

Optimal Coaddition of Realistic Images

Jim Bosch

February 17, 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 Likelihood Coadds

2.1 Likelihood for Pixel Data

We begin with an arbitrary model of the true sky $f(\mathbf{r})$, 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 I (with covariance matrix $\mathbf{C} = \mathbf{F}^{-1}$) is

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

$$= \frac{1}{2} \sum_{i,j} [I(\mathbf{x}_i) - g(\mathbf{x}_i)] F(\mathbf{x}_i, \mathbf{x}_j) [I(\mathbf{x}_j) - g(\mathbf{x}_j)] \quad (2)$$

with

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

and $\phi(\mathbf{x}, \mathbf{r})$ relates a sky value at \mathbf{r} to a discrete pixel at \mathbf{x} , encompassing the full spatially-varying point spread function (including the pixel response) as well as any photometric and astrometric transformations. That means ϕ 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 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 as

$$L = \frac{1}{2} (\mathbf{I} - \mathbf{g})^T \mathbf{F} (\mathbf{I} - \mathbf{g}) \quad (4)$$

and expand the product to

$$L = \frac{k}{2} - \mathbf{I}^T \mathbf{F} \mathbf{g} + \frac{1}{2} \mathbf{g}^T \mathbf{F} \mathbf{g} \quad (5)$$

with

$$k \equiv \mathbf{I}^T \mathbf{F} \mathbf{I} \quad (6)$$

Because the PSF acts as a low-pass filter, we can choose some grid in sky coordinates¹ on which it is well-sampled. For such a grid \mathbf{s}_i , we can reconstruct the continuous PSF for a discrete output pixel exactly using sinc interpolation:

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

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

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

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

with

$$h(\mathbf{s}) \equiv \int d^2\mathbf{r} S(\mathbf{r} - \mathbf{s}) f(\mathbf{r}) \quad (10)$$

Note that we have not assumed here that $f(\mathbf{r})$ is well-sampled, and we have made no approximations to the true PSF. The discrete version of $\phi(\mathbf{s}_i, \mathbf{x}_i)$ 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):

$$L = \frac{k}{2} - \mathbf{I}^T \mathbf{F} \phi \mathbf{h} + \frac{1}{2} \mathbf{h}^T \phi^T \mathbf{F} \phi \mathbf{h} \quad (11)$$

$$= \frac{k}{2} - \mathbf{\Psi}^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T \mathbf{\Phi} \mathbf{h} \quad (12)$$

with

$$\mathbf{\Psi} \equiv \phi^T \mathbf{F} \mathbf{I} \quad (13)$$

$$\mathbf{\Phi} \equiv \phi^T \mathbf{F} \phi \quad (14)$$

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

$$\mathbf{F} \mathbf{I} \longrightarrow \mathbf{\Psi} \quad (15)$$

$$\mathbf{F} \longrightarrow \mathbf{\Phi} \quad (16)$$

$$\mathbf{g} \longrightarrow \mathbf{h} \quad (17)$$

That is, we have defined anequivalent image Ψ with a compact and spatially invariant PSF S but a much more extended and spatially complex pixel covariance matrix Φ^{-1} ; we have transform the PSF from an operation on the model to a component of the image and its covariance matrix.

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

2.2 Coaddition

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

$$P(\{\mathbf{I}\}|f) = \prod_n P(\mathbf{I}_n|f) \quad (18)$$

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

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

Because the grid on which we evaluate \mathbf{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 $\boldsymbol{\Psi}_i$ and $\boldsymbol{\Phi}_i$, so we can just sum these:

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

$$k_c \equiv \sum_n k_n \quad (21)$$

$$\boldsymbol{\Psi}_c \equiv \sum_n \boldsymbol{\Psi}_n \quad (22)$$

$$\boldsymbol{\Phi}_c \equiv \sum_n \boldsymbol{\Phi}_n \quad (23)$$

This is clearly the same as (12), with the coadded k , $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ replacing the single-epoch ones. This coadd is a “sufficient statistic” – it allows us to exactly compute a the likelihood for any arbitrary model – for the full dataset, 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.

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

Unfortunately, this approach is likely computationally impractical in general. The simplest reason is that $\boldsymbol{\Phi}$ is enormous. For a typical 4000×4000 coadd image comprised of images with 40×40 -pixel PSFs, $\boldsymbol{\Phi}$ 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. But even in this case, the extent of the covariance matrix ($\boldsymbol{\Phi}^{-1}$) 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 are less clear; the fact that the effective PSF – now just a Sinc kernel – is the same everywhere may enable fast lookup-table methods for model evaluation.

2.3 Detection Maps

One area where Ψ and Φ 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:

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

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

$$L(\mathbf{c}, \alpha) = \frac{k}{2} - \alpha \sum_i \Psi(\mathbf{s}_i) S(\mathbf{s}_i - \boldsymbol{\mu}) + \frac{\alpha^2}{2} \sum_{i,j} \Phi(\mathbf{s}_i, \mathbf{s}_j) S(\mathbf{s}_i - \boldsymbol{\mu}) S(\mathbf{s}_j - \boldsymbol{\mu}) \quad (25)$$

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

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

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

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

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

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 :

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

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

²With 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.

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

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

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

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

We detect by applying 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].

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 Φ 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 Φ , 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 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.

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

$$\Psi(\mathbf{r}) = \frac{1}{\sigma^2} \int d^2\mathbf{x} \phi(\mathbf{r} - \mathbf{x}) I(\mathbf{x}) \quad (32)$$

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sigma^2} \int d^2\mathbf{x} \phi(\mathbf{r}_1 - \mathbf{x}) \phi(\mathbf{r}_2 - \mathbf{x}) \quad (33)$$

where we have redefined the constant point function to $\phi(\mathbf{x})$, highlighted the \mathbf{x} -dependence of \mathbf{I} , and defined the constant per-pixel variance as σ^2 . Note that the output coordinate system \mathbf{r} is the same as the input pixel grid \mathbf{x}

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.