Langmuir
Article
pubs.acs.org/Langmuir

Image Analysis with Rapid and Accurate Two-Dimensional Gaussian Fitting

Stephen M. Anthony[†] and Steve Granick*[,†,‡,§]

[†]*Department of Chemistry*, [‡]*Department of Materials Science and Engineering and* [§]*Department of Physics, University of Illinois, Urbana, Illinois 61801*

Received February 2, 2009. Revised Manuscript Received April 4, 2009

A computationally rapid image analysis method, weighted overdetermined regression, is presented for two-dimensional (2D) Gaussian fitting of particle location with subpixel resolution from a pixelized image of light intensity. Compared to least-squares Gaussian iterative fitting, which is most exact but prohibitively slow for large data sets, the precision of this new method is equivalent when the signal-to-noise ratio is high and approaches it when the signal-to-noise ratio is low, while enjoying a more than 100-fold improvement in computational time. Compared to another widely used approximation method, nine-point regression, we show that precision and speed are both improved. Additionally, weighted regression runs nearly as fast and with greatly improved precision compared to the simplest method, the moment method, which, despite its limited precision, is frequently employed because of its speed. Quantitative comparisons are presented for both circular and elliptical Gaussian intensity distributions. This new image analysis method may be useful when dealing with large data sets such as those frequently met in astronomy or in single-particle and single-molecule tracking using microscopy and may facilitate advances such as real-time quantification of microscopy images.

Introduction

The images of objects, as recorded by cameras, are most conveniently represented by two-dimensional (2D) circular or elliptical Gaussian distributions of light intensity. Some cases represent diffraction-limited point sources, for which 2D Gaussians are the most computationally tractable representation of the Airy disk, and deviations from the quality of being an Airy disk are minor in practice.[1] This is so for observations of stars and other elliptical features in digital images,[2−4] fluorescent molecules,[5−8] and quantum dots.[9] Cases involving larger objects do not involve diffraction and produce images naturally modeled using Gaussians. This is so for particle image velocimetry (PIV), colloids, and bubbles.[10−12] This paper concerns how to determine the parameters of 2D Gaussian intensity distributions in a computationally efficient fashion.

The best accuracy and precision currently comes from optimizing the parameters using least-squares iterative Gaussian fitting;[13] this is so especially at low signal-to-noise levels, but this iterative method is computationally expensive. It requires on the order of tens of milliseconds per object when using current state-of-the-art desktop personal computers. For applications such as single-particle tracking and single-molecule tracking, the total number of fits needed to analyze one data set may be on the order of 1 million, which currently can require hours of computation. Similar considerations are present in the astronomy community.

This is why, to analyze large data sets, it is common to use the centroid and moment methods,[2,12,13] which are computationally quicker but sacrifice precision. These methods do not determine all the parameters of Gaussian distributions, neither their width nor amplitude. Some other estimators have been developed that retain much of the computational efficiency of the centroid method, while giving results comparable to iterative methods; but they hold only at very high levels of signal-to-noise ratio or are otherwise limited.[11,14] They are standard in applications where the signal-to-noise ratio is high, such as PIV. For applications where noise cannot be ignored yet precision is critical, iterative methods remain standard.[5,8,13]

In the method we develop below, the computational demands approach those of the fastest methods, yet the resulting precision approaches that of least-squares Gaussian iterative fitting, even when the signal-to-noise ratio is low, and equals it when the signal-to-noise ratio is high. Even for the lowest signal-to-noise ratios, where the iterative method maintains some advantage regarding precision, our new method may represent a desirable alternative when working on large data sets, as it runs over 2 orders of magnitude faster than iterative methods.

Simulation and Methods

Image Models. Data similar to that expected from typical microscopy experiments were simulated in order to compare quantitatively the efficacy of various fitting methods. For some applications, such as PIV, a circular or elliptical Gaussian intensity distribution is almost an exact representation of the expected signal. For other applications, most notably single-molecule microscopy, the true signal is the diffraction-limited

*Corresponding author.

(1) Thompson, R. E.; Larson, D. R.; Webb, W. W. *Biophys. J.* **2002**, *82*(5), 2775–2783.
(2) Feng, Y.; Goree, J.; Liu, B. *Rev. Sci. Instrum.* **2007**, *78*, 053704.
(3) Gai, M.; Carollo, D.; Delbo, M.; Lattanzi, M. G.; Massone, G.; Bertinetto, F.; Mana, G.; Cesare, S. *Astron. Astrophys.* **2001**, *367*(1), 362–370.
(4) Gai, M.; Casertano, S.; Carollo, D.; Lattanzi, M. G. *Publ. Astron. Soc. Pac.* **1998**, *110*(749), 848–862.
(5) Huang, B.; Wang, W. Q.; Bates, M.; Zhuang, X. W. *Science* **2008**, *319*(5864), 810–813.
(6) Kao, H. P.; Verkman, A. S. *Biophys. J.* **1994**, *67*(3), 1291–1300.
(7) Moerner, W. E. *Nat. Methods* **2006**, *3*(10), 781–782.
(8) Yildiz, A.; Forkey, J. N.; McKinney, S. A.; Ha, T.; Goldman, Y. E.; Selvin, P. R. *Science* **2003**, *300*(5628), 2061–2065.
(9) Holtzer, L.; Meckel, T.; Schmidt, T. *Appl. Phys. Lett.* **2007**, *90*(5), 053902.
(10) Honkanen, M.; Saarenrinne, P.; Stoor, T.; Niinimaki, J. *Meas. Sci. Technol.* **2005**, *16*(9), 1760–1770.
(11) Nobach, H.; Honkanen, M. *Exp. Fluids* **2005**, *38*(4), 511–515.
(12) Crocker, J. C.; Grier, D. G. *J. Colloid Interface Sci.* **1996**, *179*(1), 298–310.
(13) Cheezum, M. K.; Walker, W. F.; Guilford, W. H. *Biophys. J.* **2001**, *81*(4), 2378–2388.
(14) Willert, C. E.; Gharib, M. *Exp. Fluids* **1991**, *10*(4), 181–193.

Airy spot. When pixelation and noise are considered, the Airy spot is close to being a 2D Gaussian, so that, for fitting purposes, the 2D Gaussian is an almost universally employed approximation. For simplicity, the signals simulated here were 2D circular and elliptical Gaussians. At the end of this paper, we compare them explicitly to calculations that consider the full Airy spot.

Elliptical Gaussians were generated with their major and minor axes aligned with the x and y directions of the pixelated image. This method could be trivially extended by adding an additional parameter θ, the angle at which the major axis forms with the x-axis. However, bear in mind that a common application of elliptical Gaussians is to fit the images of fluorescent particles to which astigmatism has been applied using a cylindrical lens to encode z-dimension information.[6] In such cases, the orientation of the cylindrical lens is known, and generally it is aligned with the detector. Then θ is an unnecessary parameter, and better fits are obtained using fewer parameters. In the discussion that follows, while many different aspect ratios were tested, for clarity, the majority of the analysis shown here refers to a set aspect ratio, widths of 1 and 1.5 for x and y, respectively.

In order to represent actual measurements, for each trial, a 2D Gaussian was generated at a random position not necessarily centered on a pixel. For circular Gaussians, the peak width or standard deviation of the Gaussian was fixed to 1 pixel. Additional simulations not included here reveal that results were qualitatively similar, even when the width was varied significantly. One exception of course is when the width is small enough that the intensity is limited to a single pixel, at which point no subpixel methods can be employed. Another exception is the method developed by Nobach et al.,[11] which depends strongly on peak width, and for which the 1-pixel standard deviation of the Gaussian was near the optimum anyway. Both the number of photons incident upon a pixel and the shot noise typical in charge-coupled detector (CCD) gain are Poisson processes. This noise was incorporated by selecting a random value from a Poisson distribution of mean N, where N was the number of counts for that pixel. To determine the effect of varying the signal-to-noise, the number of total counts in the pristine signal was varied, and a constant background noise, normally distributed with a standard deviation of 25, was added. Thus, there was a constant noise of 25, and additional noise where signal was present, scaling as the square root of N, N being the number of incident photons or photoelectrons. The analysis here assumes that the constant background level, which is generally easily obtainable, has already been subtracted and that the background noise level is known. Each trial was generated independently, then fit with all the fitting methods to allow direct comparison of results. Additionally, when calculating the signal-to-noise ratio, for simplicity, this article considers only the background noise.

To further mimic typical experimental conditions, the initial image was generated using more pixels than would be analyzed by the fitting methods whose efficacy would subsequently be compared. The portion to which the fit was applied was determined by selecting the brightest pixel in this image, then selecting a region of the appropriate size centered around that pixel. Some fitting methods, particularly at low signal-to-noise levels, have a tendency to localize to the center of the region selected for fitting. By selecting the region to be fit in this fashion, rather than on the true center of the signal, which is known for simulations but not experiments, the fits we obtained were representative of actual experimental constraints.

Simulations were also run with shot noise absent by applying normally distributed noise whose level was the same everywhere in the image. While not included here, these simulations showed efficacy, regarding reliability of the fit and dependence of the fit on the signal-to-noise ratio, qualitatively similar to the simulated images reported below that better approximate microscopy images.

Finally, while the 2D Gaussian is a nearly universally employed approximation of the full Airy spot pattern, simulations were also run to determine to what extent this approximation affects the precision. For these simulations, the background and shot noise were applied as usual, but this time they were applied to the Airy spot:

$$I(r) = \left(\frac{2 \cdot J_1\left(\frac{2 \cdot \pi \cdot r \cdot \mathrm{NA}}{\lambda}\right)}{r} \right)^2$$

where r is the distance from the origin, NA is the numerical aperture of the objective (0.75), λ is the wavelength of the light (570 nm), and J_1 is the Bessel function.[13] Each pixel represented 100 nm, a common experimental resolution. These simulations were only run for the two leading methods, iterative optimization, and the weighted regression method developed here.

Fitting Methods. The algorithms employed split into two categories: those that directly solve for the parameters, and those that employ iterative optimization to find their best values. Three of the methods compared here—the moment method, the regression method employed by Nobach, and the weighted overdetermined regression method we develop in this article—fall into the first category. For a circular Gaussian, four parameters suffice to describe the Gaussian: the x and y positions, the width of the Gaussian, and the amplitude. Another parameter, the total or integrated brightness of the Gaussian, frequently used in tracking algorithms, is a combination of the width and the amplitude. For elliptical Gaussians, either one or two additional parameters are needed; if the major and minor axes are aligned with the x and y coordinates, splitting the width parameter into two parameters, one for the width in each direction, suffices. If the orientation is unknown, an additional parameter must be introduced for that term.

Moment Method. The moment method is a computationally simple method to calculate the position. For each dimension, the center is given by

$$C_x = \frac{\sum_i \sum_j (x_i \cdot I_{ij})}{\sum_i \sum_j I_{ij}}$$

where x_i is the position in that dimension, and I_{ij} is the intensity of a given pixel.[12] Because of its extreme computational simplicity and reasonable accuracy, this method is one of the most widely used methods when analyzing large quantities of data.[2] Compared to the other methods compared here, the moment method does not determine exactly the same parameters. While it finds the position, it finds neither the width of the Gaussian nor its amplitude; instead, complementary methods determine related parameters. One such parameter, the total brightness of the particle, combines information from the amplitude and the widths, and is simply determined by adding up the intensity of all nearby pixels. The width is indirectly related to another parameter, the radius of gyration, and a final parameter, the ellipticity, provides a measure of whether the Gaussian is circular or elliptical, though it does not distinguish the orientation.[12] Of these, only the total brightness lends itself to direct comparison with parameters determined by the other techniques, and hence is the only one explored in this paper.

As an aside, it is helpful to be aware that the literature is inconsistent regarding the name of this method. While it is often referred to as the moment method, it is also frequently referred to as the centroid method.[13] Despite the fact that the moment method is often referred to as the centroid method, it is important to distinguish this from a similar method, also sometimes referred to as the centroid method, which is similar except that the intensity term, I_{ij}, is replaced with a constant. This paper does not employ the latter.

Nine-Point Regression. This method, developed by Nobach and Honkanen,[11] represents one of the latest in a series of methods developed for PIV. As time has progressed, methods have evolved to correct for deficiencies caused by pixel-locking, diffraction-limited lenses, and other deviations from ideal Gaussian intensity distributions. Fully described elsewhere,[11] this method applies regression to the nine point region that includes the brightest pixel, exactly solving for the elliptical Gaussian in the absence of noise. This method works admirably for the high signal-to-noise ratio problems for which it is designed, but struggles with noise. The implementation described inherently treats all Gaussians as elliptical with unknown orientation.

Iterative Optimization. The optimization method employed here is the downhill simplex method,[13] commonly used for particle tracking. The specific implementation employed here uses Matlab's fminsearch.m function. Simplex methods are both significantly simpler to implement and more commonly available, so they tend to be employed frequently. However, quasi-Newton optimization methods are almost always the most computationally efficient,[15] so, for comparison, an implementation of the quasi-Newton method, available in lsqnonlin.m, part of Matlab's Optimization Toolbox, was similarly tested, as well as the related Levenberg−Marquardt optimization. Simulation confirmed that quasi-Newton optimization is faster for this problem, frequently requiring approximately 30% less computational time, but with results otherwise identical to the downhill simplex method. Results for Levenberg−Marquardt were comparable. However, all optimization methods were found to occasionally fail to converge, returning infinite or undefined results for some parameters. Additionally, while the downhill simplex method would generate these values and continue running, the quasi-Newton and Levenberg−Marquardt methods would instead crash, terminating the simulation. The lesser computation time using the quasi-Newton method is negligible compared to the computational benefits of all the noniterative methods. Since all iterative optimization results were identical apart from minor differences in computational time, the decision was made to employ the downhill simplex method in the comparisons that follow.

We also note that iterative optimization depends somewhat upon the initial guesses of the parameters. The initial estimates employed here were the position of the brightest pixel and the intensity of the brightest pixel for that amplitude, and the standard deviation of the peak is used as proxy of the width. Experimentation revealed that even providing the actual values (before noise) altered the computation time by a factor of only 2 or less. For applications where this were significant, hybrid methods might be desirable, employing the best of the other computationally simpler methods to generate the initial estimates.

Overdetermined Weighted Regression. This method, developed in this paper, takes advantage of a computationally simpler least-squares estimator and incorporates additional features to compensate for noise. Sample implementations in Matlab are included in the Supporting Information. For the case of a circular Gaussian, the intensity of the image is described by

$$I_{xy} = A\cdot \mathrm{e}^{-((x-x_0)^2+(y-y_0)^2)/(2\cdot w^2)} + \varepsilon_{xy}$$

where I_{xy} is the intensity of the pixel, A is the peak amplitude, x and y are the coordinates of the individual pixels, x_0 and y_0 are the position of the center of the Gaussian, w is the standard deviation or width, and ε_{xy} corresponds to noise. An additional relatively constant background term may also be included; generally it is possible to determine the background intensity of an image and subtract that, a necessary step for employing this algorithm. In the absence of noise, it is possible to exactly transform this to a linear equation through use of the logarithm

$$x^2 + y^2 = (2\cdot x_0)\cdot x + (2\cdot y_0)\cdot y + (-2\cdot w^2)\cdot(\ln(I_{xy})) + (2\cdot w^2\cdot\ln(A) - {x_0}^2 - {y_0}^2)$$

which can equivalently be expressed as a linear equation with four unknowns:

$$x^2 + y^2 = a_1\cdot x + a_2\cdot y + a_3\cdot(\ln(I_{xy})) + a_4$$

In this case, we already know for each point the values of x, y, and I_{xy}. As long as more pixels are included than the number of unknowns, the system is overdetermined, and it is possible to directly determine estimates for the coefficients a_n using linear least-squares regression, after which one can substitute for the desired values.

A similar equation describes the elliptical Gaussian:

$$I_{xy} = A\cdot\exp\left[-\left(\frac{(x-x_0)^2}{2\cdot {w_x}^2} + \frac{(y-y_0)^2}{2\cdot {w_y}^2}\right)\right] + \varepsilon_{xy}$$

When transformed by the logarithm, this yields

$$\ln(I_{xy}) = \left(\frac{-1}{2\cdot {w_x}^2}\right)\cdot x^2 + \left(\frac{2\cdot x_0}{2\cdot {w_x}^2}\right)\cdot x + \left(\frac{-1}{2\cdot {w_y}^2}\right)\cdot y^2 + \left(\frac{2\cdot y_0}{2\cdot {w_y}^2}\right)\cdot y + \left(\ln(A) - \frac{{x_0}^2}{2\cdot {w_x}^2} - \frac{{y_0}^2}{2\cdot {w_y}^2}\right)$$

Again this is a linear equation, in this case with five unknowns:

$$\ln(I_{xy}) = a_1\cdot x^2 + a_2\cdot x + a_3\cdot y^2 + a_4\cdot y + a_5$$

While simple to implement, these transformations neglect the influence of noise, for which it is necessary to compensate. First, notice that noise, while approximately symmetrically distributed originally, is asymmetrically transformed when one takes the logarithm. Second, one must recognize that the various components of noise per pixel generally are constant (background noise, etc.), or else scale at a rate lower than the number of counts ($\sqrt{N}$ for Poisson shot noise). As such, pixels with greater intensity (on average, as some are higher due to noise) have a higher signal-to-noise ratio. When working on the logarithmic scale, the pixels with greater intensity have both higher signal-to-noise ratios and more symmetric noise. Therefore, it is better to assign more relative weight to the brighter pixels. Further, it is helpful to employ a threshold, such that pixels below a given intensity are not counted, particularly as it is necessary to exclude any pixels

(15) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: New York, 2007; p 1235.

with negative intensity with respect to the background, which cannot appropriately be transformed to our logarithmic scale.

The approximate mean noise level for the entire image is generally easy to obtain; for images with large regions of background, the standard deviation of the background regions suffices. Assuming this level of noise for each pixel, it is possible to determine an estimate of the noise level of each pixel on the logarithmic scale, which is then used to assign the appropriate weighting. While this weighting ignores contributions from shot noise, shot noise is only significant relative to background noise for the brightest pixels, which anyway have the highest overall signal-to-noise ratios. This then alters slightly the relative weightings of those pixels but this was found to have minimal significance.

Two additional steps were important for elliptical Gaussian distributions. When a square region of the original image was selected to be fit, for elliptical Gaussians this contains many more pixels that contain no signal. Hence, it is important to exclude contributions from points that contain only noise. To do so, we first quickly determine which rectangular subregion of the image contains some signal from the object, and restrict the algorithm to that region. Additionally, a higher-than-usual threshold is employed to minimize the contribution of remaining points with low signal-to-noise.

Results and Discussion

Convergence Failure. In order to accurately compare the methods, it must be recognized that some of the methods fail to converge upon a solution sometimes. This ratio, the fraction of instances that a given method outright failed, is plotted against the signal-to-noise ratio in Figure 1a for circular Gaussian distributions. Outright failure is considered to be when the position was off by more than 2 pixels, or the width, amplitude, or total brightness is off by more than 200% from the actual value. Infinite and undefined values are also considered to be failures. If even a single parameter was not found properly, the fit for that object using that method was disregarded.

Unsurprisingly, the moment method, due to its simplicity, demonstrates the greatest stability, continuing to generate results even at the lowest signal-to-noise ratios. Iterative optimization is the next most stable method, exhibiting failure to converge a few percent of the time for signal-to-noise ratios less than 5. Our weighted regression method is not much worse for signal-to-noise ratios larger than 5, failing marginally more often (less than 0.5% additional failure, none above a signal-to-noise ratio of 7.5). Results below a signal-to-noise ratio of 5 are worse, skyrocketing to near complete failure when the signal-to-noise ratio was less than 4. This is a consequence of the thresholds employed; at higher signal-to-noise ratios, they serve to exclude noisy points, improving the fit, but when the signal-to-noise ratio is low, not enough points are left to solve the linear regression problem. As such, no data is reported for weighted regression with a signal-to-noise ratio below 4. The method of nine-point regression fares the worst, failing more than 1% of the time even at a signal-to-noise ratio of 15, and more than 50% of the time for signal-to-noise ratios less than 5. However, as some fraction of the objects are fit at all levels, in the discussion below we report results for all levels.

When elliptical Gaussians are analyzed (Figure 1b), several differences are observed. The failure rates for the moment and iterative optimization methods are nearly the same; however, the threshold signal-to-noise ratio below which our weighted regression method has difficulty increases to roughly 8, below which the method does not do well. Neither does nine-point regression do as well as previously. In this case, the essential difference is not the switch to elliptical Gaussians, as the method inherently treats all Gaussians as elliptical, but that the method performs less well when widths deviate from their optimum values.

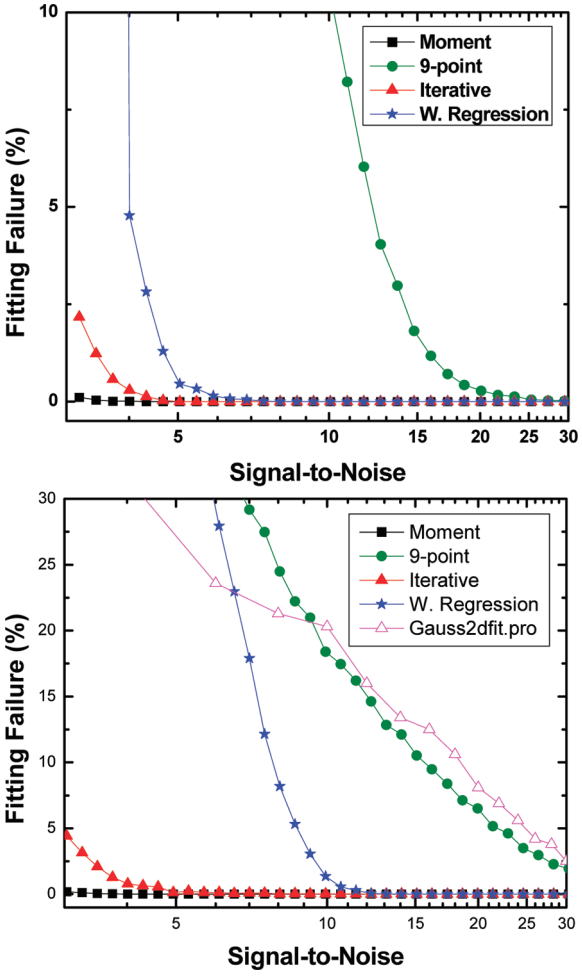


Figure 1. The fraction of objects for which fitting fails to converge appropriately for the various methods, as a function of the ratio of the signal-to-noise. Signal is considered to be the amplitude of the Gaussian intensity distribution. The background shot noise is not included in this ratio but was included in the calculations. (a) The failure fraction for circular Gaussian intensity distributions calculated using the moment method (black squares), nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). (b) The failure fraction for the same methods but using elliptical Gaussian intensity distributions. As described in the text, comparison was also made to the 2D elliptical Gaussian fitting method included in IDL, which is a commercially available program (pink open triangles). Symbols are the same in both panels.

In addition to the four methods compared in depth in this article, the 2D Gaussian fitting package in Research Systems Incorporated (RSI) IDL Version 6.3, gauss2dfit.pro, is also included because of empirical observations that the failure rate for this program is high, despite widespread usage of this program. The outputs and error messages for gauss2dfit.pro reveal that this program is an implementation of iterative optimization. When the example implementation included in IDL's online help is run, catastrophic failures occur at least 1% of the time. To provide a more direct comparison with the simulations run here, the example implementation was modified slightly to have comparable widths and image sizes. These simulations were not a direct match to the other comparisons in this paper, as they

did not include the Poisson distributed shot noise, and as the Gaussians were always perfectly centered in the image; both differences should, if anything, have decreased the number of fitting failures. Despite this, the failure rate for the built-in IDL function gauss2dfit.pro is typically the largest or nearly so. While its precision, accuracy, and computation time were not examined in depth as were the other methods in this article, there is no obvious improvement relative to the other iterative methods, and frequent catastrophic failures. At least one drop-in replacement for gauss2dfit.pro is available for IDL, which uses Levenberg−Marquardt least-squares optimization.[16]

Systematic Deviations. When systematic deviations from the true result are known to hold, one can adjust for this. Systematic deviations are shown in Figure 2 for circular Gaussians and in Figure 3 for elliptical Gaussians. So long as the signal-to-noise ratio is known, it is possible to determine this systematic deviation and to compensate for it as follows. The background noise level was already determined, so for each object, the approximate signal-to-noise level is known, as the fitting itself determines the signal. At the same time, for cases where compensation involves multiplying by a constant that depends on the signal-to-noise ratio, the uncertainty must be similarly multiplied. All precisions determined in this article have been compensated this way.

The possibility of pixel-locking is not included here, though previous analysis shows this to be a significant problem for the moment method but not for iterative optimization methods.[13] Tests on our part confirmed that pixel-locking is not significant for our weighted regression method. As anticipated on grounds of symmetry, no systematic offset is observed in position. Similarly, no systematic offset is observed in the amplitude for the three methods capable of determining it.

In contrast, bias is observed regarding the width of the Gaussians (Figure 2a) and also regarding the total brightness (Figure 2b). Bias using the iterative optimization and moment methods is comparatively negligible, while nine-point regression shows more bias, and our weighted regression method shows significantly more, particularly as the signal-to-noise ratio decreases.

Equivalent relations hold for elliptical Gaussians, shown in Figure 3, with the exception of the nine-point regression. For elliptical Gaussians, the data shown here refer to a width in x of 1 pixel, and a width in y of 1.5 pixels. As noted earlier, the nine-point regression is optimized for a width of 1 pixel and is increasingly inaccurate otherwise, as demonstrated in Figure 3b. As the total brightness (Figure 3c) depends upon the width, the effect propagates to inaccuracy regarding brightness.

Circular Gaussian Precision. In Figure 4, precision is plotted as a function of signal intensity at fixed background noise level.

Position. Figure 4a shows that subpixel resolution is easily obtained with all methods, but that the resolution depends on the analysis method. The iterative optimization method provides the best results; however, our weighted regression method achieves nearly the same level of precision. When the signal-to-noise ratio is at least 30, a level not uncommon for even single-molecule fluorescent dyes, the precision of weighted regression is only 5% worse. Even for a signal-to-noise ratio of only 20, weighted regression is only 15% off of the level of the iterative optimization. For ratios above 100, the results for the two are virtually indistinguishable. In contrast, the precision of nine-point regression is uniformly at least 50% worse, even for ratios above 500. The moment method demonstrates the worst resolution, never giving a better resolution than approximately 0.01 pixels. This is not unexpected, as the moment method suffers from flaws

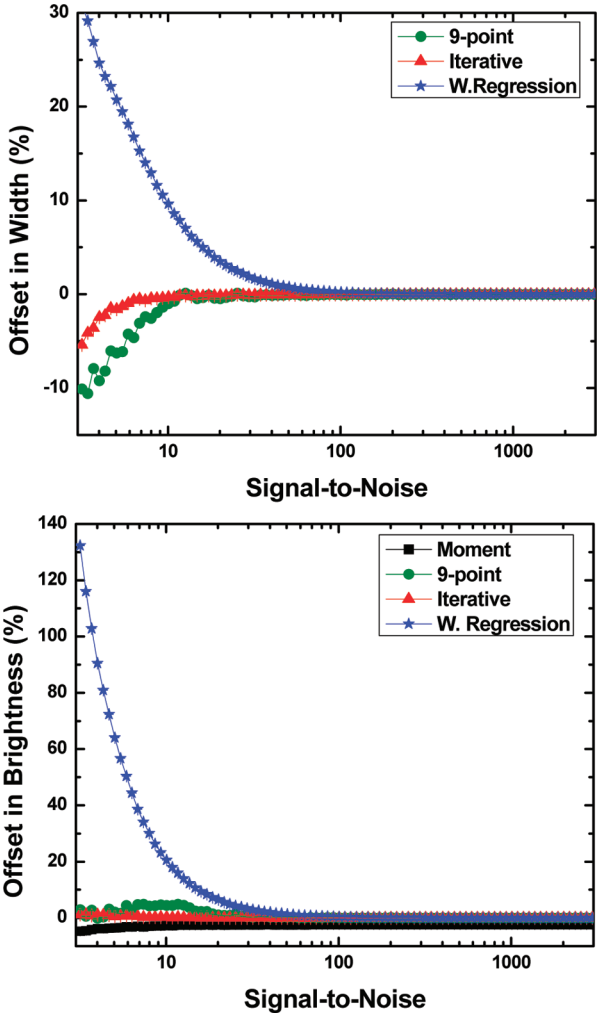


Figure 2. Offset values plotted as a function of the ratio of the signal-to-noise for circular Gaussian intensity distributions. Signal is considered to be the amplitude of the Gaussian intensity distribution. Note that once this systematic deviation has been determined, it can be compensated if desired. (a) The systematic offset in width for circular Gaussian intensity distributions obtained using the three methods that determine it: nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). (b) The systematic offset in the total brightness for circular Gaussian intensity distributions, compared using four methods: the moment method (black squares), nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). Symbols are the same in both panels.

including bias toward the center of the image, and has significant problems at low signal-to-noise.[13]

Width. Figure 4b shows the uncertainty of the width of the Gaussian as a function of the signal-to-noise. Here the comparison is slightly more subtle, with iterative optimization and weighted regression each faring well. At signal-to-noise ratios above 50, those two methods are indistinguishable. As the signal-to-noise ratio decreases, iterative optimization pulls ahead, reaching a lead of nearly 13% at a signal-to-noise ratio of 8. Beyond that, weighted regression narrows the lead, surpassing iterative optimization at a signal-to noise ratio of 5. The probable reason is that, as this behavior parallels with the cases when weighted regression begins to fail to fit some objects, the hardest to fit objects are discarded by weighted regression, whereas iterative optimization fit them but imprecisely. The method of nine-point regression is not even in contention, as it is generally at least

(16) http://www.physics.wisc.edu/~craigm/idl/fitting.html.

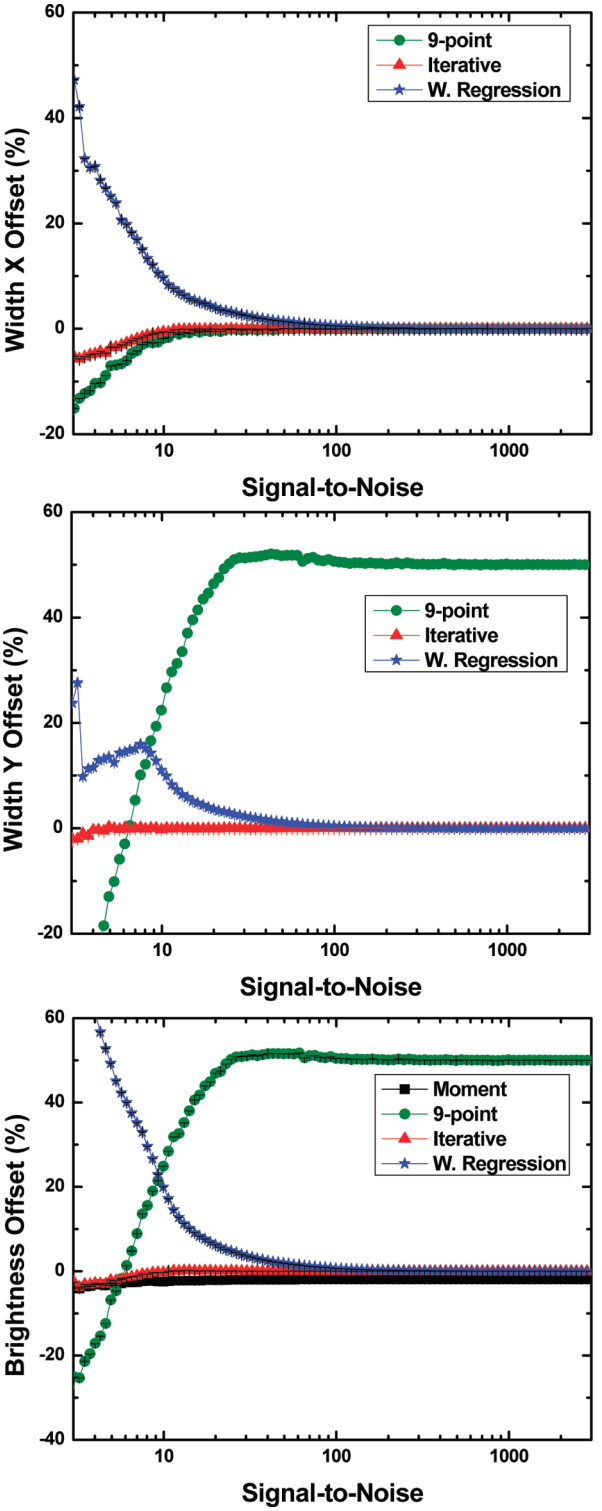


Figure 3. Offset values plotted as a function of the ratio of the signal-to-noise for elliptical Gaussian intensity distributions. Signal is considered to be the amplitude of the Gaussian intensity distribution. (a) The systematic offset in width for elliptical Gaussian intensity distributions for the three methods that determine it: nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). (b) The comparable systematic deviation in the width y. (c) The systematic offset in the total brightness for elliptical Gaussian intensity distributions for four methods: the moment method (black squares), nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). Symbols are the same in all panels.

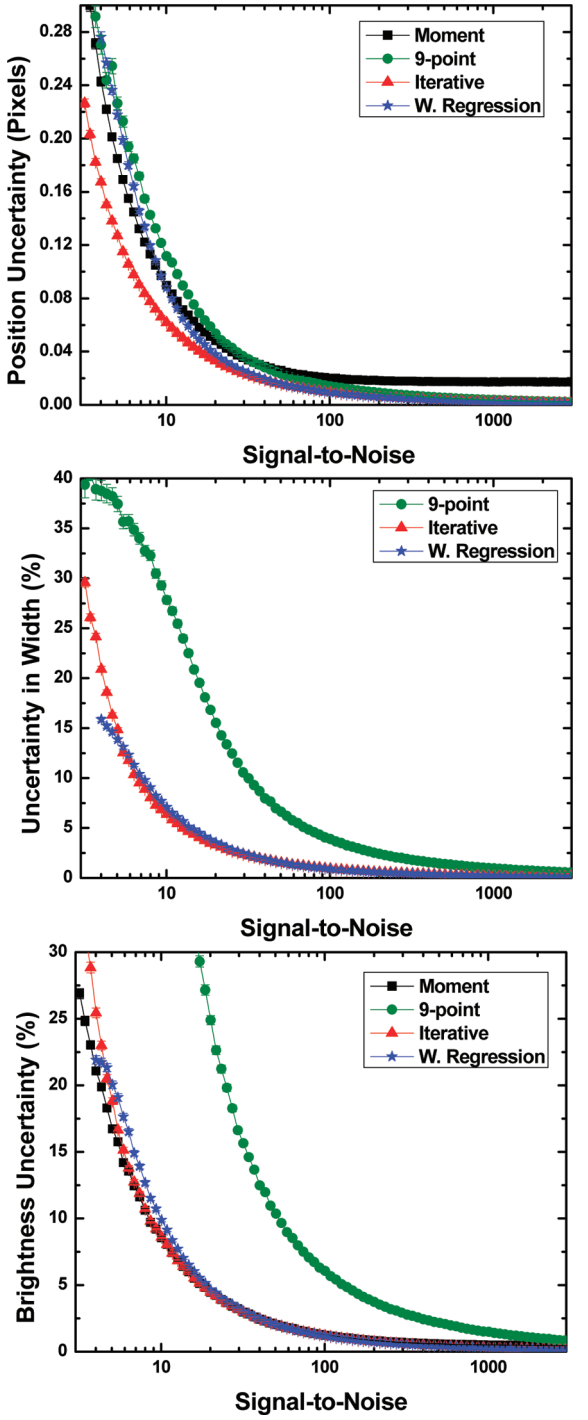
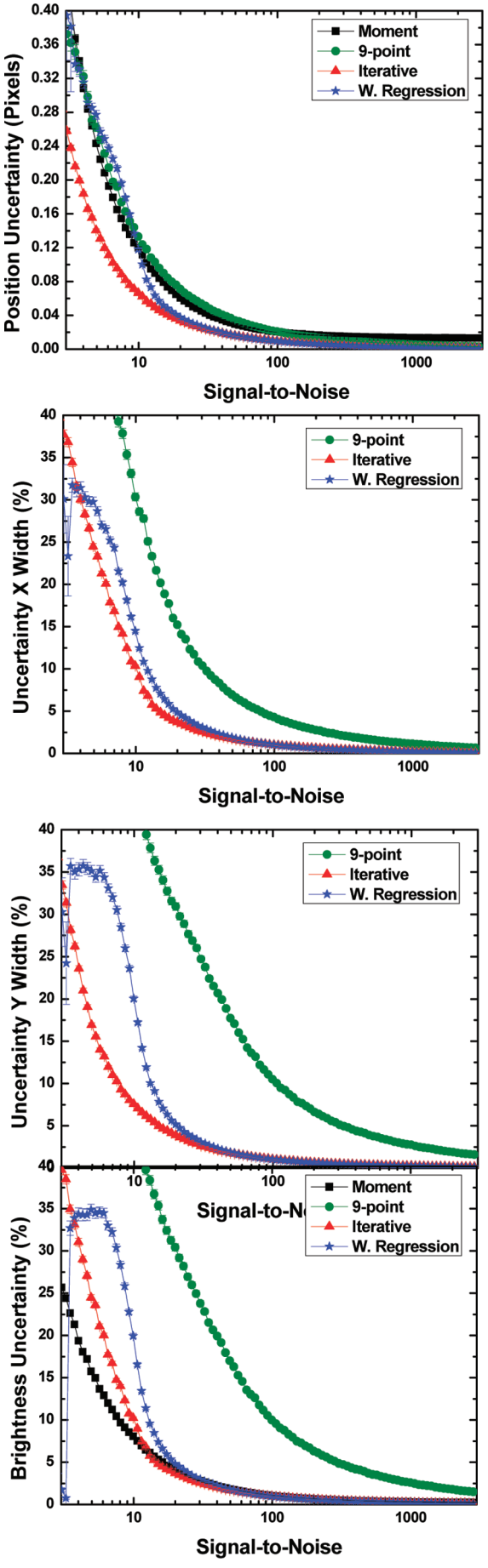


Figure 4. Precision and brightness uncertainty plotted as a function of the signal-to-noise ratio for circular Gaussian intensity distributions. The values shown here incorporate the corrections needed to compensate for systematic deviations. (a) The uncertainty in locating the center of a circular Gaussian intensity distribution for the moment method (black squares), nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). (b) The uncertainty in the determination of the width of the Gaussian intensity distribution. (c) The uncertainty in the determination of the total brightness. Symbols are the same in all panels.

4 times less precise. The moment method is not capable of determining the width of a distribution.

Total Brightness. Figure 4c shows the total intensity of the peak, which is the signal strength integrated over all pixels. In



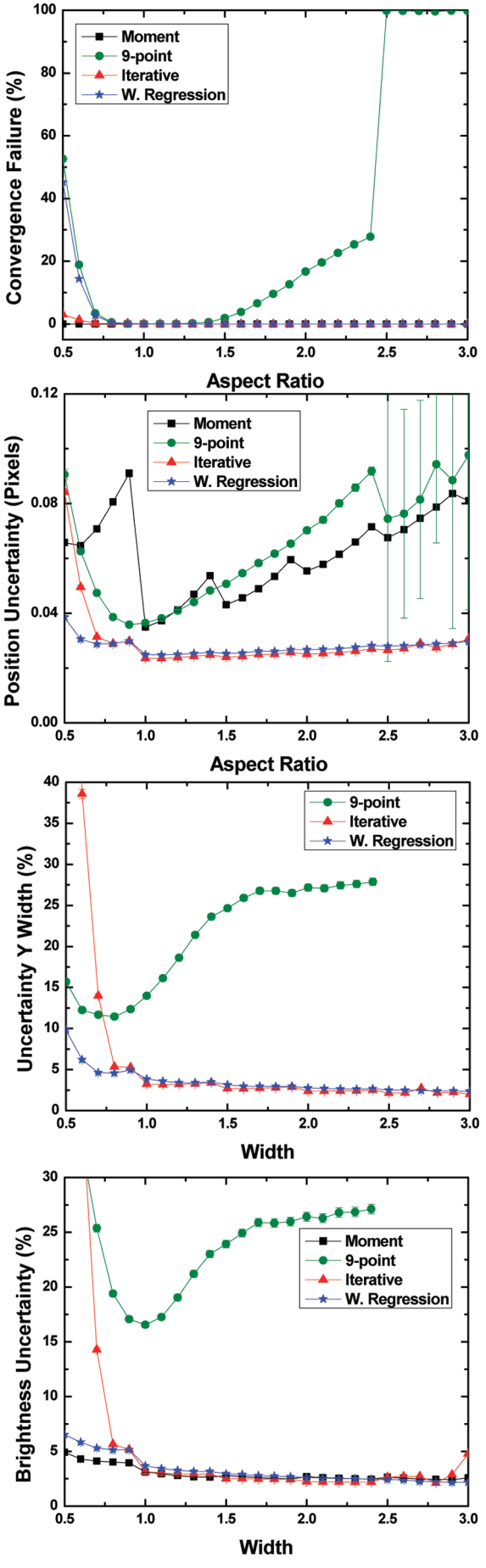
reality, the fourth independent variable is the amplitude, not the total brightness, as the total brightness depends upon both the amplitude and the width. However, the total brightness is commonly used in tracking, and for comparison purposes is more convenient, as this can be determined for all four methods, whereas amplitude as such is not determined by the moment method. While not shown separately, precision in amplitude is virtually equivalent for iterative optimization and weighted regression regardless of the signal-to-noise ratio. As such, the differences here between iterative optimization and weighted regression stem almost entirely from the difference in precision of the width. With respect to brightness, for the first time greater precision can sometimes be obtained with a method other than iterative optimization. Anywhere below a signal-to-noise of about 20, the moment method is the most precise. Above that value, both iterative optimization and weighted regression surpass it by an ever-increasing margin. As a result, since the moment method is computationally the simplest, for any signal-to-noise ratio below 20, the moment method should be run in addition to any other method employed, and the brightness from the moment method used. Nine-point regression is by far the least precise, being generally at least 5 times less accurate than the other methods.

Elliptical Gaussian Precision. Results for elliptical Gaussians, shown in Figure 5, are similar. While simulations were run for a wide range of aspect ratios, the representative results displayed here are for widths of 1 and 1.5 pixels for x and y, respectively. As usual, the weighted regression method asymptotically approaches the precision of iterative optimization as the signal-to-noise increases. The nine-point regression method reveals further troubles, due to its reliance on the width of the Gaussians being approximately 1. Recall that our weighted regression method fails for a significant fraction of objects with a signal-to-noise ratio below 8, so results obtained below that level likely only represent the cases that are easiest to fit, hence the relative improvement of weighted regression at those levels.

When examining the effect of varying the aspect ratio at a fixed signal-to-noise ratio, for the most part the relative results are independent of the aspect ratio. Results at a fixed signal-to-noise ratio of 20, with a width in x of 1 pixel and varying the width in y, are shown in Figure 6. All methods other than the simplest, the moment method, begin to break down when any width decreases below 1 pixel, sometimes from outright failing to determine the parameters and sometimes from showing rapid decrease in precision (Figure 6a). In general, as the aspect ratio increases, the precision increases slightly, as more pixels are available to be fit, but the increases in precision are comparable for most methods. The nine-point regression method, unsurprisingly, is the major exception, clearly demonstrating an optimum width of 1 pixel. Additionally, iterative optimization exhibits higher uncertainty as the aspect ratio approaches 3.

On a side note, it is also helpful to remember that, because of the definition of signal-to-noise ratio employed in this paper, peak amplitude relative to background noise, when the total number of

Figure 5. Precision and brightness uncertainty plotted as a function of the signal-to-noise ratio for elliptical Gaussian intensity distributions. (a) The uncertainty in locating the center of an elliptical Gaussian intensity distribution for the moment method (black squares), nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). (b) The uncertainty in determining the width in x, which was set at 1 pixel. (c) The uncertainty in determining the width in y, which was set to be 1.5 pixels. (d) The uncertainty in determining the total brightness. Symbols are the same in all panels.



counts is held constant but one of the widths changes, the signal-to-noise ratio varies accordingly.

Airy Spot. 2D Gaussians are nearly universally employed as substitutes for the full Airy spot pattern for purposes of image analysis, and simulations confirm that for the most part, differences are negligible. The uncertainties at low signal-to-noise ratio are virtually indistinguishable for width, position, or amplitude, regardless of whether the data being fit is a 2D Gaussian or an equivalent Airy spot, differing by only a few percent, roughly the accuracy of these simulations. Intuitively, this makes sense, as at low signal-to-noise ratios, the differences between a 2D Gaussian and an Airy spot are trivial relative to the noise level. In contrast, differences begin to manifest at higher signal-to-noise ratios of several hundred. Here, while the overall uncertainty is low, due to the high signal-to-noise ratio, the quality of the fits slowly diverges, with the methods not fitting the Airy spot quite as precisely as they do the 2D Gaussian. The differences are not very significant for the width or the amplitude, not even 20% larger even at a signal-to-noise ratio of 3000. The most significant differences are seen for position, where the uncertainty for the iterative optimization is nearly 80% greater for the Airy spot and the uncertainty for weighted regression is nearly 40% larger (Figure 7).

Computational Time. When only one single object is fit, all methods examined here run in under 20 ms on a 2.4 GHz Core 2 Duo. The processes are CPU limited, and use only one core. As such, computational time is not a consideration for some applications, if a limited number of objects are going to be fit. However, for many other applications, such as movies of how the positions of fluorescent particles change with time, it is not uncommon to fit hundreds of thousands or millions of objects and the needed computational time adds up significantly. For systems such as stochastic optical reconstruction microscopy (STORM), which often requires fitting millions of objects, Gaussian fitting is currently the most time intensive component of data analysis.[5] The improvement in computational time seen here for the weighted regression method would facilitate advances such as real-time analysis of such microscopy images.

Computational time does not depend significantly upon the signal-to-noise ratio. The moment method is the fastest, requiring 18 μs per object on average. Weighted regression and nine-point regression are comparable, requiring 83 and 104 μs, respectively. As expected, iterative optimization is costly, requiring 16 ms per object. Even when the slight performance boost that the quasi-Newton method would enjoy (roughly 30% faster) is considered, our weighted regression method is easily 2 orders of magnitude faster than iterative optimization. When considering elliptical Gaussians, the results are similar for most methods, with

Figure 6. Various quantities plotted as a function of the aspect ratio ($width_y/width_x$), at a fixed signal-to-noise ratio of 20 and width in x of 1 pixel. (a) The fraction of objects for which fitting fails to converge appropriately for the various methods for the moment method (black squares), nine-point regression (green circles), iterative optimization (red triangles), and weighted overdetermined regression (blue stars). (b) The uncertainty in the location of the center of the Gaussian intensity distribution; symbols same as above. The sawtooth pattern observed in the moment method occurs because whenever the aspect ratio increases by 0.5, the subregion of the image selected for fitting also increases. The increased uncertainty for iterative optimization and weighted regression may also correspond to the larger number of pixels used, which due to the lower signal-to-noise of many of the pixels, may effectively decrease the true signal-to-noise ratio. (c) The uncertainty in determining the width, y. (d) The uncertainty in determining the brightness. Symbols are the same in all panels.

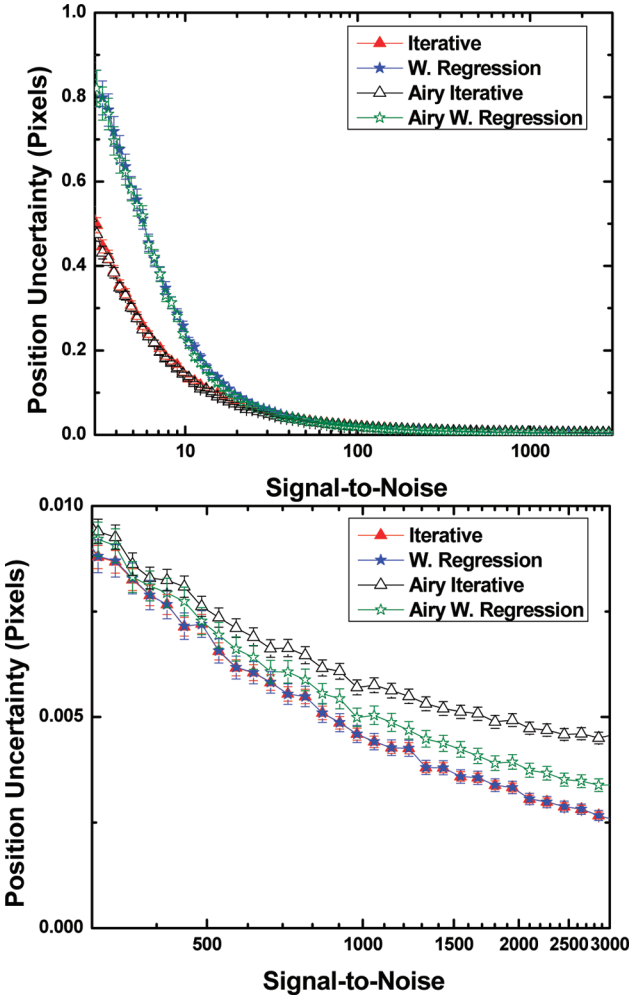


Figure 7. The uncertainty in position for simulations of both the Airy spot pattern and its equivalent 2D Gaussian approximation. (a) Abscissa is the signal-to-noise range, from 3 to 3000, for Gaussian iterative optimization (red triangles), Airy spot iterative optimization (black open triangles), Gaussian weighted regression (blue stars), and Airy spot weighted regression (green open stars). (b) Abscissa is expanded to more clearly show the region at highest signal-to-noise.

the exception that the addition of an additional parameter slows the iterative optimization significantly.

Conclusions

The method we developed here, weighted regression, represents an alternative to existing methods and offers vast improvement for a wide range of applications. Its computational efficiency is sufficient to make it a viable alternative even to the moment method for large data sets, with substantial improvements in precision. When precision is important, at all except the lowest signal-to-noise ratios, weighted regression represents a direct replacement for any other method, having equivalent or greater precision, and running more than 2 orders of magnitude faster than any method generating comparable results.

Weighted regression asymptotically approaches the precision of the leading method, iterative optimization, as the signal-to-noise ratio increases, yet even at low signal-to-noise, it does not deviate substantially from it. Further, even at low signal-to-noise, the slight loss in precision of weighted regression compared to iterative optimization may be worthwhile due to its hundred-fold reduction in computational time. If precision is of the utmost importance, weighted regression can be employed as a first-round fitting, completely determining one parameter, the amplitude, and providing near-optimum initial estimates for the other parameters. While this will still run substantially slower than weighted regression on its own, it will generally represent a 2-fold increase in speed over iterative optimization on its own.

Acknowledgment. This work was supported by the U.S. Department of Energy, Division of Materials Science, under Award No. DEFG02-02ER46019. For instrumentation, we acknowledge support from NSF-CBET-0853737, NSF-DMR-0907018, and NSF-DMR-0642573. S.M.A. acknowledges a Graduate Research Assistantship from the NSF.

Supporting Information Available: Sample Matlab implementations of the weighted regression algorithm developed in this paper. This material is available free of charge via the Internet at http://pubs.acs.org.