Spike Hunting in Galaxy Spectra

Pierpaolo Brutti, Christopher R. Genovese, Christopher J. Miller,
Robert C. Nichol, and Larry Wasserman

Abstract

In this paper we summarize analysis of galaxy spectra from the Sloan Digital Sky Survey. To each spectrum we fit non–parametrically a continuum, that is a smooth underlying trend, and a parametric "island" of spike profiles at each group (opportunely defined) of adjacent absorption or emission line locations, having, as ultimate inferential goal, the construction of reliable confidence intervals for the signed areas between each spike and the continuum. The implemented semi–parametric estimation procedure based on sieved penalized splines, has proved to be effective, computationally efficient and easy to implement.

1 Introduction.

In order to study, compare and classify stars and galaxies, astronomers need to analyze their spectra; that is, $x - y$ graphs with *wavelength* on the x–axis measured in *Ångstroms* (1 Å$= 1 \times 10^{-10}$ meters) and a measure of brightness called $flux$ (derived from binned photons counts) on the y–axis. A typical spectrum has a lot of "peaks" and "valleys" mostly identifiable as symbols of chemical elements. Each star, for example, has a different set of peaks and valleys that can be used to divide the stars into different *spectral types*. More precisely, stars give off most of their light in a certain color depending on their temperature. However, not all the light that stars give off reaches us. Depending on the star's temperature, some is absorbed by the gas in its atmosphere generating a "valley" in a particular zone of the spectrum that, for this reason, we call *absorption spike*. Sometimes, hydrogen clouds in stars for example, give off (emit) as well as absorb light. When a hydrogen cloud gives off light, it gives off light exactly the same color as the cloud absorbs. If the cloud gives off more light than it absorbs, the spectrum will have a peak instead of a valley that is an *emission spikes*. Emission lines also depend on the temperature of a star; they only appear if a star is at the right temperature.

Although highly informative, the emission and absorption spikes do not tell us the entire story. Indeed a typical spectrum results from the superposition of a smooth curve and a certain number of spikes on it. The curve itself is called a *continuum* and represents photons emitted at all energies continuously. The photons that are producing this continuum can be caused by several mechanism that are completely different than those producing photons count at the various peaks and bumps on the curve. Not only these two different kind of emission (continuum and lines) are produced differently, but they each tell us different things about the source that is emitting them.

The goal of our project consists in developing a sensible method that, given a spectrogram, is capable to estimate its continuum and statistically detect the presence of the spike at some particular locations (known a priori) in a computationally efficient way given the huge data–set, consisting on spectrograms measured from 3800Å (blue) to 9200Å (near infrared) − approximatively 4000 data–points for each spectrum − coming from the *Sloan Digital Sky Survey* (sdss[1]), an astronomical survey that takes its name from the Alfred P. Sloan Foundation, and will map

---

[1]Surf http://www.sdss.org and http://www.skyserver.org.

one–quarter of the entire sky in detail, determining the positions and absolute brightness of more than 100 million celestial objects. In this work we are mainly interested in a particular subset of this data–set consisting on galaxy spectra only.

1.1 Galaxy Spectra.

The spectra observed from galaxies are formed from a combination of something like 50 billion stars plus many molecular clouds and star–forming regions. The masters of stellar population synthesis code (i.e. building a galaxy from individual stars) are Bruzual and Charlot and we remand to their paper [15] for more information.

The analysis of galaxy spectra reveals many important properties of the celestial object we are considering. Among them the radial velocity of the galaxy, the star–formation rate, the kinematics (mass) of the galaxy, the average age and metallicity of the stellar populations. A typical spectrum, as the one in Figure 1, has three components that appear decoupled in Figure 2: the continuum, absorption and emission lines whose admissible positions are marked with blue vertical lines in Figure 1.

The strong continuum component is caused by the combination of a range of blackbody[2] emitters, mainly dense gases or solid objects which radiate heat, spanning a range in temperature. Stellar cores, for example, tend to emit light in a predominantly continuous spectrum, as do incandescent light bulbs, electric cooking stove burners, flames, fire embers, and...you! This kind of celestial objects emit radiation over a broad range of wavelengths, thus the resulting spectrum appears fairly smooth and continuous, having, as a primary commonly encountered feature, a strong break around 4000Å, caused by the blanket absorption of high energy radiation from metals in stellar atmospheres and by a deficiency of hot, blue stars (see Figure 3).

Discrete, non–continuous spectra are an observable result of the physics of atoms. Unlike a continuous spectrum source, which can radiate at an arbitrary frequency (just change the effective temperature), the electron clouds surrounding the nuclei of atoms can have very specific energies dictated by quantum mechanics. Each element on the periodic table has its own set of possible energy levels, and with few exceptions the levels are distinct and identifiable. Electrons tend to settle to the ground state, so an excited atom with an electron in a higher energy level will emit a wave of light with that exact energy to allow the electron to fall into the ground state. This energy corresponds to a specific color, or wavelength, of light, so we see a bright line at that exact wavelength. We can observe emission lines in spectra from comets, nebula and certain types of stars and galaxy.

More specifically, galaxy emission features are mainly due to gas being heated and then re–radiating energy at specific wavelengths. Young stars form within gas clouds, which they then ionize. Emission features thus reveals the presence of very hot gas and young stars, from the disks of spiral galaxies and from irregular galaxies (see Figures 4(b), and 4(c)). Key emission features, listed in Table 2, include the O[II] doublet[3] (3737Å), O[III] (4959Å and 5007Å), and the Balmer series (6563Å, 4861Å, 4340Å, 4103Å, etc.).

If light from a stellar core with a continuous spectrum encounters an atom, the wavelengths corresponding to possible energy transitions within the atom will be absorbed. The light may be re–emitted later, but as it will be re–emitted in a random direction the spectrum along the line of sight will be preferentially lacking in flux at the wavelength which corresponds to the energy transitions within the atom. We can observe absorption features in spectra from regions in space where a cooler gas lies between us and a hotter source, from stars, from planets with atmospheres, and from galaxies. Thus, for galaxies, the absorption features superimposed on the continuum are due to the absorption of atoms (metals) and molecules in stellar atmospheres, and to cold, interstellar gas clouds which siphon off radiation at key frequencies. This implies the presence of old stellar populations, which are typical found in elliptical galaxies and in the bulges of spiral galaxies (see Figure 4(a)). Key features include the Calcium H and

---

[2]An hypothetic body that completely absorbs all wavelengths of thermal radiation incident on it. Such bodies do not reflect light, and therefore appear black if their temperatures are low enough so as not to be self–luminous. All blackbodies heated to a given temperature emit thermal radiation with the same spectrum, as required by arguments of classical physics involving thermal equilibrium. However, the distribution of blackbody radiation as a function of wavelength, known as the Planck law, cannot be predicted using classical physics. This fact was the first motivating force behind the development of quantum mechanics.

[3]Some particular transitions are *forbidden* in a proper technical sense, and denoted with square brackets (e.g. O[III]). *Semi–forbidden* transitions, instead, are denoted with a single bracket (e.g. CIII]).

	Absorption Lines	Emission Lines
Caused by:	• Atoms and molecules (metals) in a star's atmosphere that absorb specific wavelengths. • Also due to cold gas in the interstellar medium which can extract energy from passing radiation.	• Caused by gas being heated and re–radiating at specific allowed wavelengths. • Stars form from gas so are often embedded in it. • Young stars ionize gas which releases radiation at specific wavelengths as it recombines.
Imply:	• Old stellar population. • Old galaxy.	• Newly formed stars. • Young galaxy.
Observed in:	• Ellipticals. • Spiral bulges.	• Spiral disks. • Irregulars

Table 1: Summary of some facts related to absorption and emission lines.

K lines (found at 3934Å and 3969Å), the G−band (4304Å), the Magnesium (5175Å) and Sodium (5894Å) lines (see Table 2).

Typical Spectra features

• Absorption		• Emission	
① Ca(H)	$= 3933.7$ Å	❶ O[II]	$= 3727.3$ Å
② Ca(K)	$= 3968.5$ Å	❷ Hδ	$= 4102.8$ Å
③ G − band	$= 4304.4$ Å	❸ Hγ	$= 4340.0$ Å
④ Mg	$= 5175.3$ Å	❹ Hβ	$= 4861.3$ Å
⑤ Na	$= 5894.0$ Å	❺ O[III]	$= 4959.0$ Å
		❻ O[III]	$= 5006.8$ Å
		❼ Hα	$= 6562.8$ Å
		❽ S2	$= 6716.0$ Å

Table 2: Symbols denoting the main spectral features of galaxy spectra together with their associated positions.

1.2 Spike Hunting: Two Applications.

The following two sections contains a couple of examples that illustrate the importance of the line detection step in the analysis of galaxy spectra.

1.2.1 Radial Velocity and Emission Red–Shift.

In the previous section we said that knowing the line positions we can recover some interesting derived quantities. Among these we mentioned the radial velocity of the galaxy. In fact most galaxy spectra are red–shifted (though some few are blue–shifted), which means their spectral features are offset compared to those measured for gasses in

the lab. More precisely we see that characteristic combinations of their lines appear systematically offset to longer wavelengths than the wavelength values measured at rest. This is interpreted similarly to a Doppler shift, and implies that the galaxies are moving away from us (receding ↦ red–shifted ↦ positive velocity) or towards us (approaching ↦ blue–shifted ↦ negative velocity).

In practice one way to measure the absolute value and direction of the shift is based on the detection of characteristic lines. In the spectrum shown in Figure 4, for example, it is easy to identify the O[II] and H emission lines, which have shifted in wavelength and then conclude that the galaxy is moving away (red–shift) at about 21,750 km/s.

Of course this procedure might be time–consuming if not automatized. For this reason the SDSS spectrographic pipeline implements a particular "hunting" routine that tries to estimate the location of strong emission lines to evaluate the radial velocity of the galaxy and, consequently, to obtain very strict bounds[4] on the positions of all the other lines (see Section 1.4).

1.2.2   High–Velocity Clouds and Galaxy Dynamic.

Since the early 1960s astronomers have thought that some galaxies, included our Milky Way, were born early in cosmic history and then evolved slowly. Today, however, evidence indicates that galaxies are continuing to grow. They cannibalize their smaller brethren and gulp down fresh gas from intergalactic space. Taking a closer look to the ongoing construction work, we see that incoming gas takes the form of *high–velocity clouds* (HVC) discovered decades ago. Only recently were some of these clouds proved to be fresh material since, observationally, they get entangled with circulating gas. These clouds come in several guises: clumps of neutral hydrogen reminiscent of intergalactic gas; a stream of gas torn out of nearby small galaxies; and highly ionized hot gas that may be dispersed throughout the intergalactic vicinity.

High–velocity clouds stymied astronomers for decades because their distances and compositions were uncertain. The only known technique to measure these properties is the *absorption–line* method. Stars and galaxies located behind HVCs act as bulbs that shine through the clouds from behind. As explained in previous sections, most of the light passes through the clouds, but a few wavelength are absorbed.

If the spectrum of a star contains absorption lines, it means a cloud must be sitting between us and the star. The distance to the star sets an upper limit on the distance to the cloud. Conversely, the lack of absorption lines, implies a lower limit on the distance to the cloud. These limits assume that other factors can be ruled out: uncertainties in the stellar distance, lack of enough heavy elements to produce detectable absorption line, and absorption line created by material within the star itself.

To determine HVC distances, the most useful light–bulbs are so called RR–*Liræ* variables and *blue horizontal branch* stars. They are numerous, their distance can be measured accurately, and few of their spectral lines overlap with those of the clouds. In principle, the absorption lines of any element could be used. To determine the heavy element content, however, the best measurements rely on spectral lines of neutral oxygen and ionized sulfur. These lines lie in the ultraviolet part of the spectrum, requiring properly equipped satellites such as the Hubble Space Telescope or Far Ultra–violet Spectroscopic Explorer (FUSE). In this case, the best light–bulbs are distant active galaxies such as quasars, because they often have featureless spectra and are brighter ultraviolet emitters than stars.

A single star or galaxy can illuminate more than one gas cloud. Each cloud moves at a different velocity, so each absorbs at a slightly different wavelength because of the Doppler effect. To distinguish the clouds requires a spectrometer with high spectral resolution, and, of course, a reliable statistical procedure to detect them properly. We remand to [1] for more information.

1.3   The Sloan Digital Sky Survey: An Intergalactic Census.

As said above, the data–set we consider here consists in a collection of galaxy spectra or, more correctly, spectro–gram, measured in five spectral bands from 3800Å(blue) to 9200Å(near infrared), coming from the spectrographic pipeline of the Sloan Digital Sky Survey (SDSS), an astronomical survey that takes its name from the Alfred P. Sloan

---

[4]These bounds are so strict that, in practice, we will assume the line locations as given in advance (see Section 2).

Foundation, and will map about one–half of the Northern sky in detail, determining the positions and absolute brightness of more than 100 million celestial objects.

As the U.S. Census Bureau collects statistical information about how many people live in the U.S., where they live, their races, their family incomes, and other characteristics that are a primary source for people trying to understand the nation, the Sloan Digital Sky Survey can be thought as a sort of celestial census, gathering information about how many galaxies and quasars the universe contains, how they are distributed, their individual properties, and how bright they are.

A *spectrogram* is, literally, a photograph of a spectrum, but the term is also more loosely used to mean any representation of a spectrum (such as a graph of light intensity versus light frequency). The device that produces a spectrogram is called a *spectrograph*. So for example, a simple prism is a spectrograph. It produces a spectrum by refracting (bending) different frequencies of light by different amounts. An ordinary compact disc is another kind of spectrograph called a diffraction grating; the closely–spaced ridges of the grating diffract different light frequencies at different angles. We could even say that a collection of rain droplets in the atmosphere forms a spectrograph, because of the rainbow they produce.

The SDSS uses a spectrograph that combines the ideas of a diffraction grating and a prism into an instrument appropriately named a *grism*. The SDSS spectrograph is designed so that the light from hundreds of galaxies and quasars can be turned into spectra at the same time. The light travels from the telescope's focal plane through optical fibers and then gets split into spectra by the *grism*. The hundreds of spectra are then photographed by a CCD camera and sent to the spectrographic pipeline for a first round of pre–processing. More precisely, astronomers select 600 distant objects at a time for spectral analysis plugging drilled aluminium discs (each hole records approximatively the light that comes from a particular celestial object) with optical fiber cables to feed light from the telescope to spectrographs. When attached to the telescope, each cable will be approximatively located at the position of a selected object in the sky (see Figure 5).

1.4   The Current Procedure.

Although not entirely satisfying, a fairly detailed description of the methodology currently implemented by the SDSS spectroscopic pipeline can be easily found on the official web–site of the survey and a brief summary of it is given below

Continuum Fitting The continuum is fit using a median/mean filter. A sliding window is created of length 300 pixels (data–points) for galaxies and stars or 1000 pixels for quasars. Pixels closer than 8 pixels (equivalent to 560km/s) for galaxies and stars or 30 pixels (2100 km/s) for quasars to any reference line are masked (i.e. not really detectable) and not used in the continuum measurement. The remaining pixels in the filter are ordered and the values between the 40th and 60th percentile are averaged to give the continuum. In some cases, the lines are fit with a cruder continuum which is given by a fifth order polynomial fit which iteratively rejects outlying points that are eventually recognized as strong lines.

Line Fitting Line fitting is performed twice in the SDSS spectroscopic pipeline. The first time it is done as part of finding characteristic emission lines in order to measure what is called an emission line *redshift*. At this stage wavelet filters are used to locate strong emission features in the spectrum. Secondly, after the redshift has been determined, all possible lines are finely searched and accurately measured again. In this paper we are mainly concerned with the latter setup even if a slight extension of our method might readily implement the former too (see Section 2). Every line in the reference list is fit as a single Gaussian, say $\mathcal{N}_1(\lambda, \sigma^2)$, on top of the continuum subtracted spectrum. Lines that are deemed close enough are fitted simultaneously as a blend. The basic line fitting is essentially based on the *Levenberg–Marquardt* non–linear least–square method (see [4]). Parameters are constrained to fall within certain values by multiplying the returned sum of squares by a steep function. The constraints are

- $\sigma \in (0.5\text{Å}, 100\text{Å})$
- if λ_0 denotes the "red–shifted" wavelength corresponding to the center of the line of interest when measured in the vacuum, then the central wavelength λ, is allowed to drift from λ_0 by no more than 450 km/sec for stars and galaxies or 1500 km/sec for quasars[5].

Hence the resulting procedure can be succinctly described as a constrained non–linear least square problem attacked using a well–known quasi–Newton optimization technique.

Detection The documentation regarding the way presence/absence of a line is being tested is quite lacking and, in addition, filled with statistically scary assertions like (quoting from the web–site): *"the emission–line confidence level has been tweaked to match our empirical success rate"*. Anyway, the most plausible option is that currently a "tweaked" χ^2 goodness of fit test is being used to do the job.

1.5 Alternative Procedures.

In a series of papers[6] professor van Dyk and the astro–statistical group at Harvard, developed an interesting, even if computationally intensive, fully parametric hierarchical Bayesian procedure to deal with low counts spectra coming from the *Chandra* X–Ray observatory. The data gathered by *Chandra*, although high–resolution, present a number of statistical challenges. For spatial data, the images are convolved with instrument characteristic which blur the image. For example, the point spread function characterizes the probability distribution of a photon's recorded pixel location relative to its actual sky coordinates. The situation is further complicated because the shape of the scatter distribution varies across the detector. Just like sky coordinates, the energy of a photon is subject to "blurring". Thus, we have to deal with three dimensional blurring of the image. Another complication involves the absorption of photons and the so–called effective area of the telescope. Depending on the energy of the photon, it has a certain probability of being absorbed, for example by the inter–stellar media between source and detector as explained earlier. The effective area, instead, is an instrumental effect but has similar impact on the data. The data are also degraded by the presence of background counts, as X–Ray photons, which arrive at the detector but do not correspond to the source of interest. The resulting model contains valuable structural information injecting the physics into the statistics but, on the other hand, it is extremely complex and characterized by large number of parameters. More in details, the structural model used to deconvolve the *Chandra* data can be written as follow

$$\boldsymbol{\lambda} = \overset{\overset{\text{censoring}}{\uparrow}}{\mathrm{P}} \quad \underset{\underset{\text{blurring}}{\downarrow}}{\mathrm{A}} \quad \overset{\overset{\text{signal}}{\uparrow}}{\boldsymbol{\mu}} \; + \; \underset{\underset{\text{background}}{\downarrow}}{\boldsymbol{\xi}} \tag{1.1}$$

and is composed by the essential ingredients listed below:

---

[5]Except for the CIV line which is allowed to be shifted by as much as 3000 km/sec.
[6]See [9],[11], [10], [12], [13] and [14].

The data: the data consists of an $L \times 1$ vector of poisson counts with independent components

$$Y_\ell \sim \text{Poisson}(\lambda_\ell) \quad \ell \in \{1, \ldots, L\},$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_L]^\mathsf{T}$ is the vector of expected counts. These data may be pixel counts from an image, counts from a set of energy channels, or counts from the set of bins constructed by crossing the image pixels with energy channels. They refer to these as the *detector counts* to distinguish them from the *idealized counts* that we would observe with a perfect instrument, i.e., an instrument without blurring, with constant effective area, and without background contamination. The idealized counts are, of course, missing data. There is no need for the number of ideal bins and detector bins to be the same, indeed they are generally not the same. Thus, the ideal unobserved data are assumed to be $J \times 1$ vector of counts.

Effective area: The effective area of the detector is quantified as a vector of probabilities, one for each ideal bin: the effective area of ideal bin j is the probability that a photon that arrives at the detector corresponding to *ideal* bin j is recorded. This results in a $J \times J$ diagonal matrix, A, with diagonal elements equal to these probabilities.

Instrument effect: Because of the instrument effect, i.e. blurring, a photon arriving at the detector, and corresponding to ideal cell j, may be recorded in one of several detector bins. Thus, if $p_{\ell j}$ is the probability that a photon corresponding to the ideal bin j is recorded by the detector in bin ℓ, we can build a $L \times J$ matrix $\mathsf{P} = [p_{\ell j}]_{(\ell,j) \in \{1,\ldots,L\} \times \{1,\ldots,J\}}$ having these probability vectors as columns.

Expected background: The expected background count in each detector cell is an $L \times 1$ vector denoted by $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_L]^\mathsf{T}$.

Expected *ideal* count: The vectors of expected detector counts, that we want to *reconstruct* or *deconvolve* from the observed data, is denoted by $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_J]^\mathsf{T}$. The μ_j's are functions of E_j, the energy of ideal energy bin j. This function has three basic components:

Continuum: Generally the continuum is modeled as a log–linear models that are linear in a transformation of energy. A power law continuum, for example, can be written $f(E_j|\boldsymbol{\theta}^C) = \log[\alpha(E_j)^\gamma]$ where $\boldsymbol{\theta}^C = [\alpha, \gamma]^\mathsf{T}$ is a vector of unknown parameters and $f(\cdot)$ is linear in $\log(E_j)$.

Spectral lines: Emission lines are modeled either as narrow Gaussian or as Lorentzian distributions or as delta functions that are added to the continuum. In statistical terms this can be formulated as a finite mixture model having as many components as the number of emission lines. So if we have K emission lines, this feature is represented as

$$g(\mathbf{p}_{j,\cdot}^E | \boldsymbol{\theta}^E) = \sum_{k=1}^K \theta_k^E p_{j,k}^E,$$

where θ_k^E represents the relative size of emission line k and $p_{j,k}^E$ is the proportion of emission line k that falls into ideal energy bin j.

Absorption features: Absorption of photons occurs independently across the photons and it is represented by a multiplicative factor $h(E_j|\boldsymbol{\theta}^A)$ that is the complement of the censoring probability, i.e. $1 - h(E_j|\boldsymbol{\theta}^A)$ is the probability that a photon with energy E_j is absorbed, and, thus, not observed. Hence, like the continuum, some absorption models can be represented by a generalized linear model, in this case with complementary log–log link and a Bernoulli model.

So, in the end, we have

$$\mu_j = \left[\delta_j f(E_j|\boldsymbol{\theta}^C) + \sum_{k=1}^K \theta_k^E p_{j,k}^E \right] \times h(E_j|\boldsymbol{\theta}^A), \tag{1.2}$$

where δ_j is the width of ideal energy bin j.

Note that even if the deconvolution stage expressed by (1.1) is extremely interesting and of unquestionable importance to the problem at hand, in this paper we will focus exclusively on the second of the two modelling steps (1.2), assuming that all the measurements are taken in the observed frame, and that no correction has to be made for the instrumental resolution.

1.6   Our Proposal: an overview.

As mentioned, the spectra are essentially a particular transformation of photon counts, recorded at various wavelengths, that embed two different structures to be "statistically" magnify: the continuum and the absorption/emission spikes. For this reason an inhomogeneous Poisson process (possibly over–dispersed) and a Gaussian process (possibly heteroscedastic) that explicitly decouple the two components in their mean function along wavelength, appear to be sensible models for low and high count spectra respectively. Given this basic idea, we have to design reasonable estimates and work out reliable confidence intervals/bands for the continuum and the spike parameters or, in other words, for the (conditional) mean function of these processes. In the present work we propose two different but interacting ways to achieve our goals:

Fully–parametric approach: assuming the continuum to be locally constant in a symmetric neighborhood of the spike center, we simply fit a physics–based parametric profile to each line or "island" of lines. The asymptotic confidence intervals for functions of the parameters (e.g. the area under a spike) are worked out explicitly by the δ-method.

Semi–parametric approach: in this case we fit non–parametrically the continuum and parametrically the spikes. Done this, we are left with two inferential options:

- build confidence bands around the continuum fit and propagate the uncertainty over the spike parameters (and consequently over functions of them).
- build confidence balls for the parametric component directly using the asymptotic normality of their (generalized) profiled–likelihood based estimates.

2   Fully Parametric Approach (FPA).

In this section we design a particular disjoint segmented regression model (see [16]) that allows for a locally (to the line positions) flat continuum upon which a sequence of physics–based parametric profile (spread) functions sit. This approach is similar to the *parametric deconvolution* setting studied in [17], [18] and [19] with the huge simplification that here, the maximum number of spread functions needed and the positions of their centers are given in advance. Note that considering the centers of some characteristic lines as unknown, that is partially coming back to the just mentioned full parametric deconvolution model[7], we might be able to estimate the red–shift of the galaxy along the line of what we said in Sections 1.4 and 1.2.1. Given the appearance of the "segments" that we shall adopt, in this section we often talk about "islands of spikes" as a graphical and hopefully effective metaphor.

We begin describing the physics behind the parametric spike profiles used in this work and defining some quantities of interests such as the so called *equivalent width*. Then we provide some details of the estimation procedure adopted and, finally, we describe how to make inference in a fast and reliable way.

---

[7]Possibly with bounds on the line positions implied by the quantum mechanic description of absorption and emission phenomena.

2.1 Profile functions.

As we said in Section 1.1, spectra can be characterized as continuum radiation, upon which are superimposed emission and absorption lines resulting from transitions between energy levels in atoms or molecules. One observes absorption lines when atoms absorb photons at a particular frequency, typically from broad–band radiation, and emission lines when they emit photons.

Energy transitions triggered by electronic excitation and de–excitation of atoms and molecules are observed primarily at optical and UV wavelengths. Transitions induced by vibrations of atoms can be observed at infrared and sub–millimeter wavelengths and those resulting from the rotation of molecules can be observed at radio wavelengths.

The detailed shape of a line profile is determined by the abundance of the element or compound producing the line, as well as the pressure and temperature of the environment. Spectral data can thus be used to determine abundances, the distribution of species with depth, and the thermal structure of a medium. They can also be used to recover the velocity distribution (kinematics), in the form of winds (atmospheres), or orbital motions (stars).

2.1.1 The Equivalent Width.

For a given frequency ν, the *absorption depth* is defined as

$$A(\nu) \triangleq \frac{F_{\text{cont}} - F(\nu_0)}{F_{\text{cont}}} = 1 - e^{-\tau(\nu)},$$

where F_{cont} is the flux density of the continuum background, $F(\nu_0)$ is the flux density at the center of the absorption line and $\tau(\nu)$ is the optical depth defined as:

$$\tau(\nu) = \tau(\nu_0) \times \phi(\nu) = \left[\int_0^{\text{L}} \mathsf{N}\alpha(\nu_0)\mathrm{d}l \right] \times \left[\frac{\alpha(\nu)}{\alpha(\nu_0)} \right]. \tag{2.1}$$

with $\tau(\nu_0)$ the optical depth at the center of the line, $\phi(\nu)$ the line shape, $\alpha(\nu)$ the extinction coefficient at frequency ν and $\alpha(\nu_0)$ the extinction coefficient at the line center.

For unresolved lines, one can measure another quantity, called *equivalent width* (EW), as follow

$$\mathsf{EW} = \int_{\mathbb{R}^+} A(\nu)\mathrm{d}\nu = \int_{\mathbb{R}^+} 1 - e^{-\tau(\nu)}\mathrm{d}\nu.$$

The equivalent width is equal to the area between the line and the continuum divided by the height at the center of the line so it correspond to the base (width) of an hypothetical rectangle having the same area and height of the line. For an absorption line, the equivalent width is equal to the width of a pure black line ($F(\nu_0) = 0$), with the same total flux as that absorbed by the line.

If the thermal structure of the medium is known, spectral line profiles can be used to derive the integrated density of the absorbing material. If the line is not resolved but the line shape is known, measurements of the equivalent width can be used. For an optically thin line, the equivalent width can be expressed as follow

$$\mathsf{EW} \approx \int_{\mathbb{R}^+} \tau(\nu)\mathrm{d}\nu = \int_{\mathbb{R}^+} \tau(\nu_0)\phi(\nu)\mathrm{d}\nu = \mathsf{N}_{\text{col}} \int_{\mathbb{R}^+} \alpha(\nu_0)\phi(\nu)\mathrm{d}\nu,$$

where $\mathsf{N}_{\text{col}} = \int \mathsf{N}\mathrm{d}l$ is the column density of the absorbing material. The equivalent width increases linearly with N_{col} as long as the material is optically thin. As the optical depth increases, the line profile becomes saturated and the equivalent width cannot continue to increase linearly with the column density. A plot of the equivalent width as a function of column density is called a *curve of growth*, and can be used to determine the abundance of an element from its observed line width. Note that for some line profiles there may be a regime in which the curve of growth is essentially flat, where the equivalent width is nearly independent of column density.

2.1.2 The *Gaussian* profile: Doppler broadening.

Due to Doppler effects, the thermal motion of absorbing atoms leads to small variations in the absorbed frequency ν. If ν_0 is the centroid frequency of the absorption line, the Doppler effect will shift the frequency toward

$$\nu = \nu_0(1 + \tfrac{v_r}{c}),$$

where v_r is the component of the velocity of the absorbing atom along the line of sight. Thus, the Doppler shift has an amplitude $\Delta\nu = \nu_0\left(\tfrac{v_r}{c}\right)$. The magnitude of this effect can be quantified recalling that the distribution of particle speeds in LTE is a *Maxwellian* and since we are interested only in the velocity distribution along our line of sight, this is just

$$\frac{\mathrm{d}n(v_r)}{n_{tot}} = \sqrt{\frac{m}{2\pi kT}} \exp\left(-\frac{mv_r^2}{2kT}\right) \mathrm{d}v_r.$$

Using the fact that

$$v_r = \frac{c(\nu - \nu_0)}{\nu_0}, \quad \mathrm{d}v_r = \frac{c\,\mathrm{d}\nu}{\nu_0},$$

the distribution function can be expressed through the number of particles capable of intercepting photons of frequency ν due to their velocity offset:

$$\frac{\mathrm{d}n(\nu)}{n_{tot}} = \frac{c}{\nu_0}\sqrt{\frac{m}{2\pi kT}} \exp\left(-\frac{mc^2(\nu - \nu_0)^2}{2\nu_0^2 kT}\right) \mathrm{d}\nu, \tag{2.2}$$

Also note that the optical depth scales such that from equation (2.1) $\tau(\nu) \propto n(\nu)$. We define the absorption line profile $\phi(\nu)$ such that $\tau(\nu) = \tau_0\,\phi(\nu)$, where τ_0 is the total absorption depth, and the profile function $\phi(\nu)$ is normalized so that $\int_0^{+\infty} \phi(\nu)\,\mathrm{d}\nu = 1$. Further, we define the *Doppler width*,

$$\Delta\nu_{\mathsf{D}} \triangleq \frac{\nu_0}{c}\sqrt{\frac{2kT}{m}}.$$

With these definitions, Equation (2.2) yields the absorption line profile function due to Doppler broadening

$$\phi(\nu) = \frac{1}{\Delta\nu_{\mathsf{D}}\sqrt{\pi}} \exp\left(-\frac{(\nu - \nu_0)^2}{\Delta\nu_{\mathsf{D}}^2}\right), \tag{2.3}$$

which is a Gaussian with full width at half maximum $(\Delta\lambda)_{1/2} = \frac{2\lambda_0}{c}\sqrt{\frac{2kT\ln 2}{m}}$.

2.1.3 The *Lorentz* profile: Natural broadening.

As a consequence of the Heisenberg uncertainty principle,

$$\Delta E \Delta t \sim \hbar,$$

the energy level of an excited state of an atom or ion can never be determined precisely because its life time Δt is finite. Thus, the line profile is broadened over a natural line width of

$$(\Delta\lambda)_{1/2} = \frac{\lambda_0^2}{\pi c \Delta t}.$$

This effect leads to a *Lorentz* profile function,

$$\phi(\nu) = \frac{\gamma}{4\pi^2} \frac{1}{(\nu - \nu_0)^2 + \left(\frac{\gamma}{4\pi}\right)^2}, \tag{2.4}$$

10

where the width γ is related to the spontaneous decay rates $A_{k,l}$ of the excited state k into all lower states l through

$$\gamma = \sum_l A_{k,l}.$$

The decay rates $A_{k,l}$ result from the quantum–mechanical treatment of the respective transitions, and can be found in standard tables.

2.1.4 Collisional broadening.

If the absorbing atom or ion responsible for an absorption line is suffering frequent collisions with other atoms or ions, the electron energy levels will be distorted. This is another mechanism leading to broadening of emission and absorption lines, called *collisional broadening* or *pressure broadening*. The magnitude of this effect depends on the frequency ν_{col} at which collisions occur. This frequency can be estimated as

$$\nu_{\mathsf{col}} = (v_{\mathsf{th}} \, n \, \sigma_{\mathsf{col}})^{-1},$$

where $v_{\mathsf{th}} = \sqrt{\frac{2kT}{m}}$ is the thermal speed of the atoms or ions, n is their density, and σ_{col} is their cross–section for collisions. The line profile resulting from collisions is a Lorentzian, just like the case of natural broadening in (2.4). Moreover, the effects of natural and collisional broadening can be combined by using an effective width

$$\Gamma = \gamma + 2\nu_{\mathsf{col}},$$

so that the resulting line profile is given by

$$\phi(\nu) = \frac{\Gamma}{4\pi^2} \frac{1}{(\nu - \nu_0)^2 + \left(\frac{\Gamma}{4\pi}\right)^2}.$$

2.1.5 The *Voigt* profile: combine Doppler, natural and collisional broadening.

All three broadening mechanisms affect the line profile. The profiles are thus a combination:

$$\phi(\nu) = \frac{\Gamma}{4\pi^2} \sqrt{\frac{m}{2\pi kT}} \int\limits_{-\infty}^{+\infty} \frac{e^{-\frac{mv_r^2}{2kT}}}{(\nu - \nu_0 - \frac{\nu_0 v_r}{c})^2 + \left(\frac{\Gamma}{4\pi}\right)^2} \, \mathrm{d}v_r,$$

which can be expressed in terms of the *Voigt* function:

$$\phi(\nu) = \frac{1}{\Delta\nu_{\mathsf{D}} \sqrt{\pi}} \, \mathbb{H}(a, u),$$

where

$$\mathbb{H}(a, u) = \frac{a}{\pi} \int\limits_{-\infty}^{+\infty} \frac{e^{-y^2}}{a^2 + (u - y)^2} \, \mathrm{d}y,$$

is called the *Hjerting* function, and

$$a \triangleq \frac{\Gamma}{4\pi\Delta\nu_{\mathsf{D}}}, \quad u \triangleq \frac{\nu - \nu_0}{\Delta\nu_{\mathsf{D}}}. \tag{2.5}$$

In a typical situation, the quantity a defined in Equation (2.5) is small, which leads to a profile where the center is dominated by the Doppler profile and the wings are dominated by the Lorentzian component.

11

2.1.6   The *Hyper*–profile.

We now introduce our last proposal: the *Hyper*–profile. The name comes from the confluent hyper–geometric function that appears in its analytic expression as well as the *Gaussian*–profile, and it quantifies a Doppler broadening effect. More specifically, emission and absorption lines in an observed spectrum of a galaxy are distorted into "spikes" due to the variable *redshift* of individuals stars in the galaxy. This distorsion is due to rotation of stars about the galactic center: the galaxy as a whole is moving away from us, but stars at one edge are moving away from us faster than stars at the center, while stars at the other edge are moving away from us slower than stars at the center.

The extra speed with which a star in a galaxy is moving toward or away from us is its *projective velocity*. It is easy to see that a curve of constant projective velocity in a galaxy is plane orthogonal to the plane of rotation, passing through the observer. We will take different planes of constant projective velocity to be parallel (see Figure 7(c) for a graphical intuition).

The shape of a spike is determined by the relative intensity of light emitted from stars at a specific projective velocity compared to the intensity of light emitted from stars with projective velocity equal to zero. According to [2], all *spiral* and many *elliptical* galaxies have a light intensity which drops off according to an "$r^{\frac{1}{4}}$–law", that is, the intensity of light at distance r units from the center of a galaxy of radius R is $c\left(R^{\frac{1}{4}} - r^{\frac{1}{4}}\right)$, where c is a constant of proportionality.

Taking $\mu = 0$ to be the actual emission frequency, the (signed) height of a spike at μ, for $x \in [-R, R]$, is[8]

$$c \times \mathsf{p_H}(x|R) = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} c\left[R^{\frac{1}{4}} - (\sqrt{x^2+y^2})^{\frac{1}{4}}\right] dy = \frac{2}{5}c\sqrt{R^2-x^2}\left[R^{\frac{1}{4}} - |x|^{\frac{1}{4}} \times {}_2\mathsf{F}_1\left(\frac{1}{2}, \frac{7}{8}, \frac{3}{2}; 1 - \frac{R^2}{x^2}\right)\right] ,$$

where ${}_2\mathsf{F}_1$ is the hypergeometric function.

$$ {}_2\mathsf{F}_1(a, b, c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!} . $$

Such a spike takes positive values in $[-R, R]$ if $c > 0$ and negative values if $c < 0$. The extreme point is at $x = 0$, where it takes the value $\frac{2}{5}cR^{\frac{5}{4}}$. The spike curve has order of contact 1 with the horizontal axis at $x = \pm R$. An example is shown in Figure 7(a).

In practice, two different approximations to the hypergeometric function are used. When fitting the spike to data, we use a Laplace approximation as described in [3]. This approximation meets the horizontal axis smoothly but is difficult to integrate. When computing the signed area between the spike and the horizontal axis, we use a Taylor approximation centered at $x = 0$. This approximation hits the horizontal axis at R with infinite slope, but about 10 terms are sufficient for a good approximation of the area, and it has an easy–to–integrate form.

2.2   Fitting "Islands" of Spikes.

For the rest of the section we will focus on the *Hyper*–profile bearing in mind that everything we say carries over to the other line profiles in a straightforward way. In our *fully parametric approach*, the mean of the data near a spike is assumed to be locally a constant, h as in Figure 7(b), that is

$$\mathsf{p}_\mu(x|\mathbf{R}, \mathbf{c}, \mathbf{h}) = \mathbf{c}\,\underbrace{\frac{2}{5}\sqrt{\mathbf{R}^2 - (x+\mu)^2}\left[\mathbf{R}^{\frac{1}{4}} - |x+\mu|^{\frac{1}{4}}{}_2\mathsf{F}_1\left(\tfrac{1}{2}, \tfrac{7}{8}, \tfrac{3}{2}, 1 - \tfrac{\mathbf{R}^2}{(x+\mu)^2}\right)\right]}_{\mathsf{p_H}(x+\mu|\mathbf{R})} + \mathbf{h}.$$

The way we fit this "winged" profile to data consists of three stages:

---

[8]See Appendix A for a detailed derivation of the profile.

1st: Fit each line separately. Given a spike centered at μ, data within 100Å of μ are used to fit the spike. If two spike centers are closer than 200Å apart, data points are used to fit the nearest spike.

2nd: Handle hydrogen doublets by fitting a pair of *discordant*[a] profiles to each line associated with hydrogen transitions.

3rd: Search for *colliding* spikes [b]. For each collision detected, fit an island with the appropriate number of spikes.

---

[a]Given two profiles $\mathsf{p}_{\mu_1}(\cdot|R_1, c_1, h_1)$ and $\mathsf{p}_{\mu_2}(\cdot|R_2, c_2, h_2)$, they are *discordant* $\Leftrightarrow \mathsf{sign}(c_1) \neq \mathsf{sign}(c_2)$ and $\mu_1 \equiv \mu_2$.
[b]Given two spike centers $\mu_1 < \mu_2$ and the associated estimates $\widehat{R}_1$ and $\widehat{R}_2$, we say that there is a *collision* $\Leftrightarrow \mu_1 + \widehat{R}_1 > \mu_2 - \widehat{R}_2$.

Given the data and the center of a spike, the MLEs $\widehat{R}$, $\widehat{c}$ and $\widehat{h}$ are approximated by minimizing the negative log–likelihood using the BFGS quasi–Newton method described in [4]. More specifically, since our islands are linear combinations of spike profiles, what we are facing is a *separable* non–linear least–square problem that can be handled more efficiently as described in Appendix B. An advantage of using this particular method over others is that it computes an approximation of the Hessian matrix of the log–likelihood evaluated at the point $(\widehat{R}, \widehat{c}, \widehat{h})$, which is used for constructing the mentioned δ–method based confidence interval for the signed area of the spike. Some results of the described procedure are illustrated in Figure 8, where a spike–collision is handled introducing an appropriate mixture of "mono–profiles", and in Figure 9, that summarizes the resolution of two different hydrogen doublets by adopting a system of "discordant" profiles.

2.3 Confidence Intervals for the Signed Area.

The desired outcome of spike–fitting is a confidence interval for the area of the spike, which we obtain via the δ–method from the asymptotic normality of MLEs. The approximate Hessian matrix mentioned above gives us an approximation $\widehat{\mathbb{I}}$ of the Fisher information matrix for the parameters (R, c, h), so that assuming we have "reached asymptopia",

$$\sqrt{n}\left[(\widehat{R}, \widehat{c}, \widehat{h})^\mathsf{T} - (R, c, h)^\mathsf{T}\right] \sim \mathcal{N}_3(\mathbf{0}, \widehat{\mathbb{I}}) ,$$

where n is the number of data points used to fit the spike.

A simple real function $g : \mathbb{R}^3 \mapsto \mathbb{R}$ which maps $(\widehat{R}, \widehat{c}, \widehat{h})$ to the signed area of the spike with these parameters is easily computed by using a Taylor approximation of the hypergeometric function mentioned above (we use about 10 terms in the approximation, see Figure 7(d)). Hence if $\widehat{\nabla}_g \triangleq \nabla g(\widehat{R}, \widehat{c}, \widehat{h})$, then using the δ–method,

$$\sqrt{n}\left[g(\widehat{R}, \widehat{c}, \widehat{h}) - g(R, c, h)\right] \sim \mathcal{N}_1\left(0, \widehat{\nabla}_g^\mathsf{T} \widehat{\mathbb{I}} \widehat{\nabla}_g\right) .$$

From this, any desired confidence interval for the signed area and the *equivalent width* can be computed.

Figure 10 shows a typical output of our FPA procedure: on the top of the picture we see the pointwise estimates of the spike parameters whereas the yellow box on the bottom contains the pointwise estimate of the area under the spike and the correspondent confidence interval at the prescribed confidence level. In this case, as highlighted by the two flags on the top–right corner, we are actually looking at a detection (the confidence interval does not contain zero) with no boundary problem, i.e. we are not at the boundary of the parameter space so that the usual asymptotics of maximum likelihood estimators hold. Note that the way we handle spike–collisions is needed for avoiding pathological situations that might cause the break down of this classical regime.

2.4 Disclaimers.

The tricky "signed area". Based on the information available at the time this work started, we assumed that to every line position we could associate, without any specific preference, an emission or an absorption spike. This

13

argument is essentially wrong as we can deduce looking more carefully at Table 2. We are working to inject this information into our procedure. Astronomers assistance would be extremely useful.

The *Hyper*–profile. It seems that there is a flaw in the derivation of the *Hyper*–profile. It is related to the meaning of R, the parameter that modulates the profile width: if R is the radius of the main radiative area of the galaxy, why its estimates varies across wavelength? The galaxy, and its radius, is always the same, is it? Once again, experts help would be needed.

3   Semi–parametric Approach (SPA).

In this section we refer exclusively to the Gaussian case, the analysis of the Poisson is in progress. For improving the fit, under this distributional assumption, we have developed an iterative semi–parametric procedure that starts treating the FPA output as a *pilot* estimate, then replaces the stepwise linear continuum by a non–parametric regression fit to the spike–subtracted spectrum, and finally re–estimates the profile parameters R and c using the continuum–subtracted spectrum.

For estimating the non–parametric component we use a sieved penalized regression technique that uses a spline basis. To make inference about the parametric component we propose two different approaches:

- **Iterative 2–stage procedure:** in this case we explicitly build a confidence "envelope" around the non–parametric fit and then propagate the uncertainty to the parametric component taking, as final confidence interval, the union of the two obtained once we consider the lower and upper bands of the envelope as the actual fit. Note that although extremely fast, this approach is pretty conservative in terms of coverage.

- **"Linearization" and Profile Likelihood:** Since we are interested mainly in making inference about functionals of the parametric component, we can consider the continuum and the variance function as (infinite dimensional) nuisance parameters and apply some results from modern semi–parametric theory (see [31]) to achieve our goal. An easiest way to establish the asymptotic normality of the weighted least square estimator for a *linearized* version of the parametric component are Theorem 2.1.2 and 2.1.3 in [32]. These results, however, are not fully satisfying since estimates of the nuisance parameters are based on a *split–sample* method to achieve independence. We are currently working to improve the results regarding independence to obtain a generalization of the results in [32].

4   Discussion.

A further approach that might prove to give better inference on the parameters consists in extending the procedure of "Curvature adjusted confidence balls", considered in [26] and [27], to the heteroscedastic case. We are examining a fast way for profiling–out nuisance parameters and build curvature adjusted confidence intervals for the parametric component. This method, unlike the ones used so far, is based on exploiting conditionally parametric models as in [28], and on using derivatives of second order and higher.

Finally, parallel to the writing of this paper, we are in the final phase of developing a MATLAB® toolbox Shot_the_Spike. Its functionalities will be accessible through a dedicated web–site. In addition, given the light system requirements also a PALM–OS® porting of Shot_the_Spike running under LyME, a freely downloadable[9] MATLAB® clone, is under development.

---

[9]See www.calerga.com.

A   Derivation of the Hyper-profile.

In Section (2.1.6) we have seen that the Hyper-profile is the solution to the following integral

$$\mathsf{p_H}(x|R) = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \left[R^{\frac{1}{4}} - \left(\sqrt{x^2+y^2}\right)^{\frac{1}{4}} \right] \mathrm{d}y, \tag{A.1}$$

and, using MATHEMATICA®[10], we obtain

$$\mathsf{p_H}(x|R) = \tfrac{2}{5}\sqrt{R^2-x^2}\left[R^{\frac{1}{4}} - |x|^{\frac{1}{4}} \times {}_2\mathsf{F}_1\left(\tfrac{1}{2},\tfrac{7}{8},\tfrac{3}{2};1-\tfrac{R^2}{x^2}\right) \right]. \tag{A.2}$$

Here we will solve (A.1) analytically eventually showing how to get back (A.2) from our handmade solution. Looking at (A.1) a bit closer we see that the main difficulty is to solve the following integral

$$\mathsf{k}(x|R) = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \left(x^2+y^2\right)^{\frac{1}{8}} \mathrm{d}y. \tag{A.3}$$

MATHEMATICA® gives us

$$\mathsf{k}(x|R) = \tfrac{2}{5}\sqrt{R^2-x^2}\left[4R^{\frac{1}{4}} + |x|^{\frac{1}{4}} \times {}_2\mathsf{F}_1\left(\tfrac{1}{2},\tfrac{7}{8},\tfrac{3}{2};1-\tfrac{R^2}{x^2}\right) \right]. \tag{A.4}$$

Lets do it by hand! We have

$$\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \left(x^2+y^2\right)^{\frac{1}{8}}\mathrm{d}y \overset{\substack{\text{symmetry about 0 of}\\ \text{the integrand } \forall\, x\in\mathbb{R}}}{=} 2\int_0^{\sqrt{R^2-x^2}} \left(x^2+y^2\right)^{\frac{1}{8}}\mathrm{d}y \overset{(\Diamond)}{=} 2\int_0^{1-\frac{R^2}{x^2}} \left(x^2-x^2t\right)^{\frac{1}{8}}\left(\mathbb{i}\,|x|\tfrac{1}{2\sqrt{t}}\right)\mathrm{d}t =$$

$$= \mathbb{i}\,|x|\cdot|x|^{\frac{1}{4}}\int_0^{1-\frac{R^2}{x^2}} t^{\frac{1}{2}-1}(1-t)^{\frac{9}{8}-1}\mathrm{d}t =$$

$$\overset{|x|\leqslant R \Rightarrow 1-\frac{R^2}{x^2}\leqslant 0}{=} \mathbb{i}\,|x|^{\frac{5}{4}}\left\{ -\int_{1-\frac{R^2}{x^2}}^0 t^{\frac{1}{2}-1}(1-t)^{\frac{9}{8}-1}\mathrm{d}t \right\} =$$

$$= -\mathbb{i}\,|x|^{\frac{5}{4}}\int_{1-\frac{R^2}{x^2}}^0 t^{\frac{1}{2}-1}(1-t)^{\frac{9}{8}-1}\mathrm{d}t = -\mathbb{i}\,|x|^{\frac{5}{4}}\mathbb{B}_{1-\frac{R^2}{x^2}}\left(\tfrac{1}{2},\tfrac{9}{8}\right).$$

In $(\Diamond)$ we have used the following substitution[11]

- $t = -\frac{y^2}{x^2} \quad\rightarrow\quad y^2 = -x^2t \quad\rightarrow\quad y = \sqrt{-x^2t} = \mathbb{i}\,|x|\sqrt{t}$,
- $\mathrm{d}y = \mathbb{i}\,|x|\frac{1}{2\sqrt{t}}\,\mathrm{d}t$,
- $y = 0 \quad\rightarrow\quad t = 0$,
- $y = \sqrt{R^2-x^2} \quad\rightarrow\quad t = 1-\frac{R^2}{x^2}$.

Hence we have obtained

$$\mathsf{k_{HM}}(x|R) = -\mathbb{i}\,|x|^{\frac{5}{4}}\mathbb{B}_{1-\frac{R^2}{x^2}}\left(\tfrac{1}{2},\tfrac{9}{8}\right), \tag{A.5}$$

---

[10]See www.wolfram.com.

[11]Since the integral is on the positive axis, we take the positive root that solves $y^2 = -x^2t$ to invert our transformation.

where $\mathbb{B}_z(a,b)$ denotes the so called *incomplete Beta function*. $\mathbb{B}_z(a,b)$ is an analytical function of z, a, and b defined in $\mathbb{C}^3$ that admits the following integral representation[12]

$$\mathbb{B}_z(a,b) : (\mathbb{C} \otimes \mathbb{C} \otimes \mathbb{C}) \mapsto \mathbb{C}$$

$$(z * a * b) \rightsquigarrow \int_0^z t^{a-1}(1-t)^{b-1}\mathrm{d}t \mid \operatorname{Re}(a) > 0.$$

Note that $\operatorname{Im}\big(\mathsf{k_{HM}}(x|R)\big) = 0$ since $\operatorname{Re}\Big(\mathbb{B}_{1-\frac{R^2}{x^2}}\big(\frac{1}{2},\frac{9}{8}\big)\Big) = 0$ being $1 - \frac{R^2}{x^2} \leqslant 0$.

Equations (A.4) and (A.5) look pretty different, do they? Well, not really. To recover (A.4) we will make use of the following four equalities

(A) $_2F_1(a,b,c;z) = \,_2F_1(b,a,c;z)$ [symmetry]

(B) $_2F_1(a,b,b+1;z) = bz^{-b}\mathbb{B}_z(b,1-a)$

(C) $\mathbb{B}_z(a,b) = \frac{1}{a+b-1}\big[(a-1)z\mathbb{B}_z(a-1,b) + (b-1)(1-z)\mathbb{B}_z(a,b-1)\big]$

(D) $\mathbb{B}_z(a-1,b) = \frac{1}{a-1}\big[(1-z)^b z^{a-1} + (a+b-1)\mathbb{B}_z(a,b)\big]$

First of all lets manipulate the hypergeometric function that appears in (A.4) to express it in terms of an incomplete Beta function:

$$_2F_1(\tfrac{1}{2},\tfrac{7}{8},\tfrac{3}{2};z) \overset{(A)}{=} \,_2F_1(\tfrac{7}{8},\tfrac{1}{2},\tfrac{3}{2};z) = \,_2F_1(\tfrac{7}{8},\tfrac{1}{2},\tfrac{1}{2}+1;z) \overset{(B)}{=} \tfrac{1}{2\sqrt{z}}\mathbb{B}_z(\tfrac{1}{2},1-\tfrac{7}{8}) =$$

$$= \tfrac{1}{2\sqrt{z}}\mathbb{B}_z(\tfrac{1}{2},\tfrac{1}{8}) = \tfrac{1}{2\sqrt{z}}\mathbb{B}_z\big(\tfrac{1}{2},(\tfrac{1}{8}+1)-1\big) = \tfrac{1}{2\sqrt{z}}\,\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}-1). \tag{A.6}$$

Now, comparing the incomplete beta function in (A.5) with the one just obtained, it seems obvious to look for a way to express a $\mathbb{B}_z(a,b-1)$ in terms of $\mathbb{B}_z(a,b)$. We will achieve this combining (C) and (D) as follow

$$\mathbb{B}_z(a,b) = \tfrac{1}{a+b-1}\Big\{(a-1)z\tfrac{1}{(a-1)}\big[(1-z)^b z^{a-1} + (a+b-1)\mathbb{B}_z(a,b)\big] + (b-1)(1-z)\mathbb{B}_z(a,b-1)\Big\} =$$

$$= \tfrac{1}{a+b-1}\big\{z^a(1-z)^b + z(a+b-1)\mathbb{B}_z(a,b) + (b-1)(1-z)\mathbb{B}_z(a,b-1)\big\} =$$

$$= \tfrac{z^a(1-z)^b}{a+b-1} + z\mathbb{B}_z(a,b) + \Big(\tfrac{b-1}{a+b-1}\Big)(1-z)\mathbb{B}_z(a,b-1),$$

hence

$$\Big(\tfrac{b-1}{a+b-1}\Big)(1-z)\mathbb{B}_z(a,b-1) = (1-z)\mathbb{B}_z(a,b) - \tfrac{z^a(1-z)^{b-1}}{a+b-1},$$

and finally

$$\boxed{\mathbb{B}_z(a,b-1) = \Big(\tfrac{a+b-1}{b-1}\Big)\mathbb{B}_z(a,b) - \Big(\tfrac{1}{b-1}\Big)z^a(1-z)^{b-1}.} \tag{A.7}$$

So we have

$$\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}-1) = \Big(\tfrac{\frac{1}{2}+\frac{9}{8}-1}{\frac{9}{8}-1}\Big)\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}) - \big(\tfrac{9}{8}-1\big)^{-1}z^{\frac{1}{2}}(1-z)^{\frac{9}{8}-1} = 5\,\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}) - 8\,z^{\frac{1}{2}}(1-z)^{\frac{1}{8}},$$

and, from (A.6),

$$_2F_1(\tfrac{1}{2},\tfrac{7}{8},\tfrac{3}{2};z) = \tfrac{1}{2\sqrt{z}}\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}-1) = \tfrac{1}{2\sqrt{z}}\Big[5\,\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}) - 8\,z^{\frac{1}{2}}(1-z)^{\frac{1}{8}}\Big] = \tfrac{5}{2\sqrt{z}}\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}) - 4(1-z)^{\frac{1}{8}} \overset{z=1-\frac{R^2}{x^2}}{=}$$

$$= \tfrac{5}{2}\Big(\tfrac{x^2}{x^2-R^2}\Big)^{\frac{1}{2}}\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}) - 4\Big(1-1+\tfrac{R^2}{x^2}\Big)^{\frac{1}{8}} \overset{|x|\leqslant R}{=} \tfrac{5}{2}\Big(\tfrac{|x|}{\partial\sqrt{R^2-x^2}}\Big)\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}) - 4|x|^{-\frac{1}{4}}R^{\frac{1}{4}} =$$

$$= \tfrac{5}{2}\Big(\tfrac{|x|}{\partial\sqrt{R^2-x^2}}\Big)\big(\tfrac{\partial}{\partial}\big)\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8}) - 4|x|^{-\frac{1}{4}}R^{\frac{1}{4}} = -\tfrac{5}{2}\Big(\tfrac{|x|}{\sqrt{R^2-x^2}}\Big)\big[\partial\,\mathbb{B}_z(\tfrac{1}{2},\tfrac{9}{8})\big] - 4|x|^{-\frac{1}{4}}R^{\frac{1}{4}},$$

---

[12]See [33].

16

that is

$$_2\mathsf{F}_1\left(\tfrac{1}{2}, \tfrac{7}{8}, \tfrac{3}{2}; z\right) = -\tfrac{5}{2}\left(\tfrac{|x|}{\sqrt{R^2-x^2}}\right)\left[\mathsf{B}_z(\tfrac{1}{2}, \tfrac{9}{8})\right] - 4|x|^{-\frac{1}{4}}R^{\frac{1}{4}}. \tag{A.8}$$

Finally, substituting (A.8) in (A.4), we obtain

$$
\begin{aligned}
\mathsf{k}(x|R) &= \tfrac{2}{5}\sqrt{R^2-x^2}\left\{4R^{\frac{1}{4}} + |x|^{\frac{1}{4}}\left[-\tfrac{5}{2}\tfrac{|x|}{\sqrt{R^2-x^2}}\left[\mathsf{B}_z(\tfrac{1}{2}, \tfrac{9}{8})\right] - 4|x|^{-\frac{1}{4}}R^{\frac{1}{4}}\right]\right\} = \\
&= \tfrac{2}{5}\sqrt{R^2-x^2}\left\{4R^{\frac{1}{4}} - \tfrac{5}{2}|x|^{\frac{1}{4}+1}\tfrac{1}{\sqrt{R^2-x^2}}\mathsf{B}_z(\tfrac{1}{2}, \tfrac{9}{8}) - 4R^{\frac{1}{4}}\right\} = -|x|^{\frac{5}{4}}\mathsf{B}_z(\tfrac{1}{2}, \tfrac{9}{8}) = \mathsf{k}_{\mathsf{HM}}(x|R).
\end{aligned}
$$

and the equivalence is proven.

B   Separable Non–Linear Least Squares.

Let $\{(x_i, y_i)\}_{i\in\{1,\dots,n\}}$ denote the data points generated by the model:

$$Y_i = \mathsf{M}(\boldsymbol{\lambda}, \boldsymbol{\theta}, x_i) + \epsilon_i = \sum_{j=1}^{p}\lambda_j f_j(\boldsymbol{\theta}, x_i) + \epsilon_i = \langle\boldsymbol{\lambda}, \mathbf{f}(\boldsymbol{\theta}, x_i)\rangle_2 + \epsilon_i \quad \forall\, i \in \{1, \dots, n\},$$

where

- $f_j : \mathbb{R}^q \times \mathbb{R} \mapsto \mathbb{R}$ are <u>given</u> functions that depend (non–linearly) on $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ and $x \in \mathcal{X} \subseteq \mathbb{R}$, with Θ and $[\boldsymbol{\theta}, x]^{\mathsf{T}} \to f_j(\boldsymbol{\theta}, x)$ $\mathcal{X}$ open sets.

- $\mathbf{f}(\boldsymbol{\theta}, x) \triangleq [f_1(\boldsymbol{\theta}, x), \dots, f_p(\boldsymbol{\theta}, x)]^{\mathsf{T}} \in \mathbb{R}^p$,

- $\boldsymbol{\lambda} \in \Lambda \subseteq \mathbb{R}^p$ are parameters that occur linearly in the model with Λ open set,

- $\boldsymbol{\varepsilon} \triangleq [\epsilon_1, \dots, \epsilon_n]^{\mathsf{T}} \sim \mathcal{N}_n(\mathbf{0}, \mathbb{I}_n)$ where $\mathbb{I}_n$ denotes the $n \times n$ identity matrix.

At this point lets introduce another bit of notation,

- Given $\mathbf{A} \in \mathbb{R}^{n\times m}$ we shall denote its j–th column with $\mathbf{A}_{(:,j)}$.

- $\boldsymbol{\psi} \triangleq [\boldsymbol{\lambda}, \boldsymbol{\theta}]^{\mathsf{T}} \in \Lambda \times \Theta \triangleq \Psi \subseteq \mathbb{R}^{q+p}$,

- $\mathbf{F}(\boldsymbol{\theta}) \triangleq [\mathbf{f}(\boldsymbol{\theta}, x_1) \cdots \mathbf{f}(\boldsymbol{\theta}, x_n)] = [f_j(\boldsymbol{\theta}, x_i)]_{(i,j)\in\{1,\dots,n\}\times\{1,\dots,p\}} \in \mathbb{R}^{n\times p} \quad \forall\, \boldsymbol{\theta} \in \Theta$, and this implies the following representation:

$$\mathbb{E}_{\boldsymbol{\psi}}(\mathbf{Y}|\mathbf{x}) \triangleq \mathsf{M}(\boldsymbol{\lambda}, \boldsymbol{\theta}, \mathbf{x}) = [\langle\boldsymbol{\lambda}, \mathbf{f}(\boldsymbol{\theta}, x_1)\rangle, \dots, \langle\boldsymbol{\lambda}, \mathbf{f}(\boldsymbol{\theta}, x_n)\rangle]^{\mathsf{T}} = \mathbf{F}(\boldsymbol{\theta})\,\boldsymbol{\lambda} \in \mathbb{R}^n,$$

- $\mathbf{F}'_k(\boldsymbol{\theta}) \triangleq \frac{\partial \mathbf{F}(\boldsymbol{\theta})}{\partial\theta_k} = \left[\frac{\partial f_j(\boldsymbol{\theta}, x_i)}{\partial\theta_k}\right]_{(i,j)\in\{1,\dots,n\}\times\{1,\dots,p\}} \in \mathbb{R}^{n\times p} \quad \forall\,(k, \boldsymbol{\theta}) \in \{1, \dots, q\} \times \Theta$,

- $\mathbf{F}'(\boldsymbol{\theta}) \triangleq \frac{\partial \mathbf{F}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \mathsf{cat}(\mathbf{F}'_1(\boldsymbol{\theta}), \cdots, \mathbf{F}'_q(\boldsymbol{\theta}), 3) \in \mathbb{R}^{n\times p\times q}$, where $\mathsf{cat}(\cdot, d)$ denotes the MATLAB® *concatenate*–operator along the d–th dimension,

17

- $\mathbf{J}_{\mathsf{M},\boldsymbol{\lambda}}(\boldsymbol{\theta}) \triangleq \frac{\partial \mathsf{M}(\boldsymbol{\lambda},\boldsymbol{\theta},\mathbf{x})}{\partial \boldsymbol{\lambda}} = \mathbf{F}(\boldsymbol{\theta}) \in \mathbb{R}^{n \times p}$,

- $\mathbf{J}_{\mathsf{M},\boldsymbol{\theta}}(\boldsymbol{\psi}) \triangleq \frac{\partial \mathsf{M}(\boldsymbol{\lambda},\boldsymbol{\theta},\mathbf{x})}{\partial \boldsymbol{\theta}} = \left[\mathbf{F}'_1(\boldsymbol{\theta})\boldsymbol{\lambda} \cdots \mathbf{F}'_q(\boldsymbol{\theta})\boldsymbol{\lambda} \right] = \left[\sum_{j=1}^{p} \lambda_j \frac{\partial f_j(\boldsymbol{\theta},x_i)}{\partial \theta_j} \right]_{(i,k)\in\{1,,n\}\times\{1,...,q\}} \in \mathbb{R}^{n \times q}$,

- $\mathbf{J}_{\mathsf{M}}(\boldsymbol{\psi}) \triangleq [\mathbf{J}_{\mathsf{M},\boldsymbol{\theta}}(\boldsymbol{\psi}) \quad \mathbf{F}(\boldsymbol{\theta})] \in \mathbb{R}^{n \times (q+p)}$,

---

- [Residuals] $\mathbf{r}(\boldsymbol{\psi}) = [r_1(\boldsymbol{\psi}), \dots, r_n(\boldsymbol{\psi})]^{\mathsf{T}} \triangleq (\mathbf{y} - \mathbf{F}(\boldsymbol{\theta})\,\boldsymbol{\lambda}) \in \mathbb{R}^n \quad \forall \boldsymbol{\psi} \in \Psi$,

- [Jacobian] $\mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi}) \triangleq \frac{\partial \mathbf{r}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = -\mathbf{J}_{\mathsf{M}}(\boldsymbol{\psi}) \in \mathbb{R}^{n \times (q+p)} \quad \forall \boldsymbol{\psi} \in \Psi$,

- [Sum of Squares] $\ell(\boldsymbol{\psi}) = \frac{1}{2}\|\mathbf{r}(\boldsymbol{\psi})\|_2^2 = \frac{1}{2}\mathbf{r}(\boldsymbol{\psi})^{\mathsf{T}}\mathbf{r}(\boldsymbol{\psi}) \in \mathbb{R}$

- $\boldsymbol{\ell}'(\boldsymbol{\psi}) \triangleq \mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi})^{\mathsf{T}}\mathbf{r}(\boldsymbol{\psi}) \in \mathbb{R}^{q+p}$,

The (local) maximum log–likelihood estimate (i.e. the least squares estimate) is then defined as:

$$\widehat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}\in\mathbb{R}^{q+p}}{\arg\min}\, \ell(\boldsymbol{\psi}),$$

and satisfied the following system of $(q + p)$ equations in $(q + p)$ unknowns:

$$\boldsymbol{\ell}'\big(\widehat{\boldsymbol{\psi}}\big) = \mathbf{0}.$$

The *Gauss–Newton method* is widely used for solving non–linear least squares problems: given an approximation $\boldsymbol{\psi}^*$ to $\widehat{\boldsymbol{\psi}}$, the next approximation is found as $\boldsymbol{\psi}^* + \boldsymbol{\delta}_{\mathsf{GN}}$, where

$$\boldsymbol{\delta}_{\mathsf{GN}} \triangleq \underset{\boldsymbol{\delta}\in\mathbb{R}^{q+p}}{\arg\min} \frac{1}{2}\|\mathbf{r}(\boldsymbol{\psi}^*) + \mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi}^*)\,\boldsymbol{\delta}\|_2^2 \triangleq \underset{\boldsymbol{\delta}\in\mathbb{R}^{q+p}}{\arg\min} \widetilde{\ell}(\boldsymbol{\psi}^*,\boldsymbol{\delta}),$$

leading to

$$\big(\mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi})^{\mathsf{T}}\mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi})\big)\boldsymbol{\delta}_{\mathsf{GN}} = \mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi})^{\mathsf{T}}\mathbf{r}(\boldsymbol{\psi}). \tag{B.1}$$

This is often modified by the *Levenberg–Marquardt strategy*, where $\boldsymbol{\delta}_{\mathsf{GN}}$ is replaced by $\boldsymbol{\delta}_{\mathsf{LM}}$ determined by (note the possible *Bayesian* interpretation as a *ridge*–adjustment to regularize an ill–posed problem):

$$\big(\mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi})^{\mathsf{T}}\mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi}) + \mu\,\mathbb{I}_{q+p}\big)\boldsymbol{\delta}_{\mathsf{LM}} = \mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi})^{\mathsf{T}}\mathbf{r}(\boldsymbol{\psi}). \tag{B.2}$$

The damping parameter (hyper–parameter) is adjusted during iteration according to how well $\widetilde{\ell}(\boldsymbol{\psi}, \boldsymbol{\delta}_{\mathsf{LM}})$ approximates $\ell(\boldsymbol{\psi} + \boldsymbol{\delta}_{\mathsf{LM}})$, see [23]. A standard way to initialize the damping parameter is the following:

$$\mu^{(0)} = \tau \cdot \max\left\{ \big[\mathbf{J}_{\mathbf{r}}\big(\boldsymbol{\psi}^{(0)}\big)^{\mathsf{T}}\mathbf{J}_{\mathbf{r}}\big(\boldsymbol{\psi}^{(0)}\big)\big]_{(i,i)} \right\}_{i\in\{1,...,q+p\}},$$

where τ and $\boldsymbol{\psi}^{(0)}$ are known values.

B.1   Reduced Parameter Set.

Instead of treating the elements in $\boldsymbol{\lambda} \in \mathbb{R}^p$ as non–linear variables, we can replace the residuals $\mathbf{r}(\boldsymbol{\psi})$ defined above with

$$\widetilde{\mathbf{r}}(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{F}(\boldsymbol{\theta})\,\boldsymbol{\lambda}(\boldsymbol{\theta})) \in \mathbb{R}^n, \tag{B.3}$$

where, for any given $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\lambda}(\boldsymbol{\theta})$ is computed as the least square solution to

$$\mathbb{E}_{\psi}(\mathbf{Y}|\mathbf{x}) \triangleq \widetilde{\mathsf{M}}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta}). \tag{B.4}$$

In particular we can express it as the solution to the normal equations,

$$\underbrace{(\mathbf{F}(\boldsymbol{\theta})^{\mathsf{T}}\mathbf{F}(\boldsymbol{\theta}))}_{\triangleq \mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{p \times p}} \boldsymbol{\lambda}(\boldsymbol{\theta}) = \underbrace{\mathbf{F}(\boldsymbol{\theta})^{\mathsf{T}}\mathbf{y}}_{\triangleq \mathbf{b}(\boldsymbol{\theta}) \in \mathbb{R}^{p}} \quad \Leftrightarrow \quad \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta}) = \mathbf{b}(\boldsymbol{\theta}). \tag{B.5}$$

With the notation above and for each $k \in \{1, \ldots, q\}$, the corresponding Jacobian matrix is given by

$$[\mathbf{J}_{\tilde{\mathbf{r}}}(\boldsymbol{\theta})]_{(:,k)} \triangleq \frac{\partial \tilde{\mathbf{r}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} = -\underbrace{\mathbf{F}'_k(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta})}_{\triangleq [\mathbf{J}_{\tilde{\mathsf{M}}}(\boldsymbol{\theta})]_{(:,k)}} - \underbrace{\mathbf{F}(\boldsymbol{\theta})\boldsymbol{\lambda}'_k(\boldsymbol{\theta})}_{\triangleq -[\mathbf{H}(\boldsymbol{\theta})]_{(:,k)}} = -[\mathbf{J}_{\tilde{\mathsf{M}}}(\boldsymbol{\theta})]_{(:,k)} + [\mathbf{H}(\boldsymbol{\theta})]_{(:,k)} \in \mathbb{R}^n \tag{B.6}$$

Deriving (B.5) then we get,

$$\mathbf{A}'_k(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\lambda}'_k(\boldsymbol{\theta}) = \mathbf{b}'(\boldsymbol{\theta})$$

where $\boldsymbol{\lambda}'_k(\boldsymbol{\theta}) \in \mathbb{R}^p$ is the k–th column of

$$\boldsymbol{\lambda}'(\boldsymbol{\theta}) \triangleq [\boldsymbol{\lambda}'_1(\boldsymbol{\theta}) \cdots \boldsymbol{\lambda}'_q(\boldsymbol{\theta})] \in \mathbb{R}^{p \times q}$$

Omitting the argument $\boldsymbol{\theta}$, then we have

$$\begin{aligned} \mathbf{A}\boldsymbol{\lambda}'_k &= \mathbf{b}' - \mathbf{A}'_k\boldsymbol{\lambda} = \\ &= \mathbf{F}'^{\mathsf{T}}_k\mathbf{y} - (\mathbf{F}'^{\mathsf{T}}_k\mathbf{F} + \mathbf{F}^{\mathsf{T}}\mathbf{F}'_k)\boldsymbol{\lambda} = \mathbf{F}'^{\mathsf{T}}_k\mathbf{r} - \mathbf{F}^{\mathsf{T}}\mathbf{F}'_k\boldsymbol{\lambda} = \\ &= \mathbf{F}'^{\mathsf{T}}_k\mathbf{r} - \mathbf{F}^{\mathsf{T}}[\mathbf{J}_{\tilde{\mathsf{M}}}]_{(:,k)}, \quad \forall\, k \in \{1, \ldots, q\}. \end{aligned} \tag{B.7}$$

From this system of equations we can find $\boldsymbol{\lambda}'_k$ and insert it in (B.6).

Remarks:

1. Note that we have replaced the matrix $\mathbf{J}_{\mathbf{r}}(\boldsymbol{\psi}) \in \mathbb{R}^{n \times (q+p)}$ with $\mathbf{J}_{\tilde{\mathbf{r}}}(\boldsymbol{\theta}) \in \mathbb{R}^{n \times q}$.

2. Note that the residual $\tilde{\mathbf{r}}(\boldsymbol{\theta})$ are <u>orthogonal</u> to the range of $\mathbf{F}(\boldsymbol{\theta})$, i.e. $\mathbb{R}^p \ni \mathbf{F}(\boldsymbol{\theta})^{\mathsf{T}}\tilde{\mathbf{r}}(\boldsymbol{\theta}) = \mathbf{0}$. This implies that $\forall\, k \in \{1, \ldots, q\}$,

$$\underbrace{[\mathbf{J}_{\tilde{\mathbf{r}}}(\boldsymbol{\theta})^{\mathsf{T}}\tilde{\mathbf{r}}(\boldsymbol{\theta})]}_{(q \times n)\ (n \times 1)}_k = \tilde{\mathbf{r}}'_k(\boldsymbol{\theta})^{\mathsf{T}}\tilde{\mathbf{r}}(\boldsymbol{\theta}) = -(\mathbf{F}'_k(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta}) + \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\lambda}'_k(\boldsymbol{\theta}))^{\mathsf{T}}\tilde{\mathbf{r}}(\boldsymbol{\theta}) = -(\mathbf{F}'_k(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta}))^{\mathsf{T}}\tilde{\mathbf{r}}(\boldsymbol{\theta}).$$

   This shows that within the reduced parameter set framework, the right–hand sides in the analogs of (B.1) and (B.2) do <u>not</u> depend on $\boldsymbol{\lambda}'_k(\boldsymbol{\theta})$. To get the correct matrix on the left–hand side, however, we have to include the term $\mathbf{F}(\boldsymbol{\theta})\boldsymbol{\lambda}'_k(\boldsymbol{\theta})$.

▲

Some special cases that lead to simple expressions for the Jacobian deserve some attention:

$q = 1$: Nothing really special here, except that (B.6) has one contribution only.

$p = 1$: In this case,

- $\mathbf{F}(\boldsymbol{\theta}) \in \mathbb{R}^n$,
- $\mathbf{F}'(\boldsymbol{\theta}) \in \mathbb{R}^{n \times q}$ (squeezing the singular dimension, see 4.),
- $\mathbf{A}(\boldsymbol{\theta}) \triangleq A(\boldsymbol{\theta})$ and $\boldsymbol{\lambda}(\boldsymbol{\theta}) \triangleq \lambda(\boldsymbol{\theta})$ are <u>scalars</u>,

- $\boldsymbol{\lambda}'(\boldsymbol{\theta}) \in \mathbb{R}^{1 \times q}$ (or in $\mathbb{R}^q$ after squeezing).

Further, $\forall\, k \in \{1, \ldots, q\}$,

$$A\boldsymbol{\lambda}'_k = \mathbf{F}'^{\mathsf{T}}_k \mathbf{r} - \mathbf{F}^{\mathsf{T}}\mathbf{F}'_k \boldsymbol{\lambda} = \mathbf{F}'^{\mathsf{T}}_k (\mathbf{r} - \mathbf{F}\boldsymbol{\lambda}).$$

so we get,

$$\boldsymbol{\lambda}' = \frac{1}{A}\mathbf{F}'^{\mathsf{T}}(\mathbf{r} - \boldsymbol{\lambda}\mathbf{F}) \tag{B.8}$$

$$\mathbf{J}_{\tilde{r}} = -\boldsymbol{\lambda}\mathbf{F}' - \mathbf{F}\boldsymbol{\lambda}'^{\,\mathsf{T}}. \tag{B.9}$$

$p \equiv q \triangleq d, \;\; f_j(\boldsymbol{\theta}, x) = f_j(\theta_j, x)$: In this case $\forall\, j \in \{1, \ldots, d\}$, the function $f_j(\cdot, \cdot)$ depends only on θ_j and x, hence $\forall\, k \in \{1, \ldots, d\}$ the matrix $\mathbf{F}'_k(\boldsymbol{\theta}) \in \mathbb{R}^{n \times d}$ has only one <u>non–zero</u> column, $[\mathbf{F}'_k(\boldsymbol{\theta})]_{(:,k)}$. For this reason, $[\mathbf{F}'^{\mathsf{T}}_k \mathbf{r}]_k$ is the only non–zero element in $\mathbf{F}'^{\mathsf{T}}_k \mathbf{r} \in \mathbb{R}^d$, and the other contributions to the right–hand side in (B.7), is

$$\mathbf{F}^{\mathsf{T}}(\boldsymbol{\lambda}_k [\mathbf{F}'_k]_{(:,k)}).$$

Further, if for any vector $\mathbf{v} \in \mathbb{R}^d$, we define the diagonal matrix $\mathbf{D}[\![\mathbf{v}]\!] = \mathsf{diag}(\mathbf{v})$, then we can write (B.7) as a matrix equation

$$\mathbf{A}\boldsymbol{\lambda}' = \mathbf{D}[\![\mathbf{F}'^{\mathsf{T}}\mathbf{r}]\!] - \mathbf{F}^{\mathsf{T}}\underbrace{\mathbf{F}'\mathbf{D}[\![\boldsymbol{\lambda}]\!]}_{\equiv\, \mathbf{J}_{\widetilde{\mathsf{M}}}}, \tag{B.10}$$

and (B.6) takes the form:

$$\mathbf{J}_{\tilde{r}} = -\mathbf{F}'\mathbf{D}[\![\boldsymbol{\lambda}]\!] - \mathbf{F}\boldsymbol{\lambda}'. \tag{B.11}$$

B.2 Computational Aspects.

For reasons of accuracy, is recommended that $\boldsymbol{\lambda}(\boldsymbol{\theta})$ (see B.4) is computed via <u>orthogonal transformation</u>: determine $\mathbf{Q}_\perp(\boldsymbol{\theta})$ with orthogonal columns so that

$$\mathbf{Q}^{\mathsf{T}}_\perp(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta}) = \mathbf{R}(\boldsymbol{\theta}),$$

where $\mathbf{R}(\boldsymbol{\theta}) \in \mathbb{R}^{p \times p}$ is <u>upper</u> triangular. Then $\boldsymbol{\lambda}(\boldsymbol{\theta})$ is determined by <u>back–substitution</u> in

$$\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta}) = \mathbf{Q}^{\mathsf{T}}_\perp(\boldsymbol{\theta})\mathbf{y}.$$

It can be shown that

$$\mathbf{A} = \mathbf{F}^{\mathsf{T}}\mathbf{F} = \mathbf{R}^{\mathsf{T}}\mathbf{R},$$

and this factorization of $\mathbf{A}$ can be used in the solution of (B.7).

As regards computational efforts per iteration step, lets consider <u>small</u> problems with <u>dense</u> matrices. First of all, lets refresh the following definitions,

- $\mathbb{R}^{n \times (q+p)} \ni \mathbf{J}_r(\boldsymbol{\psi}) = -\mathbf{J}_\mathsf{M}(\boldsymbol{\psi}) = -[\mathbf{J}_{\mathsf{M},\boldsymbol{\theta}}(\boldsymbol{\psi})\;\; \mathbf{J}_{\mathsf{M},\boldsymbol{\lambda}}(\boldsymbol{\psi})] = -\mathsf{cat}\big([\mathbf{F}'_1(\boldsymbol{\theta})\boldsymbol{\lambda} \cdots \mathbf{F}'_q(\boldsymbol{\theta})\boldsymbol{\lambda}],\; \mathbf{F}(\boldsymbol{\theta}), 1\big)$,

- $\mathbb{R}^{n \times q} \ni \mathbf{J}_{\tilde{r}}(\boldsymbol{\theta}) = -\mathbf{J}_{\widetilde{\mathsf{M}}}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta}) = -[\mathbf{F}'_1(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta}) \cdots \mathbf{F}'_q(\boldsymbol{\theta})\boldsymbol{\lambda}(\boldsymbol{\theta})] + [\mathbf{F}(\boldsymbol{\theta})\boldsymbol{\lambda}'_1(\boldsymbol{\theta}) \cdots \mathbf{F}(\boldsymbol{\theta})\boldsymbol{\lambda}'_q(\boldsymbol{\theta})]$.

Given $\boldsymbol{\lambda}$, both formulations need the evaluation of the matrices $\mathbf{F}$, $\mathbf{F}'$ and $\mathbf{J}_{\mathsf{M},\boldsymbol{\theta}}$ or its equivalent $\mathbf{J}_{\widetilde{\mathsf{M}}}$. The extra work needed to compute $\boldsymbol{\lambda}(\boldsymbol{\theta})$ as just described is about $2p^2(n - \frac{1}{3}p)$ flops, and each $\boldsymbol{\lambda}'_k(\boldsymbol{\theta})$ costs about $2p(2n + p)$ flops. Thus the <u>extra</u> work in computing $\mathbf{J}_{\tilde{r}}(\boldsymbol{\theta})$ instead of $\mathbf{J}_r(\boldsymbol{\psi})$ is about $2p(n(p + 3q) + p(q - \frac{1}{3}p))$ flops.

Some work is saved, however, in the solution of (B.2). If we form the matrix and solve the system via *Cholesky factorization*, then the cost for computing $\boldsymbol{\delta}_{\mathsf{LM}}$ is about $(q + p)^2(n + \frac{1}{3}(q + p))$ flops. With the reduced parameter set, we replace $(q + p)$ by q, and the <u>overall</u> extra cost per iteration step for the reduced algorithm is about $p(n(p + 4q) + pq - q^2 - p^2)$ flops. In many cases, this will be small compared to the effort involved in the calculation of the qp elements in $\mathbf{F}(\boldsymbol{\theta})$ and nq elements in $\mathbf{J}_{\mathsf{M},\boldsymbol{\theta}}(\boldsymbol{\theta})$. Also in the second and third special cases above, the extra work is somewhat smaller.

20

References

[1] Wakker, B.P. and van Woerden, H. *High-Velocity Clouds*. Annual Review of Astronomy and Astrophysics, 35, pp. 217-266, 1997.

[2] Bradley W. Carroll and Dale A. Ostlie. *An Introduction to Modern Astrophysics*. Addison–Wesley, 1996.

[3] Ronald W. Butler and Andrew T. A. Wood. *Laplace Approximations for Hypergeometric Functions with a Matrix Argument*. Annals of Statistics, Vol. 30, No. 4, August 2002.

[4] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[5] A. C. Davinson and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, 1999.

[6] R. Beran and L. Dünmbgen. *Modulation of Estimators and Confidence Sets*. The Annals of Statistics, 1998.

[7] B. Vidakovic. *Statistical Modelling by Wavelets*. Wiley-Interscience, 1999.

[8] J. Fan and I. Gijbels, *Local Polynomial Modelling and its Applications*. Chapman&Hall, 1996.

[9] Hans, C.M. and van Dyk, D.A. *Accounting for absorption line in high energy spectra*. In "Statistical Challenges in Modern Astronomy III", (eds. E. Feigelson and G. Babu), pp. 429-430, Springer-Verlag, New York, 2003.

[10] Sourlas, E., van Dyk, D.A., Kashyap, V., Drake, J.J. and Pease, D. *Bayesian spectral analysis of metal abundance deficient stars*. Tech. Report, 2004.

[11] Sourlas, E., van Dyk, D.A., Kashyap, V., Drake, J.J. and Pease, D. *Bayesian spectral analysis of "MAD" stars*. In "Statistical Challenges to Modern Astronomy III", (eds. E. Feigelson and G. Babu), pp. 429-430, Springer-Verlag, New York, 2003.

[12] van Dyk, D.A., Connors, A., Kashyap, V. and Siemiginowska, A. *Analysis of energy spectra with low photon counts via Bayesian posterior simulation*. The Astrophysical Journal, 548, pp. 224-243, 2001.

[13] van Dyk, D.A. and Hans, C.M. *Accounting for absorption lines in images obtained with the Chandra X-ray Observatory*. In "Spatial Cluster Modelling", (eds. D. Denison and A. Lawson), pp. 175-198, CRC Press, London, 2002.

[14] van Dyk, D.A. and Kang, H. *Highly structured hierarchical models for spectral analysis in high energy astrophysics*. Statistical Science, to appear, 2003.

[15] Bruzual, G.A. and Charlot, S. *Spectral evolution of stellar populations using isochrome synthesis*. The Astrophysical Journal, 405, pp. 538-553, 1993.

[16] Seber and Wild *Nonlinear Regression*. Wiley, 1988.

[17] Li, L., *DNA Sequencing and Parametric Deconvolution*. Statistica Sinica, 2001.

[18] Li, L. and Speed, T. P., *Parametric Deconvolution of Positive Spike Trains*. Ann. Statist., Vol. 28, No. 5, 1279-1301, 2000.

[19] Li, L. and Speed, T. P., *Deconvolution of Sparse Positive Spikes: is it ill posed?* Tech. Report, 2000.

[20] Kaufman, L. *A Variable Projection Method for Solving Separable Non–Linear Least Square Problems*. BIT, 15, 49–57, 1975.

[21] Kaufman, L. and Gay, D. *Tradeoffs in Algorithms for Separable and Block Separable Non–Linear Least Squares*. IMACS 91, Proceedings of the 13th World Congress on Computational and Applied Mathematics, Ed. R. Vichnevetsky and J. J. H. Miller, pp. 157–158, Criterion Press, Dublin, 1991.

[22] Kaufman, L., Sylvester, G. and Wright, M. *Structured Linear Least–Squares Problems in System Identification and Separable Non–Linear Data Fitting.* SIAM, J. Opt. 4 (4), 847–871, 1994.

[23] Nielsen, H. B. *Damping Parameter in Marquardt's Method.* IMM, DTU. Report IMM–REP–1999–05, 1999. Available at http://www.imm.dtu.dk/$\thicksim$hbn/publ/TR9905.ps.Z.

[24] Ruhe, A. and Wedin, P. Å. *Algorithms for Separable Non–Linear Least Squares Problems.* SIAM, Review 22 (3), 318–337, 1980.

[25] Sagara, N. and Fukushima, M. *A Continuation Method for Solving Separable Non–Linear Least Squares.* Journ. Comp. and Appl. Math. 10 (2), 157–161, 1984.

[26] Zhu Z.Y., Tang N., Wei B.C. *On Confindence Regions of Semiparametric Non–Linear Regression Models (A geometric approach).* Acta Mathematica Scientia, 20(1), pp. 68–75, 2000.

[27] Zhu, Z.Y. and Wei, B.C. *Asymptotic efficiency estimation in semi–parametric non–linear regression models.* J. Appl. Math. of Chinese Univ., 14(B), pp. 43–51, 1999.

[28] Severini, T.A. and Wong H.W. *Profile likelihood and conditionally parametric models.* Ann. Statist., 20, pp. 1768–1802, 1992.

[29] Hamilton, D.C., Watts, D.G. and Bates, D.C. *Accounting for intrinsic non-linearity in non–linear regression parameter inference regions.* Ann. Statist., 10, pp. 386–393, 1982.

[30] van der Vaart A.W. *Asymptotic Statistics.* Cambridge Unversity Press, 1998.

[31] van der Vaart A.W. *Semiparametric Statistics, Ecole d'Ete de Saint-Flour 1999.* In "Lectures on Probability Theory and Statistics", Springer Verlag, 2002. Available at http://www.math.vu.nl/sto/publications/2000-1.eps.

[32] Hardle W., Liang H. and Gao J. *Partially Linear Models.* Physica Verlag, 2000. Available at http://www.quantlet.de/scripts/plm/plm.html.

[33] Abramowitz, M. and Stegun, I. A. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, pp. 555-566, 1972.
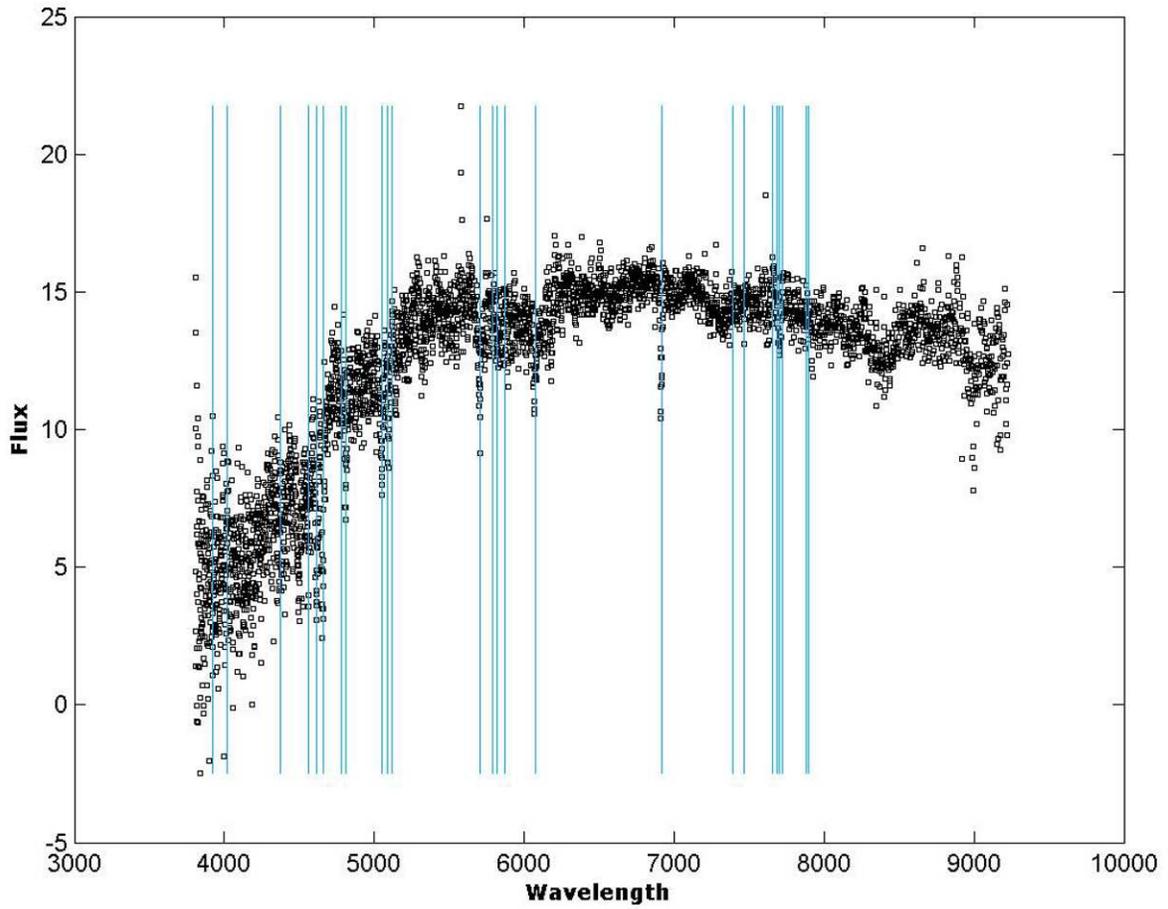
Figure 1: Plot of a spectrum coming from the SDSS with line positions superimposed.
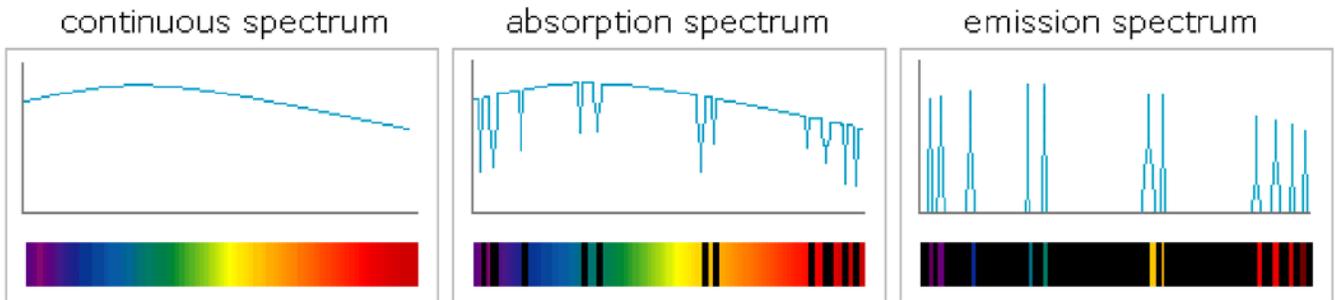
Figure 2: Example of continuum, emission and absorption features.
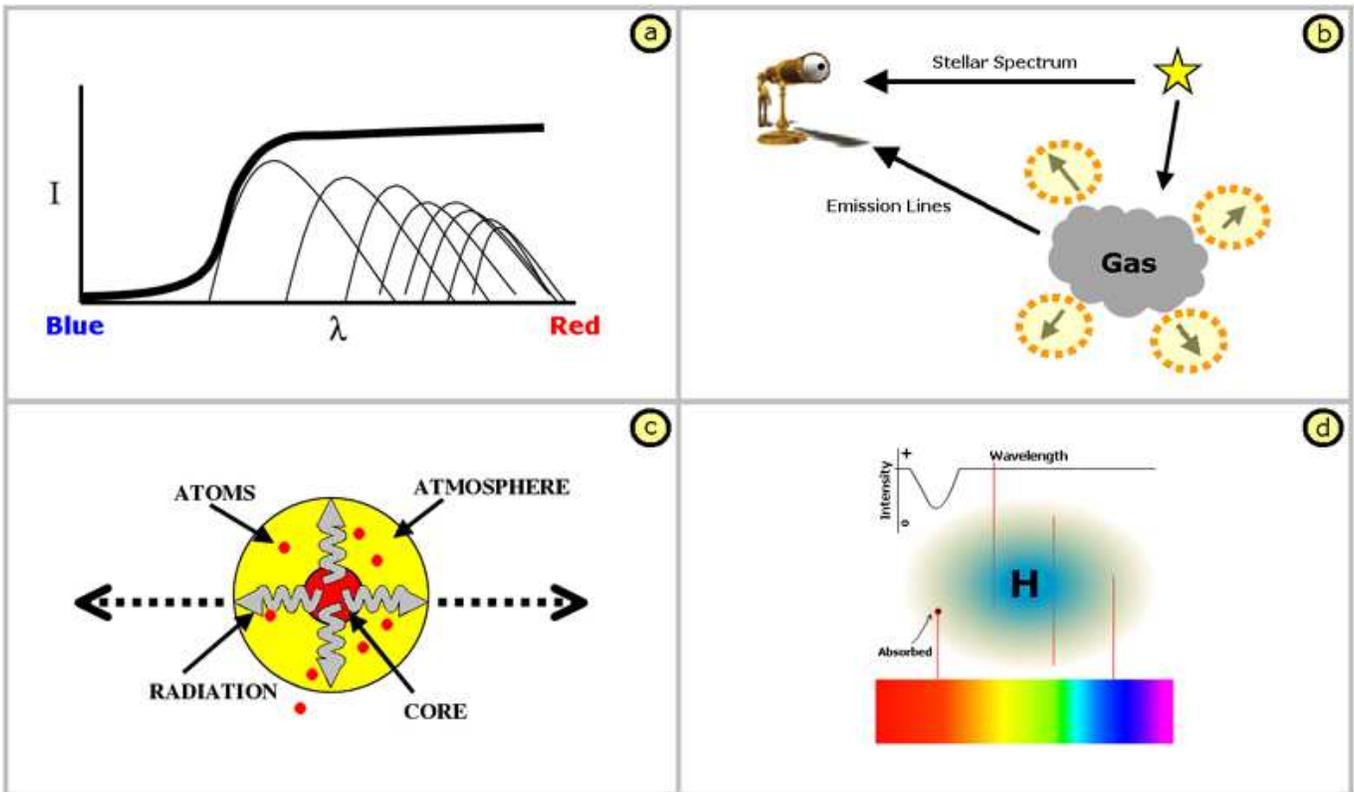
Figure 3: a) How the continuum is generated. b) Pictorial representation of the mechanism behind the presence of an emission line. c) and d) These figures exemplify two of the possible mechanisms behind the presence of an absorption line.
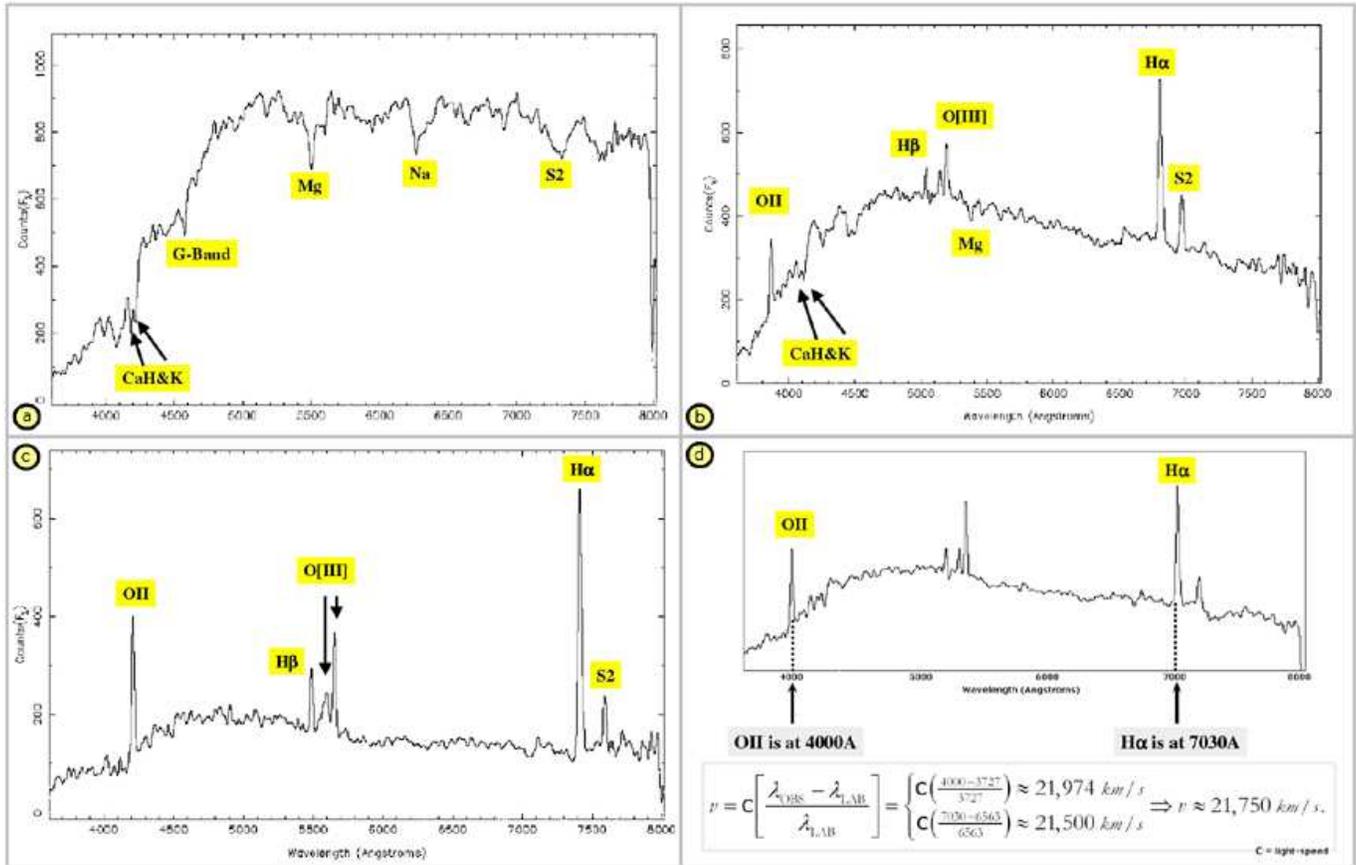
Figure 4: a)Typical spectrum of an **elliptical** galaxy. Strong absorption lines due to metals in the stellar atmospheres of a mostly low luminosity stellar population. No evidence of any emission lines and hence no young stars and no gas. b) Typical spectrum of a **spiral** galaxy. Some emission and some absorption indicating both a young and old stellar population. c) Typical spectrum of an **irregular** galaxy. A strong emission-line spectrum indicating many hot young stars heating the gas which is re-radiating at specific wavelengths depending on its the chemical composition. d) In this example we conclude that the galaxy is receding from us approximatively at $21,750$ km/s just looking at the offset of two strong emission lines.

25
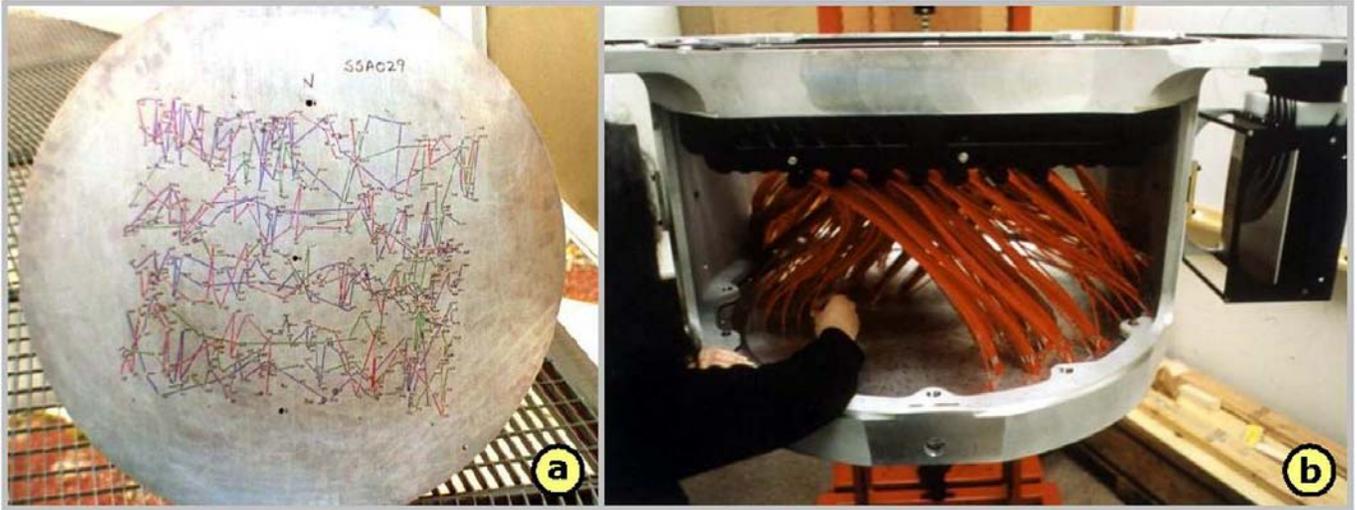
Figure 5: a) The aluminium plug plate. For each hole, there is a corresponding object in the sky waiting to have its spectrum observed. b) A plug plate that has been drilled, fitted with optical fibers, and mounted on the spectrograph assembly. The entire package will be flipped upside down and mounted to the underside of the 2.5 meter telescope at its focal plane.
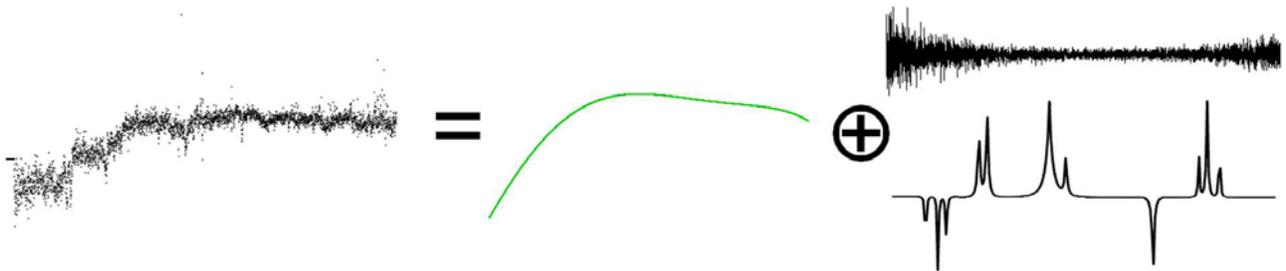
Figure 6: Basic modelling approach.
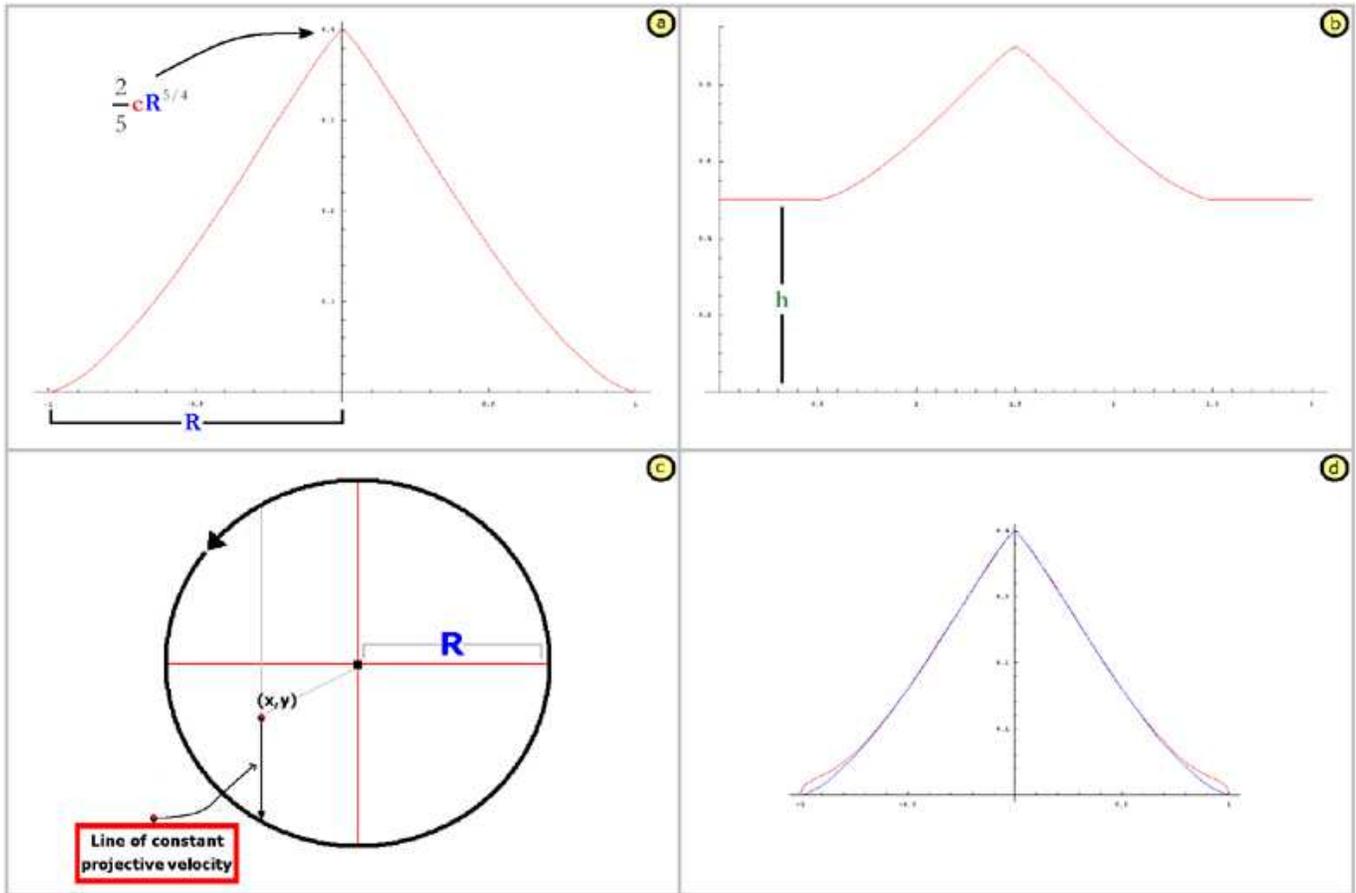
Figure 7: a) Plot of a spike centered at $x = 0$ with parameters $R = 9$ and $c = 2$. b) Plot of a spike added to a constant h, centered at $x = 1000$ with parameters $R = 6$, $c = 2$ and $h=27$. c) Cartoon of the physical reasoning behind the proposed profile function. d) Plot of the Taylor approximation for the profile function.
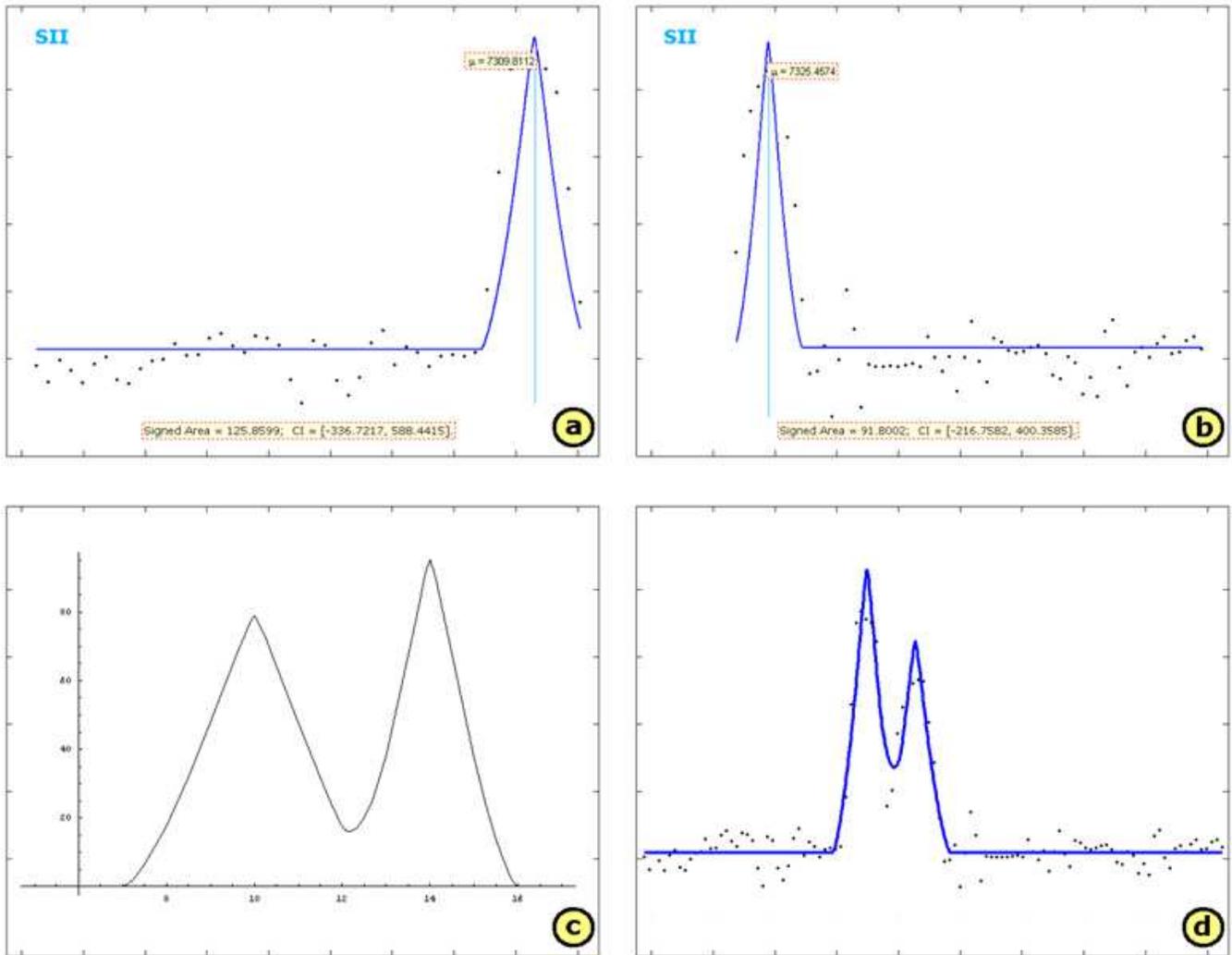
Figure 8: a) and b) These figures show two successive fitted profiles that collide, i.e. two profiles such that their central zones have a non–empty overlap. c) The mixture of profile functions that we actually fit in such a case. d) The final fit obtained that recovers the needed structure.
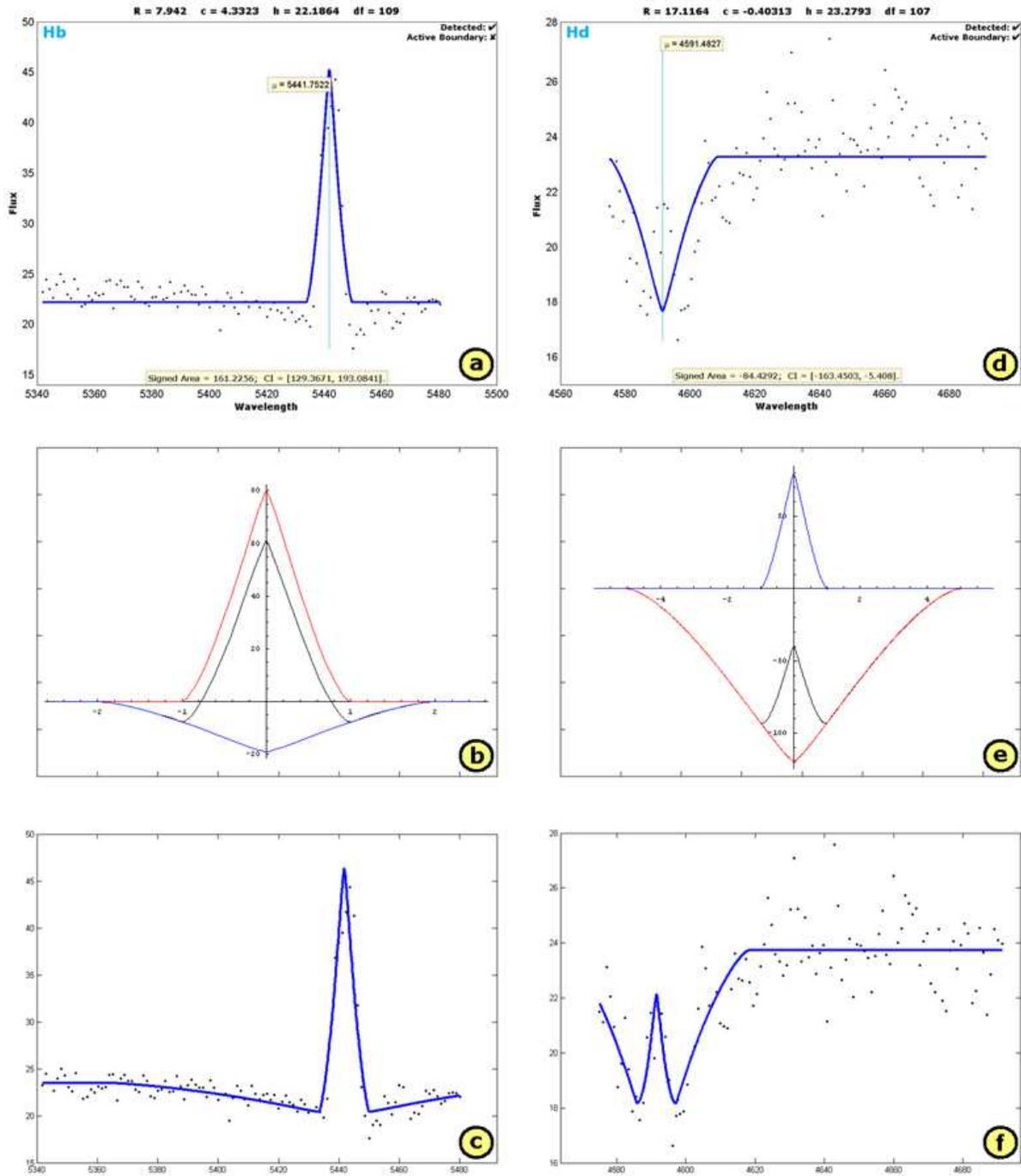
Figure 9: a) and d) These figures show the typical results obtained by fitting the usual "mono–profile" to lines associated to hydrogen doublets. b) and e) The *discordant* profiles that we will fit to hydrogen doublets. c) and f) The two final fits are obtained adopting the system of *discordant* profiles.
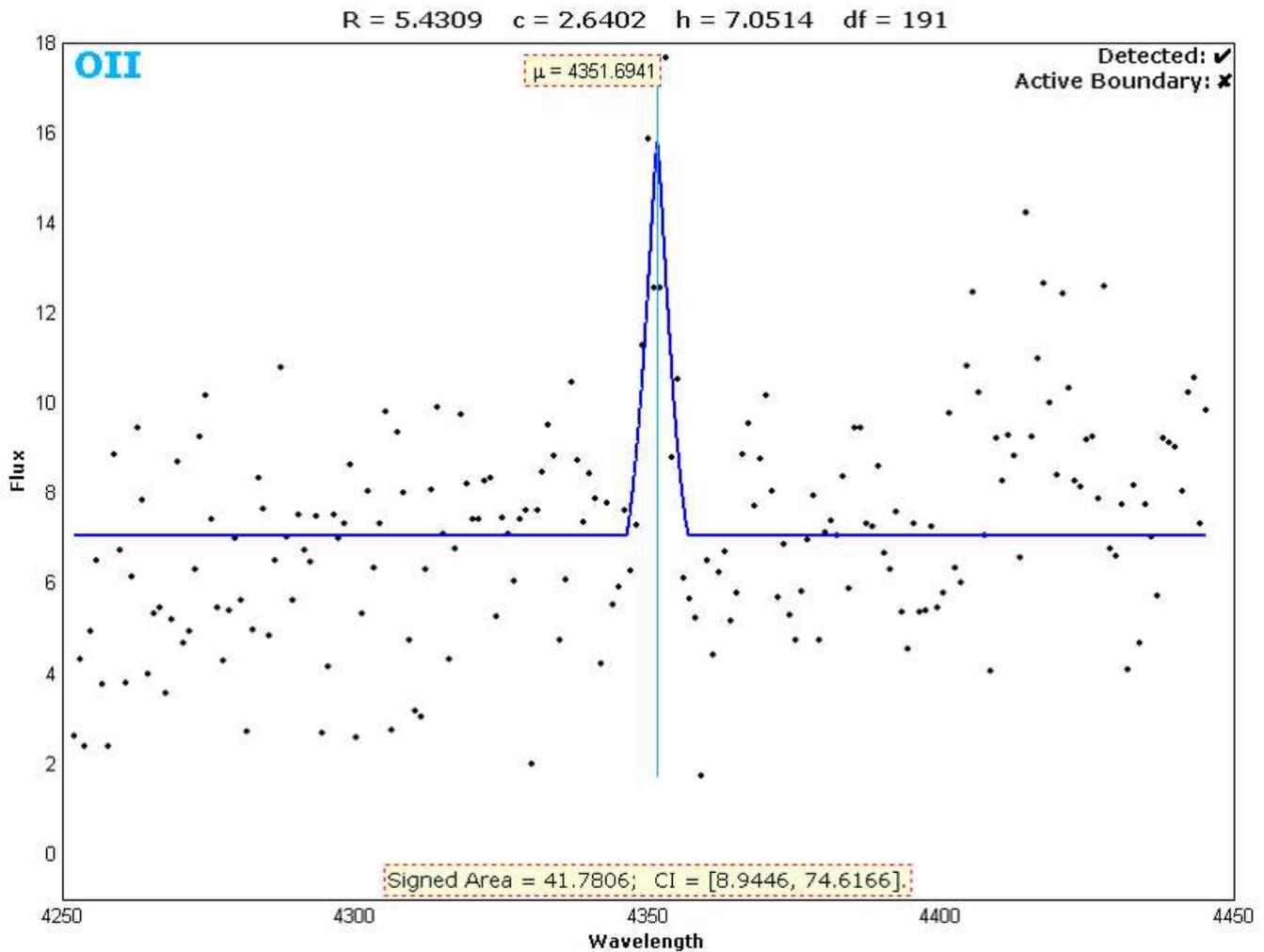
Figure 10: Typical output of our FPA procedure. The pointwise estimates of the spike parameters are on top, whereas the box on bottom contains the pointwise estimate of the area under the spike and the corresponding confidence interval. Note that in this case, as highlighted by the two flags on the top–right corner, we have a detection (the confidence interval does not contain zero) with no boundary problem, i.e. we are not at the boundary of the parameter space.