Optimal Coaddition of Realistic Images

Jim Bosch

May 9, 2016

1 Introduction

Coaddition is one of the most important steps in any astronomical image processing pipeline, but practical algorithms for coaddition either discard information or make assumptions that are at least approximately invalid for essentially all real data. As a result, some pipelines have adopted model-fitting approaches that fit simultaneously to data from all epochs [4, 5, 7], but even these must use coadds for source detection and typically use them for photometry as well. In addition, preliminary fitting on coadds will play an important role in making multi-epoch fitting computationally feasible for future surveys such as LSST.

The most common approaches to coaddition discard information. PSF-matched coaddition explicitly destroys information by degrading the PSF of the best images to match the worst, or by rejecting the worst-seeing images entirely to improve the quality of the common PSF. PSF-matched coaddition does lead to a well-defined and computationally effecient PSF, even when nonlinear statistics (for outlier rejection) are used to combine pixels. Direct coaddition, which simply averages images directly, underweights the best-seeing images as well. Nonlinear statistics cannot be used in direct coaddition, because the differing PSFs on input images mean that the value at a particular point on the sky is drawn from different distribution on different images. To model the PSF on a direct coadd, it is necessary to combine the PSF models of individual exposures using the same resampling and weighting applied to the coadd. This makes it virtually impossible for the model to be valid in regions where one or more images had masked pixels, and even a model that is only valid outside those region is a very complex data structure. It is also rare in both methods to fully capture the correlations between pixels introduced by image resampling or (for PSF-matched coadds) convolution with a matching kernel. In fact, fully accounting for the (spatially-varying) pixel covariance matrix is likely computationally infeasible.

In this paper, we will focus instead on coaddition methods that are either optimal (in the sense that they do not destroy information) or very close to optimal. These methods are not yet a part of any major processing pipeline, though in some cases the theory behind them has been common knowledge in astronomy for decades. In section §2, we introduce the likelihood formalism that ties together all of these methods, and show how this relates to the most

common approach to source detection. In §3, we extend an approach by [3] to more realistic conditions, yielding a fully general mathematical description for optimal coaddition. In §**??**, we discuss some possibilities for the numerical techniques and approximations that will be necessary to make this general approach computationally feasible.

2 Likelihood Coadds

2.1 Likelihood for Pixel Data

We begin with an arbitrary model of the true sky $f(\boldsymbol{r})$, parameterized on the sky position (two-element vector $\boldsymbol{r}$) and an arbitrary parameter vector $\boldsymbol{\theta}$. In the usual many-photon limit, we can approximate the noise as Gaussian, and hence the negative log likelihood of a single image I (with covariance matrix $\boldsymbol{C} = \boldsymbol{F}^{-1}$ and photometric scaling T) is

$$L \equiv -\ln P(\boldsymbol{I}|f) \tag{1}$$

$$= \frac{1}{2} \sum_{i,j} \left[I(\boldsymbol{x}_i) - T(\boldsymbol{x}_i)\, g(\boldsymbol{x}_i)\right] F(\boldsymbol{x}_i, \boldsymbol{x}_j) \left[I(\boldsymbol{x}_j) - T(\boldsymbol{x}_j)\, g(\boldsymbol{x}_j)\right] \tag{2}$$

with

$$g(\boldsymbol{x}_i) \equiv \int d^2\boldsymbol{r}\, \phi(\boldsymbol{x}_i, \boldsymbol{r})\, f(\boldsymbol{r}) \tag{3}$$

and $\phi(\boldsymbol{x}, \boldsymbol{r})$ relates a sky value at $\boldsymbol{r}$ to a discrete pixel at $\boldsymbol{x}$, encompassing the full spatially-varying point spread function (including the pixel response) as well as any astrometric transformations. That allows us to make no assumptions about the image being Nyquist sampled (or even regularly sampled). We've also made no assumptions about the noise (beyond Gaussianity) in using the full covariance matrix Σ, but we note that in practice this is always diagonal or very nearly diagonal, so computing the inverse is not a major concern.

We can write the likelihood in matrix notation (with $\boldsymbol{T}$ a diagonal matrix) as

$$L = \frac{1}{2} \left(\boldsymbol{I} - \boldsymbol{T}\boldsymbol{g}\right)^T \boldsymbol{F} \left(\boldsymbol{I} - \boldsymbol{T}\boldsymbol{g}\right) \tag{4}$$

and expand the product to

$$L = \frac{k}{2} - \boldsymbol{I}^T \boldsymbol{F} \boldsymbol{T} \boldsymbol{g} + \frac{1}{2} \boldsymbol{g}^T \boldsymbol{T} \boldsymbol{F} \boldsymbol{T} \boldsymbol{g} \tag{5}$$

with

$$k \equiv \boldsymbol{I}^T \boldsymbol{T} \boldsymbol{F} \boldsymbol{T} \boldsymbol{I} \tag{6}$$

2.2 Discretization

Because the PSF acts as a low-pass filter, we can choose some grid in sky coordinates[1] on which it is well-sampled. For such a grid with positions $\{s_i\}$, we can reconstruct the continuous PSF for a discrete output pixel exactly using sinc interpolation:

$$\phi(\boldsymbol{x}_i, \boldsymbol{r}) = \sum_i \phi(\boldsymbol{x}_j, \boldsymbol{s}_i)\, S(\boldsymbol{r} - \boldsymbol{s}_i) \tag{7}$$

where $S(\boldsymbol{r})$ is a normalized 2-d Sinc kernel. We can thus write the PSF-convolved model as

$$g(\boldsymbol{x}_i) = \sum_j \phi(\boldsymbol{s}_j, \boldsymbol{x}_i) \int d^2\boldsymbol{r}\, S(\boldsymbol{r} - \boldsymbol{s}_j)\, f(\boldsymbol{r}) \tag{8}$$

$$= \sum_j \phi(\boldsymbol{x}_i, \boldsymbol{s}_j)\, h(\boldsymbol{s}_j) \tag{9}$$

with

$$h(\boldsymbol{s}) \equiv \int d^2\boldsymbol{r}\, S(\boldsymbol{r} - \boldsymbol{s})\, f(\boldsymbol{r}) \tag{10}$$

Note that we have not assumed here that $f(\boldsymbol{r})$ is well-sampled, and we have made no approximations to the true PSF. The discrete version of $\phi(\boldsymbol{s}_i, \boldsymbol{x}_i)$ is a large rectangular matrix, mapping pixels on input images to the output pixel grid. Because the PSF kernel is compact spatially, it is sparse with a very predictable structure, but the number of nonzero elements is nevertheless large.

Substituting these definitions into the likelihood (5), we have (in matrix notation):

$$L = \frac{k}{2} - \boldsymbol{I}^T \boldsymbol{F} \boldsymbol{\phi} \boldsymbol{h} + \frac{1}{2} \boldsymbol{h}^T \boldsymbol{\phi}^T \boldsymbol{F} \boldsymbol{\phi} \boldsymbol{h} \tag{11}$$

$$= \frac{k}{2} - \boldsymbol{\Psi}^T \boldsymbol{h} + \frac{1}{2} \boldsymbol{h}^T \boldsymbol{\Phi} \boldsymbol{h} \tag{12}$$

with

$$\boldsymbol{\Psi} \equiv \boldsymbol{\phi}^T \boldsymbol{T} \boldsymbol{F} \boldsymbol{T} \boldsymbol{I} \tag{13}$$

$$\boldsymbol{\Phi} \equiv \boldsymbol{\phi}^T \boldsymbol{T} \boldsymbol{F} \boldsymbol{T} \boldsymbol{\phi} \tag{14}$$

Note that (12) has exactly the same form as (5), with the following substitutions:

$$\boldsymbol{T} \boldsymbol{F} \boldsymbol{I} \longrightarrow \boldsymbol{\Psi} \tag{15}$$

$$\boldsymbol{T} \boldsymbol{F} \boldsymbol{T} \longrightarrow \boldsymbol{\Phi} \tag{16}$$

$$\boldsymbol{g} \longrightarrow \boldsymbol{h} \tag{17}$$

---

[1] We are ignoring the curvature of the sky, as we are only concerned with small regions.

That is, we have defined a mathematically equivalent image Ψ with a compact and spatially invariant *mathematical* PSF S but a much more extended and spatially complex pixel covariance matrix Φ^{-1} that accounts for the *physical* PSF. This formulation of the likelihood has some advantages even for individual images:

- The new image Ψ is well-sampled and has no missing pixels, even if the original data is undersampled or irregularly sampled.

- Fitting algorithms can more easily use lookup tables and/or analytic integration to evaluate PSF-convolved galaxy models, since the mathematical PSF S is constant and analytic.

However, some important measurement algorithms – notably adaptive moments and aperture photometry – aren't defined with respect to a likelihood, and instead assume pixels are independent. These measurements cannot be performed on these equivalent-likelihood images directly. [6] use the term *proper image* to refer to images for which these sorts of algorithms are valid, defining it as an image with independent and constant noise. We consider this definition to be too restrictive – constant noise is not necessary for such algorithms to be valid, and as we will see in §3.2 imposing it as a requirement is probably undesirable.

However, the biggest disadvantage of this formulation is simply that Φ is enormous. For a single 4000×4000 LSST CCD with a 40×40-pixel PSF, Φ will have approximately 5.1×10^{10} nonzero elements (205 GB in single precision). Some of that information is certainly redundant, and a clever compression scheme may be able to reduce the size considerably.

We will ignore these computational challenges for now and return to them later in the paper.

2.3 Coaddition

So far, we've been dealing with a single image, but it is now trivially easy to extend this formalism to multiple images. The likelihood of our model f given multiple images $\boldsymbol{I}_i$ is just the product of the per-image likelihoods:

$$P(\{\boldsymbol{I}\}|f) = \prod_n P(\boldsymbol{I}_n|f) \tag{18}$$

which is equivalent to summing the (negative) log likelihoods:

$$L = \sum_n L_n = \sum_n \left[\frac{k_n}{2} - \boldsymbol{\Psi}_n^T \boldsymbol{h} + \frac{1}{2} \boldsymbol{h}^T \boldsymbol{\Phi}_n \boldsymbol{h} \right] \tag{19}$$

Because the grid on which we evaluate $\boldsymbol{h}$ is in sky coordinates, we can choose it to be the same for every input image (the only requirement is that it be

fine enough to adequately sample the most compact input PSF). This makes L linear in $\boldsymbol{\Psi}_n$ and $\boldsymbol{\Phi}_n$, so we can just sum these:

$$L = \frac{k_c}{2} - \boldsymbol{\Psi}_c^T \boldsymbol{h} + \frac{1}{2} \boldsymbol{h}^T \boldsymbol{\Phi}_c \boldsymbol{h} \tag{20}$$

$$k_c \equiv \sum_n k_n \tag{21}$$

$$\boldsymbol{\Psi}_c \equiv \sum_n \boldsymbol{\Psi}_n \tag{22}$$

$$\boldsymbol{\Phi}_c \equiv \sum_n \boldsymbol{\Phi}_n \tag{23}$$

This is clearly the same as (12), with the coadded k_c, $\boldsymbol{\Psi}_c$ and $\boldsymbol{\Phi}_c$ replacing the single-epoch ones. This coadd is a "sufficient statistic" for the full dataset (it allows us to exactly compute the likelihood for any model), with only two assumptions.

- The true sky must be the same at all epochs (no variability or motion).

- We must be able to choose a common grid on which all input PSFs are well-sampled.

Any kind of coadd is inappropriate when the former is not a reasonable approximation, and the second condition is always met by real data.

We refer to $\boldsymbol{\Psi}_c$ as a *likelihood coadd*, and $\{\boldsymbol{\Psi}_i\}$ as *likelihood images*, as these are the only images for which summing them exactly combines the information from the per-epoch likelihoods. If the PSF and noise are constant, a pixel value in one of these images is proportional to the likelihood of an isolated point source being present at the center of that pixel (see §2.4), but these conditions are only approximately met in real data.

Likelihood coadds suffer from the same computational advantages and disadvantages as the single-frame likelihoods. While the coadded quantities $\boldsymbol{\Psi}_c, \boldsymbol{\Phi}_c$ do represent a compression of the single epoch quantities $\{\boldsymbol{\Psi}_i, \boldsymbol{\Phi}_i\}$, this isn't nearly enough to make up for the increase in data size due to the transformation (16) from a diagonal covariance matrix to an extended, complex, fully-pixelized covariance matrix.

2.4 Source Detection

One area where likelihood images are practical (and in frequent use) is source detection. The traditional approach to source detection assumes a single point source model, with any background already subtracted from the image:

$$f(\boldsymbol{r}, \boldsymbol{c}, \alpha) = \alpha \, \delta(\boldsymbol{\mu} - \boldsymbol{r}) \tag{24}$$

5

where α is the flux and $\boldsymbol{\mu}$ is the position. The negative log likelihood (either on a single image or a coadd) then reduces to

$$L(\boldsymbol{c}, \alpha) = \frac{k}{2} - \alpha \sum_i \Psi(\boldsymbol{s}_i) \, S(\boldsymbol{s}_i - \boldsymbol{\mu}) + \frac{\alpha^2}{2} \sum_{i,j} \Phi(\boldsymbol{s}_i, \boldsymbol{s}_j) \, S(\boldsymbol{s}_i - \boldsymbol{\mu}) \, S(\boldsymbol{r}_j - \boldsymbol{\mu})$$

(25)

The sinc kernels S are orthogonal on the grid $\boldsymbol{s}_i$, which simplifies the above to

$$L(\boldsymbol{c}, \alpha) = \frac{k}{2} - \alpha \sum_i \Psi(\boldsymbol{s}_i) S(\boldsymbol{s}_i - \boldsymbol{\mu}) + \frac{\alpha^2}{2} \sum_i \Phi(\boldsymbol{s}_i, \boldsymbol{s}_i) \, [S(\boldsymbol{s}_i - \boldsymbol{\mu})]^2 \qquad (26)$$

This is the critical simplification: for all subsequent steps in this section, we only require the diagonal of $\boldsymbol{\Phi}$, which *is* computationally feasible to evaluate and store.

To detect sources, we'd like to simply evaluate this likelihood for $\boldsymbol{\mu} = \boldsymbol{s}_i$ – that is, consider the hypotheses that there is a point source at each of the positions where we've already evaluated $\boldsymbol{\Psi}$.[2] Because $S(0) = 1$, we then have

$$L(\boldsymbol{s}_i, \alpha) = \frac{k}{2} - \alpha \Psi(\boldsymbol{s}_i) + \frac{\alpha^2}{2} \Phi(\boldsymbol{s}_i, \boldsymbol{s}_i) \qquad (27)$$

At each pixel position, we can solve analytically for the maximum likelihood flux of a point source centered on that pixel, by setting the first derivative of L to zero and solving for α:

$$\frac{\partial L}{\partial \alpha} = -\Psi(\boldsymbol{s}_i) + \alpha \Phi(\boldsymbol{s}_i, \boldsymbol{s}_i) = 0 \qquad (28)$$

$$\bar{\alpha}(\boldsymbol{s}_i) = \frac{\Psi(\boldsymbol{s}_i)}{\Phi(\boldsymbol{s}_i, \boldsymbol{s}_i)} \qquad (29)$$

The uncertainty on the maximum likelihood flux (given the centroid) is given by

$$\sigma_\alpha(\boldsymbol{s}_i) = \left(\frac{\partial^2 L}{\partial \alpha^2} \right)^{-1/2} = \frac{1}{\sqrt{\Phi(\boldsymbol{s}_i, \boldsymbol{s}_i)}} \qquad (30)$$

and hence the signal-to-noise ratio of the flux as a function of position is

$$\nu(\boldsymbol{s}_i) = \frac{\bar{\alpha}(\boldsymbol{s}_i)}{\sigma_\alpha(\boldsymbol{s}_i)} = \frac{\Psi(\boldsymbol{s}_i)}{\sqrt{\Phi(\boldsymbol{s}_i, \boldsymbol{s}_i)}} \qquad (31)$$

We detect by appling a threshold to a *detection map* - an image of ν. In common parlance, objects with $\nu > 5$ are "5 sigma" detections. For a given significance, a single above-threshold pixel is sufficient, in contrast to the heuristic requirement for multiple pixels in *e.g.* SExtractor [1].

---

[2]With most PSFs, this will make us slightly less sensitive to point sources centered immediately between pixels, but we can account for this by just decreasing the threshold slightly.

Maximum likelihood detections are biased, however, and the significance should not be interpreted as a formal statement about the probability that the detection is real. In all realistic scenarios the prior cannot be ignored, and it is the posterior probability that must be used when calculating false detection rates. A simple unnormalized power law prior as proposed by [2] does not work. For now, we consider this problem beyond the scope of this paper.

With or without a prior, however, the fact that isolated point source detection utilizes only the diagonal of $\boldsymbol{\Phi}$ holds, making this approach to detection practical on both individual images and on coadds, even in the presence of noise or on undersampled images. Extending this approach to resolved sources or multiple sources exactly requires at least some off-diagonal elements of $\boldsymbol{\Phi}$, in addition to solving a higher-dimensional optimization problem. In practice, it is more common to use the algorithm for isolated point sources as a heuristic for detecting superpositions of sources, by identifying peaks in ν within an above-threshold region as candidates for deblending and more careful signal-to-noise evaluation. Similarly, simply binning or smoothing the ν image may result in an adequate approximation for the detection of faint extended sources.

3 Decorrelated Coadds

3.1 Factorization

While it is necessary to transform from regular images to likelihood images in order to build an optimal coadd, once a likelihood coadd has been built we can in principle do the reverse transformation to return to a "regular" coadd: one with a non-trivial PSF and a diagonal covariance matrix. This operation is known as decorrelation, and it's a matter of finding a symmetric decomposition of $\boldsymbol{\Phi}_c$:

$$\boldsymbol{\Phi}_c = \boldsymbol{A}^T \boldsymbol{D} \boldsymbol{A} \tag{32}$$

such that $\boldsymbol{A}$ is invertible, $\boldsymbol{D}$ is diagonal, and $\boldsymbol{A}$ is normalized such that

$$\sum_j A_{i,j} = 1 \tag{33}$$

for all i. This allows us to rewrite the likelihood (12) as

$$L = \frac{k}{2} - \boldsymbol{\Psi}_c^T \boldsymbol{A}^{-1} \boldsymbol{A} \boldsymbol{h} + \frac{1}{2} \boldsymbol{h}^T \boldsymbol{A}^T \boldsymbol{D} \boldsymbol{A} \boldsymbol{h} \tag{34}$$

$$= \frac{k}{2} - \boldsymbol{\Psi}_c^T \boldsymbol{A}^{-1} \boldsymbol{D}^{-1} \boldsymbol{D} \boldsymbol{A} \boldsymbol{h} + \frac{1}{2} \left(\boldsymbol{A}\boldsymbol{h}\right)^T \boldsymbol{D} \left(\boldsymbol{A}\boldsymbol{h}\right) \tag{35}$$

$$= \frac{k}{2} - \boldsymbol{I}_c^T \boldsymbol{F}_c \boldsymbol{g}_c + \frac{1}{2} \boldsymbol{g}_c^T \boldsymbol{F}_c \boldsymbol{g}_c \tag{36}$$

with

$$I_c \equiv D^{-1} A^{-T} \Psi_c \tag{37}$$

$$g_c \equiv Ah \tag{38}$$

$$F_c \equiv D \tag{39}$$

These decorrelated coadds have independent pixels, making them suitable for adaptive moments, apertures and other common source measurement algorithms that aren't derived from a likelihood. They're also much more visually similar to typical astronomical images. But they don't solve the storage problem associated with likelihood coadds: while the covariance matrix F_c is now diagonal and hence easy to store, the complexity has been transferred to the discretized PSF A.[3] Unlike the per-epoch PSFs ϕ, A is not slowly varying on large scales (due to discontinuities where the set of input images to the coadd changes), and hence it requires essentially the same storage as the likelihood covariance matrix Φ if we want to preserve all information.

Moreover, computing the factorization (32) is a hard problem – even for a relatively small patch of sky, this is a factorization of a large sparse matrix. It is also important to note that because the matrix A is the discretized PSF of the decorrelated coadd image, its inverse is a deconvolution. This is a hint that A may intrinsically be singular – or, perhaps, that solutions with nonsingular A may not be useful, as they may imply a non-local effective PSF. We do not need to compute the inverse of A directly, however; we just need to find a solution I_c to the equation

$$\Psi_c = A^T D I_c \tag{40}$$

which we can achieve by first solving for b in

$$\Psi_c = A^T b \tag{41}$$

and then setting

$$I_c = D^{-1} b \tag{42}$$

What (41) shows is that we don't really need to compute the deconvolution kernel – instead, we need to find an image b that is equal to the likelihood coadd Ψ_c when convolved by the transpose of the decorrelated coadd PSF A^T. While this is a deconvolution, it is an entirely reasonable one, because we know the likelihood coadd has effectively been convolved twice by this PSF already: once by the actual observational system, and again when we transformed from regular images to likelihood images. No noise has been added since the second convolution, so it should be entirely reversible.

---

[3]Recall that the full PSF also includes a continuous convolution with $S(r)$.

3.2 Noise Whitening

Instead of merely decorrelating the noise, we could use the same factorization (32) to go one step further and whiten the noise:

$$L = \frac{k}{2} - \boldsymbol{\Psi}_c^T \boldsymbol{A}^{-1} \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{A} \boldsymbol{h} + \frac{1}{2} \left(\boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A} \boldsymbol{h} \right)^T \left(\boldsymbol{D}^{\frac{1}{2}} \boldsymbol{A} \boldsymbol{h} \right) \tag{43}$$

$$= \frac{k}{2} - \boldsymbol{I}_w^T \boldsymbol{F}_w \boldsymbol{g}_w + \frac{1}{2} \boldsymbol{g}_w^T \boldsymbol{F}_w \boldsymbol{g}_w \tag{44}$$

and

$$\boldsymbol{I}_w \equiv \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A}^{-T} \boldsymbol{\Psi}_c \tag{45}$$

$$\boldsymbol{g}_w \equiv \boldsymbol{D}^{\frac{1}{2}} \boldsymbol{A} \boldsymbol{h} \tag{46}$$

$$\boldsymbol{F}_w \equiv \boldsymbol{\mathcal{I}} \tag{47}$$

This is no more difficult computationally, and it generates a *proper* image according to the definition of [6]. But it actually makes our coadd behave less like a typical single-epoch image, where we expect the variance in each pixel to be the approximately the square root of the pixel value. Whitening the noise forces us to track these per-pixel noise variations as part of the discretized PSF $\boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{A}$, which as a result is no longer normalized to unit sum.

3.3 Kaiser Coadds

The first development of a generally-useful (and at least theoretically practical) optimal coaddition algorithm is due to Kaiser [3], and more recently revisited by [6]. This approach is completely general with respect to the model of the sky, but it relies on several assumptions about the data:

- The input images must be Nyquist-sampled with no missing pixels (to allow discrete Fourier transforms and resampling of the image).

- The variance in each input image must be constant and uncorrelated.

- The PSF of each input image must be constant.

- The photometric scaling T must be constant for each image.

At least for ground-based optical data, these assumptions are at least approximately correct in the neighborhood of any particular source, but they are manifestly incorrect on the scale of a full sensor. However, the method also implicitly assumes that all input images are defined on the same pixel grid. While images that meet the above requirements can be resampled to the required coordinate system exactly (using sinc interpolation) or close enough to it (using e.g. Lanczos as an approximation to sinc), this resampling will in general introduce correlations in the noise and produce a spatially-varying PSF.

We will ignore these complications for now, however, which lets us rewrite (13) and (14) in the continuous limit as

$$\Psi(\boldsymbol{r}) = \frac{T}{\sigma^2} \int d^2\boldsymbol{x}\, \phi(\boldsymbol{r} - \boldsymbol{x})\, I(\boldsymbol{x}) \tag{48}$$

$$\Phi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{T^2}{\sigma^2} \int d^2\boldsymbol{x}\, \phi(\boldsymbol{r}_1 - \boldsymbol{x})\, \phi(\boldsymbol{r}_2 - \boldsymbol{x}) \tag{49}$$

where we have redefined the constant point function $\phi(\boldsymbol{x})$ – it is now a function of only one variable – and defined the constant per-pixel variance as σ^2. Note that the output coordinate system $\boldsymbol{r}$ is the same as the input coordinate system $\boldsymbol{x}$.

We can now Fourier transform both quantities (we must transform both arguments of Φ). Our treatment of the image data as continuous is justified by our assumption that the data and PSF are well-sampled, and this also allows us to use discrete Fourier transforms instead of continuous transforms in practice with appropriate padding to account for periodicity. We will continue to use continuous notation here for simplicity. The Fourier transforms are

$$\tilde{\Psi}(\boldsymbol{k}) = \frac{T}{\sigma^2} \tilde{\phi}(\boldsymbol{k})\, \tilde{I}(\boldsymbol{k}) \tag{50}$$

$$\tilde{\Phi}(\boldsymbol{k}_1, \boldsymbol{k}_2) = \frac{T^2}{\sigma^2} \delta(\boldsymbol{k}_1 + \boldsymbol{k}_2)\, \tilde{\phi}(\boldsymbol{k}_1)\, \tilde{\phi}(\boldsymbol{k}_2) \tag{51}$$

where $\delta(\boldsymbol{k})$ is the Dirac delta function, indicating that $\tilde{\Phi}$ is really only a one-parameter quantity, since it is nonzero only for $\boldsymbol{k}_1 = -\boldsymbol{k}_2$. Abusing notation, we will proceed to write it simply as

$$\tilde{\Phi}(\boldsymbol{k}) = \frac{T^2}{\sigma^2} \tilde{\phi}(-\boldsymbol{k})\, \tilde{\phi}(\boldsymbol{k}) = \frac{T^2}{\sigma^2} \tilde{\phi}^*(\boldsymbol{k}) \tilde{\phi}(\boldsymbol{k}) \tag{52}$$

where $\tilde{\phi}^*(\boldsymbol{k}) = \tilde{\phi}(-\boldsymbol{k})$ because $\phi(\boldsymbol{x})$ is real.

Because the Fourier transform is a linear operation, we can choose to sum $\tilde{\Psi}$ and $\tilde{\Phi}$ to coadd them, yielding the Fourier transform of the likelihood coadd:

$$\tilde{\Psi}_c(\boldsymbol{k}) = \sum_n \tilde{\Psi}_n(\boldsymbol{k}) = \sum_n \frac{T_n}{\sigma_n^2} \tilde{\phi}_n(\boldsymbol{k})\, \tilde{I}_n(\boldsymbol{k}) \tag{53}$$

$$\tilde{\Phi}_c(\boldsymbol{k}) = \sum_n \tilde{\Phi}_n(\boldsymbol{k}) = \sum_n \frac{T_n^2}{\sigma_n^2} \tilde{\phi}_n^*(\boldsymbol{k})\, \tilde{\phi}_n(\boldsymbol{k}) \tag{54}$$

In this formulation, it is trivial to decorrelate the noise – the variance in Fourier mode $\boldsymbol{k}$ is just $\tilde{\Phi}(\boldsymbol{k})$, and hence we can just divide the Fourier-space image by the square root of $\tilde{\Phi}$:

$$\tilde{I}_c(\boldsymbol{k}) = \frac{\tilde{\Psi}_c(\boldsymbol{k})}{\sqrt{\tilde{\Phi}_c(\boldsymbol{k})}} = \frac{\sum_n \frac{T_n}{\sigma_n^2} \tilde{\phi}_n(\boldsymbol{k})\, \tilde{I}_n(\boldsymbol{k})}{\sqrt{\sum_n \frac{T_n^2}{\sigma_n^2} \tilde{\phi}_n^*(\boldsymbol{k})\, \tilde{\phi}_n(\boldsymbol{k})}} \tag{55}$$

This appears to whitens the noise in addition to decorrelating it, without the loss of PSF normalization discussed in §3.2 – but that's just because we've already explicitly ignored per-pixel variance differences in the assumptions we made at the beginning of this section. The Fourier transform of the (normalized) PSF is then

$$\tilde{A}(\boldsymbol{k}) = \frac{\sqrt{\Phi_c \tilde{(}\boldsymbol{k})}}{\displaystyle\int d^2\boldsymbol{k'} \sqrt{\Phi_c \tilde{(}\boldsymbol{k'})}} = \sqrt{\frac{\displaystyle\sum_n \frac{T_n^2}{\sigma_n^2} \tilde{\phi}_n^*(\boldsymbol{k})\,\tilde{\phi}_n(\boldsymbol{k})}{\displaystyle\sum_n \frac{T_n^2}{\sigma_n^2}}} \tag{56}$$

While the assumptions in the Kaiser method may amount to acceptable approximations for most regions of the sky, they clearly fall apart in boundary regions where the set of input images changes discontinuously. As the number of input images increases, the fraction of the sky in such a regions increases, making this approach increasingly unsuitable as depth increases.

One way to avoid boundary problems is to build per-object coadds, in which only exposures that fully overlap an object are included. The challenge here is identifying the full extent of an object, which would require (at least) some other coadd to be built first. A detection map – as described in §2.4, a likelihood coadd in which we only need the diagonal of $\boldsymbol{\Phi}$ – could perhaps fill this role. Once that region is defined, we can build a Kaiser coadd by interpolating missing pixels (or rejecting images with them entirely) and ignoring PSF spatial variation and nonuniform noise. This will result in a slightly less optimal image – both because some images are not contributing, and because areas with nonuniform noise are not being weighted optimally. Assuming the PSF is constant over the area of a single object is standard practice even on individual epochs, and hence it shouldn't be regarded as a limitation of per-object Kaiser coadds.

4 Implementation Issues

4.1 PSF Simplification

As we have discussed, there are two major limitations of the general decorrelated coadd method described in §3.1:

- The factorization (32) is a large, expensive linear algebra problem.

- The PSF is in general different for each pixel, and is hence too large and complex to store and use effectively.

It generally makes sense to solve the second problem first: we start by specifying a form for the PSF that reduces its degrees of freedom and size to something more manageable, and that simplifies the factorization problem.

A simpler PSF does not necessarily imply approximation or loss of information if we relax the requirement that $\boldsymbol{D}$ in (32) be diagonal. In the common case where most of the coadd area is in regions where the set of input images is

constant, it makes sense to use a piecewise smooth parametrization of the PSF; this could allow $\boldsymbol{D}$ to be diagonal within those constant-input areas, while allowing it to have nonzero off-diagonal elements in boundary regions to maintain the exact factorization.

While a piecewise smooth PSF is probably best for minimizing the overall number of off-diagonal elements of $\boldsymbol{D}$, in practice we may prefer one that is is at least somewhat smooth even over boundary regions, to allow us to proceed with the approximation that the PSF is constant over the scale of individual objects. As long as the total number of elements of $\boldsymbol{D}$ is not too large, having an extended covariance matrix for some objects is not a serious problem.

We propose a model with the following form:

$$A(\boldsymbol{r}_1, \boldsymbol{r}_2) = \sum_i^M \sum_j^N \alpha_{i,j} \, B_i(\boldsymbol{r}_2 - \boldsymbol{r}_1) \, R_j(\boldsymbol{r}_1) \tag{57}$$

This is still an extremely general parametrization: B_i are basis functions for the PSF kernel at a point, R_j are basis functions for the spatial variation, and $\alpha_{i,j}$ the coefficients that combine them. Given that $\boldsymbol{A}$ is a matrix, not a continuous function, we assume above that $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are evaluated only on a discrete grid.

We will leave the details of these basis functions for future work, with the following notes:

- For the spatial basis functions $\{R_j\}$, the only real requirement is that they put an upper limit on the spatial variation in the PSF – as we have discussed, we want to be able to approximate the PSF as slowly varying over the scale of an object. As a result, Fourier or spline bases are probably good choices.

We can generate a useful kernel basis function set $\{B_i\}$ by building *Kaiser coadd* PSF models using (56) at a selection of discrete locations on the coadd image – in small regions without boundaries or bright objects, we expect the Kaiser coadd PSF to be the same as the decorrelated coadd PSFs. Ideally, these points would include at least a few points with every region with a different set of input images, but otherwise could be chosen randomly or systematically. It may make sense to then reduce the dimensionality of the basis using PCA or something similar.

4.2   Preprocessing

References

[1] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. A&AS, 117:393–404, June 1996. doi: 10.1051/aas:1996164.

[2] D. W. Hogg and E. L. Turner. A Maximum Likelihood Method to Improve Faint-Source Flux and Color Estimates. PASP, 110:727–731, June 1998. doi: 10.1086/316173.

[3] N. Kaiser. Addition of Images with Varying Seeing. Technical Report PSDC-002-011-xx, Institute for Astronomy, University of Hawaii at Manoa, Honolulu, Hawaii, July 2001.

[4] L. Miller, C. Heymans, T. D. Kitching, L. van Waerbeke, T. Erben, H. Hildebrandt, H. Hoekstra, Y. Mellier, B. T. P. Rowe, J. Coupon, J. P. Dietrich, L. Fu, J. Harnois-Déraps, M. J. Hudson, M. Kilbinger, K. Kuijken, T. Schrabback, E. Semboloni, S. Vafaei, and M. Velander. Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey. MNRAS, 429:2858–2880, March 2013. doi: 10.1093/mnras/sts454.

[5] E. S. Sheldon. An implementation of Bayesian lensing shear measurement. MNRAS, 444:L25–L29, October 2014. doi: 10.1093/mnrasl/slu104.

[6] B. Zackay and E. O. Ofek. How to coadd images? II. A coaddition image that is optimal for any purpose in the background dominated noise limit. *ArXiv e-prints*, December 2015.

[7] J. Zuntz, T. Kacprzak, L. Voigt, M. Hirsch, B. Rowe, and S. Bridle. IM3SHAPE: Maximum likelihood galaxy shear measurement code for cosmic gravitational lensing. Astrophysics Source Code Library, September 2014.