Parametric Recovery of Line-of-Sight Velocity Distributions from Absorption-Line Spectra of Galaxies via Penalized Likelihood

MICHELE CAPPELLARI

Leiden Observatory, Postbus 9513, 2300 RA Leiden, The Netherlands; cappellari@strw.leidenuniv.nl

AND

ERIC EMSELLEM

Centre de Recherche Astronomique de Lyon, 9 Avenue Charles André, 69230 Saint Genis Laval, France; emsellem@obs.univ-lyon1.fr

ABSTRACT. We investigate the accuracy of the parametric recovery of the line-of-sight velocity distribution (LOSVD) of the stars in a galaxy while working in pixel space. Problems appear when the data have a low signal-to-noise ratio or the observed LOSVD is not well sampled by the data. We propose a simple solution based on maximum penalized likelihood, and we apply it to the common situation in which the LOSVD is described by a Gauss-Hermite series. We compare different techniques by extracting the stellar kinematics from observations of the barred lenticular galaxy NGC 3384 obtained with the SAURON integral-field spectrograph.

1. INTRODUCTION

The dynamics of stars in a galaxy is uniquely defined by the distribution function and the potential in which the stars move. From the many stable equilibrium configurations for collisionless systems that can, in principle, be constructed, only some of them are actually observed in our universe. This constitutes a "fossil record" of galaxy formation. It is still not clear what observable constraints are needed to recover both the distribution function and the gravitational potential of a stellar system (but see, e.g., Dejonghe & Merritt 1992 for spherical geometry), but simple dimensionality arguments imply that this should not be possible without the knowledge of the full line-of-sight velocity distributions (LOSVD) at all spatial positions on the galaxy image on the sky. It is therefore useful to explore how the LOSVD can be recovered from the observations.

Considering galaxies as pure stellar systems, the spectrum observed at a certain sky position is a (luminosity-weighted) sum of individual stellar spectra redshifted according to their LOS velocities. If one makes the assumption that the spectrum of all stars is given by a single template, then it simply reduces to the convolution between that spectrum and the LOSVD, which can then be retrieved by solving the inverse problem, i.e., deconvolving the spectra using the template. Deconvolution is an intrinsically ill-conditioned problem that amplifies noise and measurement errors. Because of the special care required in the inversion, many techniques have been developed in the last 30 years to recover the LOSVD from the data. The evolution in the methods has been mainly driven by the gradual improvement in the observational techniques and the signal-to-noise ratio (S/N) of the data, and by the steady increase in the available computational speed.

Early methods mostly used Fourier-based techniques that allowed the LOSVD to be recovered quickly from a deconvolution process, and in some cases included techniques to reduce the effect of template mismatch (Simkin 1974; Sargent et al. 1977; Tonry & Davis 1979; Franx & Illingworth 1988; Bender 1990; Statler 1995). More recently, methods have shifted towards fitting the LOSVD directly in pixel space (Rix & White 1992; Kuijken & Merrifield 1993; van der Marel 1994; Saha & Williams 1994; Merritt 1997; Gebhardt et al. 2000; Kelson et al. 2000). The reasons for this are (1) in pixel space it becomes easy to exclude gas emission lines or bad pixels from the fit and take continuum matching directly into account; (2) current computers can accommodate the larger computational cost involved; (3) the availability of libraries with high spectral resolution stellar and galaxy spectra allows the template to be carefully matched to the observed galaxy spectrum (e.g., Emsellem et al. 2004). See de Bruyne et al. (2003) for a more detailed historical overview of the various methods.

The different techniques can be further subdivided according to whether the LOSVD is derived in a nonparametric way (in practice computed on a small set of discrete values) or parametrically, as a function of a limited number of parameters. In the latter case the Gauss-Hermite parametrization by van der Marel & Franx (1993) and Gerhard (1993) is essentially always adopted (but see, e.g., Zhao & Prada 1996 for an alternative). However, even when the LOSVD is determined nonparametrically, the Gauss-Hermite parameters are still generally used to present the result in an easily understandable way.

In this paper we study again the problem of recovering, while working in pixel space, an LOSVD described by the Gauss-Hermite parametrization. Compared to the nonparametric case,

the process and the estimation of measurement errors are simplified. In § 2 we describe the general problem. In § 3 we discuss different approaches to the extraction, we find that special care has to be taken when the LOSVD is undersampled by the data or the S/N is low, and we present a solution based on the maximum penalized likelihood formalism. In § 4 we draw some conclusions.

2. FORMULATION OF THE PROBLEM

The parametric recovery of the LOSVD in pixel space starts with creating a model galaxy spectrum $G_{\mathrm{mod}}(x)$ by convolving a template spectrum $T(x)$ by a parametrized LOSVD. Both the object and the template spectra are rebinned in wavelength to a linear scale in $x = \ln \lambda$, while usually preserving the number of spectral pixels. The best-fitting parameters of the LOSVD are determined by minimizing the χ^2, which measures the agreement between the model and the observed galaxy spectrum $G(x)$ over the set of N good pixels:

$$\chi^2 = \sum_{n=1}^{N} r_n^2, \qquad (1)$$

where the residuals are defined as

$$r_n = \frac{G_{\mathrm{mod}}(x_n) - G(x_n)}{\Delta G(x_n)}, \qquad (2)$$

where $\Delta G(x_n)$ is the measurement error on $G(x_n)$.

More specifically, the following model is adopted for the galaxy spectrum:

$$G_{\mathrm{mod}}(x) = \sum_{k=1}^{K} w_k [B * T_k](x) + \sum_{l=0}^{L} b_l \mathcal{P}_l(x) \qquad w_k \geq 0, \quad (3)$$

where T_k is in general a library of K galaxy or stellar templates, $B(x) = \mathcal{L}(cx)$ is the broadening function, where $\mathcal{L}(v)$ is the LOSVD, c is the speed of light, and $*$ denotes convolution. The $\mathcal{P}_l(x)$ are here chosen to be the Legendre polynomials of order l and account for low-frequency differences in shape between the galaxy and the templates. For each given $\mathcal{L}(v)$, the optimization of χ^2 is a bounded-variables linear least-squares problem for the weights $(w_1,\ldots, w_K, b_0,\ldots, b_L)$, which can be solved, e.g., with the specific BVLS algorithm by Lawson & Hanson (1995) or as a quadratic programming problem. Here we are interested in the determination of the parameters defining $\mathcal{L}(v)$, and in what follows we will assume that the weights of equation (3) are always optimized in this way. Multiplicative polynomials can also be included in the fit (see Kelson et al. 2000) without affecting the discussion that follows.

Following van der Marel & Franx (1993) and Gerhard (1993), it has become standard to expand the LOSVD as a Gauss-Hermite series

$$\mathcal{L}(v) = \frac{e^{-(1/2)y^2}}{\sigma\sqrt{2\pi}} \left[1 + \sum_{m=3}^{M} h_m H_m(y) \right], \qquad (4)$$

where $y = (v - V)/\sigma$ and H_m are the Hermite polynomials. With these definitions the minimization of the χ^2 in equation (1) is a nonlinear least-squares optimization problem for the M parameters $(V, \sigma, h_3, \ldots, h_M)$. Least-squares problems can be solved much more efficiently than general ones by using specific algorithms that require the user to provide the residuals r_n of equation (2) to compute explicitly the Hessian matrix of the χ^2 merit function (see, e.g., Press et al. 1992, § 15.5). Here we will use the MINPACK[1] implementation (Moré, Garbow, & Hillstrom 1980) of the Levenberg-Marquardt method for nonlinear least-squares problems.

3. DISCUSSION

In this section we compare three different approaches to the determination of the best-fitting parameters of the LOSVD in equation (4). We explain the limitations of the different methods and finally select the last one as the optimal choice. We do not address here the template mismatch issue, which we assume is minimized by the choice of an optimal library of templates in equation (3), as in Emsellem et al. (2004).

To explain the characteristics of the three methods, we will use each of them to extract a realistic but known LOSVD observed by Scorza & Bender (1995) at $r = 17''$ along the major axis of the S0 galaxy NGC 3115 (their Fig. A.2). This LOSVD (Fig. 1) is representative of the one observed in a number of galaxies and can be very well described by a double-Gaussian parametrization

$$F(v) = \sum_{j=1}^{2} I_j \exp\left[-\frac{(v - V_j)^2}{2 \sigma_j^2} \right], \qquad (5)$$

with parameters $I_1 = 0.041$, $I_2 = 0.032$, $V_1 = 48.7$ km s^{-1}, $V_2 = -77.3$ km s^{-1}, $\sigma_1 = 70.0$ km s^{-1}, and $\sigma_2 = 130.0$ km s^{-1} (the LOSVD was shifted to zero mean velocity).

The best fit to $F(v)$ using a fourth-order Gauss-Hermite series (eq. [4]) is obtained with parameters $V = 0.0$ km s^{-1}, $\sigma = 114.8$ km s^{-1}, $h_3 = -0.150$, and $h_4 = 0.036$, while the best-fitting Gaussian has $V = -2.3$ km s^{-1} and $\sigma = 118.9$ km s^{-1}.

3.1. Fitting (V, σ) First

In an ideal situation in which the LOSVD is perfectly sampled by the data, the $(h_3, \ldots, h_M)$ parameters of the LOSVD (eq. [4]) are essentially uncorrelated to (V, σ). Therefore, one expects the best-fitting parameters to change little, irrespective of whether (V, σ) are fitted first or together with the other pa-
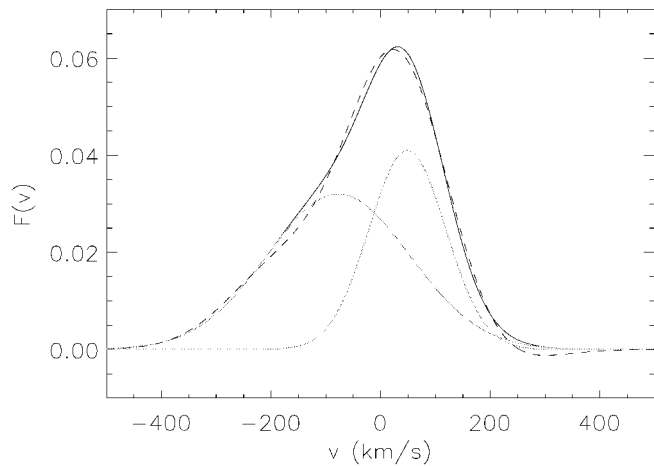

Fig. 1.—Double-Gaussian representation (eq. [5]) of the LOSVD observed by Scorza & Bender (1995) at 17″ along the major axis of the S0 galaxy NGC 3115 (*solid line*). The two individual Gaussian components are shown with the lighter lines, while the best-fitting fourth-order Gauss-Hermite parametrization is plotted with the dashed line.
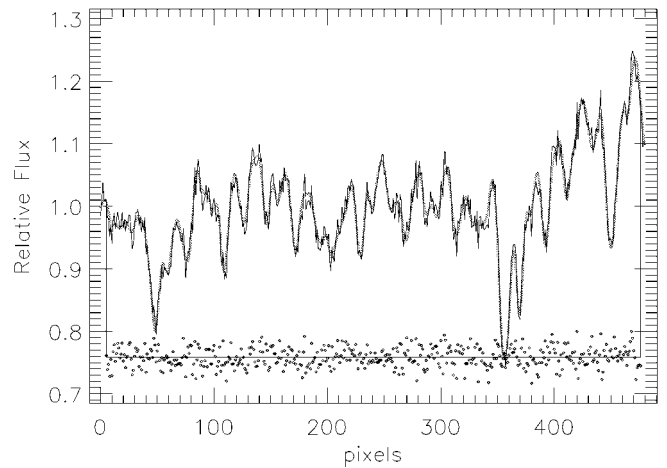

Fig. 2.—Logarithmically rebinned galaxy model spectrum (*thin noisy line*) convolved with the LOSVD of Fig. 1 and with added Poissonian noise at a S/N = 60. The dots represent the residual difference between this model and the best-fitting template (*thicker gray line*). The spectrum covers the wavelength range 4800–5380 Å, with a pixel scale of 60 km s^{-1}, and includes prominent absorption features of Hβ, Mg b, and Fe5270.

rameters. To the lowest order, the difference between the two approaches is given by eq. (10) of van der Marel & Franx (1993).

An advantage of fitting (V, σ) first is that the Gauss-Hermite parameters that one obtains coincide with the true Gauss-Hermite moments integrated over the LOSVD (see van der Marel & Franx 1993 for a detailed discussion). This also means that the value of the parameters $(h_3, \ldots, h_M)$ does not depend on the adopted number M of terms in the expansion. This precise relation between the LOSVD and the extracted kinematical parameters can be very useful for the dynamical modeling (e.g., Rix et al. 1997).

Moreover, when only a single template is adopted ($K = 1$ in eq. [3]), for each pair of (V, σ) values the best-fitting value of all the remaining parameters $(w_1, h_3, \ldots, h_M, b_0, \ldots, b_L)$ can be determined linearly. These ideas led van der Marel (1994) to design an efficient method to fit the parameters of the LOSVD in pixel space, where (V, σ) are fitted first and the remaining parameters are linearly expanded at the best-fitting (V, σ) solution.

For extracting the stellar kinematics of the 72 galaxies of the SAURON survey (de Zeeuw et al. 2002), we needed to operate in pixel space to be able to deal with the contamination due to emission-line gas present in most of the observed objects (Emsellem et al. 2004). The speed of the van der Marel (1994) pixel-fitting algorithm was an attractive feature, given the large number (~200,000) of independent spectra from the survey and the need to compute accurate errors by Monte Carlo simulations. However, the observed LOSVDs are not always well sampled by the SAURON pixel scale of $dv \approx 60$ km s^{-1}. The Nyquist critical frequency for h_4 is ~$\sigma/2$. This means that undersampling becomes a problem for the derivation of the first four Gauss-Hermite moments of the LOSVD when the ob-

served velocity dispersion is of the order of ~2 pixels (for SAURON, 120 km s^{-1}).

To test the accuracy of the recovery of the Gauss-Hermite moments using this method when undersampling becomes significant, we first created a realistic model spectrum by fitting a (logarithmically rebinned) library of Vazdekis (1999) galaxy model spectra to the average SAURON spectrum of the barred lenticular galaxy NGC 3384. This spectrum was oversampled by smoothly interpolating it on a 30 times finer spectral grid and was subsequently convolved with the LOSVD of Figure 1. It was integrated over the SAURON pixels, in the wavelength range 4800–5380 Å, and Poissonian noise was finally added to represent the simulated galaxy spectrum $G(x)$. The template $T(x)$ was obtained by integrating the original oversampled spectrum over the SAURON pixels. In Figure 2 we show an example of such a spectrum, with noise added at S/N = 60.

We subsequently tried to recover the true Gauss-Hermite moments by fitting (V, σ) first and expanding $(h_3, \ldots, h_M)$ linearly at the best-fitting (V, σ) location. To check our results we computed the true moments $(\hat{h}_3, \hat{h}_4)$, by analytically integrating over the LOSVD (eq. [8] of van der Marel & Franx 1993) using (V, σ) of the best-fitting Gaussian computed in § 3. We found $\hat{h}_3 = -0.144$ and $\hat{h}_4 = 0.029$, which, as expected, are very close to the best-fitting parameters computed before.

The result of this test, for both S/N = 60 and 600, is shown in Figure 3. Here the shape of the input LOSVD was held fixed, but the velocity scale was varied so that the σ_{in} of the best-fitting Gauss-Hermite series varied in the range 48–360 km s^{-1}, corresponding to 0.8–6 SAURON pixels, for 1000 different Monte Carlo realizations. By construction, for $\sigma_{in} = 114.8$ km s^{-1} the LOSVD reduces to the one presented
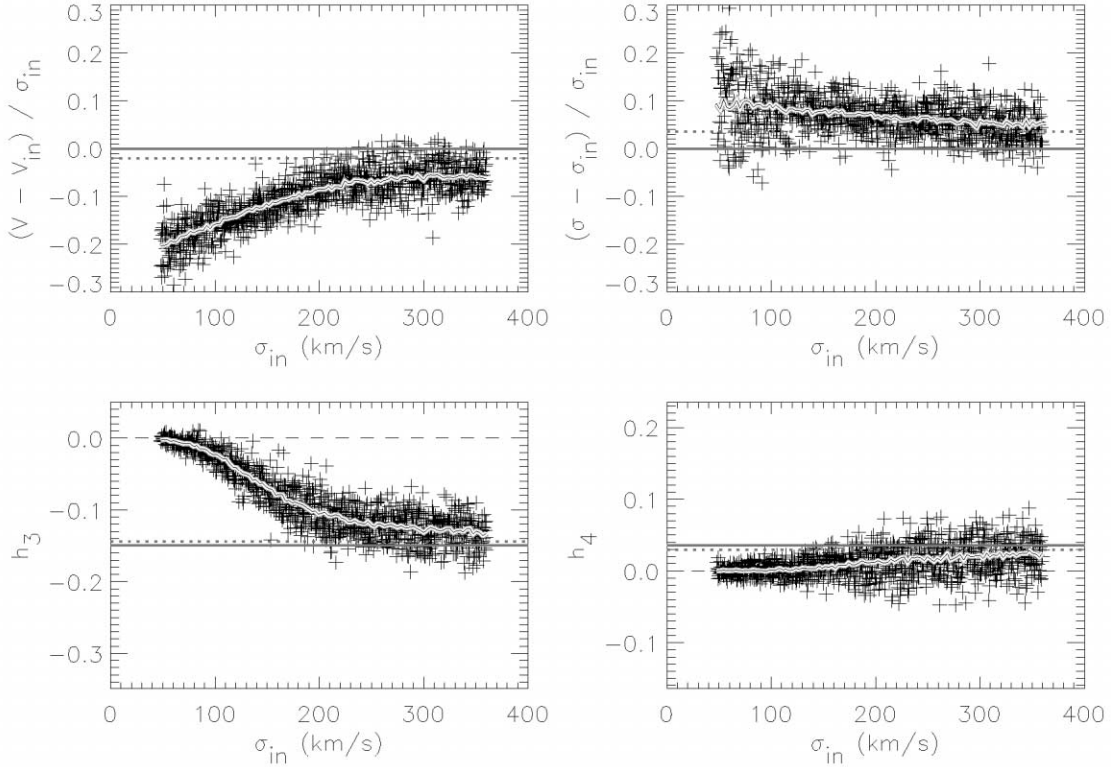
Fig. 3.—Fitting (V, σ) first. Recovery of the stellar kinematics from a model galaxy spectrum convolved with the LOSVD of Fig. 1. The shape of the LOSVD was held fixed while the velocity scale was varied so that the σ_{in} of the best-fitting Gauss-Hermite series varied in the range 48–360 km s^{-1}, corresponding to 0.8–6 pixels. In these measurements, V and σ were fitted first (nonlinearly) and only then h_3 and h_4 were fitted (linearly) with V and σ fixed to the optimal values. In each panel the thick dotted line represents the true Gauss-Hermite moments that the method is trying to recover, the thick solid line indicates the value of the best-fitting Gauss-Hermite parameters, while the crosses represent the actually measured values, as a function of the input σ_{in}, for 1000 different Monte Carlo realizations with S/N = 60. The thick noisy line shows the measured values for 100 realizations with very high S/N = 600. The recovered moments start becoming biased when $\sigma_{in} \lesssim 240$ (4 pixels), and this method becomes essentially insensitive to any deviation from a Gaussian when $\sigma_{in} \lesssim 120$ km s^{-1} (2 pixels) at any S/N.

in Figure 1. The recovery of the true moments is *strongly* biased when $\sigma_{in} \lesssim 240$ km s^{-1} (4 pixels). This can be understood by the fact that when the LOSVD is not very well sampled, an asymmetry in the profile can be compensated during the Gaussian fit by a small V-shift, while a symmetric deviation from a Gaussian is nearly equivalent to a change of σ. The bias in the recovery of the input parameters remains unchanged even for a noise-free model spectrum, although of course the scatter in the values disappears. And no improvement is observed if the template is oversampled before convolving it with a well-sampled LOSVD in equation (3). These results are representative of what we found with different realistic input LOSVDs.

What was seen in the simulation is also found on real SAURON data. The approach of fitting (V, σ) first becomes insensitive to any deviation from a Gaussian LOSVD when the observed velocity dispersion is low ($\lesssim 120$ km s^{-1}, or 2 pixels). As a practical example this method gives the misleading impression that the LOSVD in the nucleus of a low-dispersion object like NGC 3384, containing a fast-rotating disklike structure, is consistent with being essentially symmetric, as, e.g., the central LOSVD of the high-dispersion nonrotating giant

elliptical galaxy M87. This apparent similarity of the overall shape of LOSVD in the two different galaxies is an artifact of the extraction method, and we will show that it is still possible to recover, with a different method, the strong asymmetry of the LOSVD from the SAURON data of NGC 3384, even at low dispersion.

3.2. Fitting All Parameters Simultaneously

The bias present in the recovery of the Gauss-Hermite moments of an undersampled LOSVD in the previous section was mainly due to the fact that (V, σ) were not fitted simultaneously with the other LOSVD parameters. The fit should then become generally unbiased when all (V, σ, h_3, ... , h_M) parameters are varied simultaneously to optimize the fit. In this way one should recover the best-fitting parameters of the Gauss-Hermite series, which do not precisely coincide with the true Gauss-Hermite moments of the LOSVD. By definition this method will provide the lowest χ^2 in the fit, and for this reason it was also originally chosen by van der Marel & Franx (1993).

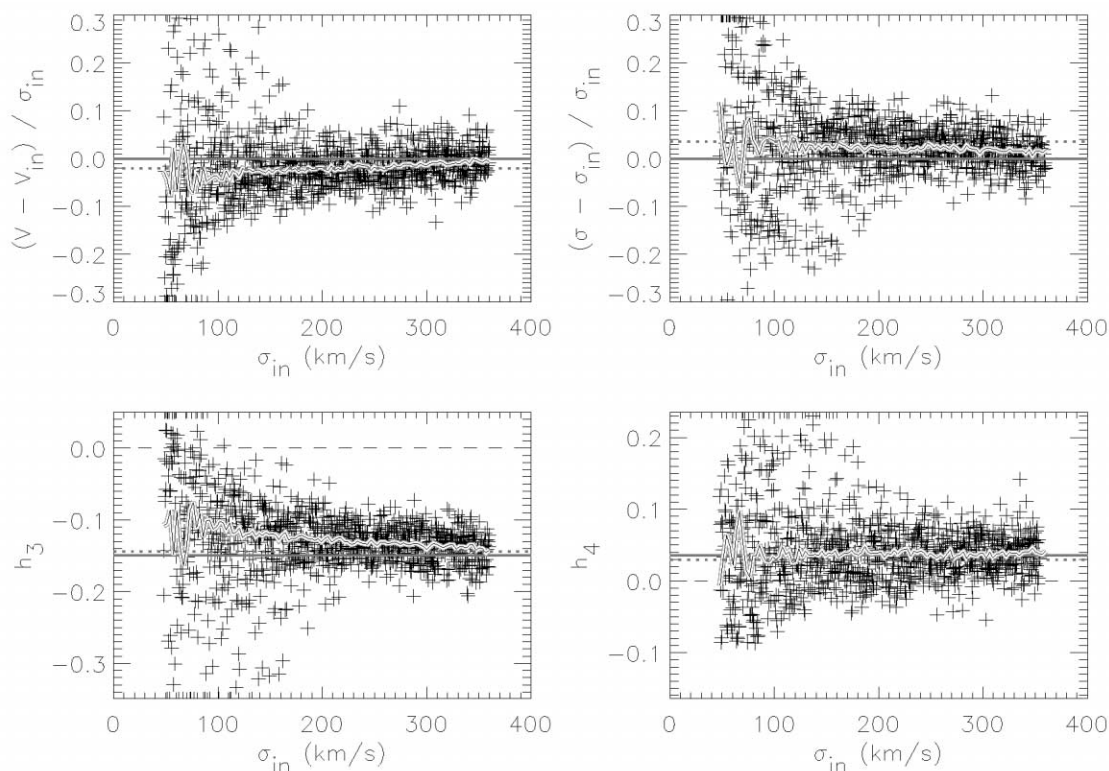In Figure 4 we repeated the experiment of the previous sec-

FIG. 4.—Same as in Fig. 3, but in this plot the four parameters (V, σ, h_3, h_4) were all fitted simultaneously (nonlinearly). Now the measurements are supposed to reproduce the best-fitting Gauss-Hermite parameters, which are indicated by the thick solid line. The measurements are generally unbiased, but no reliable h_3 and h_4 measurements can be obtained at S/N = 60 for $\sigma_{\rm in} \lesssim 120$ km s^{-1} (2 pixels). At low $\sigma_{\rm in}$ the scatter in V and σ increases because of their correlation with h_3 and h_4, respectively. Moreover, the measured σ tends to be systematically lower than the true value, while h_4 is correspondingly too high. In the case of very high S/N, the strong asymmetry of the LOSVD can be recovered, with a modest bias, down to the smallest $\sigma_{\rm in}$ values.

tion while fitting all the above nonlinear parameters together. As expected the solution is now almost unbiased and the measurements tend to be spread around the true input values for a wider $\sigma_{\rm in}$ range. However, when $\sigma_{\rm in} \lesssim 120$ km s^{-1} (2 pixels) and the LOSVD is correspondingly not well sampled, the spectrum does not contain enough information to constrain all the free parameters, and the scatter in the recovered values increases dramatically.

Moreover, at low $\sigma_{\rm in}$ the σ and h_4 values are not unbiased anymore, not being symmetrically distributed around the known input values. The reason for this asymmetry can be understood by looking at the shape of the χ^2 contours, which measure the agreement between the input spectrum $G(x)$ and its best-fitting model $G_{\rm mod}(x)$. In Figure 5 we plot the $\Delta\chi^2 = \chi^2 - \chi^2_{\rm min}$ contours for fits obtained by convolving $G(x)$ with the LOSVD of Figure 1, while keeping V and h_3 fixed to the known best-fitting values (§ 3) and varying σ and h_4 of $G_{\rm mod}(x)$. The narrow curved valley of nearly constant χ^2, for decreasing σ and increasing h_4 from the location of the minimum $\chi^2_{\rm min}$, is due to the strong correlation between these

two parameters, which is caused by the undersampling, and produces the effects seen in Figure 4.

When the errors in the measured parameters are large, one has to make a trade-off decision of scatter versus bias. Usually, when the data are unable to tightly constrain a large number of parameters, one prefers to retain in the fit only the parameters that are required to significantly decrease the χ^2. For this reason, while fitting a parametric LOSVD, it is common practice to reduce the fit to (V, σ) at low S/N: although the fit may be biased by the lack of freedom in the model, an additional flexibility would not lead to meaningful measurements, and structures and trends in the (V, σ) values can still be detected because of the decreased scatter.

Another problem with the large scatter that is present when fitting all parameters together is specific to the two-dimensional kinematical measurements obtained with an integral-field spectrograph. When presenting kinematics maps there is no easy way to also attach errors to the displayed values, and it becomes difficult to estimate from the maps when some structure is real and when it is due to noise. For this reason it would be pref-
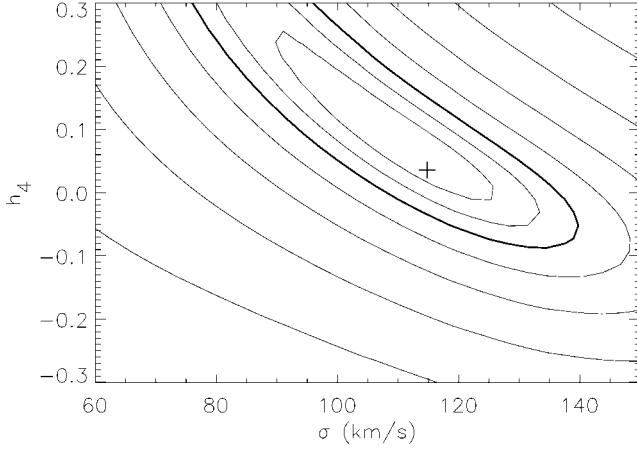
FIG. 5.—Contours of the $\Delta\chi^2 = \chi^2 - \chi^2_{\min}$, which measure the agreement between a simulated noiseless input galaxy spectrum $G(x)$ and its best-fitting model $G_{\mathrm{mod}}(x)$. As in Fig. 4, $G(x)$ was convolved with the LOSVD of Fig. 1. The $\Delta\chi^2$ is plotted as a function of σ and h_4 of the LOSVD of $G_{\mathrm{mod}}(x)$, while V and h_3 are kept fixed to the known best-fitting values (§ 3). The cross indicates the location of the known best-fitting (σ, h_4). The three lowest $\Delta\chi^2$ levels correspond to confidence levels of 1, 2, and 3 σ (*heavy line*), respectively, for a S/N = 60, while other levels are separated by factors of 2. Note the narrow valley of nearly constant χ^2 for decreasing σ and increasing h_4, which is due to the undersampling of the LOSVD. At this σ_{in} a very high S/N is needed for the fit to converge to the true minimum.

erable to display on the kinematics maps only the features that are statistically significant.

3.3. Penalized Pixel Fitting

The LOSVD of galaxies is generally well reproduced by a Gaussian (e.g., Bender, Saglia, & Gerhard 1994). For this reason, it would be preferable to use a technique to fit the LOSVD in which the solution is free to reproduce the details of the actual profile when the S/N is high, but the solution tends to a Gaussian shape when the S/N is low. This was done for the case in which the LOSVD is described by a nonparametric function by using the maximum penalized likelihood formalism (e.g., Merritt 1997), and we refer to that paper for details. Here we apply the formalism to the case in which the LOSVD is parametrically expanded as a Gauss-Hermite series. We show that this leads to a much simpler and faster implementation than for the general case.

The idea is to fit the parameters $(V, \sigma, h_3, \ldots, h_M)$ simultaneously as in § 3.2, but to add an adjustable penalty term to the χ^2 to bias the solution towards a Gaussian shape when the higher moments are unconstrained by the data so that the penalized χ^2 becomes

$$\chi^2_P = \chi^2 + \alpha\mathcal{P}. \tag{6}$$

A natural form for the penalty function $\mathcal{P}$ is given by the integrated square deviation of the line profile $\mathcal{L}(v)$ from its best-fitting Gaussian $\mathcal{G}(v)$:

$$\mathcal{D}^2 = \frac{\int_{-\infty}^{\infty} [\mathcal{L}(v) - \mathcal{G}(v)]^2 \, dv}{\int_{-\infty}^{\infty} \mathcal{G}(v)^2 \, dv}. \tag{7}$$

This penalty does not suppress noisy solutions, which are already excluded by the use of a low-order parametric expansion for $\mathcal{L}(v)$. It was shown by van der Marel & Franx (1993) that in the case in which $\mathcal{L}(v)$ has the form of equation (4), then equation (7) is well approximated by

$$\mathcal{D}^2 \approx \sum_{m=3}^{M} h_m^2. \tag{8}$$

In principle one could then define the penalty as $\mathcal{P} = \mathcal{D}^2$ and optimize χ^2_P (eq. [6]). In practice, however, this is not desirable for two reasons:

1. As explained in § 2, it is computationally much more efficient to minimize the residuals r_n (eq. [2]) than to explicitly compute the χ^2.

2. One needs a way to automatically adjust the penalty factor α according to the χ^2 of the observed fit.

We found that a simple and effective solution to these problems consists of using the following perturbed residuals as input to the nonlinear least-squares optimizer:

$$r'_n = r_n + \lambda \, \sigma(\mathbf{r}) \, \mathcal{D}, \tag{9}$$

where the variance is defined as

$$\sigma^2(\mathbf{r}) = \frac{1}{N} \sum_{n=1}^{N} r_n^2. \tag{10}$$

The qualitative interpretation of this formula is that a deviation $\mathcal{D}$ of the LOSVD from a Gaussian shape will be accepted as an improvement of the fit only if it is able to correspondingly decrease the scatter $\sigma(\mathbf{r})$ by an amount related to $\mathcal{D}$. To quantify, one can compute the objective function of the fit

$$\chi^2_P = \sum_{n=1}^{N} r_n^2 + 2\lambda\sigma(\mathbf{r})\mathcal{D} \sum_{n=1}^{N} r_n + N[\lambda\sigma(\mathbf{r})\mathcal{D}]^2. \tag{11}$$

The sum of the residuals in the second term is zero by construction, because of the fact that the weights are optimized for given $\mathcal{L}(v)$ (§ 2; this is required for this form of perturbation to work). Considering the definition of variance (eq. [10]), one can finally write
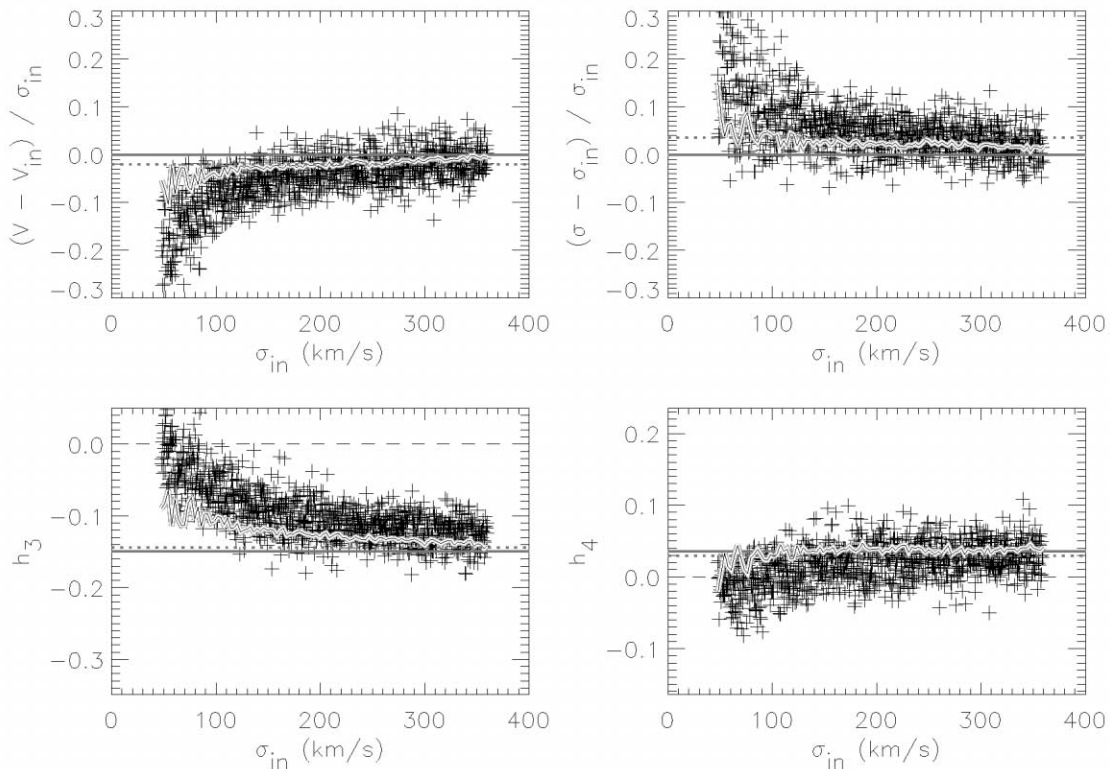
$$\chi^2_P = \chi^2(1 + \lambda^2\mathcal{D}^2), \tag{12}$$

FIG. 6.—Penalized pixel fitting. Same as in Fig. 4, but using the new penalized parametric pixel-fitting method with $\lambda = 0.7$. At S/N = 60 the bias on the parameters is now significant only when σ_{in} goes below about 120 km s^{-1} (2 pixels). But more important, the fit tends to converge to the true solution even for low σ_{in} at the higher S/N = 600.

which is of the desired form of equation (6) with $\alpha = \lambda^2 \chi^2$ automatically scaled according to the χ^2 of the fit. In practice $\sigma(r)$ in equation (9) is computed using a robust biweight estimator (Hoaglin, Mosteller, & Tukey 1983) so that equation (12) is only valid as an approximation. One can see from this formula that with $\lambda = 1$, a deviation $\mathcal{D} = 10\%$ of the LOSVD from a Gaussian (e.g., $h_3 = 0.1$ and $h_m = 0$) requires a corresponding decrease in the scatter $\sigma(r)$ of the unperturbed residuals by more than $(1 + \mathcal{D}^2)^{1/2} = 0.5\%$ to be accepted by the optimization routine as an improvement of the fit.

A statistical interpretation of equation (12) can be obtained by considering that for a good fit, $\chi^2 \sim N$. For a given λ a deviation $\mathcal{D}$ from a Gaussian needs to decrease χ^2 by more than $\Delta\chi^2 \sim N\lambda^2\mathcal{D}^2$ to be accepted, and this variation can be associated with a specific confidence level (see, e.g., Press et al. 1992, § 15.6) at which a Gaussian shape is excluded. Although this formula may serve as a guideline, we believe it is safer to test a choice of λ using simulations such as the ones presented.

A test of the application of equation (9) to the same model spectra as § 3.2 is shown in Figure 6. For the same S/N = 60 as before, using $\lambda = 0.7$, the measurements are now essentially unbiased when $\sigma_{in} \gtrsim 120$ km s^{-1} (2 pixels). For lower

input dispersion the LOSVD converges toward a Gaussian and, as expected, the bias becomes similar to Figure 3. However, the *crucial* difference with the previous case consists of the fact that for higher S/N = 600 the solution tends to converge to the known best-fitting values, because the penalty scales with the χ^2.

We mention here that alternative forms to equation (9) are possible, with the requirement that the objective function has the form of equation (6). One possibility is to include a multiplicative perturbation of the residuals:

$$r'_n = r_n(1 + \lambda\mathcal{D}^2). \quad (13)$$

In this case the objective function becomes

$$\chi_P^2 = \chi^2(1 + 2\lambda\,\mathcal{D}^2 + \lambda^2\,\mathcal{D}^4), \quad (14)$$

and if one neglects the small term containing $\mathcal{D}^4$, this becomes again of the form of equation (6), with $\alpha = 2\lambda\chi^2$. We verified that in practice this alternative form of perturbing the residuals produces virtually the same results as using equation (9) if the same α parameter is adopted. We prefer the perturbation of equation (9), given its robustness against outliers, because of
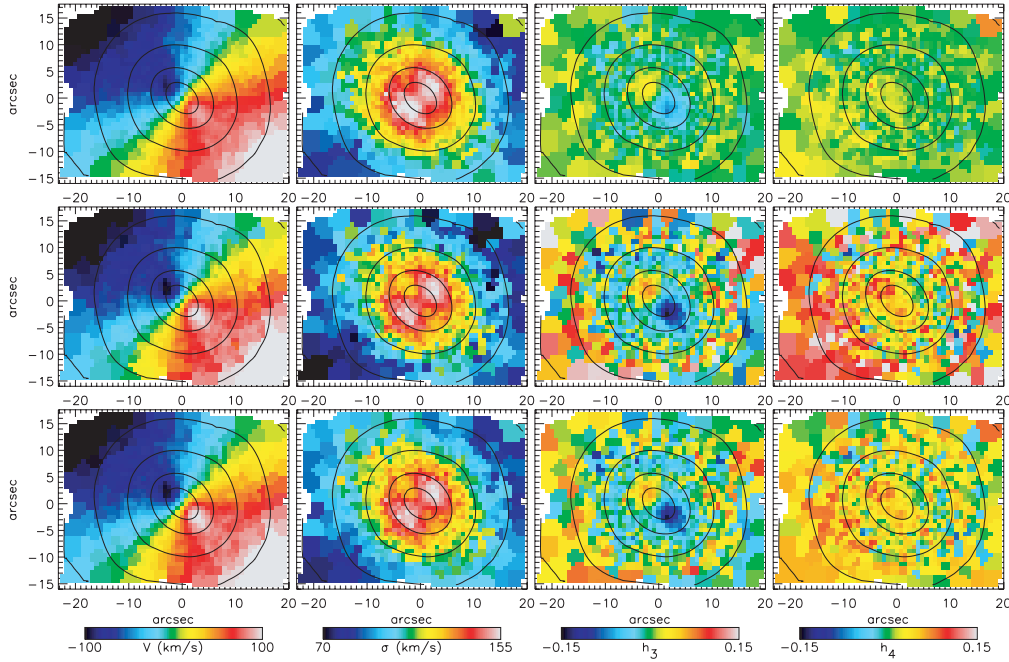
FIG. 7.—Comparison of the Voronoi 2D binned (Cappellari & Copin 2003) SAURON stellar kinematics of the barred lenticular galaxy NGC 3384 extracted with the different methods explored in this paper. *Top panels:* Kinematics extracted by first fitting (V, σ) (nonlinearly) and then expanding the (h_3, h_4) parameters at the optimal (V, σ) location. The Gauss-Hermite parameters are everywhere significantly suppressed by this method. *Middle panels:* Kinematics obtained by fitting (V, σ, h_3, h_4) simultaneously, without penalty. Due to the σ-h_4 correlation (Fig. 5) caused by the undersampling of the LOSVD, the σ is noisier and depressed, while h_4 tends to be strongly positive. The h_3 values fluctuate in the outer parts between large positive and negative values. *Bottom panels:* Kinematics measured with the new penalized pixel-fitting method (with $\lambda = 0.7$). This overcomes the problems of the two previous approaches. Only the statistically significant non-Gaussian features are preserved, otherwise the solution is smoothly reduced to a Gaussian. Contours of the reconstructed surface brightness are superimposed (in 1 mag arcsec^{-2} steps).

the use of the biweight in the computation of $\sigma(\boldsymbol{r})$. The form of equation (13) may be useful in implementations in which the average residuals are not zero for a given $\mathcal{L}(v)$.

In Figure 7 we compare the three different approaches to the recovery of the LOSVD by using them to extract the stellar kinematics from SAURON (Bacon et al. 2001) integral-field spectroscopic observations of the galaxy NGC 3384 (see de Zeeuw et al. 2002 for a description of the observations), which has a rather low velocity dispersion over the whole observed field. The data cube was spatially binned to a minimum S/N = 60 using the Voronoi two-dimensional (2D) binning method of Cappellari & Copin (2003). The same differences that we observed using synthetic-model spectra can also be seen from real data:

1. The Gauss-Hermite parameters are everywhere significantly suppressed when fitting V and σ first (Fig. 7, *top panels*).

2. The simultaneous fit of all parameters (Fig. 7, *middle panels*) tends to produce a velocity dispersion map that is noisier and has smaller values, while h_4 tends to be strongly positive. The h_3 values fluctuate in the outer parts between large positive and negative values.

3. The use of a penalty function (Fig. 7, *bottom panels*) overcomes the problems of the two previous approaches. Only the statistically significant non-Gaussian features are preserved; otherwise, the solution is smoothly reduced to a Gaussian.

We also applied the three methods to the extraction of the kinematics of all the 48 early-type galaxies of the SAURON survey (de Zeeuw et al. 2002). As expected from the simulations, the three techniques gave very different results only for galaxies like NGC 3384, which have a low-velocity dispersion. The kinematics extracted from SAURON observations of galaxies with dispersion $\gtrsim 180$ km s^{-1} over the whole observed field provided very similar result with the three methods (although by construction not identical).

3.4. Errors on the Fitted Parameters

The availability of fast computers makes Monte Carlo simulations the preferred way to estimate measurement errors for LOSVD extraction methods. This consists of repeating the full measurement process for a large number of different realization of the data, obtained by adding noise to the original spectra.

The method discussed in this paper, like many other available methods for extracting the LOSVD, makes use of some form of filtering or penalization of the solution, thus biasing the results to suppress the noise. It may be worth emphasizing that when computing Monte Carlo errors, one has to correct for the bias in the measurements introduced by the noise-suppression mechanism. Failure to correct for the bias in the kinematics can result in a significant underestimate of the confidence intervals of parameters determined by fitting dynamical models to the data.

One way to obtain proper errors from penalized methods consists of correcting the error estimates by increasing the percentile intervals by an amount given by the bias (e.g., Beers, Flynn, & Gebhardt 1990). In real measurements one does not know the actual bias (otherwise the measurements would have been corrected for it), so an estimate of the maximum expected bias can be used instead. This requires one to perform Monte Carlo simulations of the kinematic extraction to determine the maximum bias at different S/N levels and σ_{in} values.

A simpler order-of-magnitude estimate of the errors on the parameters may be obtained by noting that Figure 4 provides a more realistic representation of the errors in the measurements than Figure 6, at low σ_{in} values. In the case of the penalized pixel-fitting (pPXF) routine described in the previous section, one can then obtain the errors by setting the parameter λ in equation (9) to a very small value before running the Monte Carlo simulations. This simple approach will generally provide a conservative estimate of the actual errors.

3.5. The Algorithm

To summarize the discussion of the previous sections, the suggested algorithm for recovering the LOSVD parametrized as a Gauss-Hermite series is as follows:

1. Start with an initial guess for the parameters (V, σ), while setting the initial Gauss-Hermite parameters $h_3, \ldots, h_M = 0$.

2. Solve the subproblem of equation (3) for the weights $(w_1, \ldots, w_K, b_0, \ldots, b_L)$.

3. Compute the residuals r_n from the fit using equation (2).

4. Perturb the residuals as in equation (9) to get r_n'.

5. Feed the perturbed residuals r_n' into a nonlinear least-squares optimization routine and iterate the procedure from step (2) to fit for the parameters $(V, \sigma, h_3, \ldots, h_M)$.

3.6. Availability

An IDL routine implementing the algorithm described in this work is available online.[2]

4. CONCLUSIONS

In this paper we addressed the problem of extracting the LOSVD of the stars, parametrized using a Gauss-Hermite series, from observed galaxy spectra. Using a method that works directly in pixel space, we compared different techniques by applying them to the same problem, with both simulated and real data, and we showed that one has to pay special attention to the extraction when the LOSVD is not well sampled by the data or when the S/N is low.

We proposed that in these situations one should apply the maximum penalized likelihood formalism to extract as much information as possible from the spectra while suppressing the noise in the solution. We demonstrated that this leads to a fast and simple algorithm. We also discussed how Monte Carlo errors should be properly estimated. This pPXF method is particularly useful to extract the kinematics from integral-field spectroscopic data, since for two-dimensional maps there is no standard way to visualize errors, and one wants to be able to show only the non-Gaussian features that are statistically significant. A routine that implements the ideas of this paper is publicly available.

---

[2] See http://www.strw.leidenuniv.nl/~mcappell/idl/.

REFERENCES

Bacon, R., et al. 2001, MNRAS, 326, 23
Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100, 32
Bender, R. 1990, A&A, 229, 441
Bender, R., Saglia, R. P., & Gerhard, O. E. 1994, MNRAS, 269, 785
Cappellari, M., & Copin, Y. 2003, MNRAS, 342, 345
de Bruyne, V., Vauterin, P., de Rijcke, S., & Dejonghe, H. 2003, MNRAS, 339, 215
Dejonghe, H., & Merritt, D. 1992, ApJ, 391, 531
de Zeeuw, P. T., et al. 2002, MNRAS, 329, 513
Emsellem, E., et al. 2004, MNRAS, submitted

Franx, M., & Illingworth, G. D. 1988, ApJ, 327, L55
Gebhardt, K., et al. 2000, AJ, 119, 1157
Gerhard, O. E. 1993, MNRAS, 265, 213
Hoaglin, D. C., Mosteller, F., & Tukey, J. W. 1983, Understanding Robust and Exploratory Data Analysis (New York: Wiley)
Kelson, D. D., Illingworth, G. D., van Dokkum, P. G., & Franx, M. 2000, ApJ, 531, 159
Kuijken, K., & Merrifield, M. R. 1993, MNRAS, 264, 712
Lawson, C. L., & Hanson, R., 1995, Solving Least Squares Problems (2d ed.; Englewood Cliffs: Prentice-Hall)

Merritt, D. 1997, AJ, 114, 228

Moré, J. J., Garbow, B. S., & Hillstrom K. E., 1980, User Guide for MINPACK-1 (Argonne Nat. Lab. Rep. ANL-80-74; Argonne: Argonne National Laboratory)

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in FORTRAN 77 (2d ed; Cambridge: Cambridge Univ. Press)

Rix, H.-W., de Zeeuw, P. T., Cretton, N., van der Marel, R. P., & Carollo, C. M. 1997, ApJ, 488, 702

Rix, H.-W., & White, S. D. M. 1992, MNRAS, 254, 389

Saha, P., & Williams, T. B. 1994, AJ, 107, 1295

Sargent, W. L. W., Schechter, P. L., Boksenberg, A., & Shortridge, K. 1977, ApJ, 212, 326

Scorza, C., & Bender, R. 1995, A&A, 293, 20

Simkin, S. M. 1974, A&A, 31, 129

Statler, T. 1995, AJ, 109, 1371

Tonry, J., & Davis, M. 1979, AJ, 84, 1511

van der Marel, R. P. 1994, MNRAS, 270, 271

van der Marel, R. P., & Franx, M. 1993, ApJ, 407, 525

Vazdekis, A. 1999, ApJ, 513, 224

Zhao, H.-S., & Prada, F. 1996, MNRAS, 282, 1223