So, you do not need a fit through the data points, but to obtain the boundaries of your data, do you?

1) THE METHOD

In astronomical problems one often needs to determine the upper and/or the lower boundary of a given data set. An automatic and fast approach consists in fitting the data using a Modified Least Squares Method, where the function to be minimized, $x_i$, is defined to handle, asymetrically, the data at both sides of the boundary. For example, in the case of a set of $N$ points of coordinates $(x_i,y_i)$, and considering uncertainties only in the $Y$ axis ($\sigma$), that function can be written as

$$\chi^2 = \sum_{i=1}^{N} \frac{[y_i - f(x_i,\alpha)]^2}{\sigma_i^2},$$

where $\alpha$ is an exponent (in normal Least Squares $\alpha=2$), $y_i$ is the fitted function evaluated at $x_i$, and $\sigma_i$ is an overall weighting factor that is responsible for introducing the asymmetry in the fit. This factor is computed as explained in the Table 1. $\beta$ is the exponent that determines whether the fit is error weighted or not ($\beta=0$ to ignore errors; typically $\beta=2$ for error weighted fits), and the asymmetry factor ($\gamma$) must be greater than 1. This function can be easily minimized following a numerical strategy. The use of single polynomials can provide a good answer and is computationally very simple. An example of this polynomial boundary fit is presented in Figure 1. However, when the data exhibit rapidly changing values, a single polynomial is not always able to reproduce the observed trend. A more powerful alternative in these cases is the use of adaptive splines, which exhibit a much larger flexibility.

Figure 1: Comparison between different strategies to determine data boundaries. The gray data correspond to 100 points randomly drawn from the function $y(x)$ (line), assuming $\Delta x=0.01$ and $\Delta y=10$. The boundaries have been determined using single polynomials (blue line; upper boundary 5th degree; lower boundary 6th degree) and adaptive splines (magenta; upper boundary 4 knots; lower boundary: 10 knots). As expected, splines are more flexible and provide tighter boundaries than polynomials.

Although the last example (and the rest presented in this poster) corresponds to 1-dimensional (1D) boundaries in 2D diagrams, the method can be applied to higher dimensions (e.g. the determination of a surface as boundary for data in a 3D-parameter space), once an appropriate metric (distance) is defined in the multidimensional space.

3) ADAPTIVE SPLINES

Splines are commonly employed for interpolation and modeling of arbitrary functions. Many other authors have preferred to simple polynomials due to their flexibility. A piecewise polynomial function that is locally very simple, typically third-order polynomials (the so called cubic splines). These local polynomials are forced to pass through a prefixed number of points, which we will refer as knots. The coefficients of these polynomials are determined by imposing in addition that the first and second derivatives match at the knots (two additional conditions are required; normally they are provided by assuming that the second derivatives at the two endpoints to be zero, leading to what are normally called “natural splines”). The computation of splines is widely described in the literature (see e.g. Gerald C.F., Wheatley P.G., 1989, in Numerical Recipes (4th edition)).

The final result of a fit to splines will strongly depend on both, the number and the precise location of the knots. In order to more flexibility in the fit, Cardiel N. (1999, PhD Thesis, Universidad Complutense de Madrid) explored the possibility of setting the location of the knots as free parameters, and determine the optimal coordinates of these knots that improve the fit to the data. The solution to the problem can be derived numerically using any minimization algorithm. Here we have used DOWHILL (Nelder J.A., Mead R., 1965, Computer Journal 7, 308), which only requires to evaluate the function to be minimized (and not the derivatives), provided an initial guess to the solution is available. An implementation of the method can be found in Press et al. 1989 (Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd edition).

The software code to compute these fits is available at http://www.ucm.es/info/Astrofis/software/boundfitl

2) AN APPLICATION TO REAL DATA

A common problem when handling spectroscopic data is the determination of a “reasonable” fit to the spectra continuum. Most of the times people are happy enough just by fitting a polynomial to the general trend of the spectra, masking disturbing spectroscopic features such as important emission lines or deep absorption characteristics. A good alternative is to obtain the upper and/or lower boundary fits, either by using polynomials or adaptive splines, as explained before. In the next plots some examples of these fits to the continuum of the K0 star HD033653 are shown, in which the effects of modifying different relevant parameters (e.g., the asymmetry factor $\gamma$, the exponent $\alpha$, or the number of knots) are examined. For simplicity, the examples are not error weighted (i.e., $\beta=0$ has been assumed).

Figure 2: Examples of boundary fitting to adaptive splines. The grey points correspond to the same data displayed in Figure 1.

- Fit#1: after step#1, NITER=10.
- Fit#2: after step#2, merging colliding knots (3 and 2 in the upper boundary and 4 and 3 in the lower boundary).
- Fit#3: after step#3, NITER=10.
- Fit#4: after step#4, merging colliding knots (3 and 2 in the upper boundary and 4 and 3 in the lower boundary).
- Fit#5: after step#5, NITER=10.

Figure 3: A single polynomial boundary fit is presented in Figure 1. However, when the data exhibit rapidly changing values, a single polynomial is not always able to reproduce the observed trend. A more powerful alternative in these cases is the use of adaptive splines, which exhibit a much larger flexibility.

Figure 4: The boundaries have been determined using single polynomials (blue line; upper boundary 5th degree; lower boundary 6th degree) and adaptive splines (magenta; upper boundary 4 knots; lower boundary: 10 knots). As expected, splines are more flexible and provide tighter boundaries than polynomials.

Although the last example (and the rest presented in this poster) corresponds to 1-dimensional (1D) boundaries in 2D diagrams, the method can be applied to higher dimensions (e.g. the determination of a surface as boundary for data in a 3D-parameter space), once an appropriate metric (distance) is defined in the multidimensional space.
