Avoiding the Geometric Boundary Effect in Shear Measurement

Haoran Wang1, Jun Zhang1,2, Hekun Li1, and Zhi Shen1

1 Department of Astronomy, Shanghai Jiao Tong University, Shanghai 200240, People’s Republic of China; betaizhang@sjtu.edu.cn
2 Shanghai Key Laboratory for Particle Physics and Cosmology, Shanghai 200240, People’s Republic of China

Received 2020 December 16; revised 2021 February 18; accepted 2021 February 19; published 2021 April 9

Abstract

In image processing, source detections are inevitably affected by the presence of the geometric boundaries in the images, including the physical boundaries of the CCD, and the boundaries of masked regions due to column defects, bright diffraction spikes, etc. These boundary conditions make the source detection process not statistically isotropic and can lead to additive shear bias near the boundaries. We build a phenomenological model to understand the bias, and propose a simple method to effectively eliminate the influence of geometric boundaries on shear measurement. We demonstrate the accuracy and efficiency of this method using both simulations and the z-band imaging data from the third data release of the DECam Legacy Survey.

Unified Astronomy Thesaurus concepts: Gravitational lensing (670)

1. Introduction

To better understand the physical origin of dark matter and dark energy, a number of ongoing and planned surveys are focusing on precisely measuring the weak lensing effect, such as LSST (Abell et al. 2009), Euclid (Laureijs et al. 2011), KIDS (Hildebrandt et al. 2016), DES (Troxel et al. 2018), and HSC (Hikage et al. 2019). Accurate measurement of the cosmic shear signals remains a challenge (Hoekstra & Jain 2008; Kilbinger 2015; Mandelbaum et al. 2015). An important reason is that shear signals contain a wide range of systematic errors, leading to biased cosmological constraints. Currently, the study of systematic errors in shear measurement is still important (Mandelbaum et al. 2014, 2015; Pujol et al. 2020), including modeling bias (Bernstein 2010; Bridle et al. 2010; Voigt & Bridle 2010; Kacprzak et al. 2014), selection bias (Hirata & Seljak 2003; Martinet et al. 2019; Li et al. 2021), noise bias (Refregier et al. 2012), detection bias (Sheldon et al. 2020), etc. Some of the systematic errors are specific to the shear measurement method (e.g., model bias), and others are general problems for any weak lensing pipeline. In this work, we study the shear bias caused by the geometric boundaries in the images, a problem that we believe belongs to the second category.

There are several different kinds of geometric boundaries existing in a typical CCD image, including the physical CCD boundaries, column defects due to the failure of charge transfer along the readout direction, and masked areas due to, e.g., the diffraction spikes of bright stars. For example, Figure 1 shows part of a typical CCD image with column defects from CFHTLenS (Erben et al. 2009, 2013; Heymans et al. 2012). Because such a boundary condition is locally anisotropic, it breaks the statistical isotropy of source detection/selection, leading to shear biases that are most significant near the boundaries. The purpose of this work is to give a phenomenological model for understanding this effect, and to propose a simple solution to avoid this type of bias. This is done in Section 2. We then give a brief conclusion and discuss related issues in Section 3.

2. Boundary Problem and Its Solution

Near the boundary, a galaxy is more likely to be detected as a valid source if its orientation is parallel to the boundary. This is demonstrated in Figure 2, in which there are two groups of galaxies at two different distances from the boundary (marked by the dark thick line). The four galaxies of each distance group have the same size and shape, but different pointing directions. For the group that is closer to the boundary, it can be seen that the galaxy pointing perpendicular to the boundary (marked red) is not likely a valid source for shear measurement, as part of its shape information is missing. Only the other three sources of the same group survive. Statistically, this effect breaks the isotropy of the intrinsic galaxy shape, potentially leading to an additive bias in shear measurement. The more vertical bad columns an image has, the worse the problem is. Similarly, horizontal boundaries also lead to additive shear biases locally, but with an opposite sign. The effect is localized near the boundary, within a distance that is about the typical size of the galaxies. This type of error can in principle cause B-mode (Crittenden et al. 2002; Schneider et al. 2002) in the shear field. It does not seem possible to correct such a bias with a global additive term, as is done traditionally.

2.1. A Simple Model

To understand the problem further, we build a model to analyze the boundary effect. Since the boundaries are mostly in the horizontal and vertical directions, this effect mainly affects the $g_1$ component. The area of the affected region is determined by the length of the boundary as well as the typical size of the galaxies. Let us assume that the total number of the observed galaxies on a CCD chip is $N$, among which $N^+$ and $N^-$ are sufficiently close to the horizontal and vertical boundaries, respectively. On average, we expect a positive (or negative) additive shear bias $c$ (or $-c$) associated with every galaxy that is close enough to the horizontal (or vertical) boundaries. Thus we can get:

$$g_1^{\text{measure}} = g_1^{\text{true}} + \frac{N^+ - N^-}{N} \cdot c. \quad (1)$$

For simplicity, let us consider the case of a rectangular CCD chip of size $l_x \times l_y$, without any bad columns in the middle. Let

---

In the model-fitting methods, either on single exposure or multiple ones, images with missing parts are often used. We argue in Section 3 that in this case, there are still related anisotropic effects.
us also assume that the typical width of the affected regions along the boundaries is \( s \), which should be comparable to the typical galaxy size. We also assume that the average galaxy number density is \( n \). Then we have: \( N^{\text{ref}} \approx 2n \cdot s \cdot l_x \cdot l_y \), and therefore:

\[
g_{\text{true}}^{\text{ref}} \approx g_{\text{true}} + 2e \cdot \frac{s(l_x - l_y)}{l_x l_y}. \tag{2}
\]

2.2. Simulations

To demonstrate Equation (2), we use two types of mock galaxies: (1) random walk galaxies (called RW galaxy hereafter), which are made of point sources whose positions are generated by two-dimensional random walks, as proposed in Zhang (2008; the details are given in the next paragraph), and (2) Galsim galaxies generated by the Galsim toolkit (Rowe et al. 2015). Galsim galaxies usually have more regular shapes than RW galaxies, as shown in Figure 3.

In the RW galaxy simulation, each galaxy consists of 60 point sources generated by 60 steps within a circular region of radius \( r_i \) (9 pixels). The positions of these point sources are generated by the following algorithm:

1. The size of a single step is fixed at 1 pixel. The direction of each step is completely random.
2. The first step starts from the center of the circular region. Steps that are about to go beyond the region should restart from the center.
3. Each point source is assigned the same flux and convolved with the moffat-type PSF effect (Bridle et al. 2009). The pixel size of our simulations is set at 0′′.2, and the FWHM of the PSF for the RW galaxies is 0′′.8.

Our Galsim galaxies contain two kinds of profiles (Simard et al. 2011): the de Vaucouleurs profile and the exponential profile. For the bulge-dominated galaxies, which account for 10% of the total galaxies, we adopt only the de Vaucouleurs profile. The rest of the disk-dominated galaxies are made as a combination of these two profiles. The probability distribution of the disk scale length is derived from the fits to HSC Wide Field Planetary Camera 2 (WFPC2) observations of galaxies in the Groth strip by Simard et al. (2002), which is set as (Miller et al. 2013):

\[
P(r) \propto r \exp[-(r/a)^\alpha],
\]

where \( \alpha = 4/3 \), \( a = r_e/0.833 \), and \( r_e \) (arcsec) is decided by the \( i \)-band magnitude \( m \) via \( \ln(r_e) = -1.145 - 0.269 \times (m - 23) \). The magnitude distribution is assumed to obey \( P(m) \propto 10^{0.295m - 1.08} \), and within the range of \([20, 24]\). We limit the radius \( r \) of the galaxy in the range of \([0.1, 0.4]\) (arcsec). Two different ellipticity probability distribution functions are adopted. For bulge-dominated galaxies, we use

\[
P(e) = Ae \exp(-\alpha e - \beta e^2),
\]

with \( \alpha = 2.368 \), \( \beta = 6.691 \), and \( A = 27.8366 \). For disk-dominated galaxies, we use:

\[
P(e) = \frac{A(e[1 - \exp(\frac{e - e_{\max}}{a})]}}{(1 + e)(e^2 + e_{\max}^2)^{1/2}},
\]

with \( e_{\max} = 0.804 \), \( e_0 = 0.0256 \), \( a = 0.2539 \), and \( A = 2.4318 \), according to Miller et al. (2013).

Each Galsim galaxy is also convolved with the moffat-type PSF, with the FWHM of 0′′.6. Since the Galsim galaxies are somewhat larger than the RW galaxies before being convolved with PSF. We set their PSF size slightly smaller to prevent them from reaching the boundaries of the stamps.

To study the boundary effect, we randomly place 2500 simulated galaxy images on a 5000 × 5000 grid, with a very low background noise added. The central 4000 × 4000 region is cut out to mimic a CCD chip. Each chip is divided into many subregions of a certain size (e.g., 200 × 400, 400 × 800), separated by masks, so that each subregion can represent a geometrical boundary structure. For each boundary structure, we generate more than 30,000 such CCD chips (each of which contains about 1600 randomly distributed galaxies). Five sets of shear values are applied to sources in these CCDs: (0.02, −0.02), (0.01, −0.01), (0.00,0.00), (−0.01, 0.01), (−0.02, 0.02). We use these five pairs of shear values in the reasonable range to calibrate the shear recovery accuracy quantitatively. Each set of shear values therefore covers 6000 CCDs, containing around 10^7 galaxies. We adopt the commonly used
m and c to denote the shear bias as:

\[
\mathcal{g}_{1,2}^{\text{measured}} = (1 + m_{1,2}) \mathcal{g}_{1,2}^{\text{input}} + c_{1,2}. \tag{6}
\]

We use the Fourier_Quad pipeline for shear measurement (Zhang et al. 2015, 2016, 2019). The results for the RW galaxies are shown in Figure 4, in which we list the values of additive bias \(c_1\) as a function of the subregion sizes along the x and y directions (i.e., \(l_x\) and \(l_y\)). Due to the symmetry of the problem, we only show the results for cases with \(l_x \geq l_y\). We can see from the figure that for a fixed \(l_y\), \(c_1\) becomes larger when \(l_x\) increases, and when \(l_x\) is fixed, \(c_1\) increases with a decreasing \(l_y\). When \(l_x = l_y\), \(c_1\) is consistent with zero. These properties are all consistent with the prediction of our model in Equation (2). Note, however, that even in the case of \(l_x = l_y\), there are still local additive shear biases near the boundaries. The overall null results are just due to cancellation globally.

The parameter \(2cs\) defined in Equation (2) has a best-fit value of 408 ± 12 in the current simulation. To further check the validity of the relation in the equation, one can quantitatively check the change of the bias \(c_1\) with the ratio of the axis sizes \(l_x/l_y\) for a fixed image area \(l_x l_y\), or the other way around. The results are shown in Figure 5, which are quite consistent with Equation (2).

Another prediction of the model is that the boundary effect should be more significant for galaxies of larger average sizes. To observe this phenomenon, we generate another set of simulations with RW galaxies that are larger by 50% than those in the previous set of simulations. The results are shown in Figure 6, from which we can clearly see that \(c_1\) is larger than those reported in Figure 4 under the same condition of \((l_x, l_y)\).

We also expect the boundary effect to be more significant for more elliptical galaxies. To observe this, we repeat the simulations twice with the Galsim galaxies. Instead of using the ellipticity probability distribution function, this time, we set the galaxy ellipticities at two specific values: \(e = 0.2\) and \(e = 0.8\). The results are shown in Figures 7 and 8 accordingly. One can see that the \(c_1\) of the more elliptical cases are 10 times larger than their counterparts with smaller ellipticities.

Figure 4. Additive bias \(c_1\) for different boundary sizes \((l_x, l_y)\). All \(c_1\) values are units of \(10^{-3}\). The variances are about \(4 \times 10^{-3}\).

Figure 5. Change of the additive bias \(c_1\) as a function of \(l_x\). The upper panel is under the condition of fixed axis ratio \((l_x/l_y)\), and the lower one is for fixed total area \((l_x l_y)\). The solid curves are fittings from Equation (2) using the data shown in Figure 4. The variances of the data points are about \(4 \times 10^{-3}\).

Figure 6. Similar to Figure 4. The radii of the RW galaxies used here are larger than those used in Figure 4 by 50%. All \(c_1\) values are in units of \(10^{-3}\). The variances are about \(4 \times 10^{-3}\).

Overall, we can draw the conclusion that the additive bias due to the boundary is also influenced by the galaxy morphologies, such as their radii, ellipticities, etc. The results in this section only represent the bias levels for several special cases.

2.3. The Solution and Its Application to Real Data

As we have shown, source selection is a statistically anisotropic process near the image boundaries. The resulting shear bias cannot be treated as a universal additive bias, as it is significant only in the neighborhood of the boundaries. To avoid such a bias, a natural solution is to require all selected
sources to be at a certain distance away from the boundaries. It turns out that this condition can be easily realized by requiring that the boundaries do not intercept the image stamp. For example, in the case shown in Figure 2, we would need to remove another three stamps that are close enough to the boundary, as shown in red in Figure 9, although they are originally valid source images.

We find that this additional requirement is quite effective in removing the additive bias due to the existence of boundaries. For example, Figure 10 shows the new results of the simulations used in Figure 4 with the additional selection rule. The bias \( c_1 \) drops significantly from the level of \( 10^{-3} \) to \( 10^{-5} \).

The validity of this solution can also be tested in real data. A direct way to do so is to adopt the idea recently proposed by Zhang et al. (2019), which shows that one can use the galaxy shear estimators to restore the field-distortion signal as a way of testing the shear recovery accuracy. It can be done within the observational data itself, without requiring simulations. As in Zhang et al. (2019), here we again perform such a test with the Fourier_Quad pipeline. Our imaging data is from the third data release of Dark Energy Camera Legacy Survey (DECaLS; Dey et al. 2019). We use about 7400 calibrated z-band single exposures to do the test. Our source catalog is from Zou et al. (2019; also see Zou et al. 2017).

In Figure 11, we show the comparison between the galaxy shear and the signal induced by field distortion. The solid black line is “\( y = x \)” shown as a reference. The red square data points are the results from all valid sources confirmed by the Fourier_Quad pipeline on single exposures. In this case, one can clearly see a positive additive bias. This is due to the fact that there are many more boundary effects acting along the \( x \)-axis (defined to be the direction of R.A., which almost coincides with the direction of the longer sides of the CCD chip) than on the \( y \)-axis. The green circular points are the results achieved after adopting the new selection criterion regarding the boundary effect. This procedure removed about 5%–10% of the sources. We can see that by removing sources that are close enough to the boundaries, the additive bias in \( g_1 \) can indeed be suppressed, and the results of \( g_2 \) are hardly affected.

## 3. Conclusion and Discussion

In this paper, we have specifically studied the geometric boundary effect on the accuracy of shear measurement. The existence of boundaries due to either the CCD edges or bad columns within the images can cause locally anisotropic source detection, thereby biasing the shear measurement. This is simply because sources are more likely to be detected/selected when their orientation is parallel to the boundary. The geometric boundary effect is a selection effect related to source detection, and it causes local additive shear biases that are hard to remove globally.

We propose a model to parameterize the shear bias as a function of the side lengths of the bounded region, and find a good agreement between the model predictions and the simulation results. In our simulations, we use a large number of mock galaxies of different morphological types, including both RW galaxies and Galsim galaxies. The results of the test as well as the model prediction agree with our intuition: the boundary effects are more serious when (1) the area of the bounded region is small, (2) the ratio of the side lengths is
large, (3) the galaxies have large sizes on average, and (4) the galaxies are more elliptical. In all of the above cases, the boundaries have more impact on the selection of the galaxy sample and on the resulting shear bias.

We find that a simple way of removing such a bias is to remove the sources that are too close to the boundaries even though their isophotes of the detection threshold (e.g., 2σ–4σ above the background) do not yet touch the boundaries. This is done by requiring that no boundary marks (e.g., CCD edges, bad columns, etc.) cross the source stamp (of a predetermined fixed size). We use simulations to show that this simple procedure can indeed remove the additive shear bias related to the boundaries. We further demonstrate the existence of the boundary effect in real data, the DECaLS z-band imaging data, using the field-distortion test method proposed in Zhang et al. (2019). The results also show that our treatment can accurately remove the bias from the boundary effect.

We note that the isotropy of the source selection algorithm is also affected by the shear signal itself: two neighboring sources are prone to overlap with each other when shear is along the line connecting them (Sheldon et al. 2020). This type of effect leads to multiplicative shear biases, in contrast to the additive biases addressed in this paper. We plan to study this effect in a future work.

Since the Fourier_Quad pipeline processes shear measurement on individual exposures independently, our discussion regarding the boundary effect in this paper is quite straightforward and specific. In other shear measurement methods, e.g., model-fitting algorithms, it is often the case that multiple exposures are used together to construct the shear estimator for each galaxy. Images with missing pixels (due to boundaries for example) are often tolerated, and included in the fitting. If the images intercepted by the boundary marks are included in the fitting, their anisotropic missing-pixel distribution can still bias the shear estimators in principle. On the other hand, if such contaminated images are discarded in the fitting, the survived ones on the same exposure would tend to lie along the boundaries when they are near the boundaries, thereby causing shear bias as well. In the second case, one can consider adopting the following treatment in this paper: discarding the image according to whether the boundary marks cross its stamp, i.e., a domain of a fixed size, instead of the isophote of the detection threshold. A detailed study of this topic is, however, beyond the scope of this work.

We thank Xiaokai Chen and Dezi Liu for their help in galaxy simulation and shear measurement. The computations in this paper were run on the π 2.0 cluster supported by the Center for High Performance Computing at Shanghai Jiao Tong University. We also thank the anonymous referee for useful comments.

This work is supported by the National Key Basic Research and Development Program of China (No. 2018YFA0404504), and the NSFC grants (11673016, 11621303, 11890691, and 12073017).

**ORCID iDs**

Haoran Wang @ https://orcid.org/0000-0001-5945-8399
Jun Zhang @ https://orcid.org/0000-0003-0002-630X
Hekun Li @ https://orcid.org/0000-0002-0610-2361

**References**

Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, arXiv:0912.0201
Bernstein, G. M. 2010, MNRAS, 406, 2793
Bridle, S., Balan, S. T., Bethge, M., et al. 2010, MNRAS, 405, 2044
Bridle, S., Shawe-Taylor, J., Amara, A., et al. 2009, AnApS, 3, 6
Crittenden, R. G., Naruhrajan, P., Pen, U., & Theuns, T. 2002, ApJ, 568, 20
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Erben, T., Hildebrandt, H., Lercher, M., et al. 2009, A&A, 493, 1197
Erben, T., Hildebrandt, H., Miller, L., et al. 2013, MNRAS, 433, 2545
Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
Hikage, C., Oguri, M., Hamana, T., et al. 2019, PASJ, 71, 43
Hildebrandt, H., Viola, M., Heymans, C., et al. 2016, MNRAS, 465, 1454
Hirata, C., & Seljak, U. 2003, MNRAS, 343, 459
Hoekstra, H., & Jain, B. 2008, ARNPS, 58, 99

**Figure 11.** Test of shear recovery with the field-distortion signals using the z-band data of the third DECaLS Data Release. The left and right panels are the results of $g_1$ and $g_2$ respectively. The green and red points with 1σ error bars are the results with and without removing the boundary effect respectively.
