How to COAAD Images. I. Optimal Source Detection and Photometry of Point Sources Using Ensembles of Images

Barak Zackay and Eran O. Ofek
Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot, Israel

Received 2015 December 6; revised 2016 October 21; accepted 2016 December 28; published 2017 February 21

Abstract

Stacks of digital astronomical images are combined in order to increase image depth. The variable seeing conditions, sky background, and transparency of ground-based observations make the coaddition process nontrivial. We present image coaddition methods that maximize the signal-to-noise ratio (S/N) and optimized for source detection and flux measurement. We show that for these purposes, the best way to combine images is to apply a matched filter to each image using its own point-spread function (PSF) and only then to sum the images with the appropriate weights. Methods that either match the filter after coaddition or perform PSF homogenization prior to coaddition will result in loss of sensitivity. We argue that our method provides an increase of between a few and 25% in the survey speed of deep ground-based imaging surveys compared with weighted coaddition techniques. We demonstrate this claim using simulated data as well as data from the Palomar Transient Factory data release 2. We present a variant of this coaddition method, which is optimal for PSF or aperture photometry. We also provide an analytic formula for calculating the S/N for PSF photometry on single or multiple observations. In the next paper in this series, we present a method for image coaddition in the limit of background-dominated noise, which is optimal for any statistical test or measurement on the constant-in-time image (e.g., source detection, shape or flux measurement, or star–galaxy separation), making the original data redundant. We provide an implementation of these algorithms in MATLAB.

Key words: cosmology: observations – gravitational lensing: weak – methods: data analysis – surveys – techniques: image processing – techniques: photometric

1. Introduction

Coaddition of observations is one of the most basic operations on astronomical images. The main objectives of coaddition are to increase the sensitivity (depth) of observations, to remove signal from high-energy particle hits on the detector (sometimes called cosmic-rays), and to decrease the point-spread function (PSF) width (e.g., deconvolution from wavefront sensing, Primot et al. 1990; deconvolution from PSF, Bertin et al. 2002). One question that is common to all these tasks is what is the best method to coadd the images. This of course requires a definition of what “best” means.

Here, we assume that all observations are registered and resampled to the same grid (e.g., using Swarp; Bertin et al. 2002), and focus on the combination operation of the images. There are several common complications for image coaddition. Ground-based images are often taken under variable seeing, and transparency conditions. Even some space-based observations may suffer from such problems. For example, X-ray images commonly have non-uniform PSF across the field of view. These complications make the coaddition operation less trivial than simple (weighted) scalar addition.

Coaddition is playing, and will play, a major role in ongoing and future surveys (e.g., SDSS, York et al. 2000; PanSTARRS, Kaiser et al. 2002; PTF, Law et al. 2009, DES, The Dark Energy Survey Collaboration 2005; LSST, Iveicz et al. 2008). There are many approaches to image coaddition after the registration step. One pre-step common to many methods is to reject some of the images with the worst seeing, background, or transparency. There is no sensible prescription to which images one should reject, and it is straightforward to show that this approach leads to a loss of sensitivity compared with the maximum possible. One extreme example for this process is lucky imaging (Law et al. 2006), where one often discards up to 99% of the data frames. The next coaddition steps involves either weighted summation of the images or applying partial filtering (or deconvolution) to each image in order to homogenize the PSFs of all the images, followed by coaddition.

For example, Annis et al. (2014) recommended coadding the SDSS Stripe 82 images by weighted summation of the images, where the weights are of the form

$$w_j = \frac{F_j}{s_j \sqrt{V}},$$

and \(k = 1, m = 2,\) and \(n = 1\). Here \(w_j\) is the weight applied to the image \(i, F_j\) is proportional to the product of the telescope effective area, detector sensitivity and atmospheric transparency (i.e., flux-based photometric zero point), \(s\) is the width of the PSF, and \(V\) is the variance of all of the background noise sources (e.g., background, readout noise). On the other hand, Jiang et al. (2014) adopted a similar formula but with \(k = 1, m = 1, n = 1\). Another example for such a weighted summation was given by Fischer & Kochanski (1994), who solved a numerical optimization scheme for the optimal weights with which the images should be added.

Other approaches are also used. For example, some authors perform PSF homogenization on all the images prior to coaddition. This is done by convolving the frames with some kernel in order to ensure all the observations have the same PSF. For example, Desai et al. (2012) define a median-seeing PSF, and then for each image find the convolution kernel that will transform the image (via convolution) to a median-seeing
image, apply to all images, and then sum the images. This convolution kernel can be found using techniques outlined in Phillips & Davis (1995), Alard & Lupton (1998), and Bramich (2008). However, for images with seeing that is worse than the target seeing, the PSF homogenization operation becomes a deconvolution operation, which amplifies the noise in high spatial frequencies and creates long-range correlations in the noise. Another issue of the PSF homogenization approach is the artificial insertion of correlations between neighboring pixels. As a result, any measurement on the data becomes more involved. For example, applying a naive matched filter with the coadd-image PSF will result in additional loss of information.

In a series of several papers we will tackle the problem of optimal image coaddition. In this paper (Paper I), we will lay down the statistical formalism and assumptions we use for the coaddition problem. We construct a coaddition method that achieves the maximum possible signal-to-noise ratio (S/N) for detection of nonblended faint point-like sources in the coadded image under the assumption of constant variance Gaussian noise. In addition, we present a coaddition method that is optimal for photometric measurements of both faint and bright objects (i.e., even for sources in which the source noise is non-negligible). These coaddition products are the analogs of the linear matched filter for ensembles of images. We note that this coaddition method, optimized for source detection, was used by Masci & Fowler (2009) to reduce the WISE satellite data. However, in order to recover the sharpness of the original images, they applied deconvolution after the coaddition process.

It is noteworthy that linear matched filter images are smeared and their pixels are correlated. Therefore, general hypothesis testing and statistical measurements on them requires knowledge and consideration of the spatial covariance matrix. This makes them non-attractive for visualization and non-intuitive for performing measurements other than source detection and flux measurement (e.g., resolving binary stars or measuring galaxy shapes).

In Paper II (Zackay & Ofek 2017), we build on the results of this paper to construct a coaddition method for the special (but common) case of background-dominated noise. Equipped with the background-dominated noise assumption, we find a simple transformation that removes the pixel correlation induced by the matched-filtering step, and makes the additive noise in all pixels uncorrelated and have equal variance. Most importantly, this technique provides a sufficient statistic for any statistical measurement or decision on the data (e.g., source detection, flux measurement, shape measurement). Moreover, this method conserves the information of all the spatial frequencies, and provides a PSF which is typically sharper than the PSF of the best image in the stack we coadded. In addition, this image is practically indistinguishable from a regular astronomical observation, and any image analysis code can be applied to it unchanged. Combining all these properties together, it allows a major reduction in the amount of data products users will need in order to perform any further processing on the data (e.g., galaxy shape measurements for weak lensing). Finally, this coaddition algorithm is fast and its implementation is trivial.

---

1 The constant noise requirement can be relaxed as one can solve the problem locally in a region in which the noise variance is roughly constant.

2 A statistic is sufficient with respect to a model and its associated parameters if no other statistic that can be calculated from the same sample provides any additional information as to the value of the model parameters.

In future papers (Zackay et al. 2017, in preparation), we will apply the algorithm from Paper II for the case of rapid imaging through the turbulent atmosphere and where the PSF is either known (through observing a reference star or using a wavefront sensor) or unknown. We show that this method potentially provides better results than existing approaches for the coaddition of speckle images (e.g., speckle interferometry, Labeyrie 1970; lucky imaging, Law et al. 2006, and its suggested improvements, e.g., Garrel et al. 2012).

The structure of this paper is as follows. In Section 2, we review the statistical formalism of image coaddition and refer to some basic lemmas on how to combine random variables. In Section 3, we discuss the optimal coaddition image for faint source detection. In Section 4 we describe how to obtain the optimal coaddition image for photometric measurements, not restricted to faint sources. In Section 5, we give a simple-to-use formula to calculate the expected S/N for source detection and PSF photometry for both single images and ensembles of images. In Section 6, we suggest possible variants of this method, which, albeit sub-optimal for source detection and photometry, are resilient to particle hits, and generalize the concept of median for images. In Section 7, we demonstrate our detection algorithm on simulated and real images and show it surpasses the sensitivity of current popular methods. In Section 8 we briefly describe the code we use. Finally, we summarize the important formulae and conclude in Section 9.

---

2. Statistical Formalism and Background

Denote the $j$th measured image, in a series of background-subtracted observations of the same position, by $M_j$. Further, denote by $P_j$ the PSF of the $j$th observation.

The statistical model for the $j$th image is given by

$$M_j = (F_j T) \otimes P_j + \epsilon_j,$$

where $T$ denotes the true (background-subtracted) sky image, $F_j$ is a scalar, representing the transparency (same as in Section 1), $\otimes$ represents convolution, and the noise term $\epsilon_j$ has variance

$$V[\epsilon_j] = B_j + (F_j T) \otimes P_j,$$

where $B_j$ is the variance of all the position-independent noise—

the sum of the sky background variance, the read noise variance, and the dark current variance. For simplicity, we assume that the gain is 1 and that the counts are measured in electrons.

In the following subsections we derive several important well-known results of signal detection using hypothesis testing. The optimal detection of an attenuated signal in the presence of varying noise is reviewed in Section 2.1. In Section 2.2 we derive a general rule for weighted addition of random variables, while in Section 2.3 we derive the well-known linear matched filter solution for source detection.

2.1. Detection and Measurement of an Attenuated Signal in the Presence of Varying Noise

In this section, we review the simple problem of how to detect or measure a signal, with a known template, in data given by an ensemble of observations. We strictly assume that the signal in all observations is attenuated linearly with known attenuation coefficients, and that the noise in all the observations is Gaussian. Albeit simple, the solution to this problem is
the key toward constructing an optimal algorithm for image coaddition.

Let $X_j$ be a set of independent Gaussian variables with variance

$$V[X_j] = \sigma_j^2.$$ (4)

Given the null hypothesis, which states that there is no signal and is denoted by $\mathcal{H}_0$, we assume

$$E[X_j|\mathcal{H}_0] = 0,$$ (5)

where $E[X|\mathcal{H}]$ denotes the expectancy of the variable $X$ given that the hypothesis $\mathcal{H}$ is true. Given the alternative hypothesis, which states there is a signal, and is denoted by $\mathcal{H}_1(T)$, we assume

$$E[X_j|\mathcal{H}_1(T)] = \mu_j T.$$ (6)

Here $\mu_j$ are the known scaling factors of each observation. The log-likelihood ratio test statistic (which is proven to be the most powerful test at any given size, and therefore optimal; see Neyman & Pearson 1933) is simply given by

$$\mathcal{L} = \sum \frac{X_j^2}{2\sigma_j^2} - \sum \frac{(X_j - \mu_j T)^2}{2\sigma_j^2}. $$ (7)

For detection purposes, i.e., to reject the null hypothesis, the term $\frac{\mu_j^2 T^2}{2\sigma_j^2}$ can be ignored, as it does not depend on the data. However, this term is still relevant for flux measurements. Simplifying and absorbing the factor of 2 into the likelihood, we get

$$\mathcal{L} = \sum \frac{\mu_j T}{\sigma_j} X_j.$$ (8)

The log-likelihood ratio is linear in $T$, hence we can identify a sufficient statistic

$$S = \sum \frac{\mu_j T}{\sigma_j} X_j,$$ (9)

with a distribution that is independent of $T$, and therefore can be used to reject the null hypothesis against the alternative hypotheses $\mathcal{H}_1(T)$ for all values of $T$ simultaneously. Therefore, the optimal strategy for detection of a statistically significant signal is to calculate $S$ and check if it is above or below a certain threshold $S_\gamma$. The value of $S_\gamma$ is fixed by the desired false alarm probability $\gamma$. To evaluate this threshold, we need to calculate the expectancy and variance of $S$ given $\mathcal{H}_0$:  

$$E[S|\mathcal{H}_0] = 0,$$ (10)

Given the independence of $X_j$ and the fact that $V[X_j|\mathcal{H}_0] = \sigma_j^2$, the variance of $S$ given $\mathcal{H}_0$ is

$$V[S|\mathcal{H}_0] = \sum \frac{\mu_j^2}{\sigma_j^2} = I,$$ (11)

and the threshold is

$$S_\gamma \simeq \sqrt{-2 I \ln(\gamma) - 4}.$$ (12)

Under the same definitions, if we wish to get the most accurate estimator of $T$, we can use the fact that the likelihood of the data can be calculated for all $T$ using $S$ (Equation (9)) and $I$ (Equation (11)). This is because we can express the log likelihood of $M$ given $T$ by

$$\mathcal{L}'(M|T) = \sum \frac{(X_j - \mu_j T)^2}{2\sigma_j^2} = g(X) - ST + \frac{TT^2}{2},$$ (13)

where $g(X) = \sum \frac{X_j^2}{2\sigma_j^2}$ does not depend on $T$ and therefore cannot influence any measurement. Regardless of whether the Bayesian or frequentist approaches are used, the likelihood of the data as a function of $T$ is sufficient for any further measurement or detection process. Therefore, $S$ and $I$ are sufficient for any measurement or statistical decision on the set of statistics $\{X_j\}$.

2.2. Weighted Addition of Random Variables

Another well-known result is that the statistics $S$ and $I$ we have derived in Equations (9) and (11) naturally arise in a much more general setup, which is known as the weighted addition of random variables. Assume we have an arbitrary set of independent random variables $X_j$ and two hypotheses $\mathcal{H}_0$, $\mathcal{H}_1$ that satisfy

$$V[X_j|\mathcal{H}_0] = V[X_j|\mathcal{H}_1] \equiv V(X_j).$$ (14)

We can define the signal-to-noise ratio $)$ of a statistic $S$ with regard to the decision between $\mathcal{H}_0$ and $\mathcal{H}_1$ by

$$S/N[S] = \frac{E(S|\mathcal{H}_0) - E(S|\mathcal{H}_1)}{\sqrt{V(S)}}.$$ (15)

Next, we can ask: what is the statistic, linear in $X_j$, that maximizes this $S/N$? The solution to this problem is the following weighted mean:

$$S = \sum \frac{E(X_j|\mathcal{H}_0) - E(X_j|\mathcal{H}_1)}{V(X_j)} X_j,$$ (16)

which coincides with the previously defined $S$ (Equation (9)). Moreover, the maximal $S/N$ is $\sqrt{I}$. The detailed solution of this problem is given in Appendices A and C.

Weighted addition extends also to the case of adding estimators. If we are given a sequence of independent measurements

$$\theta_j = \mu_j T + \epsilon_j,$$ (17)

in which we assume only that $V(\epsilon_j) = \sigma_j^2$, the same set of statistics $S$ and $I$ arise, and the maximum $S/N$ estimator for $T$ is

$$T = \frac{\sum \frac{\mu_j^2}{\sigma_j^2} X_j}{\sum \frac{\mu_j^2}{\sigma_j^2}} = \frac{S}{I}.$$ (18)

This simple and rigorous fact is derived in Appendix B.

The fact that this same solution repeats under many different general contexts and assumptions allows us to derive optimal statistics for real-world astronomical problems. The first of them is the well-known matched filter for source detection in single astronomical images.
2.3. The Matched Filter Solution for Source Detection

Now that we are equipped with the simple yet robust tools for adding random variables, we can derive the standard matched filter solution for point source detection in a single image. To allow rigorous statistical analysis of source detection, we assume (throughout the paper) that the source we are trying to detect is well separated from other sources, which might interfere with its detection. Let $M$ be an image such that

$$M = T \otimes P + \epsilon,$$

where $T$ is the true background-subtracted sky image, $P$ is the image PSF, and $\epsilon$ is an additive white Gaussian noise term.

We would like to detect, or measure, a single point source at position $(x_0, y_0)$. Each pixel coordinate $(x, y)$ in the image is an independent\(^4\) information source, where the PSF scales the amount of flux each pixel would get. That is,

$$\mu(x, y) = P(x - x_0, y - y_0).$$

Using Equation (9), we can write the optimal statistic for source detection at position $(x_0, y_0)$ for a single image:

$$S(x_0, y_0) = \sum_{x,y} \frac{P(x - x_0, y - y_0)M(x, y)}{\sigma_j^2}$$

$$= \left( \frac{P \otimes M}{\sigma_j^2} \right) (x_0, y_0),$$

where $P(x, y) = P(-x, -y)$. Thus, we obtain the well-known optimal statistic for the detection of a single source in an image with additive white Gaussian noise. The solution is to convolve the image with its reversed PSF (i.e., cross-correlation of the image with its PSF). This statistic is called the linear matched filter image.

Due to its optimality, robustness, and ease of implementation and interpretation, the linear matched filter is the method of choice for many source detection tools (e.g., SExtractor; Bertin & Arnouts 1996; see a review in Masias et al. 2012).

3. Image Coaddition for Source Detection

Now that we have shown that using the tools outlined in Section 2 we get the optimal statistic for source detection in single images, we can apply these methods for source detection to an ensemble of images. Before doing that, it is worthwhile to develop an intuition on how the solution should look. Expanding the linear matched filter solution of one image to multiple images, we expect the solution to be somewhat like a three-dimensional linear matched filter summed over the image index axis, to get a statistic with the dimensions of an image. Therefore, we expect that the solution will be simply matched filtering each image with its own PSF followed by weighted summation.

For source detection, we assume that the noise is approximately Gaussian and that the variance in each pixel, excluding the variance induced by the source itself, is independent of pixel position. Since our null hypothesis is that there is no source, the noise induced by the source should be ignored. However, noise induced by neighboring sources cannot be ignored. Therefore, our assumption that the variance is constant (at least locally) in the image is true only if the image is not crowded (i.e., confusion noise can be neglected). Using these assumptions, we denote the variance of the $j$th image by

$$\sigma_j^2 = B_j = V[\epsilon].$$

Here $B_j$ is the background (or more formally the spatially invariant variance component) of the $M_j$ image. We can now apply the machinery for coaddition of random variables developed in Section 2 to the problem of image coaddition.

Let us assume that we would like to detect a point source at position $(0, 0)$. In this case the model for such a point source is $T = T_0 \delta_{(0,0)}$, where $T_0$ is the point source’s flux and $\delta_{(0,0)}$ is a two-dimensional array with 1 at position $(0, 0)$ and zero otherwise. Inserting this to the model of $M_j$ we get

$$M_j = (F_j T_0 \delta_{(0,0)}) \otimes P_j + \epsilon_j = F_j T_0 P_j + \epsilon_j.$$ (24)

Using the result from Appendix A, we can combine all the pixels from all the images as a list of independent statistics. Therefore, the optimal detection statistic for a source at position $(0, 0)$ is

$$S = \sum_{j,s} \frac{F_j P_{sj}}{\sigma_j^2} M_j(x).$$ (25)

For simplicity, $x$ represents the $(x, y)$ coordinate. Using the fact that searching a source at position $p \equiv (p_1, p_2)$ on a set of images is exactly as searching for a source at position $(0, 0)$ on the input images shifted by $-p$, we get

$$S(p) = \sum_{j,s} \frac{F_j P_{sj}}{\sigma_j^2} M_j(x - p).$$ (26)

This can be written in vectorial notation as

$$\vec{S} = \sum_j \left( \frac{F_j P_j}{\sigma_j^2} \right) \otimes \vec{M}_j,$$ (27)

where the $\sim$ sign represents the Fourier transform, and the $-$ sign denotes the complex conjugate operation. We note that by putting the bar sign over the hat sign we mean that the complex conjugate operation follows the Fourier transform operation.

It is worthwhile to note that $S$, the coaddition of filtered images, is a log-likelihood ratio image, where $S(p_1, p_2)$ is the optimal test statistic for the existence of a source at $(p_1, p_2)$. In order to find sources using $S$, we therefore have to search for local maxima. To decide if the source candidate is statistically

\(^4\)In practice the pixels maybe slightly correlated due to hardware effects in the CCD.
significant, it is more convenient to normalize $S$ by its standard deviation:

$$ S_e = \frac{S}{\sigma_S} = \frac{\sum (\frac{F_j}{\sigma_j^2}) \otimes M_j}{\sqrt{\sum_{j,x} \frac{F_j^2}{\sigma_j^2}}} . $$

(29)

The value of each local maximum divided by the standard deviation of the noise (Equation (29)) is the source significance in units of standard deviations. To get $S$ in units of flux (best estimate only in the context of background-noise-dominated source), we need to normalize differently (see Equation (18)) to get

$$ S_{\text{flux}} = \frac{\sum (\frac{F_j}{\sigma_j^2}) \otimes M_j}{\sum_{j,x} \frac{F_j^2}{\sigma_j^2}} . $$

(30)

Analyzing our solution, we can see that if we have only one image then the well-known linear matched filter is recovered. However, when we have more than one image, the optimal detection statistic is equivalent to applying a matched filter to each image separately, weighting the images by $\frac{F_j}{\sigma_j^2}$ and accumulating. This is different from the other solutions found in the literature, where you first coadd and then match filter, or perform partial filtering for homogenization, then coadd, and then match filter. The method we developed is optimal by construction, hence it will have better performance relative to other coaddition schemes. We note that with this method in hand, any additional image, no matter how poor the conditions are, will increase the sensitivity. Therefore, the rejection of images becomes a damaging operation. Because of the robustness of our optimality arguments, whenever the assumptions that we have made on the noise are true, we expect this method to surpass the performance of all popular methods found in the literature.

An important question is by how much our method improves source detection relative to other methods. The answer to this question depends on the details, like the distribution of observing conditions (e.g., seeing) and the exact method we would like to compare with. We can quantify the gain in using our method by answering a simple question: how much survey speed is being lost when a non-optimal filter is used in the case of a single image? A single matched filter cannot be adequate for multiple images with different PSFs. No matter what weights are later applied in the combination of the different matched-filtered images (all with the same filter), the $(S/N)^2$ of the other methods cannot surpass the sum of the $(S/N)^2$ in their individual pixels.

Given a single image with a Gaussian PSF with width $s$, the best filter is a Gaussian with width $s$. In Appendix D we calculate the $S/N$ loss factor, when one uses a Gaussian filter with a width $ks$. We find that the sensitivity ratio of these two cases is

$$ \frac{(S/N)_{\text{incorrect width}}}{(S/N)_{\text{optimal}}} = \left( \frac{2k}{k^2 + 1} \right)^2 . $$

(31)

This means that the reduction in survey speed (or the detection information) is

$$ \frac{(S/N)^2_{\text{incorrect width}}}{(S/N)^2_{\text{optimal}}} = \left( \frac{2k}{k^2 + 1} \right)^2 . $$

(32)

Figure 1 shows this survey speed reduction as a function of $k$. As expected, this function has a maximum of unity at $k = 1$. Using a width which is a factor of 1.2 (2) different from the actual PSF width will result in a survey speed loss of about 6% (40%). Therefore, given the typical distribution of the seeing conditions, one should expect to obtain an improvement of between a few percent and about 25% in survey speed using our method, relative to, e.g., the Annis et al. (2014) weighted summation.

4. Optimal Photometric Measurements from a Set of Observations

Another important problem in the context of astronomical image processing is measuring the flux of sources. The two leading methods for performing flux measurements are aperture photometry and PSF photometry. Both methods are well-defined on single images. When the PSF is exactly known, the method of PSF photometry is the most exact since it applies an appropriate weight to each pixel before summation. Aperture photometry is a good approximation that is often used when the star is bright, the PSF is unknown, or the measurement is done for sources that are not point-like. In this section, we present the correct way to extend photometric measurements to ensembles of observations. In Section 4.1 we discuss the extension of PSF photometry to ensembles of images, and in Section 4.2 we cover aperture photometry. We note that in the case of a uniform noise variance (i.e., faint source) and known PSF, it is both easier and more accurate to use the coaddition method presented in Paper II, and then perform aperture/PSF photometry.

An important note is that the flux measurement of a source from the images that were used for the source’s first detection is biased because of the Malmquist effect. Basically, if there is a population of sources with mean flux below the detection
threshold, some of them will have upward noise fluctuations, making them detectable. Therefore, the naive flux measurement of these sources from their discovery image may be biased. For further information, see Hogg & Turner (1998).

4.1. Extending PSF Photometry to Ensembles of Images

We define the optimal photometric measurement as a statistic $S$ such that the expectancy of the statistic is the true flux of the source and its variance is minimal. That is, $E[S] = T$, and $V[S]$ is minimal. This is equivalent to finding a statistic that maximizes the $S/N$ and normalizing it to units of flux. Because in the case of photometry we need to take into account the source noise, in this section we will work under the assumption that

$$M_j = F_j T \otimes P_j + \epsilon_j,$$

where the distribution of $\epsilon_j$ is Gaussian with variance

$$V[\epsilon_j] = B_j + F_j T \otimes P_j,$$

which does not distribute uniformly across the image, because bright stars have larger variance components. We note that we still assume that the measured source is well separated from other sources.

Following the same procedure we used for source detection, we get the maximum $S/N$ for the flux measurement. When measuring a point source at position $(0, 0)$ with true flux $T_0$, we can write

$$M_j = (F_j T_0 \delta_{0,0}) \otimes P_j + \epsilon_j = F_j P_j + \epsilon_j.$$

Again, we assume every pixel is statistically independent from the others, and therefore we can apply the previously developed machinery (Equation (18)) pixel by pixel and get the maximal $S/N$ estimator:

$$\tilde{T}_{\text{PSF photometry}} = \frac{\sum_{x} \frac{F_j P_{x,j} T_0}{F_j P_{x,j} + F_j P_{x,j} T_0} M_j(x)}{\sum_{x} \frac{F_j P_{x,j} T_0}{F_j P_{x,j} + F_j P_{x,j} T_0}}.$$

In the above formula, we ignored the small, but important, bias due to the exact subpixel location of the source. Such a bias can be dealt with by replacing the denominator of Equation (36) with

$$\sum_{x} \frac{F^2_j P_{x,j} P_{x,j}^{\Delta q}}{F_j P_{x,j} + F_j P_{x,j} T_0},$$

where $P_{x,j}^{\Delta q}$ is $P_j$, shifted by the fractional pixel shift $\Delta q$. To apply this correction, it suffices to either store the PSFs or simply calculate ahead of time the correction incurred by each subpixel shift (in a gridded table) by each PSF. In addition, in order to compute the statistic in Equation (36), we need to know the true flux $T_0$ that we are looking for. This problem could be avoided by one of the following.

1. Solving numerically for the true $T_0$ in iterations: guess $T_0 = 0$, compute $\tilde{T}_{\text{PSF photometry}}$, and then update $T_0 = \tilde{T}_{\text{PSF photometry}}$, and compute $\tilde{T}_{\text{PSF photometry}}$ again with the resulting flux from the first iteration. Since the first estimate of $T_0$ is an excellent first guess, the latter is practically converged, and more iterations are unnecessary. This solution is source specific, and therefore should be implemented only if photometry of a single source is needed.

2. Scanning the $T_0$ space logarithmically: using a slightly wrong $T_0$ does not lead to significant errors in the measurement. Since the guess of $T_0$ influences only weights that are independent of the data itself, it does not introduce a bias. The standard deviation of a statistic that is calculated with a slightly wrong $T_0$ is generally very close to the optimal value. For example, using Gaussian PSFs with width $1 < s < 10$ pixels, $1 < B < 1000$, and $1 < T < 10^6$, we find that using $T_0$, which is wrong by a factor of 2, leads to reduction of at most 1% in survey speed. This solution requires the coaddition process to produce several images ($\sim 8$, as the relevant range is from the background level up to the maximal dynamic range value. These numbers are typically only up to a factor of $2^8$ from one another).

3. Aiming for a specific $T_0$. If one knows that the targets have a specific magnitude range, it is possible to use the most relevant $T_0$.

We want to stress that the whole discussion on $T_0$ is only relevant for flux estimation of bright sources. For the detection of sources, we are not required to know $T_0$ a priori, as demonstrated in Section 3.

The calculation of this measurement statistic can be done quickly by noting that once a $T_0$ is assumed, the calculation of the measurement statistic on all image positions can be done by convolution using the fast Fourier transform (FFT).

To gain intuitive interpretation of the PSF photometry measurement statistic, it is useful to inspect the behavior of the statistic on one image. In the expression for $\tilde{T}_{\text{PSF photometry}}$, we can identify the two general approximations astronomers often make in the extreme limits of the relation between the brightness of the target and the background. On the one hand, if a target is brighter than the background on many pixels, we can see that the optimal measurement statistic we obtained is similar to aperture photometry with the optimal aperture, as in the center the weights of the pixels converge to 1. On the other hand, we see that if a target is much weaker than the background variance, the statistic converges to a matched filter on the image. When extending this to multiple images we can see that in the statistic $\tilde{T}_{\text{PSF photometry}}$ the combination of different images is rather nontrivial.

4.2. Extending Aperture Photometry to Ensembles of Images

Even when the PSF is unknown, or when the object shape for which we perform flux measurements is poorly known (e.g., a galaxy), it is still important to combine the flux measurements from different images correctly. This is because observations in the ensemble will generally have different noise properties. For each image, depending on the seeing conditions, source flux, and its PSF, one can choose the best aperture radius $r_j$ that maximizes the $S/N$ in the image (see below).

We can calculate from each image the aperture photometry measurement

$$\theta_j = M_j \otimes C_{r_j},$$

where $C_{r_j}$ is the convolution kernel of the optimal aperture $r_j$. Calculation of convolution directly (without FFT) could also be very fast if the convolution kernel is small.
where \( C_{r_j} \) is given by
\[
C_{r_j} = \begin{cases} 
1 & \text{if } |r| < r_j \\
0 & \text{if } |r| \geq r_j.
\end{cases}
\] (39)

Substituting Equation (2) into Equation (38),
\[
\theta_j = F_j T \otimes P_j \otimes C_{r_j} + \epsilon_j \otimes C_{r_j},
\] (40)

Denote the source intrinsic light distribution by \( Q \). \( Q \) satisfies
\[
\sum_x Q(x) = 1,
\] (41)

and \( T = T_0 Q \). Next, we define
\[
\mu_j = F_j T \otimes P_j \otimes C_{r_j},
\] (42)
to be the fraction of the source flux within the measurement aperture. Now, we can write
\[
\theta_j = \mu_j T_0 + \epsilon_j \otimes C_{r_j},
\] (43)
The variance of \( \epsilon_j \otimes C_{r_j} \) is
\[
V[\epsilon_j \otimes C_{r_j}] = \mu_j T_0 + \pi r_j^2 \sigma_j^2.
\] (44)

Now that the problem is cast into the same form as that solved in Section 2, we can use the maximum S/N measurement statistic (Equation (18)) to find the best estimator for aperture photometry:
\[
\tilde{\theta}_{\text{aperture photometry}} = \frac{\sum_j \mu_j \mu_j T_0 + \pi r_j^2 \sigma_j^2 \theta_j}{\sum_j \mu_j \mu_j T_0 + \pi r_j^2 \sigma_j^2}.
\] (45)

Note that if the source is very strong, the term \( \mu_j T_0 \) might dominate the noise variance \( V[\epsilon_j \otimes C_{r_j}] \). In that case, one needs to have a reasonable guess of the source flux \( T_0 \) prior to the coaddition process (see Section 4.1).

Moreover, one needs to solve for the best aperture radius \( r_j \) (that depends on \( T_0 \)). This solution can be obtained separately for each image by maximizing the squared S/N of \( \theta_j \):
\[
r_j = \arg\max_r \left( \frac{\mu_j^2(r)}{\mu_j T_0 + \pi r^2 \sigma_j^2} \right),
\] (46)

where \( \arg\max_r (f(r)) \) denotes the value of the parameter \( r \) for which the function \( f(r) \) obtains its maximum. Maximizing this could be done if one has a model for \( f(r) \), which could be deduced from the intrinsic size of the object measured and a rough PSF model.

5. Estimating the Limiting Flux and the Expected S/N of Sources

For several reasons, including S/N calculation and optimizations, it is worthwhile to have a formula for the S/N of PSF photometry over multiple images. Although this formula is easy to derive, we are not aware of a simple analytic expression for the S/N of PSF photometry in the literature for either single or multiple observations.

Following the previous section, we do not necessarily assume that the sources are weak compared to the sky, and therefore our measurement model is
\[
M_j = T \otimes P_j + \epsilon_j,
\] (47)
where \( \epsilon_j \) is a Gaussian with mean zero and variance:
\[
V[\epsilon_j] = B_j + F_j T \otimes P_j,
\] (48)
where \( B \) is the background noise and \( T \otimes P_j \) is the Poisson noise of the source.

In order to calculate the S/N we refer to the fact that \( (S/N)^2 \) is an additive quantity. That is, for any two uncorrelated statistics \( S_1, S_2 \), we can build a statistic
\[
S_{1,2} = \alpha_1 S_1 + \alpha_2 S_2
\] (49)
such that the S/N of \( S_{1,2} \) satisfies
\[
(S/N)^2[S_{1,2}] = (S/N)^2[S_1] + (S/N)^2[S_2].
\] (50)

This fact is proven in Appendix C. Using this, all that we need to do in order to calculate the total \( (S/N)^2 \) of the optimal coaddition of a source with flux \( T_0 \) in an ensemble of images is to sum the \( (S/N)^2 \) over all the pixels in all the images.

Therefore, the \( (S/N)^2 \) for flux measurement (the source noise term in the variance is irrelevant for detection purposes) is given by
\[
(S/N)^2 = \sum_j \int_{-\infty}^{\infty} T_0^2 F_j^2 P_j^2(x, y) dx dy V[M_j(x, y)]
= \sum_j \int_{-\infty}^{\infty} T_0^2 F_j^2 P_j^2(x, y) \left( B_j + F_j T_0 \right) dx dy.
\] (51)

For the sake of S/N calculation, it is a good enough approximation to assume that the PSF is a two-dimensional symmetric Gaussian. Rewriting the above with a Gaussian PSF, with width \( s_0 \), and replacing the \( x \) and \( y \) coordinates by the radial coordinate \( r \), this can be written as
\[
(S/N)^2 = \sum_j \int_0^{s_0} \left( T_0 F_j e^{-r^2/(2s_0^2)} \right)^2 \left[ (2\pi s_0^2)^2 + B_j \right] 2\pi r dr.
\] (52)

This integral has an analytic solution given by
\[
(S/N)^2 = T_0 \sum_j F_j - \sum_j 2\pi s_0^2 B_j \ln \left( 1 + \frac{F_j T_0}{2\pi s_0^2 B_j} \right).
\] (53)

In the source-noise-dominated case \( (F_j T_0 \gg 2\pi s_0^2 B_j) \), the formula simplifies to
\[
(S/N)^2 = T_0 \sum_j F_j - \sum_j 2\pi s_0^2 B_j \left[ \ln(F_j T_0) - \ln(2\pi s_0^2 B_j) \right].
\] (54)

In the background-noise-dominated case \( (F_j T_0 \ll 2\pi s_0^2 B_j) \), the \( (S/N)^2 \) becomes
\[
(S/N)^2 = \sum_j \frac{T_0^2 F_j^2}{4\pi s_0^2 B_j}.
\] (55)

We stress that this equation is correct only if the numbers \( T_0 \) and \( B_j \) are in units of electrons. Another remark is that Equation (55) is actually the exact (no approximation) \( (S/N)^2 \) for the detection of a source with flux \( T_0 \), because in this situation, the relevant variance should be exactly \( V[M_j(x, y)] \).
Therefore, by isolating $T_0$ in Equation (55) we can get a formula for the limiting magnitude for detection at a given S/N:

$$T_0 = \sqrt{\frac{(S/N)^2}{\sum_{j} f_j^2 / 4\pi^2 \beta_j^2}}. \quad (56)$$

6. Robust Coaddition

A drawback of the method presented so far is that it uses the summation operator, which is not resilient against outliers. Outliers can be a major source of noise and often require special treatment. The typical way outliers are dealt with is either using median coaddition, min/max rejection, or sigma clipping.

The summation in either Equations (26) or (27) includes weights, and in either case we may want to choose an operation that is more robust. Because of the weights in Equations (26) and (27), a simple median operation is not possible. Instead, one can use the weighted median estimator to replace the weighted average. The weighted median of a set of estimators

$$\theta_j = \mu_j T + \epsilon_j \quad (57)$$

is simply defined as the 50% weighted percentile of the normalized estimators $\theta_j$, where each of the estimators is given the weight of the inverse standard deviation. In a similar manner, one could apply sigma clipping to the estimators, removing estimators that deviate from the average estimated value of $T$ by more than 3 of its own standard deviation.

It is important to note that these operations make sense only when we measure the flux of unblended objects using the matched-filtered images. Note that since our method uses the matched-filter image as the estimator, sharp outliers like cosmic-rays or hot pixels will tend to reduce in significance (as we are using the wrong filter for their detection).

Given the problems involved in using robust estimators, we would like to suggest a different approach, which is to identify and remove sharp outliers using image subtraction, given an artifact-free image subtraction statistic. Such an image-subtraction algorithm can be found in Zackay et al. (2016). In the situation where a large number of images is coadded ($>10$), any outlier that will influence the final measurement must be easily visible in the subtraction image of any two images. Therefore, the detection and removal of any significant outliers prior to coaddition should be possible and probably constitutes the most accurate approach for dealing with artifacts. We will further discuss this in Zackay et al. (2016).

7. Tests

To verify that our method indeed works properly and to try and identify possible problems, we tested it on simulated data (Section 7.1) and real data (Section 7.2).

7.1. Simulated Data

The actual improvement of the coaddition technique depends on the distribution of the observing conditions (e.g., seeing, background, and transparency). Therefore, we use parameters that roughly mimic typical observing conditions. Obviously, for different surveys and use cases, these parameters could vary significantly.

We repeated the following simulation 100 times. We simulated twenty 1024 × 1024 pix² images, each with 2048 stars in an equally spaced grid. For simplicity, we assumed all the images have equal transparency. The seeing distribution was taken as uniform with FWHM between 1.5 and 6 pixels. The background distribution was uniform between 500 and 1900 e⁻ pixel⁻¹. The point spread functions were assumed to be circular two-dimensional Gaussians. We adjusted the brightness of the sources to be close to the detection limit in the coadded image (100 photons per image; see Section 5). The images were created aligned, so no further registration was applied.

Next, we summed the images using various coaddition schemes. The techniques we tested are (i) weighted summation of the images using the Annis et al. (2014) scheme (see Equation (1)); (ii) weighted summation of the images using the Jiang et al. (2014) method (see Equation (1)); and (iii) our method. We note that unlike Annis et al. (2014) or Jiang et al. (2014) we did not remove any images. To evaluate the detection S/N of each method, we averaged the S/N of all sources at their correct positions in the coaddition image (using Equation (29)). Given the number of sources simulated, this average is accurate to a level of about 1%. We then calculated, using the specific seeing and brightness parameters of the images, the theoretically maximal S/N for source detection (Equation (55)).

In Figure 2, we plot the distribution of the ratio of the obtained S/N with the various methods and the theoretically calculated S/N. It is clear that the method presented in this paper achieves the maximum possible S/N and that the symmetric scatter around its value is the measurement error of the achieved S/N. It is further clear that for this specific case, our method provides a ~5% higher S/N than the method of Jiang et al. (2014), and a ~10% higher S/N than that of Annis et al. (2014). These translate to a survey speed increase of about 10% over Jiang et al. (2014) and 20% over Annis et al. (2014).

7.2. Real Data

We performed several comparisons of our method with other techniques, using real data. Our test data originates from the Palomar Transient Factory (PTF10; Law et al. 2009; Rau et al. 2009) data release 2. The images were obtained under various atmospheric conditions. All the images were reduced using the

---

7. Inverse standard deviation and not inverse variance. A way to derive this is to define a symmetrical exponential loss function for the error in each image and find the value that minimizes the global loss function. This value is given analytically by the weighted median.

8. Performing a median operation on regular images will yield weird artifacts when performed on images with different PSFs. Examples for such are seeing-distribution-dependent bias, flux-dependent PSF, and the effective removal of any observation with a PSF that is different (better or worse) from the majority of the images.

9. Given the fact that weights (if calculated correctly) should take care of low-quality images, we regard this step as damaging rather than beneficial.

10. http://www.ptf.caltech.edu/ipsf
We note that the coaddition technique described in Paper II is identical (up to small numerical errors) to the method discussed in this paper. However, the method described in Paper II has several important advantages and hence should be used whenever adequate.

Next, we compared the detection \((S/N)^2\) of the sources in the various images.\(^{12}\) The reason for using \((S/N)^2\) is that it is an additive quantity that is proportional to the survey speed and to the detection information. In order to do the comparison, we run mextractor.m again in forced-detection mode. In this mode, we provide the code with a list of positions of the sources detected in the deep reference image, and the code measures the detection significance and \(S/N\) at these locations. For each of the different coaddition methods on the subset of images, we calculated the following quantity: for each star detected in the reference image, we divide its \(S/N\) measured\(^{13}\) in a coaddition image by its \(S/N\) measured in our optimally coadded image. Then, we sum the squares of these ratios, and calculate their median. This roughly gives the survey speed in each coadded image, relative to our coaddition method, and is listed in Table 2 (\(I_4\)). In these specific examples, our method provides a factor of 17% to 47% improvement over unweighted coaddition, and 8% to 18% improvement over the weighted addition methods. We note that this is sensitive to the exact seeing distribution of the observations and we estimate that the typical improvement will be between a few percent to 25%, relative to Annis et al. (2014). Finally, in Figure 4 we present, for data set 3, the relative ratio between the detection \(S/N\) of individual sources in the coadded images relative to the detection \(S/N\) in the optimal coadd image. Again, our coaddition provides better \(S/N\) both at the faint and bright

\(^{11}\) http://webhome.weizmann.ac.il/home/eofek/mplab/

\(^{12}\) Note that the detection \(S/N\) is defined without the source-noise term in the denominator (see Equation (55)).

\(^{13}\) The \(S/N\) is measured by filtering the image with its PSF, normalizing the filtered image by its own standard deviation, and reading the value at the local maximum that corresponds to the source.
Note. Selection criteria for reference images and coadd images in the various sets used for testing the coaddition methods. “Field” indicates the PTF field ID, while “CCDID” is the CCD number in the mosaic (see Law et al. 2009; Laher et al. 2014). “Type” is either Ref. for the deep reference image, or Coadd for the subset we coadded using the various methods. Note that field 100031 is in the vicinity of M51.

8. Code and Implantation Details

The formulae we present in this paper are straightforward to code. However, most of the attention in the implementation should be given to pre-processing steps and measurements of the properties of the images. This includes estimating the background mean, variance, PSF, and cosmic-ray removal.

The code that performs the coaddition technique suggested in this paper is available as part of the Astronomy and Astrophysics toolbox (available online) for MATLAB (Ofek 2014). This package can be used for all of the image-processing steps, including the de-bias, flat-field correction, and cosmic-ray removal.

We note that usual implementation details like registration, resampling to the same grid, measuring the background and variance levels, and measuring/interpolating the PSF, as always, require attention. However, the attention to the details required by this method is not different from that required for the successful application of other methods. Below we comment on several important implementation details.

Background and variance estimation: the background and variance in real wide field of view astronomical images cannot be treated as constants over the entire field of view. Therefore, we suggest estimating them locally and interpolating. To estimate the background and variance one needs to make sure that the estimators are not biased by stars or small galaxies. Our suggestion is to fit a Gaussian to the histogram of the image pixels in small regions and to reject from the fitting process pixels with high values (e.g., the upper 10th percentile of pixel values).

Estimating the transparency: the transparency $F_i$ of each image is simply its flux-based photometric zero point. However, one has to make sure that these zero-points are measured using PSF photometry rather than aperture photometry, otherwise the zero-points may depend on the seeing.

Estimating the PSF: among the complications that may affect the PSF measurement are pixelization, interpolation, and resampling. Furthermore, the PSF is not spatially constant across large fields of view (typically, changes are important on scales larger than a few arcminutes), and it may also change with intensity due to charge self-repulsion. This specifically may lead to the brighter-fatter effect (e.g., Walter 2015). We note that the fact that one needs to estimate the PSF in order to run this method should not be viewed as a drawback, as any decent method that finds sources in the image requires this step anyway.

9. Summary

We argue that popular image coaddition methods do not achieve the maximal S/N possible for source detection and photometry. We derive the optimal statistic for source detection and photometry under the assumptions that the noise is approximately Gaussian and that the target is well separated from other bright sources. We show that the optimal way to coadd images for source detection is by filtering (i.e., cross-correlating) each image with its PSF, and then summing with weights. This method is summarized by Equation (27) (or equivalently, Equation (28)). In order to find sources, we need to find local maxima in the calculated score map. The significance of these sources in units of standard deviations is given by Equation (29). We note that the computational requirements of applying the method are low, as the cross-correlation operation can be computed using the Fast Fourier transform (FFT). We also derive optimal statistics for PSF photometry (Equation (36)) and aperture photometry (Equation (45)) for an ensemble of images. Finally, we derive a formula to calculate the S/N for PSF photometry in an ensemble of images with a symmetric Gaussian PSF (Equation (53)). For statistical tasks other than source detection or photometric measurements, we refer readers to Paper II in this series, in which we provide a coaddition method that is optimal for any statistical measurement or decision, under the more restrictive (but common) assumption of background-dominated noise.

14 https://webhome.weizmann.ac.il/home/eofek/matlab/
15 We are currently using 256 × 256 arcsec\(^2\) blocks.
16 See Appendix A of Ofek et al. (2011) for a method to calculate this zero-points.
17 The run time of FFT on an image is typically an order of magnitude faster than reading/writing an image from the hard disk.
We demonstrate our coaddition method for source detection on simulated images as well as on real images. Our method increases the survey speed (for faint source detection and photometry) by a few percent to 25% over traditional methods, both in theory and in practice.

We thank Avishay Gal-Yam, Assaf Horesh, Frank Masci, William Newman, and Ora Zackay for discussions. This paper is based on observations obtained with the Samuel Oschin Telescope as part of the Palomar Transient Factory project, a scientific collaboration between the California Institute of Technology, Columbia University, Las Cumbres Observatory, the Lawrence Berkeley National Laboratory, the National Energy Research Scientific Computing Center, the University of Oxford, and the Weizmann Institute of Science. B.Z. is grateful for receiving the Clore fellowship. E.O.O. is incumbent of the Arye Dissentshik career development chair and is grateful for support by grants from the Willner Family Leadership Institute Ilan Gluzman (Secaucus NJ), Israel Science Foundation, Minerva, Weizmann-UK, and the I-Core program by the Israeli Committee for Planning and Budgeting and the Israel Science Foundation (ISF).

Appendix A
Weighted Addition of Independent Random Variables for Hypothesis Testing

Here, we will show that the same statistic that was derived in Section 2.1 as an exact solution for detection of an attenuated signal in the presence of varying Gaussian noise is the solution of a much more robust question—maximizing the $S/N$ of the weighted addition of independent random variables. Given a set of statistical variables $X_j$ with two hypotheses $H_0$ and $H_1$ with expectancies

$$E[X_j|H_0] = 0, \quad E[X_j|H_1] = \mu_j,$$

and further assuming that the variance of $X_j$ is equal under the assumption of both hypotheses, i.e.,

$$V[X_j|H_0] = V[X_j|H_1] = \sigma_j^2,$$

if we know the exact distributions of the variables $X_j$ given both hypotheses, we can use the Neyman-Pearson lemma (Neyman & Pearson 1933), which states that the optimal test statistic for the decision between $H_0$ and $H_1$ is the log-likelihood ratio test. Here, we want to construct a test statistic that we can apply even when the exact distributions are unknown, but their first and second moments are known.

Given the fact that all the statistics are assumed to be independent, it is reasonable to assume that the correct way to combine the variables is via a linear sum (if the distributions are known we can simply sum the difference in the log of the probability of observing $X_j$ for each hypothesis). That is, we are

We note that even if $E[X_j|H_0] = z_j$, it is always possible to transfer the problem to hypothesis testing on $Y_j = X_j - z_j$, in which case the expectancy given the null hypothesis is again zero.
interested in a linear statistic of the form:

$$S = \sum_j \beta_j X_j.$$  \hspace{1cm} (60)

If the set of variables we have is large, so that the distribution of a weighted sum of the variables can be approximated by a Gaussian distribution for both hypotheses, then the only important properties of the resulting sum are the expectancy of $S$ given both hypotheses, and the variance of $S$ given both hypotheses. Moreover, if the statistic has equal variance under both hypotheses, then there is one number that characterizes our ability to distinguish between the hypotheses, which is the squared $S/N$:

$$\frac{(S/N)^2}{V[S]} = \frac{|E[S|\mathcal{H}_0] - E[S|\mathcal{H}_1]|^2}{V[S]}.$$  \hspace{1cm} (61)

Thus, we want to find the linear combination that maximizes the $S/N$. i.e., find a set of weights $\beta_j$ such that the score in Equation (60) will have maximal $S/N$ (Equation (61)).

We assume that $X_j, \mu_j$, and $\beta_j$ are complex variables (this will become useful in future papers in the series). Substituting $\{0, \mu_j\}$ as the expectancies of $X_j$ given the two hypotheses, we get

$$|E[S|\mathcal{H}_0] - E[S|\mathcal{H}_1]|^2 = \sum_j |\beta_j\mu_j|^2 = |\Delta E[S]|^2.$$  \hspace{1cm} (62)

Given the hypothesis $\mathcal{H}_a$, where $\alpha \in \{0, 1\}$, the variance of $S$ is

$$V[S|\mathcal{H}_a] = \sum_j |\beta_j|^2 \sigma_j^2.$$  \hspace{1cm} (63)

In order to maximize the squared $S/N$ with respect to $\beta_j$, we equate all the partial derivatives to zero:

$$\frac{\partial (\frac{(S/N)^2}{V[S]})}{\partial \beta_j} = 0,$$  \hspace{1cm} (64)

which gives us

$$0 = 2 \frac{\partial \Delta E(S)}{\partial \beta_j} \frac{\Delta E}{V} - 2 \frac{\partial V(S)}{\partial \beta_j} \frac{\Delta E^2}{V^2} = 2 \mu_j \frac{\Delta E}{V} - 2 \beta_j \sigma_j^2 \frac{\Delta E^2}{V^2},$$  \hspace{1cm} (65)

where the overline denotes complex conjugation. Simplifying, we get the equation

$$\beta_j = \frac{\mu_j V}{\sigma_j^2 \Delta E}.$$  \hspace{1cm} (66)

Noting that multiplying all the weights $\beta_j$ by a constant factor does not change the $S/N$, we can simplify further and get

$$\beta_j = \frac{\mu_j}{\sigma_j^2} = \frac{E[X_j|\mathcal{H}_1] - E[X_j|\mathcal{H}_0]}{V_j}.$$  \hspace{1cm} (67)

Thus, we now have a general formula for the weighted addition of random variables.

We note that if the exact distributions of the random variables are known, then the optimal combination is the log-likelihood ratio test statistic. This statistic could be composed of some function of the variables themselves that is not linear.
This means that the best score will be of the form

$$\sum_j f_j(x_j).$$

(68)

What we have derived in this appendix is the best linear approximation to this score that could be derived under much less exact knowledge on the variables themselves.

### Appendix B

**Weighted Addition of Independent Estimators**

Here, we will show that the same statistic that was derived in Section 2.1 and Appendix A is the solution (up to a multiplicative constant) of another general question: what is the maximum S/N linear combination of independent estimators?

Given a set of statistics $\theta_j$ such that

$$\theta_j = \mu_j T + \epsilon_j,$$

(69)

where $\mu_j$ are known complex variables,

$$E[\epsilon_j] = 0, \quad V[\epsilon_j] = \sigma^2_j,$$

(70)

we want to find the best weighted linear combination estimator

$$\hat{\theta} = \sum_j \beta_j \theta_j,$$

(71)

such that $E[\theta] = T$ and $V[\theta]$ is minimal. Calculating the expectancy of $\theta$ we get

$$E[\theta] = \sum_j \beta_j \mu_j T,$$

(72)

while the variance of $\theta$ is

$$V[\theta] = \sum_j |\beta_j|^2 \sigma^2_j.$$  

(73)

Maximizing the $(S/N)^2$ of the estimator

$$(S/N)^2[\theta] \equiv \frac{|E[\theta]|^2}{V[\theta]}$$

(74)

is equivalent to minimizing the variance of $\theta$ with respect to a fixed $E[\theta]$. Moreover, $(S/N)^2[\theta]$ is invariant to scalar multiplication of $\theta$ because

$$(S/N)^2[\theta \beta] = \frac{|E[\theta \beta]|^2}{V[\theta \beta]} = \frac{|\beta|^2|E[\theta]|^2}{|\beta|^2 V[\theta]} = (S/N)^2[\theta].$$

(75)

Therefore, we can maximize $(S/N)^2[\theta]$ with respect to all $\beta_j$. After normalization, the minimum variance estimator is

$$\hat{\theta} = \frac{\theta}{\sum_j \beta_j \mu_j}. $$

(76)

To maximize $(S/N)^2[\theta]$ with respect to $\beta_j$, we equate all the partial derivatives to zero:

$$\frac{\partial (S/N)^2[\theta]}{\partial \beta_j} = 0,$$

(77)

which gives us

$$0 = 2 \frac{\partial E[\theta]}{\partial \beta_j} \frac{E[\theta]}{V[\theta]} - \frac{\partial V(\theta)}{\partial \beta_j} \frac{E[\theta]}{V[\theta]^2},$$

$$= 2 \beta_j \frac{E[\theta]}{V[\theta]} - \frac{2 \beta_j \sigma^2_j E[\theta]}{V[\theta]^2},$$

(78)

where the overline denotes complex conjugation. Simplifying, we get the equation

$$\beta_j = \frac{\overline{E[\theta]} V[\theta]}{\sigma^2_j E[\theta]}.$$  

(79)

Because $(S/N)^2[\theta]$ is invariant under multiplication by a constant factor, we get

$$\beta_j = \frac{\overline{E[\theta]} V[\theta]}{\sigma^2_j E[\theta]}.$$  

(80)

Adjusting the estimator to match our first requirement that $E[\theta] = T$ we can finally write

$$\hat{\theta} = \frac{\sum_j \beta_j X_j |\mu_j|^2}{\sum_j \beta_j \sigma_j^2}.$$  

(81)

### Appendix C

**S/N of the Addition of Optimally Weighted Statistics**

In Appendix A we find the best way to add statistics even when we do not know their exact distributions. Here we would like to derive a closed formula for the S/N of the resulting statistic. Denote by

$$(S/N)^2[X_j] = \frac{|\Delta E[X_j]|^2}{V[X_j]}.$$  

(82)

Write the expression for the statistic $S$, the maximum $(S/N)$-weighted addition of the variables defined in Appendix A, as

$$S = \sum_j \frac{\Delta E[X_j]}{V[X_j]} X_j.$$  

(83)

Calculating the expected difference in $S$ given the two different hypotheses, we get

$$\Delta E[S] = E[S|\mathcal{H}_1] - E[S|\mathcal{H}_0] = \sum_j \frac{|\Delta E[X_j]|^2}{V[X_j]},$$

(84)

while the variance of $S$ is

$$V[S] = \sum_j \frac{|\Delta E[X_j]|^2}{V[X_j]^2} V[X_j] = \sum_j \frac{|\Delta E[X_j]|^2}{V[X_j]^2}.$$  

(85)
Therefore, the squared S/N of $S$ is simply given by
\[
(S/N)^2[S] = \sum_j \frac{|\Delta E[X_j]|^2}{V[X_j]} = \sum_j \frac{|\Delta E[X_j]|^2}{V[X_j]} = \sum_j \frac{|\Delta E[X_j]|^2}{V[X_j]}
\]
This means that the squared S/N of a statistic, which is the optimal weighted addition of a set of statistics, is the sum of the squared S/Ns of the individual statistics. This property now allows us to easily estimate the sensitivity of observations ahead of time, with an intuitive closed formula. This (almost) same calculation applies also for optimal weighted addition of estimators, and we omit it to prevent cumbersome repetitions.

Appendix D
How Much Survey Speed is Lost When We Use the Wrong PSF?

Here we derive an analytic solution to the question of how much sensitivity is lost in the detection process when the wrong linear matched filter is used. Following previous sections, we define the measured image by
\[
M = T \otimes P + \epsilon,
\]
and assume that the source we are trying to detect is well separated from other sources. We also assume that the source we are looking for is faint relative to the background, leading to a spatially invariant noise variance:
\[
V[\epsilon(x)] = \sigma^2.
\]
For simplicity, we assume that the PSF, $P$, can be approximated as a symmetric Gaussian. We define $s_u$ as the width of the real PSF $P$, while $s_u$ is the width of the Gaussian used for match filtering. Denote by $w_i(s_u)$ the weight for the pixel $x$, when using a kernel with width $s_u$. Now, we can calculate the $(S/N)^2$ of the statistic
\[
W(s_u) = \sum_x w_i(s_u) M(x)
\]
by
\[
(S/N)^2[W(s_u)] = \frac{E[W(s_u)]^2}{V[W(s_u)]} = \left(\sum_i w_i(s_u) E[M(x)]\right)^2 \sum_i w_i^2(s_u) V[M(x)].
\]
Substituting the PSF and the match-filtering kernel with the relevant Gaussians and approximating the sums with integrals, we get
\[
(S/N)^2[W(s_u)] = \frac{T_0^2}{\sigma^2} \left(\frac{1}{(2\pi s_u^2)^2} \int e^{-|x|^2/(2s_u^2)} e^{-|x|^2/(2s_u^2)} dx_1 dx_2\right)^2
\]
where $|x|^2 = x_1^2 + x_2^2$.

The maximum of this function with respect to a given value of $s_u$ is obtained at $s_u = s_u$. When substituting $s_u = k s_u$ and dividing the $(S/N)^2$ by the maximum possible, we get
\[
(S/N)^2[W(k s_u)] = \frac{T_0^2}{\sigma^2} \left(\frac{1}{(2\pi s_u^2)^2} \int e^{-|x|^2/(2s_u^2)} e^{-|x|^2/(2k s_u^2)} dx_1 dx_2\right)^2
\]
where $|x|^2 = x_1^2 + x_2^2$.

Figure 1 shows a plot of the loss in survey speed when using the wrong PSF, as derived above. This simple exercise gives us an idea of how much sensitivity we lose when we combine images with different PSFs, because when we match-filter the coadded image, all images are match-filtered with a single PSF, which cannot simultaneously be optimal for all of them.

References
Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325
Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, ApJ, 794, 120
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bertin, E., Mellier, Y., Radovich, M., et al. 2002, adass XI, 281, 228
Bramich, D. M. 2008, MNRAS, 386, L77
Desai, S., Armstrong, R., Mohr, J. J., et al. 2012, ApJ, 757, 83
Fischer, P., & Kochanski, G. P. 1994, AJ, 107, 802
Garel, V., Guyon, O., & Baudouin, P. 2012, PASP, 124, 861
Hogg, D. W., & Turner, E. L. 1998, PASP, 110, 727
Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, arXiv:0805.2366
Jiang, L., Fan, X., Bian, F., et al. 2014, ApJS, 213, 12
Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, Proc. SPIE, 4836, 154
Labeyrie, A. 1970, A&A, 6, 85
Laher, R. R., Surace, J., Grillmair, C. J., et al. 2014, PASP, 126, 674
Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395
Masias, M., Freixenet, J., Lladó, X., & Peracaula, M. 2012, MNRAS, 422, 1674
Neyman, J., & Pearson, E. S. 1933, RSPTA, 231, 289
Ofek, E. O. 2014, Astrophysics Source Code Library, ascl:1407.005
Ofek, E. O., Frail, D. A., Breslaher, B., et al. 2011, ApJ, 740, 65
Ofek, E. O., Laher, R., Surace, J., et al. 2012, PASP, 124, 854
Phillips, A. C., & Davis, L. E. 1995, adass IV, 77, 297
Primot, J., Rousset, G., & Fontanella, J. C. 1990, JOSAA, 7, 1598
Rau, A., Kulkarni, S. R., Law, N. M., et al. 2009, PASP, 121, 1334
The Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0503436
Walter, C. W. 2015, JInst, 10, C05015
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zackay, B., & Ofek, E. O. 2017, ApJ, 836, 188
Zackay, B., Ofek, E. O., & Gal-Yam, A. 2016, ApJ, 830, 27