Scalable variational inference for super resolution microscopy

Ruoxi Sun
Columbia University

Evan Archer
Columbia University

Liam Paninski
Columbia University

Abstract

Super-resolution microscopy methods have become essential tools in biology, opening up a variety of new questions that were previously inaccessible with standard light microscopy methods. In this paper we develop new Bayesian image processing methods that extend the reach of super-resolution microscopy even further. Our method couples variational inference techniques with a data summarization based on Laplace approximation to ensure computational scalability. Our formulation makes it straightforward to incorporate prior information about the underlying sample to further improve accuracy. The proposed method obtains dramatic resolution improvements over previous methods while retaining computational tractability.

Introduction

Super-resolution microscopy techniques, such as STORM [1], PALM [2], or fPALM [3] imaging have quickly become essential tools in biology. These methods overcome the light diffraction barrier of traditional microscopy, thus enabling researchers to ask questions previously considered inaccessible (as a measure of impact, developers of these methods were awarded the Nobel prize in Chemistry in 2014). Given a sample treated with a fluorescent dye, the basic strategy is to stochastically activate fluorophores at a low rate, guaranteeing that only a sparse subset are activated at a given time. By repeatedly imaging the sample we obtain a movie wherein each frame reflects a random, sparse set of fluorophore activations. Then we exploit the sparsity of activations within each frame to localize the positions of the activated fluorophores; aggregating a long sequence of such point localizations then yields a super-resolved image (Fig. 1).

Many methods have been proposed to expand upon this basic idea, focusing upon improving localization performance within each individual frame [4]. For sparse recovery of a single frame, several modern techniques take a compressed sensing approach that exploits the true sparsity of the underlying fluorophore activations; these techniques result in a formulation as a sparse deconvolution problem [5, 6], providing scalable, fairly accurate reconstructions.

The critical message of this paper is that such standard approaches are sub-optimal because each frame is reconstructed independently, thereby discarding information that should be shared across frames. Intuitively, given $N-1$ reconstructed frames, we should have a good deal of prior information about the locations of fluorophores on the N-th frame, and ignoring this information will in general lead to highly suboptimal estimates. (This basic point has been made previously, e.g. by [7, 8]; we will discuss this work further below.)

Here, we propose a scalable Bayesian approach that properly pools information across frames and can also incorporate prior information about the image, leading to dramatic resolution improvements over previous methods while retaining computational tractability.

Model

At each frame i we observe an $L \times L$ fluorescence image $Y_i \in \mathbb{R}_+^{L \times L}$, and collect the sequences of N observed frames into the movie $Y = \{Y_i : i \in \{1, \ldots, N\}\}$. We model each observed frame Y_i as a noisy, blurred, low-resolution image,

$$Y_i \sim \text{Poisson}(AI_i); \tag{1}$$

here A is a matrix implementing convolution with a known point-spread function (PSF), scaling by the mean photon emission rate per fluorophore, and spatial downsampling; the high-resolution image $I_i \in \mathbb{R}_+^{D \times D}$ is a sparse matrix, zero except at the locations of fluorophores activated on frame i. In this application $L < D$. Below we will use the sparse repre-

Proceedings of the 20^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2017, Fort Lauderdale, Florida, USA. JMLR: W&CP volume 54.

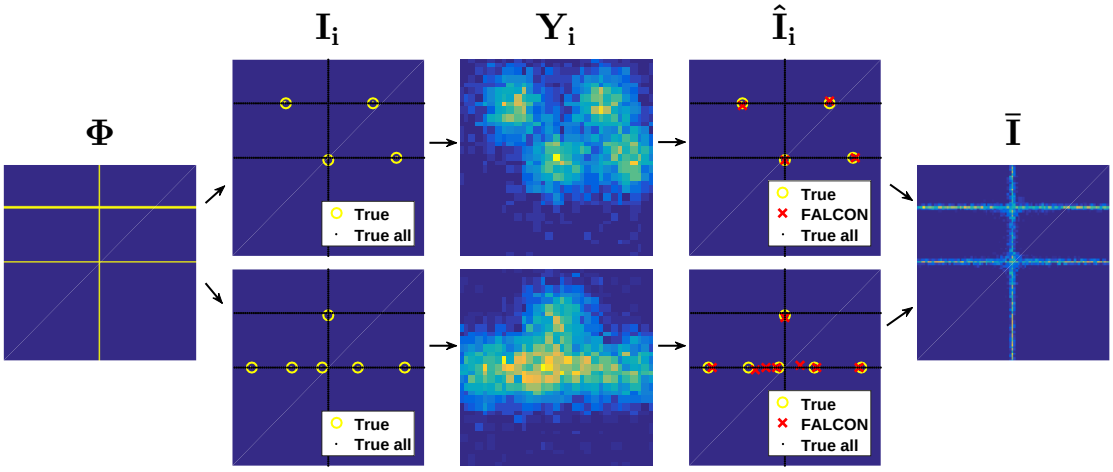

Figure 1: Overview of standard super-resolution microscopy. **Column 1**: The true fluorophore density matrix Φ. **Column 2**: I_i indicates the sparse subset of fluorophores (yellow circles) activated on frame i from Φ, which is also plotted as background (black dots); two independent sample frames shown here (top and bottom). **Column 3**: Y_i are the observed camera images on these two frames, formed by blurring and downsampling the corresponding I_i and adding Poisson noise. **Column 4**: $\hat{I}_i$ indicate the estimated locations of the active fluorophores on each frame, with the true I_i shown for comparison. The FALCON method [5] was used to compute the estimates here; note that estimator performance decreases in regions where the "bumps" in Y_i overlap significantly. **Column 5**: The standard approach to estimate $p\Phi$ is to simply average over multiple inferred frames, $\bar{I} = \frac{1}{N}\sum_i^N \hat{I}_i$.

sentation (m_i, F_i) for I_i: m_i denotes the number of active fluorophores in I_i and $F_i \in \mathbb{R}^{2m_i}$ denotes the vector of xy positions of these fluorophores. Note that multiple fluorophores can be active at the same location, so the entries of I_i are nonnegative integers; it is straightforward to extend our methods to the case that I_i can take arbitrary nonnegative real values, but we will suppress this case here for notational simplicity.

At each high-resolution pixel position (x, y), we model the activation of fluorophores by an inhomogeneous Poisson process with rate λ_{xy},

$$I_{i,xy} \sim \text{Poisson}(\lambda_{xy}), \tag{2}$$
$$\lambda_{xy} = p\Phi_{xy}, \tag{3}$$

and p is a scalar (typically under at least partial experimental control) that sets the fluorophore emission rate. The matrix $\Phi \in \mathbb{R}_+^{D\times D}$ specifies the density of fluorophores at each pixel location, and is the main object we aim to estimate; since λ and Φ are related by a constant (p), we will develop the inference methods below in terms of λ, as this leads to slightly simpler algebra.

This model can be extended to 3D [9, 10] and/or multispectral imaging [11], but for simplicity here we focus on 2D single-color imaging.

The above definitions lead to the joint probability distribution

$$P(Y, F, m, \lambda) \propto \prod_i^N \Big\{ P(Y_i|F_i, m_i)P(F_i, m_i|\lambda) \Big\} P(\lambda), \tag{4}$$

where m and F collect the N scalars m_i and vectors F_i, respectively; $P(\lambda)$ is a prior distribution on λ;

$$P(Y_i|F_i, m_i) = P(Y_i|I_i) = \prod_{x=1}^L \prod_{y=1}^L e^{-[AI_i]_{xy}} \frac{[AI_i]_{xy}^{Y_{i,xy}}}{Y_{i,xy}!}, \tag{5}$$

where we have used the equivalence between I_i and (m_i, F_i), and

$$P(F_i, m_i|\lambda) = P(F_i|\lambda, m_i)P(m_i|\lambda) \tag{6}$$
$$= \prod_{j=1}^{m_i} \left\{ \frac{\lambda_{F_i^j}}{\eta} \right\} \text{Poisson}(m_i|\eta), \tag{7}$$

where F_i^j denotes the xy position of the j-th active fluorophore in frame i, $\lambda_{F_i^j}$ is the value of the 2D function λ at location F_i^j, and we have abbreviated the normalizer $\eta = \sum_{x=1}^D \sum_{y=1}^D \lambda_{xy}$.

Inference

Now that the model and likelihoods have been defined, we can proceed to develop our estimator for the under-

lying fluorophore density image λ. We take a Bayesian approach, which requires that we approximate the posterior distributions of the unknown quantities (m_i, F_i) given the observed data Y_i. (Approximation methods are required here since this is a non-conjugate latent variable model; we cannot analytically integrate out the I_i variables.) A number of such approximation methods are available; for example, [12] recently developed MCMC methods to perform Bayesian inference in a similar model. However, these methods do not scale to the cases of interest here, where the number of frames N and pixels (D^2 and L^2) are often quite large.

Therefore we have developed a variational expectation-maximization (vEM) [13] approximate inference approach. As is standard, we need to choose a variational family of distributions q (these distributions will be used to approximate the true posterior), then write down the "evidence lower bound" (ELBO; this is a function of q and other model parameters), and then develop methods for tractably ascending the ELBO.

The most standard choice of q here (a fully factorized distribution over all latent variables, i.e., the activations I_i together with λ) does not lead to a scalable inference method, due to the very high dimensionality of $\{I_i\}$; in addition, this vanilla variational approximation is poor here because of strong posterior correlations between adjacent pixels in the I_i images. Instead, we exploit the sparse representation (m_i, F_i) for I_i: a more effective approach for approximating $p(F, m|Y, \lambda)$ was to use a simple point estimate for m (discussed below) and then, conditionally on m, a factorized ("mean-field") approximation for $p(F|m, Y, \lambda)$. Thus we approximate

$$P(F|\lambda, Y, m) \approx q(F) = \prod_i^N \prod_j^{m_i} q_{ij}(F_i^j). \tag{8}$$

We have factorized across frames i and active fluorophores j within each frame; here each $q_{ij} \in \mathbb{R}_+^{D\times D}$ is a probability density on the $D \times D$ grid that summarizes our approximate posterior beliefs about the fluorophore location F_i^j. In practice each q_{ij} will be extremely sparse, with very compact support, as we will discuss further below (Fig 2E).

The ELBO is given by:

$$\mathcal{L}(\lambda, q(F)) = \int q(F) \ln \frac{P(F, \lambda, Y|m)}{q(F)} dF \tag{9}$$

Our goal is to maximize $\mathcal{L}(\lambda, q(F))$ with respect to the distributions q_{ij} and image λ.

We will use a coordinate-ascent approach in which we update one q_{ij} or λ at a time; as discussed below, after one more approximation each update step can be computed cheaply (and parallelizes easily), and empirically only a few coordinate sweeps are necessary for convergence to a local optimum.

Laplace Approximation

Computing each q_{ij} update directly requires the computation of an $L \times L$ sum over the observed data image Y_i and several $D \times D$ sums over the other factors $q_{ij'}$, and since we have to compute these updates repeatedly, it is important to reduce the computation time in this inner loop. We have found that we can effectively summarize the data in each frame by using a conditional Laplace approximation to the likelihood. Specifically, we approximate

$$P(Y_i|F_i, m_i) \propto \mathcal{N}(F_i|\hat{F}_i, \hat{\Sigma}_i), \tag{10}$$

where the left hand side is the Poisson likelihood from eq. 5 and the right hand side denotes a multivariate normal density over $F_i \in R^{2m_i}$, with mean

$$\hat{F}_i = \arg\max_{F_i} P(Y_i|F_i, m_i) \tag{11}$$

and covariance inverse to the Fisher information J_i,

$$\hat{\Sigma}_i = [J_i]^{-1} = [-\nabla^2_{F_i} \ln P(Y_i|F_i, m_i)]|^{-1}_{F_i=\hat{F}_i}. \tag{12}$$

This Gaussian approximation to the Poisson likelihood is well-known to be accurate in the high-information regime where a sufficient number of photons are observed; see [14] for further discussion, and supplementary Fig. 8 for empirical evaluations of this approximation in the context of our simulations. (However, note that this Laplace approximation is not equivalent to assuming a Gaussian noise model with constant variance for Y_i; the Poisson noise model used here is significantly more accurate and consistent with the physics of shot noise.)

As we will see in the next subsection, this approximation allows us to replace the expensive sums noted above with evaluations of a much simpler $2m_i$-dimensional quadratic form. $\hat{F}_i$ and $\hat{\Sigma}_i$ serve as approximate sufficient statistics for Y_i, drastically reducing the size of the data that needs to be touched per iteration. In fact, the observed Fisher information matrix J_i is sparse - if fluorophores j and j' are sufficiently distant (more than a couple PSF widths apart) then $J_{i,(j,j')} = 0$, and this can be used to further speed up the computation. In practice, we compute J_i via automatic differentiation [15] and locally optimize eq. 11 numerically using an efficient initializer discussed further below.

The Laplace approximation also provides a convenient initialization for the q_{ij}'s: we simply set each q_{ij} to be the marginal (Gaussian) density of F_i^j in eq. 10, with

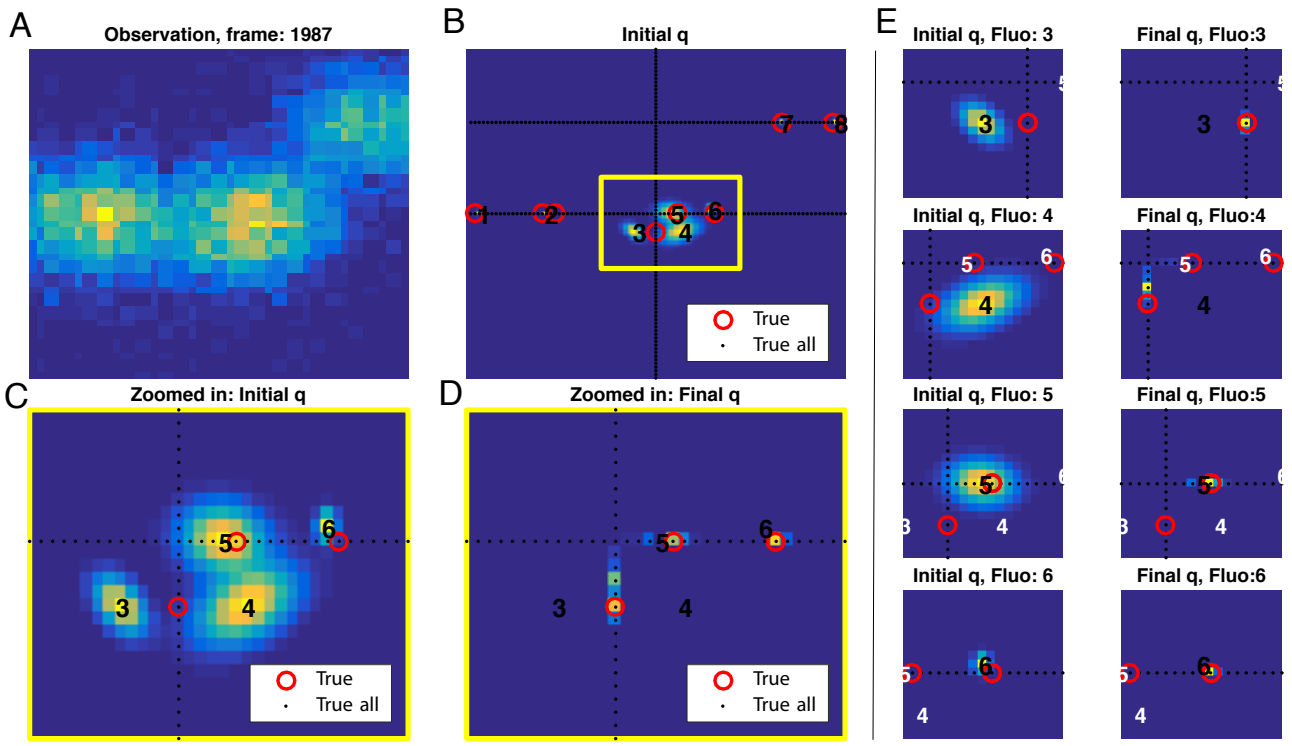

Figure 2: **Updating the factors q_{ij} in the E step**. **(A)** A single simulated observation frame Y_i. **(B)** 8 superimposed initial q_{ij} distributions for frame i, computed via Laplace approximations with means $\hat{F}_i^j$. True active fluorophores in I_i are labeled as red circles; true λ indicated by black dots; numbers indicate each $\hat{F}_i^j$ (ordering is arbitrary). Fluorophores 1,2,7, and 8 are relatively spatially isolated, with correspondingly large Fisher information (see supplementary Fig. 8) and so their initial q distributions are highly concentrated (and cannot even be seen beneath the red circles). In contrast, the closely overlapping PSF's of fluorophores 3,4,5, and 6 lead to broad initializations of q. **(C)** Zoom of yellow region outlined in **B**. **(D)** Final q_{ij}'s estimated by the vEM algorithm (same region as in panel **B**). Note that these have converged onto the region of positive λ (despite not having access to the ground truth λ), and the four original estimated fluorophore distributions have essentially converged near the 3 true active fluorophores in this region. **(E)** Further zoom showing details of each q_{ij} in **D**. The locations of other $\hat{F}_i^j$'s are indicated by white numbers. Left column: initial q_{ij}'s; right column: final q_{ij}'s. Again, the numbers indicate the $\hat{F}_i^j$ locations, which correspond to the peaks of the initial q_{ij}'s. Note the significant differences between the initial and final q_{ij}'s.

q_{ij} set to zero for all pixels sufficiently distant from $\hat{F}_i^j$.

Variational EM Algorithm

$$\ln P(F, \lambda, Y|m) \approx \sum_i^N \Big\{ \ln \mathcal{N}(F_i|\hat{F}_i, \hat{\Sigma}_i) + \ln \prod_j^{m_i} \frac{\lambda_{F_i^j}}{\eta} + \ln \text{Poisson}(m_i|\eta) \Big\} + \ln P(\lambda) + const. \quad (13)$$

Now we can put the pieces together and derive our vEM algorithm. The first step is to expand the ELBO eq.9, plugging in our factorized q, the Laplace approximation eq.10, and the likelihood eq.7 to arrive at

The vEM algorithm alternates between an E step (in which we optimize the ELBO wrt each q_{ij}, with λ and all the other q's held fixed), and an M step (in which we optimize the ELBO wrt λ with all the q's held fixed).

M step:

$$\hat{\lambda} = \arg\max_{\lambda} \mathcal{L}(\lambda, q(F)) \tag{14}$$

$$= \arg\max_{\lambda} \ \mathbb{E}_q[\ln P(F, \lambda, Y|m)] \tag{15}$$

$$= \arg\max_{\lambda} \sum_{i}^{N} \sum_{j=1}^{\hat{m}_i} q_{ij} \odot \ln \lambda - N\eta + \ln p(\lambda), \tag{16}$$

with $\odot$ denoting pointwise multiplication. If we use a flat prior for λ, the $p(\lambda)$ term can be dropped, and if we abbreviate $Q = \sum_i^N \sum_j^{\hat{m}_i} q_{ij}$, we have the solution $\hat{\lambda} = Q/N$. (Recall that $\lambda, \hat{\lambda}$, and $Q \in \mathbb{R}_+^{D\times D}$.) This is a natural generalization of the MLE for a discretized inhomogeneous Poisson process.

E step:

$$q_{ij} = \arg\max_{q_{ij}} \mathcal{L}(\hat{\lambda}, q) \tag{17}$$

$$\propto \exp\left\{ \mathbb{E}_{q_{\setminus ij}} \left[\ln P(F, \hat{\lambda}, Y|m) \right] \right\} \tag{18}$$

$$\propto \exp\left\{ \mathbb{E}_{q_{\setminus ij}} \left[\sum_{n}^{N} -\frac{1}{2}(F_n - \hat{F}_n)^T J_n (F_n - \hat{F}_n) \right.\right.$$

$$\left.\left. + \sum_{n}^{N} \sum_{k=1}^{m_n} \ln \hat{\lambda}_{F_n^k} \right] \right\} \tag{19}$$

$$\propto \hat{\lambda}_{F_i^j} \exp\left\{ -\frac{1}{2}(F_i^j - \hat{F}_i^j)^T J_i^{jj} (F_i^j - \hat{F}_i^j) \right.$$

$$\left. - \sum_{k\neq j}^{m_i} (F_i^j - \hat{F}_i^j)^T J_i^{jk} (\mu_{ik} - \hat{F}_i^k) \right\}; \tag{20}$$

here we have abbreviated $q_{\setminus ij} = q/q_{ij}$, J_i^{jk} is the 2×2 block of J_i corresponding to fluorophores j and k, and $\mu_{ik} = E_{q_{ik}} F_i^k$. Note that in the end, due to the Laplace approximation, the q_{ik}'s only enter the update above via their means, and that the updated q_{ij} is simply proportional to a Gaussian factor multiplied by $\hat{\lambda}$.

Finally, note that the effective support of each q_{ij} tends to shrink compared to the initialization (and this increasing sparsity can be readily exploited computationally); this makes sense, because our initialization (from the Laplace approximation) is based only on the likelihood of a single frame Y_i — when we incorporate the information from other frames (via $\hat{\lambda}$) the approximate posterior q_{ij} tends to become more concentrated. See Fig. 2 for an illustration.

Extensions and further details

In the developments above we have deferred several questions. How do we estimate m_i, the number of active fluorophores in each frame? How do we initialize the optimization problem eq.11 in the Laplace approximation for the likelihood? How do we make use of prior information $P(\lambda)$ in the M-step?

For the first two tasks mentioned above we exploit preexisting solutions. Specifically, we have found that the FALCON [5] method provides fairly good preliminary estimates of both the number and the location of fluorophores in each frame i; the former is used as m_i and the latter are used to initialize the optimization in eq. 11.

One of the major benefits of a Bayesian approach is that we can easily incorporate prior information about parameters of interest - in this case, λ. In principle it is possible to incorporate various sources of prior information about λ, but here we restrict our attention to the simplest case: in many cases the true underlying λ is known to be sparse, and we can exploit this fact to improve our estimates significantly. (Note that this sparsity constraint on λ is in addition to the fact that the images I_i are sparse, a fact that we have already exploited repeatedly. Also note that standard super-resolution approaches exploit the sparsity of each I_i — but since they simply average over the estimated $\hat{I}_i$ to obtain $\hat{\lambda}$, previous approaches have not attempted to further exploit the sparsity of λ.) An effective and computationally trivial approach is to apply the standard L1 "soft threshold" operator [16] to Q in the M-step (eq.16):

$$\hat{\lambda} = S_c(Q)/N = S_c\left[\sum_{i}^{N} \sum_{j}^{m_i} q_{ij}(F_i^j) \right]/N \tag{21}$$

where we define $S_c[x] = \max(x - c, 0)$; the threshold c can easily be chosen to achieve an a desired level of sparsity (typically set by prior knowledge, though cross-validation could be used here instead).

When active fluorophores are well-isolated in the image (i.e., the "bumps" corresponding to each active fluorophore are sufficiently distinguishable) then FALCON's estimates are typically accurate, and the corresponding entries of the Fisher information matrix J_i are large. However, when the bumps overlap then the Fisher information can decrease significantly (see supplementary Fig. 8 for an illustration) and the accuracy of m_i and the nearby fluorophore estimates $\hat{F}_i^j$ decrease. In this case we can achieve significantly improved accuracy by exploiting information from other frames, via the estimated $\hat{\lambda}$.

Thus the full algorithm proceeds as follows. To initialize we run FALCON on each frame and compute the Laplace approximation, then run vEM (restricting attention to the $\sim 50\%$ of frames on which fluorophore activation was sparsest, to improve localization accu-

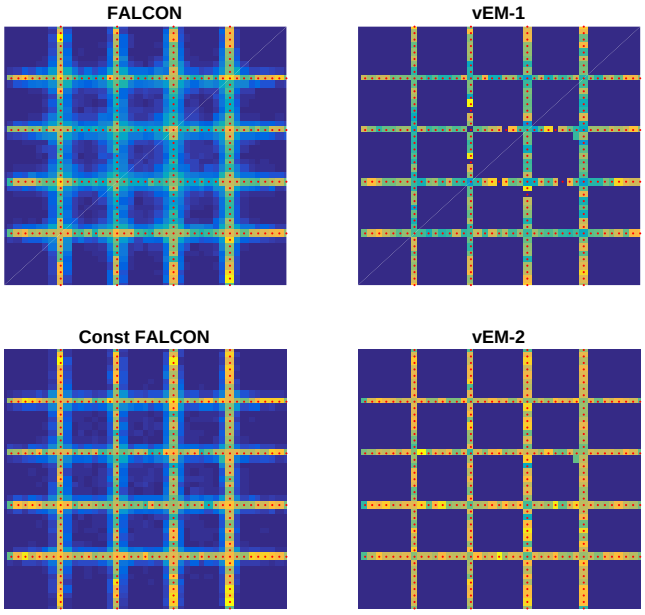

Figure 3: **Illustrating the algorithm steps on a simulated example**. Upper left: FALCON estimate given 5000 frames of data. Upper right: output after first run of vEM, based on the 2000 sparsest frames. This estimate is used to constrain FALCON to obtain a more accurate preliminary estimate $\hat{F}$ and $\hat{m}$, with all 5000 frames (lower left), and our final estimate using all 5000 frames after a second vEM run is shown in the lower right. The grid shape of the true underlying simulated image (red dots) is recovered essentially perfectly in the lower right; estimation noise (averaged over frames) blurs the true grid shape significantly in the left panels.

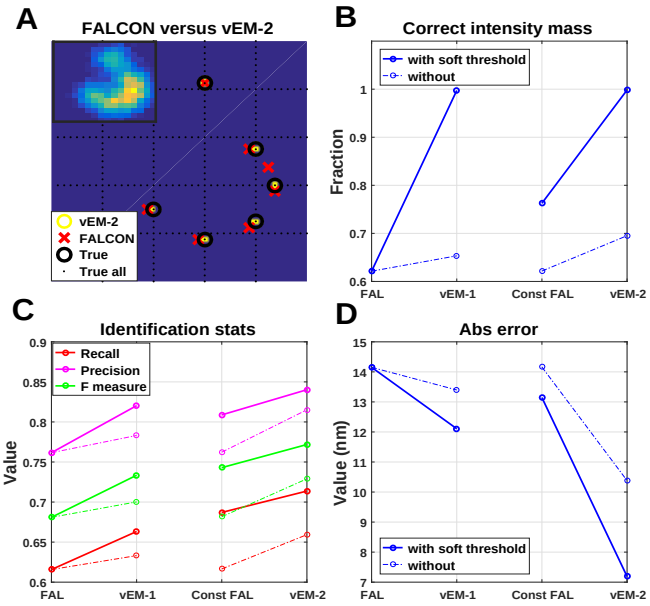

Figure 4: **Evaluating performance of each algorithm step**. **(A)** Estimates of active fluorophore locations in a single frame. Dotted black line indicates location of remaining (inactive) fluorophores. Note that vEM-2 estimates are more accurate than FALCON estimates. Algorithm steps and all simulation details follow Fig. 3, except in the vEM-1 step we compute results over all 5000 frames (not just 2000 frames, as in Fig. 3 upper right), for apples-to-apples comparison against the vEM-2 results here. Inset: observed data Y_i for this frame. **(B)** The percentage of $\hat{\lambda}$ contained within the true support after each algorithm step. Note that vEM leads to significant improvements over FALCON; applying soft-thresholding in the M-step also provides significant improvements. **(C)** Recall, Precision and F-measure of identified fluorophores (solid: soft-thresholded; dashed: no soft-thresholding), and **(D)** mean absolute error of fluorophore location estimates. In both cases, similar trends as in **(B)** are visible.

racy). Then we rerun FALCON incorporating information from the preliminary $\hat{\lambda}$ estimate to improve the estimates $\hat{F}$ and $\hat{m}$. FALCON uses an L1-penalized regression approach to obtain preliminary estimates of I_i from Y_i; it is straightforward to include a weighted L1 term where the weight is inversely proportional to $\hat{\lambda}$ to encourage the FALCON output to localize near regions of high $\hat{\lambda}$ (and to eliminate some spurious location estimates). Then we can use the resulting updated $\hat{\lambda}$-constrained FALCON estimates of the fluorophore locations to re-initialize eq. 11 on the subset of frames where the preliminary FALCON and vEM results disagree (updating these m_i as well), and proceed with further vEM iterations. This procedure can in principle be iterated, though we find in practice that one outer iteration typically suffices. In the inner loop, we found that just 5 vEM iterations were sufficient. See Fig. 3 for an illustration of each algorithm step's output.

Results

Figures 3-6 detail simulated comparisons between FALCON, a state-of-the-art super-resolution algorithm [5], and the vEM algorithm developed here. The

simulated image was a simple grid pattern; full simulation details are given in the Appendix. In Fig. 3 it is clear that the variational EM algorithm recovers the true grid support in this simulated example more accurately than does the FALCON algorithm. Fig. 4 quantifies the performance of the new proposed algorithm following each step illustrated in Fig. 3. Specifically, we examine the proportion of fluorophores whose estimated positions were recovered on the correct support of the true underlying grid image (panel B); the frame-by-frame precision and recall (and F measure, defined as the harmonic mean of precision and recall) of individual fluorophore estimates (panel C); and the frame-by-frame absolute error of individual fluorophore estimates (panel D). In each panel, we see that vEM leads to significant improvements over FALCON; applying soft-thresholding in the M-step also leads to significant improvements.

Figure 5 quantifies these results further, and adds comparisons to other competitive algorithms in the literature. Again the conclusion is that the vEM approach provides significantly more accurate estimates at little computational cost. Supplementary Figures 9 and 10 in the appendix show that this conclusion holds fairly uniformly over a wide range of PSF widths and average fluorophore densities, respectively.

Figure 6 provides a visual summary of one of the critical points of this paper: as the number of observed frames N increases, the vEM estimator continues to improve, and by $N = 5000$ is able to recover the true support of the underlying grid image with almost perfect accuracy. FALCON, on the other hand (as well as other approaches that estimate each frame independently), outputs estimates that appear blurry, due to noise in the estimated fluorophore locations, averaged over many frames — and this effective blur (and resulting loss of resolution) does not decrease asymptotically as N increases, since unlike vEM, FALCON does not exploit information from the $(N-1)$ other frames to improve estimation of individual frames.

Finally, Figure 7 shows a comparison of FALCON vs vEM applied to real data (see Appendix for full details). In this case the ground truth image is not available for comparison, but nonetheless the results are consistent with the simulated results described above: vEM leads to a sharper, better-resolved image than does FALCON.

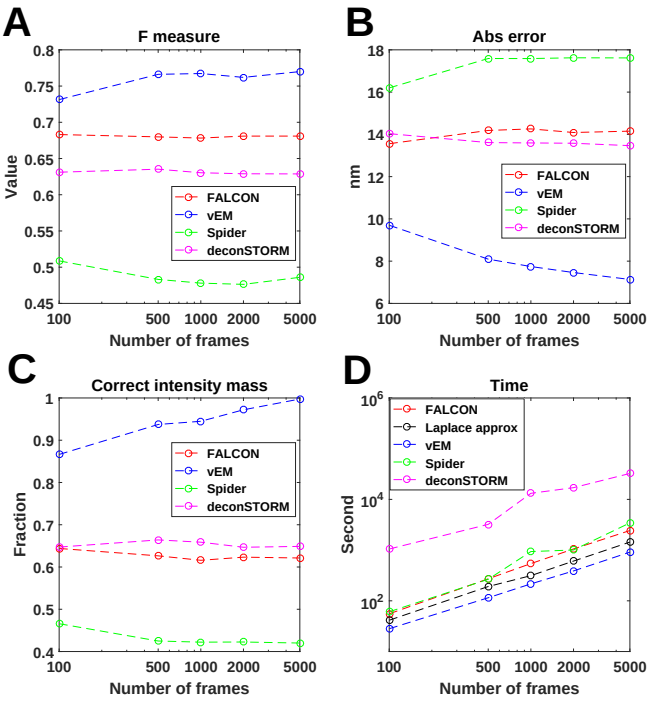

Figure 5: **Evaluation of vEM in comparison with FALCON [5], deconSTORM [7], and SPIDER [17] as a function of the number of observed frames** N. **A-C**: F-measure, mean absolute fluorophore estimation error, and fraction of fluorophore mass recovered correctly on the true underlying grid computed as in Fig. 4. The vEM approach outperforms the other state-of-the-art algorithms on all of these metrics. **D**: Computational time of each algorithm step. Our full algorithm runs FALCON (red curve), computes the Laplace approximation (black curve), then iterates vEM to convergence (blue curve), then repeats the whole process on at least a subset of frames, so overall speed is $\sim$ 2x slower than FALCON overall. The deconSTORM algorithm is relatively much slower here.

Discussion

We have introduced scalable Bayesian methods for improved estimation in super-resolution microscopy. By further extending the reach of these critical imaging methods, our approach can significantly impact a variety of biological applications. The hybrid vEM / Laplace-approximation / sparse-representation approach developed here is more generally applicable in other hierarchical sparse signal model applications [12]. Our methods exploit the insight that sharing information across image frames significantly improves accuracy — and this effect grows more powerful as the number of frames N increases.

Similar points have appeared previously in the super-resolution microscopy literature, notably in [7] and [8].

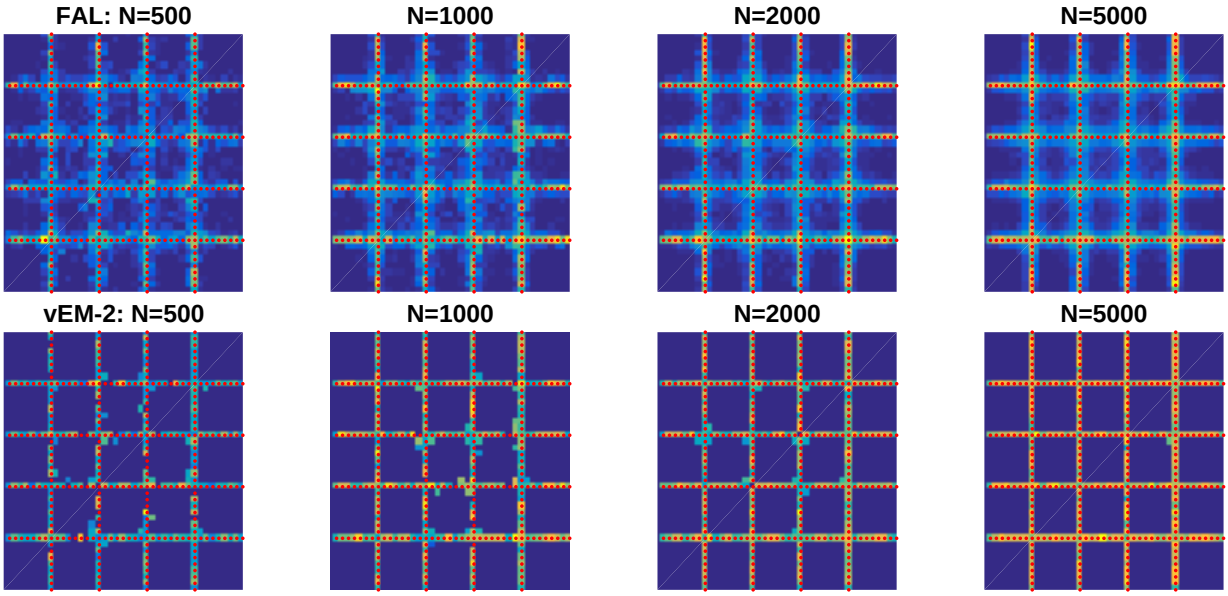

Figure 6: **Estimates with varying number of observed frames N**. Estimated $\hat{\lambda}$ images output by FALCON (upper panels) and vEM (lower panels) with 500, 1000, 2000, and 5000 simulated frames (with imaging parameters such as PSF width and fluorophore density p held fixed over all frames). Red dots indicate the ground truth grid. The grid recovery accuracy of the vEM algorithm continues to improve with N — recovering the underlying grid structure nearly perfectly when $N = 5000$ — but the estimation noise-induced blur in the FALCON estimate does not decrease with N.

The methods introduced in [8] are seldom used in practice on large-scale imaging data, due to prohibitive computational expense. The vEM methods we have introduced here are much more scalable (Fig. 5D); indeed, we were unable to obtain good results from the method used in [8] in a reasonable amount of computational time ($>$ 1 day) and so we did not show comparisons against this method here (see appendix for further discussion).

The deconSTORM method described in [7] (see also Fig. 5) attempts to improve upon simple Richardson-Lucy deconvolution by incorporating local information about the survival of active fluorophores from one frame i to the next $(i + 1)$. Our approach is orthogonal: we share information between frames I_i *globally*, through $\hat{\lambda}$. As we discuss in the appendix ("Markov model"), the vEM framework extends easily to handle these local correlations between fluorophores at frames i and $i + 1$. We observed that although incorporating these local correlations can slightly improve the recovery of individual fluorophores (Fig.11), the local Markov model does not significantly qualitatively improve the accuracy of the final estimated $\hat{\lambda}$ (Fig.12).

A final interesting and important direction for future work would be to extend some of the methods developed here to the case where the fluorophores are moving from frame to frame, in the context of single-particle tracking experiments [18].

vEM code is available here: `https://github.com/SunRuoxi/vEM`

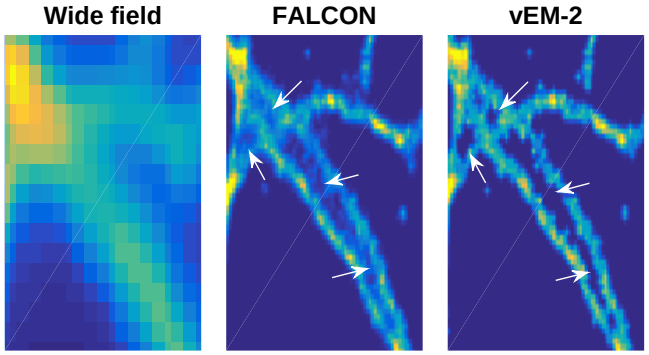

Figure 7: **Analysis of real tubulin image data**. Final resolved images output by FALCON and vEM-2 with 5000 frames. Wide field image, with all fluorophores turned on simultaneously, is given in first panel. Note that FALCON image is blurrier than the vEM-2 image, especially in areas of high fluorophore density, e.g., where multiple tubulin branches are close together, as noted by white arrows.

References

[1] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793–796, 2006.

[2] Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.

[3] Samuel T Hess, Thanu PK Girirajan, and Michael D Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272, 2006.

[4] Daniel Sage, Hagai Kirshner, Thomas Pengo, Nico Stuurman, Junhong Min, Suliana Manley, and Michael Unser. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature methods*, 2015.

[5] Junhong Min, Cédric Vonesch, Hagai Kirshner, Lina Carlini, Nicolas Olivier, Seamus Holden, Suliana Manley, Jong Chul Ye, and Michael Unser. FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data. *Scientific reports*, 4, 2014.

[6] Lei Zhu, Wei Zhang, Daniel Elnatan, and Bo Huang. Faster STORM using compressed sensing. *Nature methods*, 9(7):721–723, 2012.

[7] Eran A Mukamel, Hazen Babcock, and Xiaowei Zhuang. Statistical deconvolution for superresolution fluorescence microscopy. *Biophysical journal*, 102(10):2391–2400, 2012.

[8] Susan Cox, Edward Rosten, James Monypenny, Tijana Jovanovic-Talisman, Dylan T Burnette, Jennifer Lippincott-Schwartz, Gareth E Jones, and Rainer Heintzmann. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature methods*, 9(2):195–200, 2012.

[9] Bo Huang, Wenqin Wang, Mark Bates, and Xiaowei Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 319(5864):810–813, 2008.

[10] Hazen Babcock, Yaron M Sigal, and Xiaowei Zhuang. A high-density 3d localization algorithm for stochastic optical reconstruction microscopy. *Optical Nanoscopy*, 1(1):1, 2012.

[11] Mark Bates, Graham T Dempsey, Kok Hao Chen, and Xiaowei Zhuang. Multicolor super-resolution fluorescence imaging via multi-parameter fluorophore detection. *ChemPhysChem*, 13(1):99–107, 2012.

[12] Michel A Picardo, Josh Merel, Kalman A Katlowitz, Daniela Vallentin, Daniel E Okobi, Sam E Benezra, Rachel C Clary, Eftychios A Pnevmatikakis, Liam Paninski, and Michael A Long. Population-level representation of a temporal sequence underlying song production in the zebra finch. *Neuron*, 90(4):866–876, 2016.

[13] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *arXiv preprint arXiv:1601.00670*, 2016.

[14] Eran A Mukamel and Mark J Schnitzer. Unified resolution bounds for conventional and stochastic localization fluorescence microscopy. *Physical review letters*, 109(16):168102, 2012.

[15] Matthew J Weinstein and Anil V Rao. Algorithm: Adigator, a toolbox for the algorithmic differentiation of mathematical functions in matlab using source transformation via operator overloading.

[16] Francis Bach, Rodolphe Jenatton, Julien Mairal, Guillaume Obozinski, et al. Convex optimization with sparsity-inducing norms. *Optimization for Machine Learning*, 5, 2011.

[17] Siewert Hugelier, Johan J de Rooi, Romain Bernex, Sam Duwé, Olivier Devos, Michel Sliwa, Peter Dedecker, Paul HC Eilers, and Cyril Ruckebusch. Sparse deconvolution of high-density super-resolution images. *Scientific reports*, 6, 2016.

[18] Zhaolong Shen and Sean B Andersson. Tracking multiple fluorescent particles in two dimensions in a confocal microscope. In *CDC*, pages 6052–6057. Citeseer, 2009.

[19] Alex Small and Shane Stahlheber. Fluorophore localization algorithms for super-resolution microscopy. *Nature methods*, 11(3):267–279, 2014.