Chapter 1

Astronomical Classification and Parameterization

"Classification provides order to the impenetrable thicket of forms."


Classification provides a fundamental characterization of a survey data base and forms a vital step before proceeding to any further scientific study or investigation. Astronomical classification is mainly done on the basis of the nature of the spectra of the emitted radiation. The radiation, which spreads over a wide range of wavelengths, ranging from the $\gamma$-ray to the radio wavelength, also undergoes the process of scattering and absorption by the intermediate medium it encounters on its way to the observer. The signatures of these imprints are more enhanced in the ultra-violet (UV) wavelength than in the optical, infrared or other window of the electromagnetic radiation. This turns into a bonanza to probe the interstellar medium (ISM) or intergalactic medium (IGM). The UV ranges of spectral features are also particularly interesting to study the evolutionary properties of galaxies because the UV is
1 Stellar spectral classification

Sensitive to spectral signatures of various contributors such as the warm interstellar medium (ISM) and the hot stellar populations like early O/B stars, Wolf-Rayet stars or hot late type stars [18-20]. This chapter presents a brief general overview on the stellar spectral classification, the galaxy morphology classification and on the parameterization of interstellar extinction. A literature survey is also presented on the work so far done for the automated analysis of spectroscopic and photometric database.

1.1 Stellar spectral classification

Stellar spectral classification has played a major role in the development of modern astronomy. It was Joseph Fraunhofer in Munich who first observed solar and stellar spectra using a glass prism placed in front of a small theodolite telescope in 1814. Later around 1855, Bunsen and Kirchhoff identified these lines as produced by absorption of the light by atoms present in the atmosphere of the Sun and the Earth [21]. These are the two milestones from which the science of astrophysics was born. Starting right from the determination of position of the astronomical objects, to the estimation of their physical characteristics, all the information can be extracted using the spectra of emitted electromagnetic radiation from them.

Attempts have been made to classify stars by looking at the color or at the pattern and width of the spectral lines. However, the classification of Stellar spectra that took place at Harvard College Observatory in Massachusetts, U.S.A, during 1885 turned out to be the most significant one. In this case the spectra was classified on the basis of the absorption lines and the resulting classification is known as the Henry Draper Catalogue, in memory of the U.
S. astronomer Henry Draper. With the understanding of the fact that the absorption process, and hence the appearance of the lines in the spectra, is governed by the temperature of the source, this catalogue actually turned out to be a one-dimensional temperature based classification scheme [21].

The modern classification system in use is the MKK system or (updated) MK system. The system was introduced in 1943 by (and is named after the initials of) William Morgan, Phillip Keenan and Edith Kellman of Yerkes Observatory. It is also known as the Yerkes spectral classification, which is basically an extension of the Harvard classification system, with the addition of luminosity as the second dimension to the classification scheme. The luminosity classes are indicated by the Roman numerals 0, I, II, III, IV, V, VI, and VII representing respectively the Hypergiants, Supergiants, Bright giants, Normal giants, Sub giants, Dwarfs (main sequence), sub dwarfs, and white dwarfs. The classification is defined by a set of standard stars and is based on the visual appearance of the spectra.

1.1.1 Physical basis of Stellar Spectral Classification

A star emits radiation ranging from the Ultra-Violet (UV) to the infrared (IR) wavelength of the electromagnetic spectrum. The source of radiation is the self-sustained nuclear fusion at the hot interior (~1Kev) that keeps the star balanced against the gravitational collapse. At such high temperatures the core of the star is composed of nearly completely ionized plasma. As one moves in a radially outward direction, the temperature decreases and the plasma becomes partially ionized towards the outer edge the star. The radiation generated by the nuclear reaction is transported along the temperature gradient before finally getting emitted to the atmosphere from the photosphere, which
Stellar spectral classification

is a surface at an optical depth of $\tau = 2/3$. Depending on the local conditions
prevailing in different layers, the radiation encounters electron scattering, and
free-free, bound-free absorptions by ions. However, in exceptional cases of cool
stars the radiation emitted may also contain the contribution from the chromo-
spheric and the corona region of the star. In any case, the local thermodynamic
equilibrium (LTE) is a reasonable assumption at the surface of the star, the
surface behaving like a perfect black body. Hence the energy distribution of
the radiation can be approximated by blackbody spectrum characterized by
the surface temperature, $T_s$, superimposed by different absorption and emis-
sion lines. The primary source of information about the stars is the analysis
of these lines in the spectrum [22].

The observed spectrum of a stellar object depends mainly on the tempera-
ture, and to a lesser extent on the pressure, chemical abundances and on the
intervening medium between source and observer. The one that is most influen-
tial in determining the shape of the spectral feature is the pressure broadening
of the spectral lines. For most of the normal stars the average density lies
approximately in the range between $10^{-8}$ to $10^6$ times that of the Sun, with
the smaller stars having higher densities [23]. Although the mean density is
not the only factor which determines the pressure at the outer layer but it is
the major factor. The smaller the star the greater is the pressure at outer layer
giving rise to high surface gravity. When the pressure is higher (as in a giant
star), the individual atoms are packed more closely. The electronic properties
of atoms will affect each neighbor’s internal energy levels. The atoms will be
able to absorb in a wider range of wavelengths about the mean wavelength.
In a small star, like the Sun, the gas pressure must be higher to maintain
balance against gravity. Here the atoms are even more closely packed and
neighboring atom's electronic properties are even more affected. Hence the absorption features will appear even wider. The second dimension of the MK system is the pressure or the surface gravity factor that accounts for appearance of the absorption lines. The narrowest line spectra are designated as 0 (hyper giant) through I to VII. For stars of a given temperature, narrow lines correlate with low-pressure atmospheres, large stellar radii, and hence a high luminosity. Because this line-width classification correlates with luminosity, the designation I to VII is termed as the luminosity class. The impact of the intervening medium on the spectrum will be discussed separately in section §1.3.

1.2 Galaxy classification

Galaxies are massive systems of stars, gas, dust and dark matter bound together gravitationally as a single physical entity. Until the early 20th century, it was believed that the Milky Way was the only such structure existing in the universe. The identification of other galaxies, or independent stellar systems, goes back to 1924, when Edwin Hubble found Cepheid variables in the nearby galaxies Messier 33 (or M33 also known as Pinwheel Galaxy) and Messier 31 (or M31 also known as Andromeda Galaxy). Using the period-luminosity relationship of Cepheid variable stars, Hubble determined the distance to M31 to be around 750 kpc and that it had a diameter larger than that of the Milky Way. The numbers established that these were objects outside the confines of the Milky Way. Later more thorough studies showed that besides Milky Way, there actually exists a large number of such galaxies in the Universe. When looked through the optical band, galaxies are found to have two components,
a central bulge and a disk component, although some galaxies may have one component only. These galaxies exhibit a wider range of properties, from giant galaxies of $10^{13} M_\odot$ and size over 150,000 pc in diameter to dwarf galaxies of $10^6 M_\odot$ and about 1,000 pc in size.

The galaxy classification was first attempted by Sir William Herschel, and his son, Sir John Herschel. The most common classification scheme in use today is the Hubble Classification scheme by Edwin Hubble who realized that the vast majority of galaxies have only a small number of shapes. A classification based on their optical appearance or morphology is limited to three fundamental types of galaxies - ellipticals (E), spirals (S), and irregulars (Irr). In Hubble classification the first two types are distinguished by rotational symmetry about dominating non-stellar nuclei, and the latter lacks the trait [24]. Ellipticals appear very smooth and spherical, and the luminosity fades smoothly from bright nuclei to indefinite edges. Depending on the ellipticity, elliptical galaxies are farther classified as $E0, E1, ..., E7$ etc, where the notation $E_n$ is the ellipticity calculated as the ratio of the semi-major axis to the semi-minor axis, $(b/a) = (1 - n/10)$. Hubble defined three criteria to distinguish among spirals: (a) relative size of the unresolved nuclear region, (b) extent to which the arms are unwound, and (c) degree of resolution in the arms. Galaxies whose nuclear region is very extended, arms are tightly wound and unresolved are termed as the Sa galaxies. The intermediate group, denoted as Sb, have large central regions and thin open arms, or a smaller nuclear region and closely coiled arms. Finally, galaxies with inconspicuous nuclei and highly resolved, patchy, widely opened arms are termed as the Sc galaxies. Hubble classification of galaxies can be conveniently represented in a sequence bearing the structure of a tuning fork, known as the Hubble tuning.
1.2 Galaxy classification

Figure 1.1: Galaxies of different morphological structures arranged in Hubble sequence. Taken from http://cas.sdss.org/dr6/en/proj/advanced/galaxies/images/TuningFork.jpg

fork sequence (see fig 1.1). Later on a class of galaxies, roughly intermediate between ellipticals and early-type Sa spirals, was introduced in this classification scheme. They are known as the lenticular galaxies S0s. S0 galaxies lack spiral arms and has a low-surface brightness. Although originally based solely on optical appearance, however, the appearance is also closely correlated with other physical properties of galaxies. For example, it is found that as one moves along the sequence E-Sa-Sb-Sc-Ir, the integrated color and the spectral types of the galaxies increase monotonically [25–27]. As a matter fact, this led Hubble to the conclusion that the sequence represents the evolutionary history of galaxies starting at the left end of the diagram (fig 1.1) and evolving to the right. However, it is now established that there is no physical basis for such belief of evolution in the galaxy morphology. Besides the Hubble scheme, there
also exists some other schemes proposed by different groups later on. Some examples of such schemes are schemes as proposed by de Vaucouleurs [26], Elmegreen, Van den Bergh [28–30], and Morgan [31, 32].

1.3 Extinction and scattering

The regions between the stars in the Galaxy is not empty but filled with gas, dust, magnetic field and charged particles. They are collectively known as the interstellar medium (ISM). The role of interstellar matter in astronomy is threefold. First, it provides the source of materials for the formation of new stars. The material, in turn is partially replenished by mass loss from aging stars. Second, direct observation of interstellar matter provides information about the Galaxy that complements evidence obtained from studying stars. Finally, the presence of interstellar material reddens the light by dramatically affecting the observation along the line of sight to an object.

About 98% of the ISM is in the form of gas, mainly consisting of neutral hydrogen atom (HI) and helium (HeI) atom. However, in the dense regions, H$_2$, CO, CN gas exist in the form of molecular clouds. The neutral gas is usually cold and the electron generally stays in its ground state. However, protons and electrons can have spin. These spins produce magnetic field such that when the spins are aligned the ground state has a slightly higher energy than when the spins are opposed. This is called hyperfine splitting of the ground state of hydrogen and results in the emission of 21 cm radio waves from HI clouds. On the other hand the molecular gas emits by rotation rather than direct transition of their electrons. The radiation emitted tends to be in the micro-wave band.
Interstellar dust is mostly composed of carbon and silicate and are produced in the envelopes around the red super giant stars. Stellar winds and the planetary nebula phase throw these elements into the ISM. They comprise about 1% of the total volume of the ISM. Yet it effects the radiation coming from the astronomical objects substantially. The interaction mainly takes place by the processes of dust grain scatter and absorbed radiation. The radiation scattered by the dusty object can propagate in any direction and has the same wavelength as the incident one. The fraction may be very small but it never gets wiped out. On the other hand, the radiation absorbed by a dust grain gets transformed into its thermal energy and then gets re-radiated at some wavelength longer than the absorbed wavelength. The observational manifestations of the dust-radiation interaction are extinction and polarization.

Suppose, light coming from a distant astronomical object is intervened by a dust cloud. Considering the dust geometry to be spherical of radius $a$, distributed with number density $n_d$ per unit volume in a cylindrical column of length $L$ and unit cross-sectional area along the line of sight from the object, the reduction in the intensity of the source, due to extinction, while passing...
through an elemental column of length $dL$ is

$$\frac{dI}{I} = -n_dC_{ext}dL$$

(1.1)

where $C_{ext}$ is the extinction cross-section. Integrating equation 1.1 over the entire path-length,

$$I = I_0 \exp(-\tau)$$

(1.2)

Here $I_0$ is the intensity at $L = 0$ and

$$\tau = \int_0^L n_dC_{ext}dL = N_dC_{ext}$$

(1.3)

is the optical depth of extinction.

The extinction, corresponding to the attenuation of the intensity of radiation of the distant object on passing through the foreground dust cloud, contains the contributions from both the scattering and the absorption process. So, $Extinction = scattering + absorption$ and the extinction efficiency is equal to the sum of corresponding efficiency factors for scattering and absorption. The scattering efficiency parameter, $Q_{sca} = C_{sca}/\pi a^2$, and absorption efficiency parameter, $Q_{abs} = C_{abs}/\pi a^2$, are function of two quantities, a dimensionless size parameter $X = 2\pi a/\lambda$ and a composition parameter which is the complex refractive index, $m = n - ik$, of the grain materials. The quantities $n, k$ are called the optical constants and are functions of wavelength. Extinction, thus, can potentially reveal information about the composition and size distribution of the grains. Moreover, the optical properties of interstellar dust.
1.3 Extinction and scattering

vary along different directions. Such variation may be produced by a number of physical processes, such as grain growth by coagulation, size-dependent destruction and selective acceleration of small grains by radiation pressure in stellar winds etc. The variation along two different stars are more clearly exhibited at ultraviolet wavelength. Thus, the spatial variation of the extinction in UV may also reveal the degree and nature of dust grain processing occurring in the ISM [34].

The observational technique to measure the extinction is the pairing of stars of identical spectral type and luminosity class but having unequal reddening, and comparing their energy distributions.

Let us consider one such pair of stars \((a_1, a_2)\) situated at distance, of \(d_1\) and \(d_2\) respectively, expressed in parsec unit. Let \(m_{1\lambda}\) be the apparent monochromatic magnitude of the non-reddened star with absolute magnitude \(M_{1\lambda}\). Then,

\[
m_{1\lambda} = M_{1\lambda} + 5\log d_1 - 5
\]

Similarly, if \(m_{2\lambda}\) is the apparent monochromatic magnitude of the reddened star with absolute magnitude \(M_{2\lambda}\), then, the light from this star will be dimmed at each wavelength by interstellar matter by a factor \(A_\lambda\),

\[
A_\lambda = -2.5 \log \left( \frac{I}{I_0} \right)
\]

so that,

\[
m_{2\lambda} = M_{2\lambda} + 5\log d_2 - 5 + A_\lambda
\]
1.3 Extinction and scattering

Subtracting equation 1.4 from 1.6, we get

\[ A_\lambda = m_{2\lambda} - m_{1\lambda} + 5\log \left( \frac{d_2}{d_1} \right) \]

\[ = \Delta m_\lambda + 5\log \left( \frac{d_2}{d_1} \right) \quad (1.7) \]

Here \( M_{1\lambda} = M_{2\lambda} \), both stars \( a_1 \) and \( a_2 \) being of same spectral type and luminosity. Thus one can get \( A_\lambda \) by subtracting the monochromatic magnitudes of the two stars, provided the ratio of their distances is known. Hence, the quantity \( \Delta m_\lambda \) may be used to represent the extinction parameter \( A_\lambda \).

Equation 1.7 can be normalized to find a universal curve of \( A_\lambda \) versus \( \lambda \), irrespective of the couple of stars chosen. One can write a general expression by eliminating the constant term in the equation, using two standard wavelengths \( \lambda_1 \) and \( \lambda_2 \), such as

\[ E_{\text{norm}} = \frac{A_\lambda - A_{\lambda_2}}{A_{\lambda_1} - A_{\lambda_2}} = \frac{\Delta m_\lambda - \Delta m_{\lambda_2}}{\Delta m_{\lambda_1} - \Delta m_{\lambda_2}} = \frac{E(\lambda - \lambda_2)}{E(\lambda_1 - \lambda_2)} \quad (1.8) \]

where \( E(\lambda_1 - \lambda_2) \) is the color excess of the star. This normalization is done using the fact that the ratio, \( R_\nu = A_V/E(\lambda_B - \lambda_V) \), between interstellar absorption at the effective wavelength of the \( V \) filter, \( A_V \), to the color excess \( E(\lambda_B - \lambda_V) \) of the star, is nearly constant [35]. The value of \( R_\nu \) ranges between 2.2 and 5.8 for line of sight along which UV extinction has been measured. It has a mean value of 3.1 for diffuse interstellar medium [34]. Here \( \lambda_B \) and \( \lambda_V \) correspond to the \( B \) and \( V \) pass bands in the Johnson filter system (see section §3.2 for further information on Johnson system) so that equation 1.8 takes the
Thus equation 1.9 can be used to deduce the absolute extinction $A_\lambda/A_V$. It is to be noted that in deriving equation 1.9, we have assumed the magnitudes at the monochromatic wavelength in equations 1.4 and 1.6. However, in practical
1.3 Extinction and scattering

cases these magnitudes have to be replaced by the magnitudes at the effective wavelengths of the filter system.

Figure 1.3 shows the mean Galactic interstellar extinction curve in UV due to Seaton [36]. The curve plots $E(\lambda - \lambda_B)/E(\lambda_B - \lambda_V)$ vs the wavelengths in the range from 1200Å to 3200Å. The plot adopts a value of 3.1 for the mean ratio $R = A_V/E_{B-V}$. The absolute value of $R$ is calculated as the value of the ordinate at which $1/\lambda$ tends to zero [37]. The most remarkable aspect of this extinction curve is a broad, roughly symmetric peak in the mid-ultraviolet centered at 2175Å. The full width at half maximum (FWHM) of the peak is about 480Å [38]. This is known as the "2175Å bump" which was first discovered in 1960 [39, 40]. Seaton’s mean extinction law is commonly used to correct the fall in the observed flux (eqn1.5) for the presence of the Galactic dust:

$$ I = I_0(\lambda)10^{-0.4A_\lambda} $$  (1.10)

where, $A_\lambda = E(\lambda - \lambda_V) = \left[\frac{E(\lambda - \lambda_V)}{E(\lambda_B - \lambda_V)}\right]E(\lambda_B - \lambda_V)$.

The observational analysis of the extinction provides us with “in depth” knowledge, while polarization provides the “in breadth” information [41]. Polarization reveals the distribution of the dust and their alignment in the Galaxy. It should be mentioned that the directional starlight polarization depends on the orientation of the plane-of-sky component of the interstellar magnetic field, $B_{\perp}$. Thus the data of the polarimetric surveys along with other observations like Zeeman splitting of 21 cm neutral hydrogen line can reveal the magnetic field structure at different scales. However, it is also to be noted that the polarization features have been discovered in only a few directions in the Galaxy,
Although the 2175Å bump is a common attribute to all extinction curves [41].

To summarize, the study of the interstellar extinction and polarization, together with the constraints from the cosmic abundances, allows one to estimate the chemical composition, size and shape of dust grains. The observational quantities needed for these information are the intensity, color and polarization of the scattered light. Additionally the orientation angle of the interstellar linear polarization may also provide the projected direction of alignment, i.e. the projected direction of the interstellar magnetic field $B_\perp$.

### 1.4 Automatized astronomical classification and analysis

The methods of automated spectral classification may be divided as, (a) criteria evaluation by quantitative measurements, (b) pattern recognition and cross-correlation methods, and (c) mixed procedure. The techniques have been used by different groups to automate the astronomical classification and analysis over the years.

The pattern recognition and cross-correlation method have been first successfully used by Kurutz [42]. He implemented the method for the analysis of the MK classification of low (14Å) resolution spectra extended over 3500-7500 Å [43]. Gulati et al. [6] was one of the first groups to apply ANN for the stellar classification. They classified digitized databases of optical and UV spectra by using conventional metric distance minimization methods and Artificial Neural Networks. The group has also determined the UV reddening for the O and B type stars using ANN [13]. Two other contemporary groups working in this application were Hippel et al. [7] and Weaver et al. [9]. Hippel et al. used a
two-layer network to classify stars in the range B3 to M4 with the optical spectra (3800 - 5150Å). Weaver et al. classified A-type stars in the infra-red (5800 - 8900 Å) at 15Å resolution using the equivalent width of absorption lines or the full spectra as the ANN input in two different attempts. They also determined reddening for A type stars to an accuracy of 0.05 in $E(B - V)$. Coryn Bailer-Jones used ANN and principal component analysis (PCA) for analyzing stellar spectra [44] in the wavelength range of 3800 - 5200 Å and published a catalog of spectral indices of 684 stars based on the spectra observed with the coudé feed instrument at the KPNO. Vieira et al [4] classified IUE spectra in UV (150 - 3200Å) range using a very high complex neural network architecture. However, they omitted the spectral range 1950 - 2350Å because of the low signal to noise ratio (S/N) in the region.

In galactic astronomy, Odewahn et al. [45] pioneered the use of ANN based scheme for automated segregation of stars and galaxies based on point-spread function (PSF) fitting. Mahonen et al.[46] used SOM (self organizing maps) based neural network using the CCD images directly. Mähönen et al [47] also introduced another method based on fuzzy set reasoning. Philip et al (2002) used the difference boosting neural network for the star-galaxy classification of NOAO Deep Wide Field Survey (NDWFS) Qin et al (2003) also demonstrated the use of spectra for the same purpose using RBF (Radial Basis Function) neural network, in the wavelength range of 3800-7420Å.

From the above discussion it is clear that most of the attempts towards the automatized classification and analysis of the astronomical data are based on spectroscopic data in the optical or infrared region. Some of the reported works are in the ultra violet region, however, these are for selected star types only. Therefore there is a need for further work to incorporate other stellar
types. Moreover, the use of the band integrated data in ANN, for classification and analysis purpose, has not yet been reported, this is particularly relevant since such data is expected from the upcoming Indo-Israeli satellite mission TAUVEX.

Some space missions also monitor the sky to look for new transient events. Temporal analysis of such events provide information on the size and geometry of the emitting region. Establishing that two lightcurves measured in two different energy bands are correlated with each other is an important temporal analysis for various kinds of astrophysical sources. The detection and measurement of the level of correlation constrains the number of radiative processes active in the source and can be used to validate (or rule out) physical models based on spectral analysis. Phase and time-lags detected for correlated light-curves can provide further insight into the geometry and size of the emitting region [48]. Often in these applications, the light curves available for analysis are of short length and have measurement errors. The true temporal behaviour of a source can only be established if there are robust estimates of the errors on the cross-correlation, phase and time-lags [49]. The standard method to estimate the error on the cross-correlation involves dividing the light curves into several equal segments and finding the cross-correlation for each. Then the net cross-correlation is given by the average of the different segments and the variance is quoted as an error [50]. For example, this technique is implemented by the function “crosscor” of the high energy astrophysics software HEASOFT*. The method is reliable only if the light curves can be divided into a large number of segments (>> 10) and each segment is sufficiently long and not dominated by measurement errors. The temporal behaviour of many

*http://heasarc.gsfc.nasa.gov/docs/software/lheasoft/
astrophysical systems depends on the time-scales of the analysis and hence by using this method, one loses information on the behaviour of the system on time-scales comparable to the length of the original data. Moreover, there does not seem to be any established way by which this method can be extended to get an estimate of the time-lag between the light curves and its error.

These deficiencies can be overcome by using a Monte Carlo technique where one simulates a large number of pairs of light curves having the same assumed temporal properties and with the same measurement errors as the original pair [51]. The results of the original pair can be compared with the simulated ones to ascertain the confidence level of the cross-correlation and time-lag. The simulated light curves should take into account the stochastic fluctuations of the light curves and not just the measurements errors. Indeed, when the light curve is sampled unevenly and with measurement errors changing in time, the Monte Carlo technique may be the only way to obtain reliable estimates [52]. However, the Monte Carlo technique is numerically expensive and hence not practical for analysis of a large set of data. Moreover, the results depend on the subjectivity of the assumed temporal properties of the system. For example, to ascertain the errors on an observed cross-correlation and time-lag value, the simulations are generally done with the assumption that these are the true intrinsic values. Similar assumptions have to be made on the shape of the power spectra of the light curves. In the literature, there is an analytical estimate for the cross-correlation known as Bartlett's equation [53]. This method is available in the "crosscorrelation" function in the IMSL numerical libraries\(^\text{1}\). The error as measured by this method is accurate only when the complete knowledge of the cross-correlation and auto-correlation functions is

\(^{1}\text{http://www.vni.com} \)
available. The effectiveness of the method for short duration light curves is not certain. Moreover, the error estimate does not naturally translate into error estimates for the phase and time lag between the light curves. Because of these reasons Bartlett’s equation is not often used in astronomical contexts.

In the following chapters of this thesis, we have gradually developed ANN based automated scheme for classification and parametrization of large scale data set. In the penultimate chapter, we have developed an analytical expression of error for the temporal analysis of short duration light curves from astrophysical sources.