Image Improvement and Restoration in Optical Time Series

Yash Gondhalekar$^1$, Margarita Safonova$^2$, Archana Mathur$^3$ and Snehanshu Saha$^4$

$^1$CSIS, BITS Pilani, K. K. Birla Goa Campus, Goa, India
$^2$Indian Institute of Astrophysics, Bangalore, India
$^3$Nitte Meenakshi Institute of Technology and Management, Bangalore, India
$^4$CSIS and APPCAIR, BITS Pilani, K. K. Birla Goa Campus, Goa, India

ABSTRACT

Globular clusters (GCs) are considered strong candidates for hosting free-floating planets. However, since they are not bound to a star, they are undetectable by any conventional planetary detection methods: transit, radial velocity, or direct imaging. Gravitational microlensing, which causes transient brightening of background stars by passing foreground masses, is, on the other hand, an established method of detecting planets and proves promising for application in GCs. By applying differential photometry on the time-series images of GCs, we can extract variability events, build light curves, and inspect them for the presence of microlensing events. However, instrumental anomalies and varying observing conditions over a long observational campaign result in the distortion of the stellar PSF, which affects the subtraction quality and leads to false-positive transient detection. We use the Scaled Gradient Projection (SGP) iterative image reconstruction algorithm to restore stellar shapes in our data. We investigate using a more flexible divergence measure, the $\beta$-divergence, as opposed to the commonly used Kullback–Leibler divergence in SGP, adapting parameter $\beta$ to the data. An extensive empirical study comparing SGP performance under these two objective functions is carried out using several physically-motivated metrics such as Full-Width-at-Half-Maximum (FWHM), ellipticity, and radial profiles. Our experiments on star stamps, extracted from real-sky GC field images, suggest that $\beta$-divergence not only improves the restoration quality but also enhances flux conservation compared to the KL divergence. We find that using $\beta$-divergence in SGP is a promising approach in astronomical image restoration.

Key words: methods: numerical – techniques: image processing – globular clusters: individual: (NGC 6205)

1 INTRODUCTION

Globular clusters have always been a working laboratory for observers, providing prolific hunting grounds for exotic objects, whether it is elusive ‘redback’ millisecond pulsars (Zhang et al. 2022), free-floating planets (FFPs) (Safonova et al. 2016), or intermediate-mass black holes (IMBHs) (Miller & Hamilton 2002), to name a few. However, the dedicated planetary search campaigns produced null results (Gilliland et al. 2000; Weldrake, Sackett & Bridges 2007, 2008; Nasimbeni et al. 2012), and intensive searches for GCs IMBH through their accretion signatures (see, e.g., Wrobel et al. (2019)) also did not yield any results to date. Planets in GCs are proposed to exist as FFPs (Hurley & Shara 2002). Moreover, in the top 20 dense clusters, FFP number may exceed the number of stars by a factor of ~100 (e.g., Fregeau et al. (2002); Hurley & Shara (2002)). The traditional detection methods, such as transit or radial velocity, will not work for the FFPs, and their direct imaging is also not possible due to large distances – the closest GC to Earth is M4 (NGC 6121) at ~2 Kpc away. Attempts to detect the central IMBHs are also not direct and present enormous observational difficulties due to the severe crowding of stars in GC cores. Gravitational microlensing (ML) – a transient brightening of a background star due to gravitational deflection of light by foreground mass(es), on the other hand, is a method that may work ideally for the case of GCs. In globular clusters, distances and velocities of the lenses and sources are well constrained, breaking the usual mass-distance degeneracy. Since no background stars are detectable within the highly crowded core, any discovered event will be due to a star in the globular cluster. The probability of lensing is high in the dense environment of a GC, and different mass distributions in the centers (binary or singular central IMBH, assembly of low-mass stellar remnants) would result in different lensing signatures. GCs are also ideally suited for gathering CCD photometry, providing measurements for thousands of stars in a single frame, maximizing the temporal coverage, and increasing the probability of detection (Safonova et al. (2016) and refs. therein). Safonova (2010) proposed the ML method to search for the central cluster’s IMBH, where observations are required to be conducted over a few years due to the long time scales of central IMBH ML events. Our group monitored a set of selected GCs looking for ML signatures of possible
central IMBH (Safonova and Stalin 2010). It continued with the ML search for FFPs (Safonova et al. 2016), accumulating hundreds of time-series frames of about 20 Galactic GCs. To extract the photometric variability, we employ the differential imaging analysis (DIA) method, which showed excellent results when applied to large data sets, and was shown to be sensitive to ML events even for very faint source stars (Alcock et al. 2000).

Successful differential photometry of crowded fields requires images to be of relatively uniform quality. Modern DIA techniques usually consider the effects of variable atmospheric extinction and exposure times. The methods even work better as the crowding increases because more pixels contain information about the PSF difference in denser fields. However, if a time series is taken over the long baselines, other sources of noise can creep in, such as e.g., PSF distortion, which DIA usually cannot handle even for moderately elongated PSFs. Such unfortunate scenarios mainly occur due to either strong winds during the exposure or bad focusing in the telescope.

One of the most critical steps in image subtraction is to derive a kernel that matches the PSF and background variations between a chosen reference image and the test image. The shape of the kernel depends on the shape of the stars on the test image. It has been verified that the kernel can model deviations from the ideal circular shape, even if the stars on the test image are elongated or blurry (Alard & Lupton 1998). However, we note that this is true only up to very mild distortion levels. In cases of higher levels of distortion of star shapes on the test images, the subtraction results are often sub-optimal, leaving unwanted residuals around the star location.

We cannot entirely rely on the kernel derivation process in such scenarios. Since such images adversely affect the image subtraction process and hence the detection of transients, they are usually discarded from the dataset so that only good images remain (see, e.g., Sec. 2.1 in Servillat et al. (2011)). However, when the observing program is over a long baseline and the exposures are taken at a low cadence, each image becomes vitally important as it is impossible to repeat them. Moreover, since we want to extract as much information as possible about the rare ML events, simply discarding them could lead to a significant data loss. As a result, we need to restore stellar shapes from the affected images in the time series to use them in subtraction and further inspection, keeping the information loss at a minimum.

The distortions in stellar shapes could be seen in the form of triangular, diamond, elliptical, or other shapes (we do not include the uniform spread in PSF here since DIA packages can model them using an optimal kernel solution). Such poor images affect the image subtraction process, leading to 1) false-positive detection of transients, and 2) large-scale noise structure in the subtracted image. Here, we aim to restore the stars’ circular shapes while preserving their flux using iterative statistical methods for single-image deconvolution.

The SGP algorithm, initially proposed in Bonettini, Zanella, & Zanni (2009), has proven promising for astronomical image deconvolution with prior constraints in recent years. SGP has been studied on various astronomical sources ranging from point sources in open clusters to extended objects like nebulae (Prato et al. 2012) or to restore motion-blurred star images obtained from a star sensor (Wang et al. 2018). SGP-based blind deconvolution was also studied in Jia et al. (2017), and an improved SGP method along with a PSF estimation algorithm was proposed in Wei & Bai (2015). In this paper, our main focus is to test the efficacy of SGP with a flexible divergence measure, the $\beta$ divergence, as opposed to the commonly used KL divergence and, further, to allow the parameter $\beta$ to learn from the data. Through a comprehensive empirical study of SGP with both KL and $\beta$-divergence using several quantifiable metrics, we show that, due to its more flexible nature, $\beta$-divergence shows improved restoration results in many cases, and also exhibits better flux conservation than KL divergence. We suggest that $\beta$-divergence is a promising alternative to KL divergence for image restoration in real observational data.

2 METHODS

2.1 Dataset and Data Reduction

The dataset used for analysis and experiments in this paper consists of time-series images of the Messier 13 (NGC 6205) globular cluster, henceforth M13, obtained in 2008–2015 period on the 2-m Himalayan Chandra Telescope (HCT) of the Indian Institute of Astrophysics (IIA), Leh, Ladakh, LAO, located at 4500 m above sea level. As the lensing curves are achromatic, to distinguish from variations due to other phenomena, the observations were carried out in two filters (I and V bands) several times a night (when possible) in short exposures to avoid saturation of bright cluster stars. This study is focused on 244 2K×2K I-band images taken with the Himalayan Faint Object Spectrograph and Camera (HFOSC) mounted on the HCT. HFOSC is equipped with a Thompson CCD of 2048×2048 pixels with a pixel scale of $0.65/296$ pix, equivalent to a total field of view (FOV) of $\sim 10' \times 10'$. The readout noise, gain, and readout time of the CCD is 4.87 $\epsilon$, 1.22 $\epsilon$/ADU, and 90 sec, respectively. The typical seeing values observed in the images from the dataset are 6 – 8.5 pixels.

All images were subjected to the usual image reduction process (bias subtraction, flat-fielding, and cosmic rays removal) using the IRAF (Tody (1986); Tody (1993)) scripts developed by our group. Flat fields were constructed from dithered images of the twilight sky, and any star images were removed by combining flats in each band using a median filter. Illumination and fringe corrections were not required for our dataset because the detector’s field of view is small, and the dark current is negligible since the CCD is cooled down to $\sim 100^\circ$C. To remove the cosmic rays, we employed the IRAF task crmedian which uses a median filtering approach to replace cosmic ray pixel values with the median value. We also developed a utility script to automate removing bad bias and flat frames.

2.2 Difference Image Analysis Package (DIAPL)

Traditional profile-fitting codes like the DAOPHOT (Stetson 1987) and DoPHOT (Schechter et al. 1993) have been used

1 IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the National Science Foundation.
in the past for stars’ variability search. Such approaches iteratively subtract brighter stars to reveal fainter stars which work well only up to moderately crowded fields. However, in highly crowded fields used typically in ML searches, even the very bright stars could be blended due to which traditional PSF-fitting is sub-optimal. Moreover, in such settings, if a particular variable star is not present on the reference frame, it could lead to a severe loss of information, especially given that ML events are infrequent.

It is suggested to use the image subtraction technique to mitigate this problem by instead leveraging time-series images of a crowded field for inspecting stars for variability. The PSF of the stars on the test image, which need to be differentiated, is preserved after the subtraction, and one could perform photometry on the difference images (Wozniak 2000). It also alleviates the problem of multi-PSF fits due to profile blending that routinely occurs in observations of densely crowded fields. Alard & Lupton (1998) suggested an image subtraction technique to find an optimal kernel solution while also fitting differences in the background level between the reference and test frames, to obtain good quality subtracted images that take into account information from even the most crowded fields to devise a kernel. Using differential photometry on the time series of GC images, one could build the light curves, extract intrinsic variability events and inspect them for the presence of microlensing events.

The DIAPL package\(^2\) is an efficient implementation of the Optimal Image Subtraction (OIS) method described by Alard & Lupton (1998) and Wozniak (2000). It matches the reference (generally a mean stack of best-seeing frames) and test frames, matches the background between the two frames, and devises a kernel solution to subtract the degraded version of the reference frame from the test frame. In practice, we found it efficient in handling overlapping stellar profiles in densely crowded fields, making it useful for our study with good results, provided the images are sufficiently good quality.

### 2.3 Modeling the Point Spread Function

PSF models are an excellent mathematical representation of the shapes of the stars and allows us to better represent and quantify star shapes instead of using the stars themselves (Berry & Burnell 2005). Here we use an analytical approach to model the PSFs by using the PSF coefficients output from

\(^2\) We used the DIAPL code for Difference Image Analysis (DIA) (Wozniak 2000) as modified by Wojtek Pych. The package, along with its documentation, can be found at [https://users.camk.edu.pl/pych/DIAPL/](https://users.camk.edu.pl/pych/DIAPL/).
DIAPL. DIAPL first finds candidate PSF stars in the input frame (using the sfind program) and then calculates the PSF model parameters, such as the raster size of the PSF model, $x$ and $y$ width scales of the PSF, number of Gaussian functions used to build the PSF model, degree of polynomial describing local and spatial dependent shapes, etc., using the getpsf program. DIAPL calculates the PSF model on the reference frame for image subtraction by default. We divided the full frame into $4 \times 4$ subframes to eliminate possible spatial dependence of the PSF, modified the getpsf program, and extended it to find a PSF model for each subframe of each original frame by storing the relevant PSF coefficients in a binary file and using those to develop a PSF model in the form of a 2-D matrix. In Appendix A, we describe the procedure to convert a functional PSF model to a matrix representation.

### 2.4 Quantification of PSF shapes

We use the PSF models to distinguish between good and bad original images. We opt for a set of concrete criteria for the segregation of PSF shapes instead of entirely relying on subjective visual classification that is both time-consuming and error-prone.

Since the subframe PSFs did not show significant variation across the field of view (due to the small FOV and similar distortion across the entire field), we use the mean of subframe PSFs to construct a single PSF model for the whole frame, which is used for analysis further in this section. We note that slight ellipticity in the shapes of stars does not pose a problem, since the image subtraction kernel can be modeled appropriately and is flexible to a certain extent. We aim here to select the frames with either high ellipticities, or of weird and unknown shapes.

We define a heuristic as follows: we require knowing a few visually good examples of good and bad PSFs that the human eye can immediately recognize. These would be used for the automated classification of all PSF models. We designate good PSFs as "Class 1" and bad PSFs as "Class 2". We found the slope of the least-squares fitted line to the radially averaged 1-D power spectra of the PSF matrices to be a sufficiently good indicator for distinguishing between good and bad seeing PSFs for our purpose. Using the mean of slopes of power spectra of visually good examples from Class 1 and Class 2, we classified all PSF matrices into Class 1 if their power spectrum slope lies within one standard deviation of the slopes of all visually selected good PSFs. Since it is acceptable not to select slightly distorted PSFs, as mentioned above, we compute the Mean Structural Similarity Index (MSSIM) between the PSF with the lowest Full-Width-Half-Maximum (FWHM) from the dataset and the current PSF and repeat this for all calculated PSF models. We empirically decided to use a threshold of 0.95. If MSSIM $\geq 0.95$, we do not consider it a bad PSF even though it did not have a rapidly decaying power spectrum. This also ensures that we exclude near-circular but wider PSFs. This procedure classified 29 out of the 244 images, including images from the visually selected samples, to be considerably distorted, and we only consider stars from these images for restoration.

Figure 1 shows subframes extracted from the original images in the top panel, corresponding to the varying PSF degradation levels. It also shows the corresponding subtracted subframes obtained using the DIAPL package's pipeline.

### 2.5 Extraction of starcutouts and flux criterion

Before investigating the applications of SGP, we would like to briefly introduce some aspects. It has been observed that while the restoration algorithms can be applied to star-field images containing multiple stars from simulated images, their application on real star-field images is not straightforward, in that it tends to suppress the dimmer stars and enhances the brighter stars (Wu & Barba 1998). In SGP, it is speculated that the projection onto the non-negative orthant could be responsible for the suppression of the fainter stars for which Prato et al. (2012) suggested a possible remedy: to decrease the tolerance used in the KL divergence stopping criterion. To alleviate any possible issues due to this problem, we here apply SGP on square-sized star cutouts extracted from the original GC images. To extract star cutouts, we leverage the photutils package's Cutout2D class. The cutout size can be set such that the cutout serves as a minimal bounding box surrounding the star, padded with a few pixels to ensure full containment of the stars' wings. The local background level is estimated using a larger region around the star before extracting the cutouts.

SGP, at the least, requires the observed image (in our case, the extracted star cutout), the estimated PSF model, and the estimated scalar background level as input. For background estimation, we use the Background2D class from photutils and use the median value of the background map to yield a scalar background level. Despite the visual insignificance of spatial variation of the PSF over the full field, we utilize the sub-frame estimated PSF model for restoring stars observed in that sub-frame. The flux is estimated as $\sum I - N \cdot bkg$, where $N$ is the total number of pixels in the observed star cutout, $I$, with a background level, $bkg$. We also note that since we focus on stars that are distorted in shape, measuring flux by integrating a Gaussian function fitted to the star is unreliable since the fit would tend to have high residuals.

### 2.6 Deconvolution methods theory

The equation for an astronomical image acquired using a CCD can be given by:

$$g = Af + b + \eta,$$

where $g$ is the observed degraded and noisy image from a telescope, $A$ is the PSF, $f$ denotes the undegraded noise-free image of the object that is unknown and to be estimated, $b$ denotes the background level, and $\eta$ is the additive read-out noise. More specifically, modeling of the observed image can be described by a mixed Poisson-Gaussian noise (Bertero, Boccacci, & De Mol 2021). In many practical cases, the additive read-out noise component is often modeled by Poisson noise (Snyder et al. (1994); Bertero et al. (2009); Anconelli et al. (2005)) or simply ignored for simplification (approaches using least-squares methods, that assume additive noise models, on images corrupted with Poisson noise also exist (Vio, Bardsley & Wamsteker 2005)), unless the images have very low photon count in which case taking into account the readout noise might be important. In our case, the images have been bias-subtracted, so we do not model the readout noise.
It is a well-known fact that direct solutions to estimate \( f \) are often unacceptable due to the ill-conditioned nature of the restoration problem (Molina et al. 2001). We also verify this fact by noting that the condition numbers of the obtained PSF matrices in our case are much larger than one; they lie roughly in the range \( 10^{17} - 10^{24} \), which signifies that the matrices are ill-conditioned and near-singular. In such cases, its inverse calculation is prone to significant errors, and it is suggested to use iterative restoration methods instead of finding direct inverse solutions (in a maximum likelihood setting, for example, we would like to move closer to the maxima of the likelihood function iteratively and as such it is the-the advantages since minimizing the KL divergence is equivalent to the traditional divergence measure used in SGP and has great background-added version of the partially reconstructed im-

As such, it is suggested to use iterative restoration methods instead of finding direct inverse solutions (in a maximum likelihood setting, for example, we would like to move closer to the maxima of the likelihood function iteratively and as such it is the-

Note that while the authors in the above references considered the case \( \beta \geq 1 \), the definition can be extended for all real values of \( \beta \), as described in Févotte & Idier (2011). It is a family of functions parametrized only by a single parameter \( \beta \) that controls the trade-off between the robustness and efficiency of the estimators of parameters. The special cases of \( \beta = 2, 1, 0 \) correspond to the Euclidean distance, the generalised KL divergence, and the Itakura-Saito divergence, respectively. Specific values of \( \beta \) can be used if one has complete knowledge of the noise model: \( \beta = 2 \) can be used in cases of Gaussian noise, \( \beta = 1 \) for Poisson noise, and \( \beta = 0 \) for multiplicative Gamma noise (Févotte & Cemgil 2009).

A neat feature of this class of divergence is that it smoothly connects commonly-known distance measures described above. A common application of \( \beta \)-divergence has been in non-negative matrix factorization (Kompass (2007); Févotte & Idier (2011)). Another peculiar feature of the class of \( \beta \)-divergences is the robustness to outliers which was exploited in various applications (Basu et al. (1998); Mihoko & Eguchi (2002); Akrami et al. (2022)).

As stated above, astronomical images acquired from CCD consist of a combination of Gaussian and Poisson noises and hence a Poisson noise assumption is not perfect. Although this assumption is valid in many cases, we hypothesize that a value of \( \beta \in (1, 2] \), or some value around \( \beta = 1 \), might serve as a better loss function in some cases due to the complicated nature of noise in real observational images.

Here apart from using \( \beta \)-divergence instead of KL divergence in SGP, we exploit the parameterized nature of \( \beta \)-divergence by adapting \( \beta \) to the data using gradient descent, for further improvement and investigate whether a suitable \( \beta \), learned from the data, can outperform the theoretically-rationalized choice of KL divergence. The recipes needed are: calculating the \( \beta \)-divergence of the blurred and background-added version of the reconstructed version from the original image, its derivative defined as:

\[
\nabla J(f; g) = (Af + b)^{\beta-1} - A^T g(Af + b)^{\beta-2},
\]

and the derivative of \( \beta \)-divergence with respect to (w.r.t) \( \beta \), for updating the parameter \( \beta \). Only the last recipe is a new one and the other two are simply the \( \beta \)-divergence equivalent of KL divergence as used in the original SGP algorithm. It is important to note that the \( \beta \)-divergence is convex w.r.t \( Af + b \) for \( \beta \in [1, 2] \) (Févotte & Idier 2011), so one can restrict \( \beta \) within this range if convexity is desired. However, we here perform no such restriction. We also note that adapting \( \beta \), as done here, is not necessary, in which the case one must appropriately set \( \beta \). Further details of the SGP algorithm, including the modification to update the parameter \( \beta \) and a few other changes, are described in Appendix B.
2.7.1 Termination criterion

There are several stopping rules that can be used depending on the problem, to prevent amplification of noise during the iterations (see Section 3.3 of Prato et al. (2012) for a review of some rules). For point sources in simulation studies, SGP can generally be pushed to a convergence (Prato et al. 2012) (see Bertero et al. (2004) for a similar observation in the case of the RL algorithm).

Here, we terminate based on the convergence of the objective (or the data fidelity) function such that the iteration is stopped when:

\[ |J(f^{k+1};g) - J(f^k;g)| \leq tol \cdot J(f^k;g), \]

where \( tol \) is the tolerance level which we set to \( 10^{-4} \). Apart from this, we also set a maximum iteration limit to 500 which provides regularization.

3 EXPERIMENTS AND RESULTS

3.1 Setup and evaluation metrics

We demonstrate the capabilities of the SGP algorithm, leverage several metrics to quantify the restoration quality, and compare the KL divergence and \( \beta \)-divergence variants of SGP. All experiments were conducted using Python 3.8.10 on a computer equipped with Intel(R) Core(TM) i3-1005G1 CPU processor at 1.20 GHz.

We test the algorithms on a set of star cutouts. To ensure minimal selection bias, we randomly select a few stars from each image obtained using the procedure described in Section 2.4, which gives us a total of 170 star cutouts, each of \( 30 \times 30 \) pixel size. This allows us to test the algorithms under varying conditions (low and high photon count) and star shapes. We use the PSF model for the subframe where that star is located. Moreover, the PSF models obtained from the description in Section 2.3 were not centered, and hence we center the PSF matrix using scikit-learn’s KernelCenterer class before restoration to minimize centroid shifts in the restored image.

The initialization of the restored images is set to be the same as the input star cutout for all algorithms. It is known that even if SGP has a lot more parameters than RL, extensive experiments have led to an optimization of the parameters such that no specific parameter tuning is needed irrespective of the application (Prato et al. 2012). For this reason, we do not perform any SGP parameter tuning and use the default values described earlier (Prato et al. 2012) noted that any choice of the step length parameter \( \alpha \) belong to \([\alpha_{min}, \alpha_{max}]\) is a valid choice, and that \( \alpha \) can be tuned inside this interval to optimize the reconstruction performance; however we have not experimented with different values of \( \alpha \). It has been found that including the flux constraint does not remarkably improve the convergence rate of SGP (see Fig. 4 in Bonettini, Zanella, & Zanni (2009)). However, due to photometry considerations, we have necessarily used the flux conservation constraint. It is also important to note that our implementation here does not include boundary effect correction (see Bertero & Boccacci (2005) and Bertero et al. (2013) for discussion on this issue).

For experiments here, we randomly choose the initial value of \( \beta \) uniformly in the range \([0.95, 1.05]\) 30 times, and select the initial \( \beta \) that yields the lowest FWHM and ellipticity among all the tried \( \beta \) values. The reason for a narrower interval is because we did not find any significant improvement in the results by choosing larger \( \beta \) values (in fact, the performance got worse for larger \( \beta \) values such as 1.2 or 1.5, a similar point was noted in Basu et al. (1998) where larger \( \beta \) values were found to be less efficient). Moreover, we sometimes found improved results by using an initial \( \beta \) value < 1. Hence, the extension of the interval below 1 (note that, as stated in Section 2.7, convexity is guaranteed only for \( \beta \in [1,2] \), but here we also use \( \beta \) values < 1 since we empirically found it to also give good results). Empirically, we found it beneficial to terminate the iterations using \( \beta \)-divergence early (we set the maximum iteration limit to 10), since otherwise the restored results looked visually bad. The reason why \( \beta \)-divergence, in our formulation here, might require less iterations to generate plausible and better reconstructions than KL divergence is because along with the reconstructed image, the parameter \( \beta \) also gets updated, which might help achieve faster convergence. In general, our experiments suggest that early-stopping in the case of \( \beta \)-divergence is a crucial aspect to achieve good results.

For calculating the metrics to quantify restoration quality, we have used the SourceCatalog class from photutils (Bradley et al. 2022) that conveniently yields several source properties such as the FWHM, ellipticity, and flux.

We also use radial profiles of the original and restored stars as a metric for comparison. For calculating the radial profile of a star, we design radial bins and calculate the number of bins lying inside each bin, weighted by the data values. We use the 1-Wasserstein distance (hereafter, 1-WD) to evaluate the restoration quality. For this, we compute the 1-WD by fitting two different 1-D Gaussian functions to the original and restored star’s profile and compare the 1-WD between the radial profile and the fit before and after restoration. The Gaussian is represented as \( f(x) = A \exp \left(-\frac{(x-x_0)^2}{2\sigma^2}\right) \), where \( A \) is the amplitude, \( x_0 \) and \( \sigma \) are the mean and the standard deviation of the Gaussian function, respectively. The fit is done in the least-squares sense using the iterative Levenberg-Marquardt procedure, for which we use the LevMarLSQFitter class from astropy (Astropy Collaboration et al. 2013, 2018). We set the amplitude \( A \) to be 0.8 times the maximum pixel value of the star cutout whose profile needs to be calculated (Zucker et al. 2018), and set the standard deviation using the relationship between FWHM and the 1-\( \sigma \) Gaussian standard deviation, \( \sigma = \frac{\text{FWHM}}{2\sqrt{2\ln 2}} \) since we already calculated the FWHM of the original and restored stars. We note that by fitting a Gaussian, we have implicitly assumed the ideal star profile is a Gaussian function.

3.2 Restoration quality comparison

Figure 2 shows a few examples of the restoration results. It can be visually observed that the restored stars using \( \beta \)-divergence are more compact than using KL divergence. Smaller 1-WD values when using \( \beta \)-divergence (except in (b)) also provides an indication that the restored stars in these four cases are much closer to a Gaussian profile as compared to KL divergence. As we will see later, this observation is consistent across a large set of star cutouts.
Figure 2. Four different examples of comparison between the restored star cutouts using KL divergence and $\beta$-divergence. The bottom sub-panels in each panel shows the corresponding radial profile and the 1-d Gaussian fit. The 1-Wasserstein distance is also shown for each case, where lower values mean the profile is closer to a Gaussian profile.

We now perform a comprehensive comparison of the performance of SGP with KL divergence and $\beta$-divergence. As stated earlier, the algorithms require the observed image (star cutout in our case), scalar background level estimated from the region around the star, and the PSF model as inputs. Figure 3 shows the comparison between the different approaches using four metrics: flux, FWHM, ellipticity, and 1-WD between the radial profile and the corresponding fitted 1-d Gaussian profiles, across all the extracted star cutouts. It shows that $\beta$-divergence overall provides benefit over the KL divergence. (a) and (b) also indicate that benefits of using $\beta$-divergence over KL divergence is more apparent when the FWHM and ellipticity of the original star is high, which shows that $\beta$-divergence is slightly more robust to different star shapes than KL divergence. This is a fair expectation since the parameter, $\beta$, is allowed to learn from the data, providing a more flexible loss measure. Overall, in $\sim$67.06% cases, the 1-WD using $\beta$-divergence was lower than when using KL divergence, which signifies a slight overall improvement in the radial profiles of stars using $\beta$-divergence.

3.3 Flux conservation comparison

Flux conservation is paramount for reliable photometric studies on the restored stars. This makes flux conservation an important aspect to consider while comparing the two approaches. We conducted an image-by-image flux comparison, and design a criterion to further quantify the flux conservation capability for both, KL and $\beta$-divergence: if the restored star’s flux lies within the error region of 5% of the original star’s flux, then we consider that restoration to be successful in terms of flux conservation, i.e. if $\text{flux}_{\text{restored}} = \text{flux}_{\text{original}} \pm 0.05 \times \text{flux}_{\text{original}}$, then the flux conservation is satisfactory (an error region is required since the background estimation also has some uncertainties). We found that $\sim$87.65% stars using KL divergence and $\sim$94.12% stars using $\beta$-divergence show a satisfactory flux conservation based on the above-defined criterion. We speculate this enhanced flux conservation when using $\beta$-divergence is due to the increased flexibility of the loss function.

4 DISCUSSION AND CONCLUSIONS

Image restoration in astronomy was considered to be a luxurious field until an "impossible" mistake of spherical aberration was identified in the primary mirror of the Hubble Space Telescope (HST) in 1990 (Molina et al. 2001). Since then, much attention and research has been carried out to develop novel techniques for restoring astronomical images. Considering the importance of novel image restoration approaches in astronomy, it is of great interest to experiment with new computationally feasible and practically applicable approaches. Undoubtedly, our scientific analyses will be much
Figure 3. (a) and (b) denotes the FWHM and ellipticity difference plotted against the FWHM and ellipticity of the original star, respectively. (c) and (d) plot the kernel density estimate (KDE) of the flux difference between the restored and original star and 1-Wasserstein distance (1-WD) between the 1-d radial profiles of the restored and original star, respectively. Negative slopes in the fitted lines in (a) and (b) indicate that stars restored using $\beta$-divergence have a lower FWHM and ellipticity than using KL divergence. In (c), the flux difference distribution is narrower and does not contain a heavy tail, unlike the case of KL divergence. In (d), the 1-WD distribution is slightly shifted towards zero for $\beta$-divergence indicating a slight advantage; however the benefits of $\beta$-divergence near tails is less apparent.

more reliable if we could eliminate undesired situations such as bad atmospheric conditions and irregular cadences. Image restoration serves this purpose instead of removing such affected images from the dataset.

In this paper, we performed single image deconvolution where the PSF is known. This work was primarily inspired by the recent advancements in the applications of the SGP algorithm, which is theoretically and empirically proven to be more efficient than the famous RL deconvolution algorithm on a wide range of astronomical images such as nebulae, galaxies, and open star clusters. Due to its effective strategies for improved convergence, SGP has been recently studied extensively to test its plausibility as a possible replacement for RL.

Apart from testing the original SGP algorithm, we also examined using a broader class of divergence measure, the $\beta$-divergence, which encompasses KL divergence as a special case. We note that the changes we experimented in this paper do not modify the SGP algorithm itself, but instead show the capability of SGP under a more flexible loss function. Moreover, while most previous works test the applications of SGP on simulated images where the noise and blurring conditions are fully controlled, we test it here on real astronomical images, without any control over the image quality and observational conditions, and ensured minimal selection bias to select stars for restoration. Through extensive experiments on a set of distorted star cutouts and comparing using several performance metrics, we show that $\beta$-divergence can provide fine improvement to using the KL divergence criterion in the SGP algorithm in many cases. This is the first study, to the best of our knowledge, to test the SGP algorithm using a variety of quantifiable, scientifically-motivated metrics. More detailed studies on a variety of astronomical objects and observational conditions would be required to fully understand the properties of $\beta$-divergence in practice. Overall, using flexible and robust divergence measures such
as the $\beta$-divergence might be a promising line of research to improve current image restoration pipelines.

The strategy described here to update the parameter $\beta$ for making it adapt to the data is in essence a stochastic gradient descent procedure since the updates are made for each image individually. Looking at this from a machine learning point of view, it could be an interesting line of research for restoration of entire globular cluster images and image analysis post-restoration is ongoing, and the results will be reported in a separate communication.

ACKNOWLEDGEMENTS

This research made use of Photutils, an Astropy package for detection and photometry of astronomical sources (Bradley et al. 2022). We thank PRISMA, a project of the Italian Ministry of University and Research, grant 2008T5KA4L, for making their SGP code publicly available. M.S. acknowledges the financial support by the Department of Science and Technology (DST), Government of India, under the Women Scientist Scheme (PH), project reference number SR/WOS-A/PM-17/2019.

DATA AVAILABILITY

The codes associated with this manuscript are available at the following website: https://github.com/Yash-10/beta-sgp/. We also make the data (consisting of M13 globular cluster processed images used in this study) publicly available\(^3\).

REFERENCES

Akrami, H. et al., 2022. Knowledge-Based Systems, 238, 107886. https://doi.org/10.1016/j.knosys.2021.107886.

Alard, C. & Lupton R. H., 1998. ApJ, 503, 325. doi:10.1086/305984

Alcock C., Allsman R. A., Alves D. R., Axelrod T. S., Becker A. C., Bennett D. P., Cook K. H., et al., 2000, ApJ, 541, 734. doi:10.1086/309484

Anconelli, B., Bertero, M., Boccacci, P., Carlibet, M., Lanteri, H., 2005. A&A 430(2), 731. DOI: 10.1051/0004-6361:20041493

Astropy Collaboration, Price-Whelan A. M., Sipőcz B. M., Gignther H. M., Lim P. L., Crawford S. M., Conseil S., et al., 2018, AJ, 156, 123. doi:10.3847/1538-3881/aabe4f

Astropy Collaboration, Robitaille T. P., Tollerud E. J., Greenfield P., Droettboom M., Bray E., Aldcroft T., et al., 2013. A&A, 558, A33. doi:10.1051/0004-6361/201322068

Barzilai, J. and Borwein, J. M., 1998. IMA Journal of Numerical Analysis, 8(1), 141–148. https://doi.org/10.1093/imanum/8.1.141

Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C., 1998. Biometrika, 85(3), 549. http://www.jstor.org/stable/2337385

Berry, R. and Burnell, J., 2005, The Handbook of Astronomical Image Processing, 2nd ed., Richmond, VA: Willmann-Bell.

Bertero, M., Anconelli, B., Boccacci, P., et al., 2004. The Large Binocular Telescope: Challenging Problems in Image Restoration. In Proc. ADA-III (Astronomical Data Analysis-III) Conference. Sant’Agata sui Due Golfi, Italy; 29 April - 1 May 2004, eds. F. Murtagh, G. Longo, J.-L. Starck, and V. Di Gesu’, 2004 eWIC Conferences. DOI:10.14236/euwic/ADA-III2004.19

Bertero, M. and Boccacci, P., 2005. A&A, 437(1), pp. 369–374. https://doi.org/10.1051/0004-6361:20052717.

Bertero, M., Boccacci, Patrizia D., G. & Vicidomini, G., 2009. Inverse Problems, 25, 123006. 10.1088/0266-5611/25/12/123006.

Bertero, M., Boccacci, P., Prato, M. and Zanni, L., 2013. European Astronomical Society Publications Series, 59, 325.

Bertero, M., Boccacci, P. and De Mof, C., 2021. Introduction to Inverse Problems in Imaging. CRC Press. 2nd ed. https://doi.org/10.1201/9781003032755

Bonettini, S., Zanella, R., Zanni, L., 2009. Inverse problems, 25(1), 015002. doi:10.1088/0266-5611/25/1/015002

Bradley L., et al., 2022, astropy/photutils: 1.4.0, doi:10.5281/zenodo.6385735

Dai, Y.-H. & Fletcher, R., 2006. Math. Program, 106, 403. 10.1007/s10107-005-0595-2.

Eguchi, S. and Kano, Y., 2001. Robustifying maximum likelihood estimation. Tokyo Institute of Statistical Mathematics, Tokyo, Japan, Tech. Rep.

Févotte, C. and Idier, J., 2011. Neural Computation, 23(9), 2421. doi:10.1162/NECO_a_00168.

Févotte, C. and Cemgil, A. T., 2009. In Proc. 2009 17th European Signal Processing Conference, pp. 1913-1917.

Freguei J. M., Joshi K. J., Portegies Zwart S. F., Rasio F. A., 2002. ApJ, 570, 171. doi:10.1086/339576

Gilliland, R. L., Brown, T. M., Guhathakurta, P., Sarajedini, A., Milone, et al., 2000. ApJL, 545, L47. doi:10.1086/317334

Hurley, J. R., Shara, M. M., 2002. ApJ, 565, 12511. doi:10.1086/373921

Tody D., 1986, SPIE, 627, 733. doi:10.1117/12.968154

Tody D., 1993, ASPC, 52, 173

Tody D., 1993, ASPC, 52, 173

Jia, P., Sun R., Wang W., Cai D., Liu H., 2017. MNRAS, 470, 1950. doi:10.1093/mnras/stx1336

Kompas, R., 2007. Neural computation, 19(3), 780. https://doi.org/10.1162/neco.2007.19.3.780

Lim, J. S., 1990. Two-dimensional signal and image processing. Englewood Cliffs, NJ: Prentice Hall, 1990, 710 p.

Lucy, L. B., 1974. AJ, 79, 745. doi:10.1086/111605

Mihoko, M. & Eguchi, S., 2002. Neural computation, 14, 1859. 10.1162/089976602760128045.

Miller M. C., Hamilton D. P., 2002. MNRAS, 330, 232. doi:10.1046/j.1365-8711.2002.05112.x

Molina, R., Nunez J., Cortijo, F. J. and Mateos, J., 2001. ISPM, 18, 11. doi:10.1093/imanum/8.1.141

Nascimento, V., Bedin, L. R., Piotto, G., De Marchi, F., Rich, R. M., 2012. A&A, 541, A144. doi:10.1051/0004-6361/201118655

Prato, M., Cavicchioli, R., Zanni, L., Boccacci, P., Bertero, M., 2012. A&A, 539, A133. doi:10.1051/0004-6361/201118681

Richardson, W. H., 1972. JOSA, 62, 55

Safonova, M. & Stalin, C. S., 2010. New Astr., 15, 450.

Safonova, M., 2010. Searching for IMBHs in globular clusters. In Proc. 2nd WSEAS international conference on Nanotechnol-
APPENDIX A: GENERATING A PSF MATRIX

In summary, DIAPL’s procedure to calculate the PSF model parameters is as follows: First it reads the stellar coordinates output from the `find` program along with a bad pixel mask file. Then it rejects stars that lie close to each other or near the edges. It also performs an isolation test for the stars. For this, it only selects stars that are above the NSIG_DETECTσ level and rejects any star whose flux is contaminated due to crowding. We set NSIG_DETECT to 3. On these well-isolated candidate stars, an initial symmetric circular PSF fit is made for a few iterations, followed by a final fit for another few iterations. The candidate stars might still be affected by cosmic rays or blending of profiles with nearby stars. Hence, during the final fit, a sigma clipping procedure on the light distribution of the candidate star is performed to mitigate such undesirable scenarios. All of the parameters required during the whole process are user-defined. The `getpsf` routine calculates a set of PSF vector coefficients which are used to build a PSF matrix. We produce PSF matrices of size 30 × 30 pixels. For this, we need to calculate the pixel value at each location, \((x, y)\), of the defined region, where \(x, y \in [-hw, +hw]\), and \(hw\) denotes the half-width of the PSF model, set to 15 for our case, resulting in a 30 × 30-pixels PSF model. We start with a zero pixel value for each location in the raster, and keep on accumulating values in a pixel based on the NGAUSS and NDEG_LOCAL parameters, denoting the number of Gaussians used to build the PSF model (two in our case), and the degree of the polynomial used to describe the PSF shape, in our case set to two. The PSF model is a sum of two Gaussians, with the first Gaussian describing the core and the second describing the wings of the star. The second Gaussian is set to be \(\approx 0.548\) times wider than the first, which is specified by the SIGMA_INC parameter. It is beneficial to model both of them separately, since both have different shape and statistics, for example, the wings of any star is buried in photon noise and the distance at which this happens is different for fainter and for brighter stars. After setting all the parameters, we reproduced DIAPL’s `psf_core` script from the `phot` program in Python to calculate the values of the entries of the PSF matrix. Figure A1 shows the PSF matrices generated by this procedure. As mentioned in the main text, we generate PSF coefficients and, thus, the PSF matrix for each subframe separately without accounting for spatial variations.

APPENDIX B: SGP ALGORITHM DETAILS

The SGP algorithm, with a modification to update \(\beta\), is shown in Algorithm 1, and \(\beta\) updation is shown in line 13. As seen in line 15, we also schedule the learning rate using an exponential decay schedule given by

\[
\eta_i = \eta_0 \cdot e^{-k},
\]

where \(\eta_0\) is the initial learning rate, \(k\) is the exponential decay parameter, and \(\eta_i\) the learning rate at any iteration, \(i \geq 1\).

We reproduce a Python implementation of the MATLAB code of SGP\(^4\) for single-image deconvolution proposed by Prato et al. (2012). We have modified the flux-conservation step to handle pixel saturation during the projection step alongside ensuring non-negativity of pixels: denoting the pixel counts above which the pixel is said to be saturated by \(c\), we set the condition: \(x_i = \min(x_i, c)\) so that we ensure no pixel is deemed saturated in the restored image. There are several reasons we incorporate this condition. First, if we do not account for saturation, photometry on the restored image would flag the star saturated and hence exclude them for any further analyses, which is undesirable. Second, after restoration the same flux is spread among smaller number of pixels (since the restored star would have a lower FWHM than the original star). Hence it is likely that a pixel exceeds the saturating condition if the pixels in the original image were already near saturation.

Choices for updating the scaling matrix include a diagonal matrix that approximates the inverse of the Hessian matrix, \(\nabla^2 J(x)\) or a diagonal matrix that can be used to rewrite the RL method, in which the latter is computationally less expensive (Bonettini, Zanella, & Zanni 2009). We use the latter option where the updating rule becomes,

\[
d_k = \min\{L_2, \max\{L_1, x_k^e\}\},
\]

\(^4\) http://www.unife.it/prin/software
where $L_1$ and $L_2$ are the lower and upper bounds, respectively, on the elements of the scaling matrix. We use the appealing choice for the bounds, that adapt themselves to the data, described in Prato et al. (2012).

As far as the projection step for flux conservation is concerned, we must solve a non-negative and linearly constrained strictly convex quadratic programming problem for flux conservation. Several linear-time projection algorithms exist in the literature (see Bonettini, Zanella, & Zanni (2009) for references). We use the secant-based approach suggested by Dai and Fletcher (2006) that has shown good performance. For updating the step length parameter, we alternate between the two Barzilai & Borwein step length (BB) rules (Barzilai & Borwein 1998) only after the first 20 iterations, as suggested in Prato et al. (2012) (see Sec. 3 in Bonettini, Zanella, & Zanni 2009 for more discussion). This effective strategy also makes the choice of initial $\tau$ less important for convergence (Bonettini, Zanella, & Zanni 2009). In step 3 of Algorithm 1, the subscript + in the projection operator $P$ denotes the closed convex set containing $x$ that satisfies the flux conservation and the non-negativity constraints on $x$.

The SGP parameters are set as follows: $\beta_0 = 0.4$, $\gamma = 10^{-4}$, $M = 1$, $\alpha_{min} = 10^{-5}$, $\alpha_{max} = 10^5$, $M_0 = 3$, $\tau = 0.5$, $\alpha_0 = 10$, where all values except $\alpha_0$ are taken from Prato et al. (2012), and use a maximum of 1000 flux conservation projections. Moreover, $M = 1$ implies that the line-search strategy reduces to the standard monotone Armijo rule (Bonettini, Zanella, & Zanni 2009), and we have used it since we did not find any significant improvements by using a non-monotone strategy. Hyperparameters due to the inclusion of beta divergence are set as follows: initial learning rate for updating $\beta$, $\eta = 10^{-3}$, exponential learning rate schedule parameter, $k = 0.1$. The procedure we used to set the initial value of $\beta$ is described in Section 3.1.

**APPENDIX C: MORE VISUAL COMPARISON**

Here we provide an extended visual comparison of SGP with KL and $\beta$-divergence, as a continuation to Figure 2 in Section 3.2. The restored results shown here further demonstrate visually that $\beta$-divergence can outperform KL divergence.
**Figure C2.** Extended visual comparison of Figure 2 (Section 3.2) between 16 different restored star cutouts using KL divergence and $\beta$-divergence.