A METHOD OF CONTINUUM ESTIMATION IN SPECTRAL CLASSIFICATION EXPERIMENTS

J. PELT
Tartu Astrophysical Observatory

ABSTRACT: A flexible method of continuum or pseudo-continuum estimation for digitally recorded spectra is presented. The properties of the algorithm are evaluated considering the requirements for spectral classification experiments.

1. Introduction

There are many methods to estimate the level of continuum in digitally recorded spectra, most of them requiring human intervention. In automatic data reduction systems the operator can test various methods to be used or adjust algorithm parameters only at the time of experiment setup. For the full data set under study, the processing must proceed automatically, to avoid subjective judgements and sometimes even "trimming" of data to fit some preset goals in spectral classification. The precision and objectivity of quantitative classification heavily depends on method of continuum estimation.

The definition of the continuum, even for the physical spectrum observed without any distortion, is not a trivial task. For more or less hot stars, we can treat the continuum as a certain physical reality. For late type stars however, only the so-called "pseudo-continuum" can be considered.

In this paper we try to put some objectivity into the term "pseudo-continuum" in the context of quantitative spectral classification. We propose a particularly useful technique to compute the pseudo-continuum and present a method to evaluate it by using artificially generated data.
2. Definition of the continuum in stellar classification experiments

In the automatic spectral classification experiment we have a sample of stars with physical spectra \( Y_n(\lambda), n = 1, 2, \ldots, N \). If \( C_n(\lambda) \) denotes the true physical continuum for each star, then the set of quantities:

\[
q^n_K = \int_{-\infty}^{\infty} \frac{C_n(\lambda) - Y_n(\lambda)}{C_n(\lambda)} d\lambda
\]

in which \( \omega_k(\lambda), k = 1, 2, \ldots, K \) are the appropriate weight functions, can be combined in the descriptor vectors

\[
Q^n = \begin{bmatrix} q^n_1 & q^n_2 & \cdots & q^n_K \end{bmatrix}
\]

and used in the following analysis by various classification and pattern recognition methods. However, the physical spectra \( Y_n(\lambda) \) are not directly observable and, consequently, the vectors \( Q^n \) can be used only in theoretical work.

The observed and digitally recorded spectra \( X_n(i), i = 1, 2, \ldots, L \) are the modulated (by interstellar absorption, atmospheric extinction, instrumental transmission and photographic response functions), smeared (by instrumental profile and atmospheric seeing conditions), corrupted (by photographic noise and readout errors) and sampled versions of \( Y_n(\lambda) \).

Exact recovery of absorption band strengths \( q^n_K \) from \( X_n(i) \) is impossible and even estimates with explicit error bounds can be computed with great effort, if at all. Therefore, we must consider some approximations.

We assume that the distortions caused by different photographic responses and varying dispersions along the \( \lambda \) - axis have been removed (as much as it is possible) by the techniques of spectral data reduction (Ardeberg et al. 1975 and references therein). This allows us to define another set of quantities for rectified spectra \( X_{rec}^n(i) \):

\[
q^n_K = \frac{\int_{-\infty}^{\infty} \frac{C_n(\lambda) - X_{rec}^n(\lambda)}{C_n(\lambda)} d\lambda}{\int_{-\infty}^{\infty} \frac{C_n(\lambda) - X_{rec}^n(\lambda)}{C_n(\lambda)} d\lambda}
\]
where $\omega_k(l)$ are again certain weights for separating different absorption bands and the sequences $C^n(l)$ are the so-called "pseudo-continua". Our basic assumption is that the vectors:

$$Q^n = \{ q^n_1, q^n_2, \ldots, q^n_K \}$$

for a carefully computed set $C^n(l)$ can be treated as being slightly distorted by approximation errors, one-to-one, continuous and coordinate-wise monotonic mapping of set $Q^n$:

$$\hat{Q}^n = F(Q^n) + E^n$$

(here $E^n$ denotes mapping errors). If this assumption is valid, then the separation of different spectral classes or estimation of physical parameters can be carried out in the space of estimates $\hat{Q}^n$ (for the spectrum under study, as well as for the spectra in the training set of the classification scheme). In traditional methods (Ardeberg et al. 1975, Malyuto and Pelt 1981, Hill et al. 1982) the observed spectra $X^n(l)$ are fully rectified (demodulated, filtered and s. u.) to get estimates of the real physical spectra $Y^n(\lambda)$, and after that, the vectors $Q^n$ are computed. As a result, all reduction errors are also absorbed in the values of $Q^n$.

The choice of values for the pseudo-continuum $C^n(l)$ is not a trivial task, e.g. if we take $C^n(l) = X^n(l), l=1,2,\ldots,L$ we get zero values for all bands $q^n_k$, and the mapping $F$ is not definitely one-to-one. On the other hand, if we take $C^n(l) = C_0^n$ (constant continuum level), then all effects of different modulations are also present in the values of $q^n_k$ and it can be only accidentally so, that monotonicity of $F$ is observed. Thus, we must compromise between the "absolute" flexibility of the first method and "absolute" stiffness of the second.

The problems related to the weight systems $\omega_k(l)$ are not so severe. As a rule, we can ignore the effects of smearing and sampling, and consequently, the weights $\omega_k(l)$ can be computed as the sampled versions of continuous weights $\omega_k(\lambda)$. Nevertheless, for narrow spectral bands where
the width of the smearing profile and the widths of bands are comparable, we must consider a certain "leakage" of spectral energy from one band to another. In this paper, however, we are not going to discuss these matters in detail and leave it for future work.

We define the pseudo-continuum in the following way. First we construct an algorithm for estimating the continuum applicable for early type stars. This algorithm contains a set of free adjustable parameters. By using the Monte-Carlo type statistical experiments we pick up the best combination of parameters in the sense of one-to-oneness, continuity and monotonicity of the mapping operator \( F \), stability against noise and modulations. Then the "best" combination is used in spectral classification experiments.

3. Basic algorithm

Our algorithm is based on a simple analytical model for recorded spectra:

\[
X(l) = C(l) + I(l) + N(l), \quad l = 1, 2, \ldots, L
\]

where \( I(l) \geq 0 \) for the given region. Consequently, the algorithm can be used only in the analysis of absorption spectra.

At the moment we neglect the noise term \( N(l) \) and consider the recorded spectrum as a sum of the smooth continuum \( C(l) \) and a more erratic line spectrum \( I(l) \). The smoothness of \( C(l) \) means that its Fourier transform is effectively band-limited. Contrarily, the energy in the Fourier transformed line spectrum \( I(l) \) is distributed more uniformly. Thus, two spectral components of recorded spectra can be characterized by different energy distribution in the transform domain and by nonpositiveness of the line spectrum in the wavelength domain. This information can be used to separate them.

Suppose we can construct a linear filter \( W(l) \) which is positive (\( W(l) \geq 0 \) for every \( l \)), so that all the frequency components of the continuum \( C(l) \) pass through the filter without distortion and that in the "stop band" it effectively removes a part of energy from the line spectrum.
Applying the filter to the original recorded spectrum we get the first estimate for the continuum

\[ C^1(l) = C(l) + W(l) * I(l) \]

in which the asterisk denotes convolution (filtering) operation. Because of the positiveness of \( W(l) \) and the nonpositiveness of \( I(l) \), the estimate \( C^1(l) \) is strongly below the real continuum level. This is the "smooth" part of our algorithm.

Some parts of recorded spectra exceed the computed estimate \( C^1(l) \). Consequently, we can construct a new recorded spectrum without losing a bit of the real continuum:

\[ X^1(l) = \max\left\{ X(l), C^1(l) \right\} \]

This is the "clip" part of our algorithm. It is evident that \( X^1(l) \) can be once again presented as the sum of two components

\[ X^1(l) = C(l) + I^1(l) \]

where \( I^1(l) \) is the clipped version of the line spectrum. By iteratively applying the "smooth" and "clip" steps we get a sequence of estimates \( C^i(l), i=1,2,... \) apparently converging to \( C(l) \).

Such "smooth" and "clip" algorithms are also used in picture processing (for background elimination) and in spectral data reduction (Peat and Pemberton 1970, Malyuto and Pelt 1981). Here the most important problem is, that no such filter \( W(l) \) exactly exists, and it must be approximated in some way or other. Secondly, the real recorded spectrum is corrupted by noise and our assumptions are not exactly valid. Therefore, some practical experiments and qualitative or quantitative evaluation of different variants proposed must be carried out.

The "clipping" operation is equally simple in every possible realization of the algorithm, however, the "smooth" part can be implemented in very different ways. Various regression schemes have been proposed (Peat and Pemberton 1970, Hill et al. 1982). Some authors suggest that

We divide the observed spectrum into LCONT equal intervals and approximate the current estimate of the continuum by the third degree polynomial in each interval. The best approximates (in the sense of least sum of squares) can be computed using subroutines from the monograph of de Boor. However, the approximation must be carried out for every iteration step. Consequently, some optimizations can be made by removing some computations out of main iteration loop. We are not going to discuss these optimizations in detail. An interested reader can receive a copy of algorithm listing from the author.

When applying the algorithm to the observed spectra, we were annoyed to see that relative approximation errors in the underexposed areas of the spectrum were rather large. To overcome the problem, we adopted a two-step procedure. At the first step, we computed the rather smooth approximation to the spectral distribution function, and divided the recorded spectrum by it (see also Malyuto etc. 1986). After that we found iteratively the continuum estimate and finally multiplied it by smooth spectral distribution function. As a result, we obtained a uniform distribution of relative approximation errors. The smooth spectral distribution function was computed by the same basic spline approximation routine as the continuum. Only the number of approximation intervals (and with it the degree of smoothness) was varied.

We can see that the whole algorithm depends effectively on three parameters: the number of intervals for computing the spline approximation for the spectral distribution function LDIST, the number of intervals for computing the spline approximations for the continuum LCONT and number of iterations NITER of "smooth" and "clip" steps. For stars with well defined continua, we can stop the iteration depending on the convergence of the algorithm. For late type stars, for which only pseudo-continuum can be computed, however, the number of iterations must be considered as the basic parameter of the iteration process.

At the first stage of our investigation, we plotted out a rather extensive set of continuum estimates and tried to choose between the
parameters by visually inspecting and evaluating the plots. However, we were not satisfied by this process, because no consensus was obtained between the different spectroscopists in our observatory about the "true" or "best" continuum. To put some objectivity into the parameters determining process, we formulated the basic assumptions for the spectral classification experiment. Then we evaluated various parameter sets starting from these assumptions in the Monte-Carlo type experiment.

4. Algorithm evaluation

One-to-one correspondence of the mapping operator $F$ cannot be evaluated directly because of lack of the original $\mathcal{O}^n$ values. However, with artificially generated local disturbances in the observed spectra we can, more or less consistently, estimate resulting changes in the $\mathcal{O}^n$ values and evaluate qualitatively properties of the various sets of basic parameters. We selected the sets of values LDIST, LCONT, NITER for which monotonicity in $\mathcal{O}^n$ values against gradual distortion by added artificial absorption lines was observed and for which the $\mathcal{O}^n$ values were in reasonable limits (particularly - we avoided the negative band strengths).

For the selected sets of parameters, we adopted the following procedure of quantitative evaluation: we distorted observed spectra by multiplying them by various waves $D_r^r(l), r=1,2,\ldots, R$ (see fig. 1): 

$$X_r^r(l) = D_r^r(l) \cdot X_{rec}^r(l)$$

and after estimating the pseudo-continuum $C_r^r(l)$ for each distorted spectrum we computed the $\mathcal{O}_r^r$ estimates. Theoretically, the quantities $\mathcal{O}_r^r$ cannot depend on the distortions $D_r^r(l)$. Practically, because of several approximations, they do, and we can estimate their variability by computing appropriate standard deviations. In Fig.2 you can see a typical result of such an experiment: standard deviations are plotted against parameter LCONT, other two parameters are set to LDIST=3 and NITER=50, respectively. Fig.2 displays typical behavior of inner variability in $\mathcal{O}_r^r$ values depending on the number of approximation nodes LCONT. In the region of LCONT=6-8 the dependence starts to decrease monotonically until it will be equal to zero. The result is consistent with our
Figure 1. Five distortion curves used in Monte-Carlo experiments.

Figure 2. Standard deviations of pseudo-continua (arbitrary units) in experiment with artificially added noise and in experiment with distortions. Dependence on smoothing parameter LCONT.
earlier expectations as the large value for \( \text{LCONT} \) means more flexibility for the pseudo-continuum, which approximates more and more closely the recorded spectrum. At the limit, they coincide, the standard deviations become equal to zero and (regrettably) the values of \( \hat{\alpha}^n \) also converge to zero. Thus, we must compromise between the inner variability and informativity of \( \hat{\alpha}^n \) values. A sample of evaluation results is presented in Table 1, where standard deviations (in arbitrary units) are given against various sets of basic parameters. These sets that were not adopted in the first stage of selection process are marked by zero entry in the table.

### Table 1

<table>
<thead>
<tr>
<th>NITER</th>
<th>LDIST</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9.33</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.61</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.55</td>
<td>2.44</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0</td>
<td>6.71</td>
<td>4.62</td>
<td>4.88</td>
<td>7.46</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>9.55</td>
<td>4.19</td>
<td>4.93</td>
<td>5.32</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.26</td>
<td>5.34</td>
<td>4.85</td>
<td>1.70</td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>8.31</td>
<td>5.32</td>
<td>4.05</td>
<td>3.87</td>
<td>6.16</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>7.38</td>
<td>3.66</td>
<td>4.04</td>
<td>4.46</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3.49</td>
<td>4.40</td>
<td>4.04</td>
<td>1.43</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>7.77</td>
<td>4.80</td>
<td>3.82</td>
<td>3.39</td>
<td>5.55</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>6.72</td>
<td>3.37</td>
<td>3.60</td>
<td>4.03</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3.19</td>
<td>4.01</td>
<td>3.75</td>
<td>1.39</td>
</tr>
</tbody>
</table>

After that, we evaluated various sets of basic algorithm parameters against corruption by artificially generated noise. Once again we computed the corrupted spectra:

\[
\hat{\alpha}^n_c(1) = \hat{\alpha}^n_{rec}(1) + \alpha \cdot e(1)
\]
for various levels $\alpha$ and various realizations of normally distributed random numbers $\epsilon(1)$ and estimated the inner variability of computed vectors $\hat{Q}_C^n$. In Fig.2 you can see a typical result of such an experiment. The standard deviations in inner variability decrease for the small values of $LCONT$ and afterwards increase more or less monotonically. Some of the evaluation results are given in Table II.

Table II

<table>
<thead>
<tr>
<th>NITER</th>
<th>LDIST</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.97</td>
<td>1.16</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.02</td>
<td>0.94</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.36</td>
<td>1.20</td>
<td>1.101</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0</td>
<td>0.90</td>
<td>0.59</td>
<td>0.54</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.85</td>
<td>0.75</td>
<td>0.86</td>
<td>0.97</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.82</td>
<td>0.86</td>
<td>1.12</td>
<td>1.14</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>2</td>
<td>0.68</td>
<td>0.50</td>
<td>0.60</td>
<td>1.08</td>
<td>1.53</td>
<td>1.49</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.75</td>
<td>1.06</td>
<td>0.78</td>
<td>1.12</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0.92</td>
<td>0.91</td>
<td>1.11</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>0.69</td>
<td>0.74</td>
<td>1.12</td>
<td>1.03</td>
<td>1.18</td>
<td>1.01</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.75</td>
<td>0.87</td>
<td>0.86</td>
<td>1.27</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1.03</td>
<td>1.66</td>
<td>1.21</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>

Combining the results from all previous evaluations we selected the most appropriate values for program parameters in our classification experiment. In Figs.3-5 you can see the pseudo-continua for three different stars computed with the "best" values for $LDIST$, $LCONT$, $NITER$. 

104
Figure 3. Measured spectrum and computed pseudo-continuum for HD 206859.

Figure 4. Measured spectrum and computed pseudo-continuum for HD 203504.
5. Conclusion

For the ongoing project of fully automatic quantitative spectral classification system in Tartu Observatory we have developed a rather flexible routine to estimate the continuum or pseudo-continuum in digitally recorded spectra. A large set of numerical experiments has been carried out to evaluate qualitatively and quantitatively the properties of the estimation algorithm. As a result, it was found that we must compromise to meet various demands and to be consistent with various assumptions originating from the theoretical setup of the classification experiment. It seems to us now, that the two-stage processing scheme, where, in the first stage, the parameters of the algorithm are "fine tuned" to fit the particular data set and, in the second stage, the algorithm is used "blindly" for the whole set of spectra, is acceptable both practically and theoretically.
Acknowledgements

The author expresses his gratitude to V. Malyuto, T. Shvelidze and L. Hanni for acting as experts in the qualitative evaluation of computed continua. T. Shvelidze also took part in the debugging process of algorithms.

References


Peat, D.W., Pemberton A.C., 1970 Observatory, 90 141.

Author's address:

J. Pelt
Estonian SSR Academy of Sciences
W. Struve Astrophysical Observatory
Toravere
SU - 202444 Tartu
USSR