Fundamental Plane Relations

What the Tully-Fisher relation is for spirals, the Fundamental Plane relation is for ellipticals. Like Tully-Fisher, the technique is a relation between a distant-dependent quantity of the galaxy (luminosity or size), and a distance-independent quantity (the velocity of the galaxy’s stars). But there are a few differences.

If you start with the virial theorem

\[ \frac{M}{R} \propto v^2 \]  \hspace{1cm} (10.01)

assume that all galaxies have the same mass-to-light ratio (\( \mathcal{L} \propto M \)) and assume that all galaxies have the same surface brightness, \( \Sigma = \mathcal{L}/R^2 \), then it is easy to show that the

\[ \mathcal{L} \propto v^4 \]  \hspace{1cm} (10.02)

where \( v \) is the motion of the particles within the galaxy. In the case of spirals, (10.02) is called the Tully-Fisher relation, and \( v \) is the velocity of the gas in the galaxy’s disk. In the case of elliptical galaxies, (10.02) is called the Faber-Jackson relation, and \( v \) is actually \( \sigma \), the velocity dispersion of the stars close to the nucleus (where a high signal-to-noise spectrum is obtainable).

Of course, all galaxies do not have the same surface brightness. So if you take \( \Sigma = \mathcal{L}/R^2 \) and substitute that into the virial theorem (keeping the mass-to-light ratio assumption), then a relation between luminosity, surface brightness, and velocity dispersion is the natural consequence, i.e.,

\[ \mathcal{L} \propto \sigma^4 \Sigma^{-1} \]  \hspace{1cm} (10.03)
Observations of real ellipticals show that when $L$ is the total $B$-band luminosity of the galaxy, $\sigma$ is the central velocity dispersion, and $\Sigma$ is the galaxy surface brightness measured at a radius that contains 1/2 the total light from the galaxy ($R_e$), the coefficients in (10.03) are not 4 and $-1$, but 2.7 and $-0.7$. This is the elliptical galaxy fundamental plane relation.

Note that the elliptical galaxy fundamental plane has some similarities to the HR diagram’s luminosity-temperature-period relation. Recall that we described the locations of stars on the HR diagram as positions in a luminosity-temperature-mass 3-space. Stars were not distributed randomly in this 3-space, but inhabited specific regions: the main sequence is a line in 3-space, while the giant branch and white dwarf cooling regions are planes. Similarly, elliptical galaxies are not distributed randomly in luminosity-surface brightness-velocity dispersion space; instead, they are confined to a plane – the fundamental plane.

There is one other similarity with the HR diagram. As we learned when we studied Cepheids, pulsation period could be substituted in for one of the axes of the HR diagram 3-space to form a new 3-space. Thus, we created a period-luminosity-temperature relation. Similarly, the fundamental plane has other projections. There is a relationship between the luminosity of a galaxy and its mean metallicity. (Large galaxies are more metal-rich.) So the strength of the magnesium absorption line ($\text{Mg}_2$) can replace one of the axes. Or, since stellar line-blanketing depends on metallicity, galaxy color can be substituted in. In fact, there are a plethora of elliptical galaxy properties that correlate with each other (luminosity, metallicity, velocity dispersion, color, amount of UV (shortward of 1500 Å) light, specific number of planetary nebulae, etc.). A fundamental plane exists for each of these parameters.
Note, however, that (10.03) is not always used directly. Dimensionally, \( \mathcal{L}/\Sigma^2 \) has the units of size, squared. Thus, one group (called the Seven Samurai) combined these two variables, and took a square root to create \( D_n \), which is a characteristic size. Specifically, their relation is

\[
D_n \propto \sigma^{1.2} \tag{10.04}
\]

where \( D_n \) is the diameter defined by a circular aperture centered on the galaxy that encloses a mean blue surface brightness of 20.75 mag per square arcsec. This is called the \( D_n-\sigma \) relation. What it is, in effect, is the optimal projection of the \( \mathcal{L}-\Sigma \) plane into a single dimension.

The fundamental plane has quite a large amount of scatter, up to \( \sim 20\% \). In part, this may be caused by the reliance on central velocity dispersion, which may be affected by local conditions (say, a central black hole or non-isotropic orbits). However, the relation does have one big advantage over Tully-Fisher. Elliptical galaxies are almost always found in clusters. Thus, you don’t have to make any assumptions about the space density of these objects, or whether your magnitude-limited sample is biased. Since all the ellipticals really are at one distance, objects aren’t continually being scattered into and out of your sample. Moreover, since the ellipticals are together, a rich cluster will have many objects, so you can improve your distance estimate by \( \sqrt{N} \).
Brightest Cluster Galaxies

The use of first-ranked galaxies as distance indicators goes back at least as far as the 1950’s. The premise of the technique is that the brightest galaxy in all rich clusters is a standard candle. If true, the technique has the obvious advantage that, since you are observing the brightest galaxies in the universe, you can make measurements at extreme distances. However, it does have a few technical and scientific drawbacks.

The first deals with a specific type of bias called the Scott effect. As you look out farther into space, the volume of your survey grows. (Under simple Euclidean geometry, it would grow as $4\pi r^2 dr$.) Thus, there is more of a chance of finding an exceptionally rich cluster at large distance than at small distance. Since a rich cluster will have more galaxies than a poor cluster, this means that there is more of a chance of a galaxy populating the bright-end tail of a luminosity function. Thus, at large distances, you are more likely to find a truly exceptionally bright galaxy. This can bias your distance measurements.

Another technical difficulty involved in observing BCG’s at high redshift is the problem of measuring the *total* luminosity of a galaxy. There is no hard limit to the edge of a galaxy, and normally, one only observes light within a specific aperture. Thus, one measures an *isophotal* magnitude, *i.e.*, the light contained within a certain aperture (say, out to where the galaxy surface brightness becomes fainter than 26 mag per sq. arcsec). However, at high-redshift, cosmological surface brightness dimming will cause you to measure a smaller fraction of the galaxy. You must therefore try to correct and/or extrapolate the galaxy profile to get a *total* magnitude for the galaxy.
Another problem which occurs when observing at high redshift is that the rest wavelength of the observing bandpass has shifted. Consider an observation in the $V$ bandpass, which is centered at $\sim 5500$ Å and is roughly 800 Å, from 5100 Å to 5900 Å. When used on a galaxy located at $z = 1$, your detector will be observing the galaxy’s flux from 2550 Å to 2950 Å. Thus, not only are you observing the galaxy’s UV light, instead of its optical light, but your bandpass has become narrower by a factor of $1 + z$. In order to correct for this effect, you must either use a different filter matched to the redshift of the galaxy, or you must model the flux distribution of the galaxy, and predict the correction you must apply based on the galaxy’s color. (By color, we mean the magnitude difference between the galaxy as seen in the observer’s frame and the galaxy as it emitted in the rest frame.) This term is called the K-correction. It obviously depends on the type of galaxy being observed (star-forming spirals will have different K-corrections than old-star ellipticals).

Finally, when looking out to large redshift, you have to deal with the effects of stellar evolution. Galaxies at high redshift should appear brighter than similar galaxies today, since their main sequences will extend to brighter stars (as will their giant branch).

Nevertheless, when observations are made to small ($z < 0.05$) redshift, it appears that BCG’s are reasonable standard candles ($\sigma \sim 0.25$ mag). But there still may be a Scott effect.
**Supernovae**

The enormous brightness of supernovae make them tempting as standard candles. However, it has only been without the last couple of years that their potential has been realized.

The spectra of supernovae fall into many categories (see below), but physically there are only two types, Type Ia and Type II. Both can be used for distance purposes, but the analysis of Type Ia is much simpler (and believeable).

**SUPERNova CLASSIFICATION**

**MAXIMUM Light SPECTRA**

```
SN II

Light Curve Shape
Maximum Light Continuum

II L  II P  SN 1987A  SN 1987K

SN I

Si / no Si

Ia  He / no He  Ib  Ic
```

**LATE SPECTRA ~ 6 MONTHS (SUPERNEBULAR)**

```
SN II

O / H

(H, O, Ca)

SN I

O / no O

SN 1987K  Ib, c  Ia

(H, Ca)  (O, Ca)  (Fe, Co)
```
Type II Supernovae: Expanding Photosphere Method

Type II (and Types Ib and Ic) supernovae are seen only in star-forming galaxies, and are thus thought to come from massive stars which have evolved to iron in their core. When iron is produced, there is no more energy to sustain the star: the core collapses, and the star explodes. This explosion produces a variety of species (including large amounts of the light CNO elements), but also a large amount of Ni$^{56}$. Ni$^{56}$, however, is radioactive, and it very quickly beta-decays to Co$^{56}$ (half-life of 6.1 days) and then to Fe$^{56}$ (half-life 77.12 days). The electrons which are emitted during these decays carry away a lot of energy (1.72 MeV for Co$^{56}$; 3.58 MeV for Fe$^{56}$). These electrons collide with the surrounding material and heat it up. It is this energy input that is responsible, in large part, for the supernova light curve. The differences in the observed spectra presumably have to do with the different compositions of the stars’ atmospheres when they explode, and the energy input from other radioactive species. (For instance, some stars may have lost all the hydrogen in their photosphere due to mass loss on the giant branch.)

Because Type II supernovae come from stars with very different masses, they are not standard candles. However, one can derive their distance using the Expanding Photosphere Method (EPM). (This is just a fancy name for Baade-Wesselink.) Note, however, that figuring out how the color-temperature of a 10,000 km s$^{-1}$ expanding photosphere relates to the velocity of the observed absorption lines is extremely non-trivial. But at least, for Type II objects, hydrogen dominates the spectra, so you can make a first order approximation that the atmosphere is pure hydrogen and neglect radiative transfer in all the millions of individual metal absorption lines.
Type Ia Supernovae as Standard Candles

Type Ia supernovae are found mostly in star-forming galaxies, but also appear in ellipticals and other old stellar populations. Because these objects show no hydrogen absorption in their spectrum, the generally (but not universally) accepted model for these objects is that they come from carbon/oxygen white dwarfs, which, via accretion or coalescence with another white dwarf, get bumped over the Chandrasekhar limit. In the catastrophic collapse which follows, virtually the entire star is changed to \(^{56}\text{Ni}\) which then decays to \(^{56}\text{Fe}\) to produce a characteristic SN Ia light curve.

If one plots apparent supernova magnitude at maximum vs. host galaxy redshift, one sees that there is only a bit of scatter (~0.36 mag) about the line. This is consistent with the hypothesis that all Type Ia are identical, i.e., they are a standard candle. (Such an interpretation fits in with the idea that SN Ia all come from objects which have just gone over the Chandrasekhar limit.) Under the standard candle hypothesis, the scatter in the Hubble diagram is all due to observational errors, and all one needs to do is calibrate the relation to obtain the distance to a single, nearby galaxy that has hosted a Type Ia event. For a long time, this one event was supernova 1937C, which occurred in the nearby irregular galaxy IC 4182, and was measured by Baade and Zwicky. This calibration produced a Hubble Constant that was near \(H_0 = 50\) km s\(^{-1}\).
In the early 1990’s, there was a great increase in the number of supernovae with well observed light curves. With these data, it became apparent that not all Type Ia supernovae are identical: there is a relation between SN Ia’s absolute luminosity at maximum, and the time it takes the supernova to fade. (Recall that a similar type of relation exists for novae. In the supernova case, however, bright SN fade slower.) With this improvement, SN Ia has become an excellent standard candle, which can be seen out to $z \sim 1$. Because SN Ia are rare, not many have gone off in galaxies with known distances. Thus, their zero-point calibration remains weak. However, many, many distant supernovae are being found each year, and through these, it is possible to look for deviations in the Hubble Constant and measure $q_0$.

Note that EPM (or Baade-Wesselink) is virtually impossible to use with SN Ia, because, by definition, they do not contain hydrogen. Thus, the radiative transfer calculation in the atmosphere must keep track of many, many atomic transitions.