Decomposing the host galaxy from high-$z$ QSOs using principal component analysis

Wei He$^{1,2}$ and Yi-Ping Wang$^1$

$^1$ National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; whe@nao.cas.cn
$^2$ School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China

Received 2019 March 8; accepted 2019 April 30

Abstract High resolution deep imaging from space and adaptive optics techniques with large ground-based facilities have enabled studies examining faint host galaxies of high redshift quasi-stellar objects (QSOs). However, the related image processing techniques, especially for a precise point-spread function (PSF) reconstruction and characterization of the host galaxy light profiles, have yet to be optimized. We present here the scientific performance of a principal component analysis (PCA) based PSF subtraction of the central bright point source of high redshift QSO images, as well as further characterization of the host galaxy profile by directly fitting a Sérsic model to the residual image using the Markov Chain Monte Carlo (MCMC) algorithm. With a set of reference PSF star images which represent interleaving exposures between the QSO imaging, we can create an orthogonal basis of eigen-images and restore the PSF of QSO images by projecting the QSO images onto the basis. In this way, we can quantify the modes in which the PSF varies with time by a basis function that characterizes the temporal variations of the reference star as well as the QSO images. To verify the algorithm, we performed a simulation and applied this method to one of the high-$z$ QSO targets from Mechtley et al. We demonstrate that the PCA-based PSF subtraction and further modeling of the galaxy’s light profile using MCMC fitting would sufficiently remove the effects from central dominating point sources, and improve characterization ability for the host galaxies of high-$z$ QSOs to the background noise level which is much better than previous two-component fitting procedures.

Key words: quasars: general — galaxies: active — galaxies: high-redshift — methods: data analysis

1 INTRODUCTION

Quasars appear to be isolated star-like mysterious objects in early optical images. Only since the middle of the 1990s has the superior spatial resolution of the Hubble Space Telescope (HST) not only helped to reveal the host galaxies of quasars but has also provided reliable measurement of the masses of the central black holes for dozens of nearby galaxies. A scaling relation between the black hole masses and the masses of their host spheroid has been deduced for quasars and local galaxies using various space and ground-based observations and dynamical models (McLeod & Rieke 1994; Kormendy & Richstone 1995; Magorrian et al. 1996; Faber et al. 1997; Slik & Rees 1998; Wang & Biermann 1998; Wang et al. 2000). Due to the scattered nuclear light, measurements of the quasar host components, which are usually faint and of small angular size, are still technically challenging. Some success in detecting the hosts of quasars out to $z \sim 2$ was achieved with the NICMOS infrared camera on the HST (Kukula et al. 2001; Ridgway et al. 2001; Peng et al. 2006). However, obtaining reliable measurements of luminosities and scalelengths for the host galaxies is extremely difficult, and it suffers from the bias of nonoptimal point-spread function (PSF) subtraction methods. The reliable detection of host galaxy light around a luminous quasar is even trickier at high-$z$, due to the expected overwhelming dominance of the nuclear light, the small angular size of any reasonably anticipated host galaxy, and the faintness of any potential host galaxy relative to the scattered nuclear light in the wings of the PSF.

In order to precisely characterize the quasi-stellar object (QSO) host galaxy properties, we need to properly estimate the Strehl ratio and flux of the PSF and subtract the light contribution from the central bright point source in the QSO images. This is because a slight difference between the shape of the QSO and the reference star’s PSF could create artificial residual, and therefore cause errors
in the QSO host galaxy measurements. In this case, we usually select at least one good PSF reference star which would match the conditions between the quasar and the PSF reference star observations as much as possible. In previous studies, separation of nuclear and host galaxy light was firstly performed using scaled PSF subtraction, where the PSF reference star was observed to characterize the PSF, and the PSF subtraction was achieved by scaling the flux of the averaged PSF reference star images to the QSO central peak intensity and then subtracting the scaled PSF from the QSO images (Lehnert et al. 1999; Sanchez et al. 2004). This method usually implies an over-subtraction of the nuclear PSF and provides only a lower limit on the host flux, though it probably ensures exclusion of all nuclear light from the subtracted residual.

There are a few two-component models which directly fit the QSO image using a point source and a Sersic profile representing the surface brightness distribution of the stellar population of the host galaxy, such as GALFIT (Peng et al. 2002, 2006). However, these models have fundamentally covariant parameters, and hence there are limitations to the currently-existing fitting algorithms for the two-dimensional (2-D) surface brightness distribution, when applied to the point source subtraction problem. For example, the fluxes of the two components are necessarily correlated since the total flux of the two-component model must match the measured flux from the QSO image being fitted. Unfortunately, a way to quantify the degeneracy had not been provided. Moreover, the host galaxy is much fainter than the central point source which dominates the central light profile of the QSO image, and thus a simultaneous modeling would suffer from this issue and cause a peculiar solution. Such a problem would be more severe for luminous QSOs at high-z, due to the small angular size and faintness of any potential host galaxy which is hidden in the glare of bright nuclear light.

On the other hand, much effort has been made over recent years in high contrast imaging research, in order to directly detect and characterize stellar companion objects, i.e., exoplanets and debris disks. For a full high contrast performance, image post-processing techniques are necessary and complementary to starlight suppression coronagraphic optics and related observing procedures, such as angular and simultaneous spectral differential imaging (ADI and SSDI respectively) as well as other methods. The locally optimized combination of images (LOCI) developed by Lafrenière et al. (2007) and principal component analysis (PCA: Amara & Quanz 2012; Soummer et al. 2012) are currently the two main image post-processing techniques, where the LOCI algorithm modeling local stellar PSF structure in every image has proven to be effective at finding faint companions. However, the LOCI algorithm is usually an extension of ADI and relies upon multiple matched PSFs and is thus too costly in terms of observing time for high-z quasar studies. Nevertheless, PCA models the global PSF structure and the PSF variation in time by identifying the main linear components of the variation, and thus has advantages for the study of faint extended objects which overlap with the central PSF.

In this paper, we propose applying a PCA-based PSF subtraction to decompose a host galaxy from a QSO image, where the faint extended host galaxy is overlapped with the wing of the central PSF. With a set of reference PSF star images which were interleaving exposures between the QSO imaging, we can create an orthogonal basis of eigen-images and restore the PSF of the QSO images by projecting the QSO images on the basis. In this way, we can quantify the modes in which the PSF varies with time by a basis function that characterizes the temporal variations of the reference star as well as the QSO images. After removal of the central bright point source, we directly fit a Sersic model to the residual image using a Markov Chain Monte Carlo (MCMC) algorithm to characterize the host galaxy properties. To test this approach, we perform a simulation and apply this algorithm to one of the high-z QSO samples from Mechtley et al. (2016). We demonstrate that PCA-based PSF subtraction, and further modeling of the host galaxy using MCMC fitting, can sufficiently remove the effects of the central dominating point sources, and improve the characterization ability of the host galaxies of high-z QSOs with respect to the background noise level.

The outline of this paper is as follows. In Section 2, we will introduce the PCA method of decomposing a faint source from the wing of the central PSF of a QSO image. Details about the simulations are presented in Section 3. A discussion on the simulation is provided in Section 4. Application of this method on observed data from the HST is made in Section 5.

2 METHODOLOGY

2.1 Overview of the PCA Algorithm

The key challenge in analyzing high-z QSO images is to decompose the faint source from the central bright point source, which could be modeled as a convolution between a delta function and the PSF. The PSF can be considered a stochastic process of the state of the observing condition, as well as a combination of telescope and instrument acquiring the data. Moreover observations of reference stars have been carried out under similar observing conditions as the QSO target. So, the PSF of the target could be modeled by analyzing the PSF of the reference stars. An effective way to model the PSF is to decompose it into a set of basis functions. According to Soummer et al. (2012) and Amara
Finally, as the eigen-PSF basis has been procured, projection of the QSO image onto the eigen-PSF basis would approximately restore the point source structure in the image. Also the QSO image $Q$ can be composed of an astronomical PSF $P$ and faint host galaxy $G$, which means $Q = P + G$. Subtracting this projection from $Q$ would output the final residual image $F$ for the modeling process,

$$F = Q - \sum_{m=1}^{m_{\text{klip}}} \langle Q, Z_m \rangle Z_m$$

$$= (P - \sum_{m=1}^{m_{\text{klip}}} \langle P, Z_m \rangle Z_m) + (G - \sum_{m=1}^{m_{\text{klip}}} \langle G, Z_m \rangle Z_m).$$

It is expected that the eigen-PSF basis could help to restore the point source as much as possible, so the term $(P - \sum_{m=1}^{m_{\text{klip}}} \langle P, Z_m \rangle Z_m)$ could be ignored in the modeling process. In fact, the impact of this algorithm on the galaxy profile is the projection of $G$ onto the retained eigen-PSFs $Z_m \{m=1...m_{\text{klip}}\}$ which have been considered and compensated for in Equation (4). As is well known, $G$ can be modeled by a parametric model which describes the host galaxy morphology profile. So by fitting with the final residual image $F$, we could estimate the parameters of the host galaxy profile.

2.3 Characterization of QSO Host Galaxies

2.3.1 Sérsic model

With the aim of describing galaxy structure from 2-D information contained in the image, parametric models represented by particular functions are chosen to fit the galaxy light distribution. The Sérsic model is a particularly useful and flexible kind of profile for galaxy fitting, which is defined by

$$I(r) = I_e \exp\{-\kappa[(r/r_e)^{1/n} - 1]\},$$

where $r_e$ is the ‘effective radius’ (hereafter radius), which is the half-light radius of the galaxy. The Sérsic index $n$ describes the profile shape of the light distribution. Together with positions $x$ and $y$, ellipticity and position angle, these seven free parameters define the morphology profile of the host galaxy. While $(n = 4)$ gives profiles defined by de Vaucouleurs’ law, which are approximations of typical early-type galaxies, $(n = 1)$ represents the exponential light-profiles of galactic disks.

2.3.2 The MCMC method

With the PCA-based PSF subtraction algorithm and parametric model of host galaxies that we described above,
we adopted the MCMC algorithm which is an alternative method for effective modeling and parameter estimation to fit the residual images after PSF subtraction. Given a set of observations \( y \), and a model described by a set of parameters \( \theta \), inferences can be made about the probability distribution of the parameters based on sampling in the parameter space

\[
P(\theta | y) = \frac{P(\theta) P(y|\theta)}{P(y)}.
\]

\( P(\theta|y) \) is the posterior probability distribution, which describes the probability distribution of the model parameters given the data, while \( P(\theta) \) is the prior probability distribution of the model parameters, determined from physical principles and previous knowledge. A preliminary study with non-uniform prior distributions confirms that the posterior maximum and confidence values could significantly vary for low signal-to-noise ratio (S/N) data. However, for simplicity, we used uniform prior distributions for all the parameters for the tests in this paper. \( P(y|\theta) \) is the likelihood function which tells the probability of the observed data under a set of model parameter values. \( P(y) \) is taken as a proportionality constant since the prior probability distribution of the observed data does not change here with different parameter values.

In this paper, the observed data \( y \) in Equation (6) are the final residual image \( F \) after PCA-based PSF subtraction from the simulated QSO image in Equation (4), and the model data \( M(p) \) in Equation (7) are the residual image of the Sèrsic model of the host galaxy which is processed by the same subtraction procedure as the simulated QSO image. Thus the conditional probability \( P(y(p)|\theta) \) can be calculated through,

\[
P(y(p)|\theta) = \frac{1}{\sqrt{(2\pi\sigma^2(p))}} \exp \left( \frac{y(p) - M(p)}{2\sigma^2(p)} \right).
\]

The variance \( \sigma(p) \) is the noise estimated from the target images. The likelihood function \( P(y|\theta) \) is the joint probability product of the individual pixel probabilities. With the MCMC method, one could search through the complex parameter space to recover the distribution of posterior probability for the light profile parameters. In this process, once the Markov chain has converged, we will sample the posterior to map the joint posterior probability of those parameters.

3 SIMULATION

To verify the PCA-based PSF subtraction and the MCMC fitting algorithm, we created a set of QSO images in order to check how well our method could perform in recovering the parameters describing light profiles of galaxies.

3.1 Simulation Setup

Firstly, we obtained the imaging data of the QSO UM402 which was observed using Subaru/IRCS with adaptive optics (AO) on 2003 Sept. 17–19. The preliminary data have been analyzed and published in Wang et al. (2015). Observations of the PSF reference star were nested in the QSO imaging, in order to monitor the temporal variability of the PSF in QSO imaging. We adopted a series of 19 images of the PSF star which was observed on 2003 Sept.17 to be the basis of our simulation. We cut out the innermost 200×200 pixels (radius~4.6 arcseconds) centered on the star and all the images have been sky subtracted. We assumed that the PSF variation of the QSO imaging could be represented well by observations of the reference star, since the PSF reference star is always observed under conditions close to those of the target as we have described. To mimic the real observations of the QSO and the reference PSF star, we divided those 19 images of the reference star into two groups, which are named the ‘\( psf_{\text{target}} \)’ and the ‘\( psf_{\text{star}} \)’ groups. The ‘\( psf_{\text{star}} \)’ group is now designated as the set of reference images \( S \) in Equation (1), and its median image is defined as \( psf_{\text{star}} \). We regarded the median image of the ‘\( psf_{\text{target}} \)’ as \( psf_{\text{target}} \), which will be used as the PSF to create the simulated QSO image. Instead of a direct separation of these 19 images into two groups, we selected every other image of the reference star images in a time sequence and assigned them to each group in order to balance the PSF variation.

To simulate an individual noise-free galaxy accurately, we followed Häussler et al. (2007) to adopt a hybrid approach to solve the problem of allocating flux on pixels and this was done by creating the images following the Sèrsic model by a factor of 10 or 100 bigger, and then rebinning the image while holding the total flux constant. To increase accuracy, the inner parts of our simulated profiles (the central 60×60 pixels of our 200×200 image size) have been oversampled by a factor of 10, while the very inner parts (20×20 pixels) are oversampled by a factor of 100. Then this projected Sèrsic profile will be convolved with the \( psf_{\text{target}} \) to simulate the host galaxy of the QSO.

We created the QSO image by adding the \( psf_{\text{target}} \) to the center of the simulated host galaxy. To make it consistent with the actual situation, we adopted the magnitude of UM402 as the magnitude of the central point source and clipped the sky background of the UM402 target image and added it to the simulated QSO image. Finally, there are eight parameters applied to simulate the QSO image, \( mag_{\text{diff}} \) (magnitude of galaxy − magnitude of central point source), \( radius \), orientation, axis ratio, Sèrsic index, \( x \) and \( y \) positions (identical center with point source) and sky level.
In our simulation, as we mainly focused on recovery of the brightness, radius and Sérsic index of the host galaxy, we set the rest of the parameters to be reasonable and fixed values. In this work, we only simulated two typical galaxy light profiles: purely exponential profiles ($n = 1$, disk galaxies) and the de Vaucouleurs luminosity profile ($n = 4$, bulge galaxies). As the full width at half maximum (FWHM) of the PSF image is about 5 pixels, we chose the effective radius of the galaxy to be 1 and 2 times the PSF FWHM since most host galaxies with high-$z$ QSOs can be small in scale, so their radii are 5 and 10 pixels. Moreover, the $mag_{\text{diff}}$ between host galaxy and point source was set to three different levels, 2.57, 1.82 and 1.06, and the host galaxy light profile indicated by the faintest level would be close to the sky noise level.

After the whole set of simulated QSO images was acquired, we applied the PCA-based PSF subtraction and MCMC fitting procedure to them. For better understanding the impact of PSF mismatch, we also created a set of images with only simulated galaxies (without a central bright point source), and ran an identical procedure on them as that of simulated QSO images. Comparison with the simulated results is given in the next section.

To evaluate the simulated images, we adopted the definition of S/N from Häussler et al. (2007)

$$S/N = \langle \rho \rangle \ast \left([\langle \rho \rangle + \rho_{\text{sky}}]/\sqrt{\langle \rho_{\text{sky}} \rangle + \sigma_{\text{sky}}}\right)^{-1/2},$$

where $\langle \rho \rangle$ is the average count rate for galaxy pixels within the radius, $\langle \rho_{\text{sky}} \rangle$ is the average count rate for a central point source within the radius, $\rho_{\text{sky}}$ is the background flux within a pixel and $\sigma_{\text{sky}}$ is the uncertainty of the background sky estimate. Since in this work we examine a faint high-$z$ QSO image, the sky noise is mostly dominant. For the set of simulated images, the S/N values of the galaxies with either disk or bulge morphologies are nearly comparable, and the galaxies with $radius = 5$ pixels have slightly higher S/N than those with $radius = 10$ pixels.

### 3.2 Results of Simulation

In Figures 1 and 2, we show the recovery results for the $mag_{\text{diff}}$ and $radius$ of the host galaxy with astronomically representative model parameters. The results were from the posterior probability distribution of MCMC fitting within 1-sigma error. We noticed from the simulations that there is a certain amount of deviation between the $psf_{\text{target}}$ which is used to create simulated the QSO images and the PSF of QSO images which is reconstructed using the eigenvectors from the ‘$psf_{\text{star}}$ group’. The error in reconstruction (mainly PSF mismatch) could still lead to recovery deviation of the galactic light profiles. Without the effect of the PSF mismatch, the $mag_{\text{diff}}$ and $radius$ could be perfectly recovered for the galaxy only images in Figure 2, except for the faintest bulge galaxies. Because the PCA-based PSF subtraction could not eliminate the bad effects of the PSF mismatch, this would cause the recovered parameter $mag_{\text{diff}}$ of the simulated galaxies to be smaller than the input ones.

We noticed from the simulations that the QSO host galaxies of the disk morphology and with effective radius = 5 pixels show better recovery of the parameter radius than those of radius = 10 pixels, since they are more concentrated and have better S/N. However, for the bulge galaxies, recovery of the parameters $mag_{\text{diff}}$ and radius has slightly larger deviation from the input values than that of disk galaxies in the right panel of Figure 1. The similarity between the concentrated light profile of the bulge and the PSF would cause some information on the light profile to be subtracted during the PSF subtraction procedure and make its recovery hard for sky noise dominated images.

Moreover, we present the recovered Sérsic index from our simulations for the six simulated QSO images and galaxy only images in Table 1. The same as for $mag_{\text{diff}}$ and radius, the Sérsic indices of the galaxy only images were recovered well except for the faintest bulge galaxies. For both disk and bulge galaxies, the results for QSO images were still quite close to those of the galaxy only images and suffer limited impact from the PSF mismatch.

### 4 DISCUSSION

We have presented the simulation results in Section 3.2. It is obvious that PCA-based PSF subtraction could circumvent the problem of unknown flux for a central bright point source which is a serious weak point of the traditional two-component fitting method. Besides, restoring the PSF of QSO images using a PCA algorithm is much better and more realistic than simply adopting an analytic function in the 2-D fitting. To better understand these issues, we sent the same set of simulated QSO images to GALFIT for fitting. However, we noticed that GALFIT could not give any reasonable results for most cases, and we had to allocate fixed values for Sérsic index or radius for the fitting process, otherwise GALFIT would report that a possible numerical problem has been detected. It is easy to understand that it is harder to recover the parameters for host galaxies with lower surface brightness and higher Sérsic index. Although our simulations using a PCA-based method show similar trends for the fitting results, we actually do not need to fix the parameters before the fitting process. This is because the PCA-based PSF-subtraction has eliminated the PSF mismatch well before the fitting of the light profiles and the MCMC algorithm would ensure that the parameter space could be explored as much as possible, there-
Recovery of the magnitude and size of the host galaxies. We consider both disk and bulge morphologies \((n = 1 \text{ and } n = 4)\). For each morphology, we run simulations for galaxies with \(\text{radius} = 5, 10\) pixels, and \(\text{mag}_{\text{diff}} = 2.57, 1.82, 1.06\), with six cases in total. The red symbols are the input parameters for the simulations and the blue symbols with error bars are recovered ones. The same symbols in red or blue color are for the input or recovered parameters of the same host galaxies.

Fig. 2  Same as Fig. 1, except that the ‘target image’ is only a simulated galaxy without a central point source.

<table>
<thead>
<tr>
<th>(\text{mag}_{\text{diff}})</th>
<th>(\text{radius})</th>
<th>(n = 1) (galaxy only)</th>
<th>(n = 4) (galaxy only)</th>
<th>(n = 1)</th>
<th>(n = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.57</td>
<td>5</td>
<td>[0.85, 1.09]</td>
<td>[1.51, 4.15]</td>
<td>[0.81, 1.13]</td>
<td>[3.58, 5.52]</td>
</tr>
<tr>
<td>2.57</td>
<td>10</td>
<td>[0.75, 1.45]</td>
<td>[2.00, 4.32]</td>
<td>[1.30, 1.67]</td>
<td>[2.10, 6.48]</td>
</tr>
<tr>
<td>1.82</td>
<td>5</td>
<td>[0.61, 1.17]</td>
<td>[2.46, 3.37]</td>
<td>[1.06, 1.23]</td>
<td>[2.81, 4.48]</td>
</tr>
<tr>
<td>1.82</td>
<td>10</td>
<td>[0.99, 1.12]</td>
<td>[2.36, 3.81]</td>
<td>[1.08, 1.46]</td>
<td>[2.23, 3.89]</td>
</tr>
<tr>
<td>1.06</td>
<td>5</td>
<td>[0.80, 1.00]</td>
<td>[4.05, 4.53]</td>
<td>[0.87, 1.17]</td>
<td>[3.18, 3.92]</td>
</tr>
<tr>
<td>1.06</td>
<td>10</td>
<td>[0.79, 1.05]</td>
<td>[3.53, 4.12]</td>
<td>[0.93, 1.03]</td>
<td>[3.87, 4.36]</td>
</tr>
</tbody>
</table>

Before, the characterization ability of the host galaxy could be improved to the background noise level. In fact, we also ran GALFIT with fixed PSF magnitude for the fitting process and we noticed that this really improves the reliability of the recovered galaxy parameters. We now understand that the problem of determining the flux for central point source is a serious weak point for 2-D fitting methods, while PCA-based PSF subtraction and further modeling using MCMC algorithm is much better in this respect. Besides GALFIT, we also ran the two-component MCMC fitting procedures for the same set of simulated QSO images. However, we found that keeping the flux of the central bright point source as a free parameter would make it difficult for the fitting process to explore the whole parameter space and the PSF mismatch problem still exists.

In order to understand the effects of PSF mismatch on the recovery of the galaxy light profile, we compare the PCA-based PSF subtraction method with the traditional method where the flux of the \(\text{psf}_{\text{star}}\) is simply aligned and scaled to match the value of the central five pixels of the \(\text{psf}_{\text{target}}\) image, and the scaled \(\text{psf}_{\text{star}}\) image would be finally subtracted from the QSO image. The contours of the residual images from PCA-based PSF subtraction ('\(\text{psf}_{\text{star}}\) group' as the reference library) and tra-
Fig. 3 Comparison of the residual map between the PCA-based PSF subtraction and the traditional method. The $x$-axis and $y$-axis are in arcsec. We found that for the innermost part of the residual image, the PCA-based PSF subtraction method eliminates the effects of PSF mismatch better than the traditional method. As the radius becomes larger, the difference between these two methods gradually diminishes to the background noise level.

Furthermore, we exchanged the two data sets ‘psf\_target group’ and ‘psf\_star group’. In this case, the ‘psf\_target group’ will be employed as the reference star library, while the median image of the ‘psf\_star group’
will be taken as the ‘psf target’. We executed the same PSF subtraction procedures as before and plot the results in Figure 5. From comparing Figure 3 and Figure 5, we ascertain that the PSF mismatch is really caused by the difference between the two datasets, and the PCA-based subtraction method would minimize this kind of difference. Also we again ran the simulation procedures as we did before, and we noticed that the recovered parameters for the galaxy light profile are quite close to the results which we display in Figures 1, 2 and Table 1. These comparisons and results demonstrate that the PCA-based PSF subtraction and MCMC fitting procedure could maximize the parameter space explored, and optimize the recovery of the galaxy morphology profile with only the effects of noise caused by background and limited PSF mismatch.

Another issue that would lead to deviation of recovery results is whether the signals of the host galaxies are sufficiently significant over the PSF mismatch and sky noise. We directly compare the signal of the set of simulated host galaxies with the noise level at different radii as depicted in Figure 4, in order to understand the reliability of the recovered parameters for different kinds of host galaxy light profiles. In Figure 4, we only examined four of the faintest galaxies from the set of simulated galaxies ($m = 20.5$, $m = 21.0$, $m = 21.5$, and $m = 22.0$).
two disk galaxies and two bulges), and plot their count rate distribution against the PSF mismatch level. Unlike the disk galaxies, the count rate of the bulge galaxies is much more concentrated in the center part. Therefore, PSF mismatch becomes the more substantial noise source in the host galaxy recovery procedure within the central 5-pixel range, and this makes the recovery of parameters for bulge galaxies parameters harder than that of disk galaxies.

5 APPLIED EXAMPLE - SDSSJ094737.70+110843.3

To further verify this PSF subtraction method and the modeling of the host galaxies, we applied it to one of the 19 quasar targets from Mechtley et al. (2016) to decompose the host galaxy from this high-z QSO. Though all the 19 quasar targets from Mechtley et al. (2016) were decomposed with simultaneous fitting of the central point source and the underlying host galaxy profile was based on Bayesian modeling method, only the target SDSSJ094737.70+110843.3 was presented with detailed results of posterior probability distribution for different parameters in figure 2 from Mechtley et al. (2016). So, this target is suitable for testing with our method. The quasar at $z = 2$ was observed with the WFC3 infrared channel using the F160W filter, and the integration time was 1597 s. Each observation was split into four exposures, and dithered using the standard four-point half-pixel box pattern.

The data processing procedures were elaborated on in Mechtley et al. (2016). We began with individual flat-fielded flux-calibrated exposures retrieved from the HST archive. The four exposures for each pointing were combined using the AstroDrizzle software package (Koekemoer et al. 2013) with an output plate scale of 0.060 per pixel and a pixfrac parameter of 0.8. Also, we generated variance maps by copying the WFC3 ‘ERR’ arrays into the standard image arrays, and re-ran the drizzle process using the same parameters and weighting scheme. The variance maps are required by Bayesian analysis since the count rate of the central point source for the bright QSO dominates over the sky background. Therefore, underestimated errors could cause problems for fitting the galaxy morphological parameters.

For various reasons, it was chosen to build a library of empirical PSF models from the WFC3 archival data which contain bright stars with high S/N near the center of the field of view, where all the quasar targets were observed. We adopted exposures for three PSF stars which were selected from two quasar host galaxy study programs, and the images of these PSF stars could provide accurate count rate determination in their cores. In the same way as for the quasar target, we created drizzled images and variance maps of these stars using the specified plate scale and weighting scheme.

We applied our PCA-based PSF subtraction and MCMC fitting procedure to this target and the posterior probability distribution is shown in Figure 6. Though the parameters we adopted are slightly different from the parameters from Mechtley et al. (2016), we could still compare the key information on the light profile distribution. The recovered radii of the host galaxy from both our procedure and Mechtley et al. (2016) are consistent, while the Sérsic index from our calculation is about 0.2 more than that from Mechtley et al. (2016), and the $mag_{diff}$ is about 0.35 lower than the Sérsic mag — PSF mag in figure 2 of Mechtley et al. (2016). To examine the reason why there is a small gap between our results and those in Mechtley et al. (2016), we did the same calculation using the two-component MCMC fitting as Mechtley et al. (2016) did for this target. We found that the recovered results from both the PCA-based PSF subtraction and MCMC fitting procedure and this two-component MCMC fitting procedure are almost identical. Thus, we assumed that the difference mainly comes from the PSF stars which we selected and processed, where we only adopted three suitable PSF stars from the HST library, while Mechtley et al. (2016) had iteratively searched for the best PSF star from the library in the fitting procedure.

6 SUMMARY

In this paper, we have described a PCA-based PSF subtraction method, as well as further modeling of the galaxy light profile using an MCMC algorithm. For QSO host galaxy studies, this method can overcome the problem of unknown brightness for the central point source and minimize bad effects from the PSF mismatch. Therefore, this would improve the ability to characterize the underlying host galaxy of high-z QSOs to the background noise level.

To verify the scientific performance of this algorithm and compare with other methods, we have created a set of mock images of high-z QSOs, and the morphological parameters (such as radius, magnitude and Sérsic index) of their underlying faint host galaxies would cover an observationally reasonable range. According to the simulations in which the PCA-based PSF subtraction and MCMC fitting were applied to the mock images to recover the galaxy parameters, we found that our procedure has advantages over simultaneous fitting of the central point source and the underlying host galaxy profile since this method could largely utilize information from the PSF library to eliminate the error introduced by PSF mismatch. We also found that our method has better performance for disk galaxies than for bulge galaxies, since the light profiles of bulge galaxies mimic the PSF itself to some extent and this would lose information on the galaxy light profile during the PCA-based PSF subtraction procedure.
The PSF subtraction method that we adopt here is general and widely applicable to studies that require removal of the bright central point source. In the case of high-

Acknowledgements This project/publication was made possible through the support of a grant from the John Templeton Foundation and the National Astronomical Observatories, Chinese Academy of Sciences. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation or the National Astronomical Observatories, Chinese Academy of Sciences. WH acknowledges the Subaru team and the National Natural Science Foundation of China (Grant No. 11890694). YPW acknowledges the National Natural Science Foundation of China (Grant Nos. 10173025, 10673013, 11073031 and 11653005) and the National Basic Research Program of China (973 Program, TG 2000077602).

This research made use of Astropy\textsuperscript{1}, a community-developed core Python package for astronomy (The Astropy Collaboration et al. 2013, 2018). MCMC sampling made use of EMCEE (Foreman-Mackey et al. 2013) and MCMC corner plots made use of CORNER\texttt{.PY} (Foreman-Mackey 2016).

References

Kormendy, J., & Richstone, D. 1995, ARAA, 33, 581
Wang, Y. P., He, W., Yamada, T., et al. 2015, RAA (Research in Astronomy and Astrophysics), 15, 673

\textsuperscript{1}http://www.astropy.org