Automated Spectrophotometry

G. I. THOMPSON

Royal Observatory, Edinburgh

Starting from a digitized and calibrated stellar spectrum—what can we most usefully do with it?

Traditionally a spectrum consists of two parts, a continuum and some lines. These have been variously defined in the past, so perhaps another definition may be allowed. Over large ranges the two components can be distinguished quite sharply in the spatial frequency plane. The continuum contains

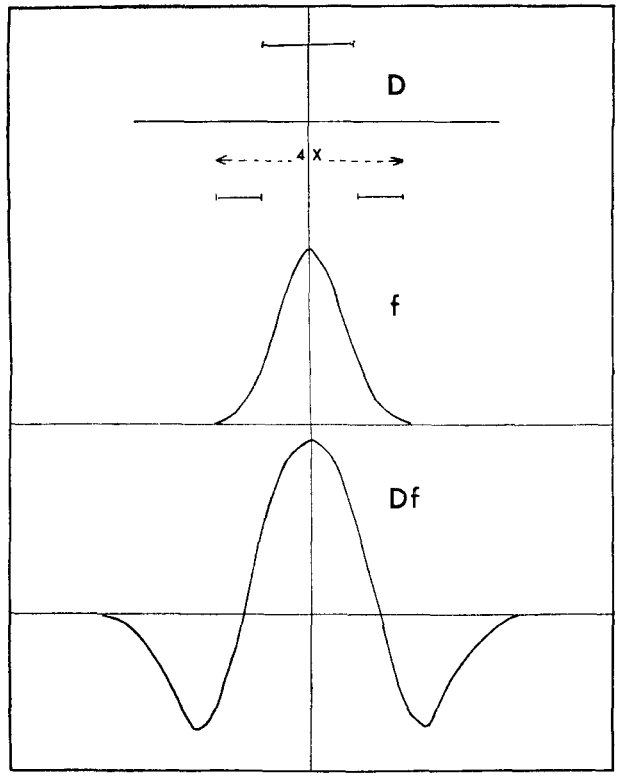

Fig. 1

only low frequencies, whereas in the line the information is contained in the high frequencies. In the word "continuum" is included the instrumental absorption functions, emulsion sensitivity, wide interstellar and atmospheric bands, and even the wide wings of very strong lines. From the word "lines", on the other hand, the strongest lines are excluded; these may contain intermediate spatial frequencies in their core regions. They are relatively infrequent and so may be treated as special cases.

The wide gaps between the characteristic frequencies of line and continuum allow us to separate them fairly effectively. One method is to form the convolution of the spectrum with a function such as that at the top of Figure 1, where the distance 4X is roughly the width of a line. Evidently any constant or linear trend is annulled by such a convolution and any curvature at least reduced by it. A line is transformed into the function at the bottom of the diagram. The effect of such a convolution on a spectrum is shown in Figure 2. The lower curve is a G0V type spectrum at a modest dispersion, 15Å/mm. Most of the continuum curvature is due to vignetting. Hα is just off the diagram and its wings may be contributing to the downward trend, although this is unlikely at G0V. The convoluted spectrum is at the top. All the curvature has disappeared; only a function oscillating about zero is left, while the lines stand up smartly, the least blended showing off their elegant side-lobes.

https://doi.org/10.1017/S0252921100032450 Published online by Cambridge University Press

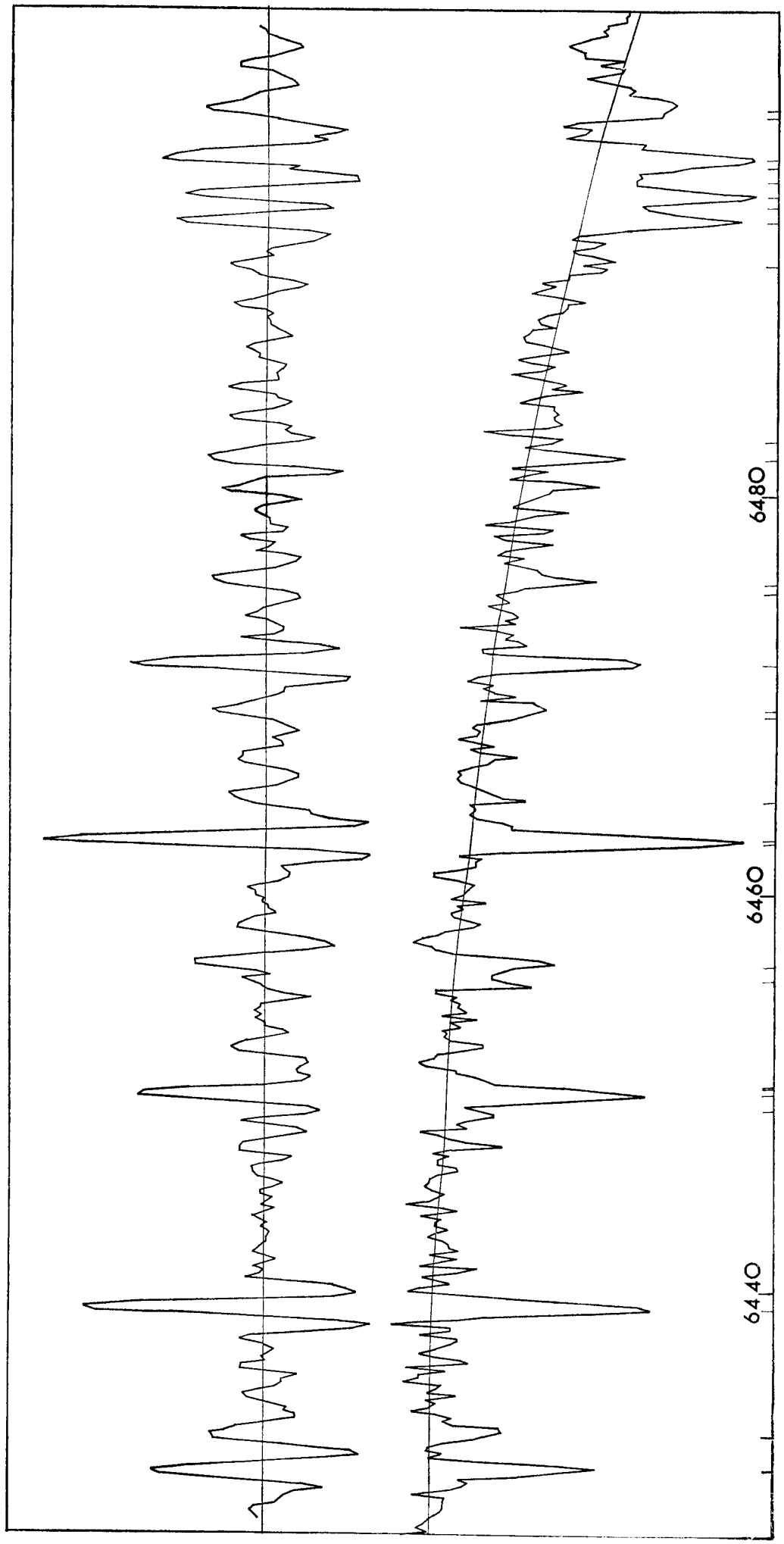

Fig. 2

https://doi.org/10.1017/S0252921100032450 Published online by Cambridge University Press

The first thing which can be done with such a function is to find where the lines are. A line is now defined as a peak sufficiently high above zero. This is much easier to program than finding a dip below a variable background. I have done this now for many spectra and have tested the accuracy achieved by assigning a wavelength to the least blended peaks and setting up a dispersion formula. The residual about this curve is a measure of the accuracy. It turns out to be 2 or 3 μm for moderately strong lines, which is at least as good as a human measurer can do. This process has now been applied to about 100 spectra and sufficient data has accumulated to demonstrate such second order effects as radial velocity stretching and the fact that the accuracy is perceptibly poorer when the lines are wider due to rotational broadening.

The second thing that can be done with the convoluted spectrum is to measure the line strength. The height of a peak above the zero line is proportional to the line strength. The significant thing about this remark is that the strength is found without explicitly determining the position of the low-frequency continuum.

To convert the height to an equivalent width in milliangstroms it is necessary to know the line profile. Here we specialize somewhat further and suppose that all the lines have the same profile. This will be true at moderate dispersion where the profile is dominated by the instrument, or rotational velocity, or both. It is possible to find the mean shape by forming the autocovariance function of the convoluted spectrum. Among other things this contains the autocovariance function of the mean profile. The "other things" include side peaks due to line blending and the autocovariance of the noise. In practice it proves possible to get rid of these "other things" and so the profile is determined. I have found with my 100 spectra that it is possible to detect rotational velocities in the region 0–20 km/sec by this means.

In principle blending introduces no difficulties. At one end of Figure 2 there is a group of apparently three lines and it is evident that the height of the three corresponding peaks in the convoluted spectrum must contain contributions from the other lines. When the profile is known it is possible to give these contributions as a factor times the line strength, and so it is possible to set up linear equations between the height of the peaks and all the line strengths in the blend, which can be solved in the usual way. This can be extended to larger groups, and in practice blends containing 50 lines are encountered and can be coped with.

I have written a program to about this stage of development and used it on 100 spectra of 25 F and G type main sequence stars. This has been in the green and red where line crowding is not serious. The results, compared across the 100 spectra, are reasonably satisfactory. The details are given in a forthcoming "*Publication of the Royal Observatory, Edinburgh*".

The program is specialized to attempt to measure one useful parameter per line. Useful here means that the signal to noise ratio exceeds unity. Future extensions must be to the case where we are allowed to measure more than one useful parameter per line, or are forced to measure fewer than one.

The first is the case of higher resolution, where it should be possible to obtain some information on intrinsic profile for at least some of the lines. Probably the natural thing to do would be to convolute the spectrum with two functions of the same shape as that in Figure 1, one being about as wide as a line and one much narrower. The first measures the total strength, the other the core strength. Their difference and the core strength should be reasonably independent, and if they were both useful must contain information on the intrinsic profile. This is not quite the form in which a theoretician would express his results, but it is up to him to adapt to provide a comparison with observations.

The second extension is not primarily to lower resolution, but to the case of more intense line crowding. I believe my program will deal with any resolution, no matter how low, provided the lines remain sufficiently isolated. Crowding may be present intrinsically, *e.g.* a solar type spectrum in the blue, or induced by a large rotational velocity which widens all lines.

For the crowded line case the difficulty is that there is not sufficient information to give a useful measure for the strength of most lines. If the signal to noise ratio is usually unity or less, what can we usefully measure?

Study of the simplest possible case, two interacting lines, gives a clue. When they are widely separated, it is possible to measure both strengths A_1 and A_2 and also to compute their expected errors $\sigma(A_1)$, $\sigma(A_2)$ due to noise, and it is assumed that both signal to noise ratios, $A_1/\sigma(A_1)$, $A_2/\sigma(A_2)$ are large enough that these measures are useful. As the lines move closer both $\sigma(A_1)$ and $\sigma(A_2)$ start to increase rapidly, and eventually a stage is reached when the S/N ratios decrease to the point where the measures have no value. When the lines exactly superpose the errors become infinite. We should not expect to be able to measure them separately, but we should expect to be able to measure their sum. This is the clue to what can be usefully measured, for it can be shown that the linear equations mentioned

previously, that are set up to solve for the strengths, can be rewritten to solve, not for A_1 and A_2, but for $A_1 + A_2$ and $A_1 - A_2$. Also it transpires that $\sigma(A_1 + A_2)$ remains comparable to $\sigma(A)$ for an isolated line, while $\sigma(A_1 - A_2)$ becomes very large. In other words, when the pair are close together we give up the attempt to measure two parameters, and rather we measure only one, $A_1 + A_2$, while $A_1 - A_2$ has such poor S/N ratio that it is discarded. This can be generalized to the case of many lines. Suppose there is a group of 20, but only sufficient information for one or two parameters with useful S/N ratio. We can estimate some linear combination of their strength, for example, the total strength and the sum of the first ten less the sum of the second ten. Probably the best choice will depend on the distribution of the lines within the blend.

This defines the sort of measurements we can make on crowded spectra. The next question is what to do with the results once found. The sum of two equivalent widths is useless for curve of growth applications, for example. On the other hand the total equivalent width between two given wavelengths is a parameter with a status more reminiscent of a colour than an equivalent width. A colour is a measure of the relative amount of light in two regions and has proved its usefulness in statistical and classification investigations. A summed equivalent width will presumably have to be used in the same way. We must adjust our ideas of why we measure spectra to fit such an approach.

DISCUSSION

D. W. LATHAM: How do you display these large masses of data—as you are processing them with the computer, or do you let the computer make all the decisions as to which lines to measure, when it has encountered a blend, etc.?

G. I. THOMPSON: I have in the past just printed out a list of peaks where the computer finds lines, and correlated these manually to wavelengths, to set up a dispersion curve. This is the only manual interaction, and I'm busily trying at the moment to get rid of this stage as well.

P. J. TREANOR: I should like to stress the importance of these quantitative measures of blended lines. Visual estimates of such blends have played an important part in classical systems of spectral classification (*e.g.* the ratio of the G-band to Hγ), and it is important to put such criteria on a quantitative basis.

J. G. DAVIES: A similar problem in radio astronomy has been tackled by fitting one or more copies of the expected line-profile to the recorded spectrum. The centre frequency, width, and amplitude of each line, and the number of lines used, are varied until the remaining r.m.s. error is minimized. This is a much more general process, and probably applicable to this part of your problem as well.

D. W. LATHAM: I have tried to apply Davies' kind of analysis to high-dispersion spectra and I found that I did not have sufficient independent information elements in the blend to cause the iteration to converge, except very slowly. A typical situation requires a solution for 6 or 9 unknowns from perhaps 15 data points.

G. I. THOMPSON: I do assume that I know what lines are going to be present in the blend.