which kind of problem.
— *Julian H. Krolik 1992*

1. From Irregular to Regular Sampling

Many astronomical observations consist of irregularly sampled data series, e.g., spectra from spectrographs with non-linear wavelength scales, time series of light curves or radial velocity observations. It is well known that the analysis of irregularly spaced data sets is more complicated than that of regularly spaced ones. Therefore, as a promising strategy, one might consider resampling a given irregularly sampled data series onto a regular grid, in order to use conventional tools for further analysis.

Resampling always requires some form of interpolation or, in the presence of noise, *estimation*, which effectively allows the construction of an "underlying" continuous function representing the discrete data. The goal is to use an interpolation/estimation method which preserves the relevant information as much as possible. Therefore the assumptions on which the interpolation/estimation method rests must not preclude (or predetermine) the results of the subsequent analysis.

Constructing a continuous representation of an irregularly sampled data set may have a virtue of its own, since structure present in the data is often more clearly displayed by a continuous curve than by the scattered, clumped original data points, and may therefore more readily be picked up by the eye.

2. Interpolation Methods in Current Use

Numerous interpolation/estimation schemes exist, many of which generalise to the case of irregular sampling. Methods may be distinguished according to whether they are conceptually simple or complex, linear or non-linear, direct or iterative, and exact or approximate. Furthermore they may accommodate different statistical weights, respect a potentially existing non-negativity constraint, or allow an estimate of the interpolation error.

2.1. Direct Fourier Transform

This method (Scargle 1989) is applicable to data with arbitrary sampling. It allows to explicitly compute individual Fourier coefficients. The basic formulae are generalized to the case of unequal statistical weights. The method is equivalent to least-squares estimation of a single-frequency harmonic. By computing the Fourier coefficients at all required discrete frequencies and then Fourier-transforming back, a kind of interpolation can be obtained.

2.2. Compound Fourier Transform

The compound Fourier transform (FT) method (Meisel 1979) has been developed for gapped data. It uses a stack of individual discrete FTs. The exposition of the method is quite convoluted, and (therefore?) the method seems to be rarely used.

2.3. Matrix Inversion

Kuhn (1982) considers the irregular-to-regular resampling problem, and discusses two methods for solving the resulting matrix equation. One method is apparently applicable to jittered sampling, where the interpolation problem can be treated as a perturbation problem. Small off-diagonal terms in the interpolation matrix permit an iterative solution of the matrix equation. The other, direct method is applicable to regular sampling with missing data, and exploits the fact that in this case the interpolation matrix is circulant. Swan (1982) discusses the interpolation problem as an inverse problem in the presence of noise. The irregular sampling geometry makes the interpolation matrix ill-conditioned, leading to noise-amplifications, when straightforwardly applied to the data. Swan discusses two regularization strategies.

2.4. Least-Squares Estimation

This method, also known as minimum variance estimation, is a linear method applicable to arbitrary sampling patterns. For a signal with additive Gaussian noise it is a maximum likelihood estimation method. The LS-method permits the estimation of the probable interpolation error, and in its generalized formulation allows the inclusion of statistical weights. Barning (1963) devised an iterative scheme for the simultaneous LS-fitting of multiple sinusoids to data without weights. Ferraz-Mello (1981) advanced a formula for a weighted LS-fit (linear regression) of a single sinusoid model to zero-mean data. Gram-Schmidt orthogonalization is used to decouple the two sine and cosine terms of which the model is composed. He also considers the case when more than one harmonic is present in the data. Ferraz-Mello recommends an iterative scheme of subtracting previously fitted individual components ("pre-whitening"). Rybicki & Press (1992) present a general formulation of the LS-estimation method for arbitrary sampling. An interpolation error estimate is provided.

2.5. Pre-Whitening and CLEAN Deconvolution

This method is applicable to a regularly sampled data series with missing data. Gray & Desikachary (1973), a year before the publication of CLEAN, employ an iterative scheme in the Fourier domain, where the Fourier transformed window pattern is identified in the direct Fourier transform of the data and subtracted. The required corresponding phase is estimated from the data. In a similar fashion Roberts et al. (1987) use a generalized, complex version of the CLEAN deconvolution method, to deconvolve the direct ("dirty") Fourier transform of the data, using the window function as "beam". CLEAN differs from Gray & Desikachary's method by using a smaller "gain" factor for the subtractions.

2.6. Autoregressive Maximum-Entropy Interpolation

This method by Fahlman & Ulrych (1982), applicable to a regularly sampled data series with missing data segments ("gaps"), attempts to fill the data gaps. Gap-filling starts with estimating an initial autoregressive (AR) model of order less than the length of the shortest known data segment. A maximum Burg-entropy algorithm is used to predict unknown data values. A new AR model is estimated from the entire data series, with known and predicted values, and iteration proceeds until convergence. The original method was improved by Brown & Christensen-Dalsgaard (1990) via inclusion of a bandpass-filtering preprocessing step. Despite the wide use of maximum entropy methods elsewhere in astronomy (particularly aperture synthesis imagery), ME interpolation seems not to have caught on widely (cf. Krolik 1992).

2.7. Polynomial Interpolation

This method (Groth 1975) is applicable to data with arbitrary sampling. A set of orthonormal polynomials is constructed over the set of irregular sampling locations. The interpolation is obtained as a linear superposition of these base polynomials. Statistical weights may be incorporated.

2.8. Trigonometric Polynomial Interpolation

This method is applicable to arbitrarily sampled band-limited, periodic (deterministic) data. It is based on the fact that a band-limited, periodic function can be exactly reconstructed from a sufficient number of irregularly spaced samples, using an explicit closed-form formula based on trigonometric polynomials (Adorf 1993). The method, originally devised by Cauchy in 1841, constructs a set of orthogonal basis functions over the set of irregular sampling locations. The method's linearity permits an estimate of the interpolation error. Interpolation with trigonometric polynomials may become an invaluable tool for high-fidelity interpolation of high signal-to-noise data.

2.9. POCS Interpolation

The POCS (projection onto convex sets) based interpolation method (Adorf 1993) is a conceptually simple, iterative scheme, directly applicable to regularly sampled observations with missing data, where the underlying continuous signal has a known (or assumed) fundamental period. The POCS method attempts to fill the data gaps by constructing a bandpass filtered version of the data series. The procedure iterates two "projection" operators: the first "completes" the original data series by filling gaps with arbitrary values, then applies a bandpass-filter; the second re-substitutes known data values. The POCS process provably converges, if the intersection of the convex sets is non-empty.

References

Adorf, H.-M. 1993, in Proc. 5th ESO/ST-ECF Data Analysis Workshop, eds. P. Grosbøl et al. (Garching, ESO), p. 191

Barning, F. J. M. 1963, Bull. Astron. Inst. Netherlands, 17, 22

Brown, T. M., & Christensen-Dalsgaard, J. 1990, ApJ, 349, 667

Fahlman, G. G., & Ulrych, T. J. 1982, MNRAS, 199, 53

Ferraz-Mello, S. 1981, AJ, 86, 619

Gray, D. F., & Desikachary, K. 1973, ApJ, 181, 523

Groth, E. J. 1975, ApJS, 29, 443

Krolik, J. H. 1992, in Statistical Challenges in Modern Astronomy, eds. E. D. Feigelson & G. J. Babu (New York, Springer-Verlag), p. 349

Kuhn, J. R. 1982, AJ, 87, 196

Meisel, D. D. 1979, AJ, 84, 116

Roberts, D. H., Lehár, J., & Dreher, J. W. 1987, AJ, 93, 968

Rybicki, G. B., & Press, W. H. 1992, ApJ, 398, 169

Scargle, J. D. 1989, ApJ, 343, 874

Swan, P. R. 1982, AJ, 87, 1608