

# Optimal Coaddition of Realistic Images

Jim Bosch

February 15, 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 efficient 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 extend a common method for single-frame

source detection to a fully-general and exact (but computationally impractical) coaddition scheme. In section sec:kaiser-coadds, we discuss a variant originally developed by [3] and recently tested by [6] that solves the computational challenge for (unrealistically) simple images. In section sec:decorrelated-coadds, we generalize Kaiser’s approach to realistic images, resulting in an algorithm that may be computationally feasible (but still extremely challenging) with suitably clever numerical techniques.

## 2 Detection Maps

### 2.1 Likelihood for Pixel Data

We begin with an arbitrary model of the true sky  $f(\mathbf{r}, \boldsymbol{\theta})$ , parameterized on the sky position (two-element vector  $\mathbf{r}$ ) 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  $\mathbf{I}$  is

$$L(\boldsymbol{\theta}) \equiv -\ln P(\mathbf{I}|\boldsymbol{\theta}) \quad (1)$$

$$= \frac{1}{2} [\mathbf{I} - \mathbf{g}(\boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{I} - \mathbf{g}(\boldsymbol{\theta})] \quad (2)$$

with

$$g_i(\boldsymbol{\theta}) \equiv \int d^2 \phi(\mathbf{x}_i, \mathbf{r}) f(\mathbf{r}, \boldsymbol{\theta}) \quad (3)$$

and  $\phi(\mathbf{x}, \mathbf{r})$  relates a sky value at  $\mathbf{r}$  to discrete pixel  $\mathbf{x}_i$ , encompassing the full spatially-varying point spread function (including the pixel response) as well as any photometric and astrometric transformations. That means  $\phi$  is not normalized to one, but we’ve also made 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  $\boldsymbol{\Sigma}$ , but we note that in practice this is always diagonal or very nearly diagonal, so computing the inverse is not a major concern.

By expanding this product, we can focus on the terms that are independent of the model  $f$ :

$$L(\boldsymbol{\theta}) = \frac{1}{2} \mathbf{I}^T \boldsymbol{\Sigma}^{-1} \mathbf{I} - \mathbf{I}^T \boldsymbol{\Sigma}^{-1} \mathbf{g}(\boldsymbol{\theta}) + \frac{1}{2} [\mathbf{g}(\boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{g}(\boldsymbol{\theta})] \quad (4)$$

$$= \frac{k}{2} - \int d^2 \mathbf{r} f(\mathbf{r}, \boldsymbol{\theta}) \Psi(\mathbf{r}) + \frac{1}{2} \int d^2 \mathbf{r}_1 \int d^2 \mathbf{r}_2 f(\mathbf{r}_1, \boldsymbol{\theta}) f(\mathbf{r}_2, \boldsymbol{\theta}) \Phi(\mathbf{r}_1, \mathbf{r}_2) \quad (5)$$

with

$$k \equiv \sum_{i,j} I_i I_j [\Sigma^{-1}]_{i,j} \quad (6)$$

$$\Psi(\mathbf{r}) \equiv \sum_{i,j} \phi(\mathbf{x}_i, \mathbf{r}) I_j [\Sigma^{-1}]_{i,j} \quad (7)$$

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) \equiv \sum_{i,j} \phi(\mathbf{x}_i, \mathbf{r}_1) \phi(\mathbf{x}_j, \mathbf{r}_2) [\Sigma^{-1}]_{i,j} \quad (8)$$

## 2.2 Discretization

As we have defined them thus far,  $\Phi$  and  $\Psi$  are continuous functions, but we can straightforwardly sample them on a discrete regular grid in  $\mathbf{r}$ :

$$\hat{\Psi}_i \equiv \Psi(\mathbf{r}_i) \quad (9)$$

$$\hat{\Phi}_{i,j} \equiv \Phi(\mathbf{r}_i, \mathbf{r}_j) \quad (10)$$

$$(11)$$

It is always possible to construct a grid  $\mathbf{r}$  such that  $\Psi$  and  $\Phi$  are Nyquist-sampled, because the PSFs of the input images ensure that both quantities have a maximum spatial frequency. This allows us to reconstruct the continuous functions exactly using sinc interpolation:

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = \sum_{i,j} \hat{\Phi}_{i,j} S(\mathbf{r}_i - \mathbf{r}_1) S(\mathbf{r}_j - \mathbf{r}_2) \quad (12)$$

$$\Psi(x) = \sum_i \hat{\Psi}_i S(\mathbf{r}_i - \mathbf{r}_1) \quad (13)$$

where  $S(\mathbf{r})$  is a normalized 2-d Sinc kernel. We can now rewrite the full negative log-likelihood as

$$L(\boldsymbol{\theta}) = \frac{k}{2} - \sum_i \hat{\Psi}_i h_i(\boldsymbol{\theta}) + \frac{1}{2} \sum_{i,j} \hat{\Phi}_{i,j} h_i(\boldsymbol{\theta}) h_j(\boldsymbol{\theta}) \quad (14)$$

$$= \frac{k}{2} - \hat{\Psi}^T [\mathbf{h}(\boldsymbol{\theta})] + \frac{1}{2} [\mathbf{h}(\boldsymbol{\theta})]^T \hat{\Phi} [\mathbf{h}(\boldsymbol{\theta})] \quad (15)$$

$$(16)$$

with

$$h_i(\boldsymbol{\theta}) = \int d^2\mathbf{r} S(\mathbf{r}_i - \mathbf{r}) f(\mathbf{r}, \boldsymbol{\theta}) \quad (17)$$

This equation has exactly the same form as (5), with the following substitutions:

$$\mathbf{I} \longrightarrow \hat{\Psi} \quad (18)$$

$$\Sigma^{-1} \longrightarrow \hat{\Phi} \quad (19)$$

$$\phi \longrightarrow S \quad (20)$$

That is, we have defined a statistically equivalent image  $\Psi$  with a compact and spatially invariant PSF  $S$  but a much more extended and spatially complex pixel covariance matrix  $\Phi$ .

### 2.3 Coaddition

So far, we've been dealing with a single image, but it is now trivially easy to extend to multiple images. Because  $L$  is linear in  $\hat{\Phi}$  and  $\hat{\Psi}$ , and our grid is chosen to be in sky coordinates, we can simply sum the  $k$ ,  $\hat{\Psi}$ , and  $\hat{\Phi}$  quantities from each image, and then reinterpret (16) as the negative log likelihood of the full ensemble of images. This relationship is exact, which means  $k$ ,  $\hat{\Psi}$ , and  $\hat{\Phi}$  are a sufficient statistic, and no information is lost.

Unfortunately, this approach is likely computationally impractical in general. The simplest reason is that the covariance matrix  $\Phi$  is enormous. For a typical  $4000 \times 4000$  coadd image comprised of images with  $40 \times 40$ -pixel PSFs,  $\Phi$  will have approximately  $5.1 \times 10^{10}$  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. But even in this case, the extent of the covariance matrix would make it more difficult to use the likelihood in this form, mostly because common algorithms such as aperture photometry and weighted-moment shapes implicitly assume independent pixels. The computational tradeoffs for forward-modeling algorithms is less clear; the number of pixels on which a model must be evaluated will increase, even if the PSF is approximated by a Lanczos kernel instead of a full sinc function. But the fact that the PSF is the same everywhere may enable fast lookup-table methods for model evaluation.

### 2.4 Source Detection

One area where  $\Psi$  and  $\Phi$  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:

$$f(\mathbf{r}, \mathbf{c}, \alpha) = \alpha \delta(\mathbf{c} - \mathbf{r}) \quad (21)$$

where  $\alpha$  is the flux and  $\mathbf{c}$  is the position. The negative log likelihood then reduces to

$$L(\mathbf{c}, \alpha) = \frac{k}{2} - \alpha \sum_i \hat{\Psi}_i S(\mathbf{r}_i - \mathbf{c}) + \frac{\alpha^2}{2} \sum_{i,j} \hat{\Phi}_{i,j} S(\mathbf{r}_i - \mathbf{c}) S(\mathbf{r}_j - \mathbf{c}) \quad (22)$$

The sinc kernels  $S$  are orthogonal, which simplifies the above to

$$L(\mathbf{c}, \alpha) = \frac{k}{2} - \alpha \sum_i \hat{\Psi}_i S(\mathbf{r}_i - \mathbf{c}) + \frac{\alpha^2}{2} \sum_i \hat{\Phi}_{i,i} [S(\mathbf{r}_i - \mathbf{c})]^2 \quad (23)$$

To actually detect sources, however, we'd like to simply evaluate this likelihood for  $\mathbf{c} = \mathbf{r}_i$  - that is, consider the hypotheses that there is a point source at each

of the positions where we've already evaluated  $\hat{\Psi}$ .<sup>1</sup> Because  $S(0) = 1$ , we then have

$$L(\mathbf{r}_i, \alpha) = \frac{k}{2} - \alpha \hat{\Psi}_i + \frac{\alpha^2}{2} \hat{\Phi}_{i,i} \quad (24)$$

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  $\alpha$ :

$$\frac{\partial L}{\partial \alpha} = -\hat{\Psi}_i + \alpha \hat{\Phi}_{i,i} = 0 \quad (25)$$

$$\bar{\alpha}(\mathbf{r}_i) = \frac{\hat{\Psi}_i}{\hat{\Phi}_{i,i}} \quad (26)$$

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

$$\sigma_\alpha(\mathbf{r}_i) = \left( \frac{\partial^2 L}{\partial \alpha^2} \right)^{-1/2} = \frac{1}{\sqrt{\hat{\Phi}_{i,i}}} \quad (27)$$

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

$$\nu(\mathbf{r}_i) = \frac{\bar{\alpha}(\mathbf{r}_i)}{\sigma_\alpha(\mathbf{r}_i)} = \frac{\hat{\Psi}_i}{\sqrt{\hat{\Phi}_{i,i}}} \quad (28)$$

We detect by applying a threshold to an image of this significance  $\nu$ . In common parlance, objects with  $\nu > 5$  are “ $\sigma$ ” 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].

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, it is important to note that the likelihood for isolated point sources utilizes only the diagonal of  $\hat{\Phi}$ , making this approach to detection practical on both individual images and on coadds. Extending this approach to resolved sources or multiple sources exactly requires at least some off-diagonal elements of  $\hat{\Phi}$ , addition to solving a higher-dimensional optimization problem. In practice, it is more common to use the algorithm for isolated point sources to detect superpositions of sources, by identifying peaks in  $\nu$  within an above-threshold region as candidates for deblending. Similarly, simply binning or smoothing the  $\nu$  image may result in an adequate approximation for the detection of faint extended sources.

<sup>1</sup>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.

### 3 Kaiser Coadds

### 4 General Decorrelated Coadds

## 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.