SUPER-RESOLUTION IMAGE RECONSTRUCTION FOR HIGH-DENSITY 3D SINGLE-MOLECULE MICROSCOPY

Jiaqing Huang[1,3], *Mingzhai Sun*[3], *and Yuejie Chi*[1,2]

[1]Department of Electrical and Computer Engineering, [2]Department of Biomedical Informatics
[3]Davis Heart and Lung Research Institute
The Ohio State University, Columbus, OH 43210, USA

ABSTRACT

Single-molecule localization based super-resolution microscopy achieves sub-diffraction-limit spatial resolution by localizing a sparse subset of stochastically activated emitters in each frame. Its temporal resolution, however, is constrained by the maximal density of activated emitters that can be successfully reconstructed. The state-of-the-art three-dimensional (3D) reconstruction algorithm based on compressed sensing suffers from high computational complexity and gridding error due to model mismatch. In this paper, we propose a novel super-resolution algorithm for 3D image reconstruction, dubbed TVSTORM, which promotes the sparsity of activated emitters without discretizing their locations. Several strategies are pursued to improve the reconstruction quality under the Poisson noise model, and reduce the computational time by an order-of-magnitude. Simulation results are provided to validate the favorable performance of the proposed algorithm.

Index Terms— Super-resolution microscopy, 3D image reconstruction, optimization

1. INTRODUCTION

The fluorescence microscopy has found numerous applications in the biological field. However, due to the optical diffraction, the resolution of a conventional fluorescence microscopy is limited to the Rayleigh limit, 0.61λ/NA, where λ is the wavelength of emission light and NA is the numerical aperture of the objective lens.

In the past few decades, several novel imaging techniques have been developed to break the diffraction limit by over an order of magnitude both in the lateral and axial directions [1, 2, 3, 4]. Among these techniques, single-molecule based super-resolution techniques, such as stochastic optical reconstruction microscopy (STORM) [3] and photo-activated localization microscopy (PALM) [4], improve the spatial resolution significantly by activating and localizing a sparse subset of emitters within the nanometer scale in each frame. Huang et. al. [5] extended STORM to three-dimensional (3D) imaging based on optical astigmatism. By placing a weak cylindrical lens into the optical path, the ellipticity of the emitter's point spread function (PSF) varies along the axial direction, making it possible to differentiate emitters at different axial locations. The final super-resolution image is then constructed by superimposing the reconstructed emitter locations from all frames. Therefore, the temporal resolution is limited by the number of frames needed to acquire a super-resolution image, making it desirable to develop image reconstruction algorithms that can handle higher emitter density per frame.

Compressed sensing based reconstruction (CSSTORM) [6, 7, 8] has been proposed as a robust and high performance algorithm for high-density super-resolution image reconstruction which can be applied for both 2D and 3D imaging. For each frame, CSSTORM first imposes a fine-grained grid to model the locations of activated emitters as a sparse signal in a discrete dictionary, of which the image on the camera becomes linear measurements, then solves a sparse recovery problem based on ℓ_1 minimization to invert the emitters' locations. However, this introduces an inevitable mismatch between the true continuous-valued location of the emitter, and its estimated location on the discrete grid [9]. To reduce the artifact by the mismatch error, the grid needs to be fine enough which results in an extremely large dictionary, making the computation very expensive. Moreover, heuristic post-processing steps are typically added to CSSTORM to enhance performance. Another algorithm called MempSTORM has been recently proposed to solve 2D super-resolution image reconstruction [10] based on 2D spectrum estimation techniques in the signal processing literature. MempSTORM directly estimates the continuous-valued location of the activated emitters without any discretization by transforming the image to the spectral domain. It is demonstrated to handle the same level of emitter density as CSSTORM with much faster computational speed. However, MempSTORM cannot be readily extended to 3D super-resolution image reconstruction. In these works, the noise is modeled as an additive bounded or Gaussian noise.

In this paper, we propose a novel super-resolution algorithm, dubbed TVSTORM, for 3D image reconstruction under the Poisson noise model, which is more appropriate for photon count data. The camera image is treated as an observation drawn from a Poisson distribution whose parameters are determined by the 3D PSF profile and the locations of activated emitters. The reconstruction is performed by minimizing a negative Poisson log-likelihood penalized by the total-variation norm [11] of the point source signal composed of the activated emitters. The total-variation norm can be viewed as a continuous analog of ℓ_1 norm for finite-dimensional vectors to promote emitter sparsity without discretizing their locations. We solve this optimization efficiently following a variant of the alternating descent conditional gradient method in [12]. Specifically, in each iteration, TVSTORM first selects a new point source and adds it to the current estimate, whose location is determined by a first-order linearization of the Poisson log-likelihood function over a coarse grid, and then refines the estimate of all included point sources by gradient descent using backtracking line search. Through numer-

This work is supported in part by NSF under the grant CCF-1527456 and by ONR under the grant N00014-15-1-2387.

ical experiments, TVSTORM demonstrates an order-of-magnitude improvement on the computational cost over CSSTORM due to the elimination of optimizing over a fine-grained grid. It also shows significant improvement on the localization accuracy over CSSTORM in terms of detection rate, false discovery rate and precision, without adding post-processing steps.

The rest of the paper is organized as follows. Section 2 describes the problem formulation of super-resolution image reconstruction. We present our proposed algorithm in Section 3 and numerical experiments are provided in Section 4. Finally, we conclude in Section 5.

2. PROBLEM STATEMENT

We begin by introducing the imaging system of single-molecule localization microscopy. In each frame, a sparse subset of emitters are activated. Let $\theta^{(i)} = [\theta_x^{(i)}, \theta_y^{(i)}, \theta_z^{(i)}, \theta_I^{(i)}] = [\theta_L^{(i)}, \theta_I^{(i)}]$ be the parameters for the i^{th} emitter, where $\theta_L^{(i)} = [\theta_x^{(i)}, \theta_y^{(i)}, \theta_z^{(i)}] \in \mathcal{S}$ are the coordinates in x, y and z dimensions, respectively,

$$\mathcal{S} = \Big\{\theta_L = [\theta_x, \theta_y, \theta_z] \Big| \theta_x \in (x_{\min}, x_{\max}),\ \theta_y \in (y_{\min}, y_{\max}), \theta_z \in (z_{\min}, z_{\max})\Big\},$$

and $\theta_I^{(i)} > 0$ denotes its intensity. Let $\Theta = \{\theta^{(1)}, \theta^{(2)}, ..., \theta^{(P)}\}$ be the set of parameters, where P is the total number of emitters. We can write the set of activated emitters $\chi = \chi(x, y, z|\Theta)$ as a sparse superposition of point sources, given as

$$\chi = \chi(x, y, z|\Theta) = \sum_{i=1}^{P} \theta_I^{(i)} \delta(x - \theta_x^{(i)}, y - \theta_y^{(i)}, z - \theta_z^{(i)}), \quad (1)$$

where $\delta(x - x_0, y - y_0, z - z_0)$ is a Dirac measure located at (x_0, y_0, z_0). For notational convenience, we also use $\chi(\Theta)$ to denote $\chi(x, y, z|\Theta)$. We denote the admissible set of χ as $\mathcal{G} = \{\chi = \chi(\Theta)|\Theta = \{\theta^{(i)}\}_{i=1}^{P}, P \in Z^+, \theta_L^{(i)} \in S, \theta_I^{(i)} >= 0, 1 <= i <= P\}$. Due to diffraction, the point source signal χ is low-pass filtered by the PSF of the microscopy before forming the image, whose shape is given as a 2D Gaussian function with the ellipticity determined by the location along the z direction:

$$K(x, y|x_0, y_0, z_0) = \frac{1}{2\pi\sigma_x(z_0)\sigma_y(z_0)} e^{-\left(\frac{(x-x_0)^2}{2\sigma_x(z_0)^2} + \frac{(y-y_0)^2}{2\sigma_y(z_0)^2}\right)}, \quad (2)$$

where $\sigma_x(z_0)$ and $\sigma_y(z_0)$ are the standard deviations in the x and y directions, given as [5]:

$$\sigma_x(z_0) = \sigma_{x,0}\sqrt{1 + \sum_{i=2}^{4} A_{x,i}\left(\frac{z_0 - c_x}{d_x}\right)^i}, \quad (3)$$

where $\sigma_{x,0}$, $A_{x,i}$, $i = 2, 3, 4$, c_x and d_x are known parameters of the defocusing function [7]. The expression for $\sigma_y(z_0)$ can be defined similarly.

The expected photon rate received at the $(u, v)^{\text{th}}$ camera pixel, denoted as $\mu(u, v)$, can be written as a convolution between the PSF in (2) and the point source signal in (1), integrated over the area of a pixel:

$$\mu(u, v|\chi) = \int_{u-\frac{1}{2}}^{u+\frac{1}{2}} \int_{v-\frac{1}{2}}^{v+\frac{1}{2}} (K * \chi)(x, y)\, dxdy, \quad (4)$$

where $*$ denotes convolution, and

$$(K * \chi)(x, y) = \sum_{i=1}^{P} \theta_I^{(i)} K(x, y|\theta_x^{(i)}, \theta_y^{(i)}, \theta_z^{(i)}). \quad (5)$$

Therefore, (4) can be rewritten as

$$\begin{aligned} &\mu(u, v|\chi) \\ &= \sum_{i=1}^{P} \frac{\theta_I^{(i)}}{4} \left[Q\left(\frac{u - \theta_x^{(i)} + \frac{1}{2}}{\sqrt{2}\sigma_x(\theta_z^{(i)})}\right) - Q\left(\frac{u - \theta_x^{(i)} - \frac{1}{2}}{\sqrt{2}\sigma_x(\theta_z^{(i)})}\right) \right] \\ &\times \left[Q\left(\frac{v - \theta_y^{(i)} + \frac{1}{2}}{\sqrt{2}\sigma_y(\theta_z^{(i)})}\right) - Q\left(\frac{v - \theta_y^{(i)} - \frac{1}{2}}{\sqrt{2}\sigma_y(\theta_z^{(i)})}\right) \right], \end{aligned} \quad (6)$$

where $Q(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

The number of photons hitting the camera at the $(u, v)^{\text{th}}$ pixel, denoted as $y(u, v)$, follows an independent Poisson distribution with the parameter $\mu(u, v|\chi)$, given as

$$\Pr(y(u, v) = z|\chi) = \frac{\mu(u, v|\chi)^z e^{-\mu(u,v|\chi)}}{z!}, \quad z \in \mathbb{Z}^+. \quad (7)$$

Denote the camera image as $y = \{y(u, v)\}$. The objective of super-resolution is to estimate the point source signal $\chi(\Theta)$, given the observed image y.

3. PROPOSED APPROACH

In this section, we describe the proposed TVSTORM algorithm for high-density 3D super-resolution imaging under the Poisson noise model explained in Section 2.

3.1. Theoretical framework

We first define the loss function, $\ell(y|\chi)$, as the negative Poisson log-likelihood of observing y given χ. According to (7),

$$\begin{aligned} \ell(y|\chi) &= -\log\left(\prod_u \prod_v \Pr(y(u, v)|\chi)\right) \\ &= \sum_u \sum_v \left(\mu(u, v|\chi) - y(u, v)\log(\mu(u, v|\chi))\right) + C, \end{aligned}$$

where C is a constant that does not depend on χ. As suggested in [13], for Poisson log-likelihood, it is advantageous to introduce a small offset $0 < \beta \ll 1$ to improve stability, where we modify $\ell(y|\chi)$ as

$$\ell(y|\chi) = \sum_u \sum_v \left(\mu(u, v|\chi) - y(u, v)\log(\mu(u, v|\chi) + \beta)\right). \quad (8)$$

On the other hand, we wish to motivate the underlying sparsity of the activated emitters. To this end, we define the total-variation norm of the point source signal χ, $\|\chi\|_{\text{TV}}$, whose measure-theoretic definition can be found in [11]. The total-variation norm can be seen as a generalization of the ℓ_1 norm for finite-dimensional vectors to the continuous space without imposing a discrete grid for the location of the emitters. For the point source signal in (1), $\|\chi\|_{\text{TV}} = \sum_{i=1}^{P} \theta_I^{(i)}$.

We propose to seek the point source signal χ that minimizes the loss function $\ell(y|\chi)$ penalized by its total-variation norm, given as

$$\hat{\chi} = \underset{\chi \in \mathcal{G}}{\text{argmin}}\ \ell(y|\chi) + \epsilon\|\chi\|_{\text{TV}}, \quad (9)$$

where ϵ is a regularization parameter that controls the trade-off between the fidelity to the observation and the sparsity of the emitters. We denote this approach as TVSTORM.

3.2. Implementation

Unfortunately, the algorithm (9) is in general non-convex for 3D imaging due to the likelihood term, and challenging to solve. Nonetheless, we develop an iterative algorithm that can be regarded as a variant of the alternating descent conditional gradient method in [12]. The description is given in Algorithm 1. TVSTORM is an iterative algorithm, where in each iteration, a new point source is first selected and added to the current estimate of χ, and then the estimate of χ is refined by gradient descent using backtracking line search. The algorithm stops when the intensity of the most recently added point source falls below a given threshold.

Algorithm 1 TVSTORM

Input Parameter: threshold γ
$t \leftarrow 0$
$\hat{\Theta}^{(0)} \leftarrow \varnothing$
$\hat{\chi}^{(0)} \leftarrow \chi(x, y, z|\hat{\Theta}^{(0)})$
repeat
    ▷ **SELECT**
    $\hat{\theta}_L^{(t+1)} \leftarrow \underset{\theta \in \mathcal{S}_{\text{coarse}}}{\operatorname{argmin}} \left\langle \frac{\partial \ell(y|\hat{\chi}^{(t)})}{\partial \mu(\hat{\chi}^{(t)})}, \mu(\chi(x, y, z|\theta)) \right\rangle.$
    $\hat{\Theta}^{(t+1)} \leftarrow \hat{\Theta}^{(t)} \cup \{\hat{\theta}_L^{(t+1)}\}$
    ▷ **REFINE**
    $\hat{\Theta}^{(t+1)} \leftarrow \text{REFINE}(y, \hat{\Theta}^{(t+1)})$
    $\hat{\chi}^{(t+1)} \leftarrow \chi(x, y, z|\hat{\Theta}^{(t+1)})$
    $t \leftarrow t + 1$
until $\hat{\theta}_I^{(t)} < \gamma$

Let $\hat{\chi}^{(t)} = \chi(x, y, z|\hat{\Theta}^{(t)})$ be the estimate of the emitter object at the t^{th} iteration, where $\hat{\Theta}^{(t)}$ represents the parameters of the point sources in $\hat{\chi}^{(t)}$. At the $(t+1)^{\text{th}}$ iteration, the SELECT step aims to add one point source (with parameter $\hat{\theta}^{(t+1)}$) to the emitter object of the previous iteration, $\hat{\chi}^{(t)}$, by minimizing the first-order Taylor series approximation of $\ell(y|\hat{\chi})$ around $\hat{\chi}^{(t)}$, which is

$$\begin{aligned} &\ell\left(y\middle|\hat{\chi}^{(t)} + \chi\left(x, y, z|\theta\right)\right) \\ \approx &\ell(y|\hat{\chi}^{(t)}) + \left\langle \frac{\partial \ell(y|\hat{\chi}^{(t)})}{\partial \hat{\chi}^{(t)}}, \chi(x, y, z|\theta) \right\rangle. \end{aligned} \tag{10}$$

Let $\mu(\chi) = \{\mu(u, v|\chi)\}$ denote the noise-free image generated from χ. It can be shown that [12]

$$\left\langle \frac{\partial \ell(y|\hat{\chi}^{(t)})}{\partial \hat{\chi}^{(t)}}, \chi(x, y, z|\theta) \right\rangle = \left\langle \frac{\partial \ell(y|\hat{\chi}^{(t)})}{\partial \mu(\hat{\chi}^{(t)})}, \mu(\chi(x, y, z|\theta)) \right\rangle, \tag{11}$$

which does not depend on the intensity. Therefore, the location of the new point source is selected as

$$\hat{\theta}_L^{(t+1)} = \underset{\theta \in S_{\text{coarse}}}{\operatorname{argmin}} \left\langle \frac{\partial \ell(y|\hat{\chi}^{(t)})}{\partial \mu(\hat{\chi}^{(t)})}, \mu(\chi(x, y, z|\theta)) \right\rangle, \tag{12}$$

where S_{coarse} is a coarse grid over $\mathcal{S}$. We only require a coarse grid since the locations will be refined afterwards.

The REFINE step aims to find the maximum likelihood estimate of $\chi^{(t+1)}$ with the number of point sources fixed by minimizing the loss function using iterative gradient descent over $\hat{\Theta}^{(t+1)}$. For each parameter $\theta \in \hat{\Theta}^{(t+1)}$, we first find the direction that decreases the loss function by calculating the partial derivative of the loss function with respect to θ, whose expressions can be derived using (6) and (8). The step size is then determined using backtracking line search to speed up convergence. Described fully in Algorithm 2, this step constitutes mainly the computational cost of TVSTORM.

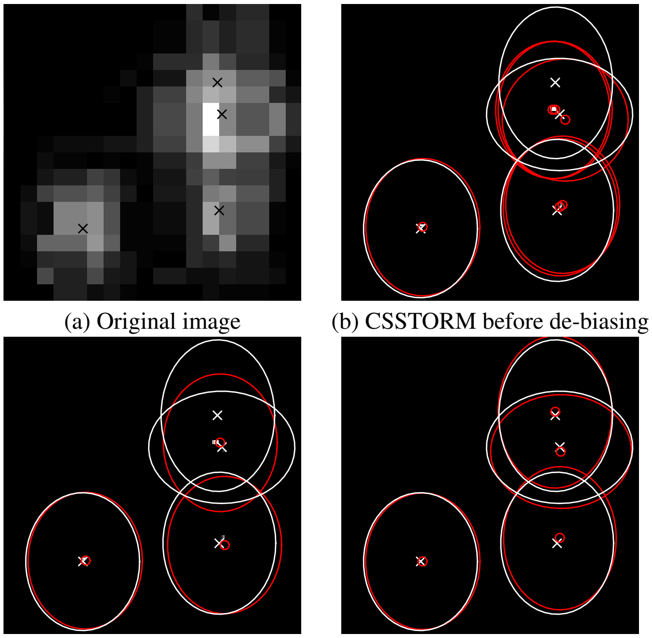
Algorithm 2 REFINE(y, Θ)

Input Parameters: $\alpha_0, \tau \in (0, 1), c \in (0, 1)$
repeat
    for every $\theta^{(i)}$ in Θ **do**
        for every $\theta_j^{(i)}$ in $\theta^{(i)}$ **do**
            $\alpha \leftarrow \alpha_0$
            $\widetilde{\Theta} \leftarrow \Theta$
            repeat
                $\widetilde{\theta}_j^{(i)} \leftarrow \theta_j^{(i)} - \alpha \frac{\partial \ell(y|\chi(\Theta))}{\partial \theta_j^{(i)}}$
                $\alpha \leftarrow \alpha \times \tau$
            until $\ell(y|\chi(\widetilde{\Theta})) \le \ell(y|\chi(\Theta)) - \alpha c \| \frac{\partial \ell(y|\chi(\Theta))}{\partial \theta_j^{(i)}} \|_2^2$
            $\Theta \leftarrow \widetilde{\Theta}$
        end for
    end for
until convergence

4. NUMERICAL EXPERIMENTS

We first examine the reconstruction quality using CSSTORM and TVSTORM on a single frame. We generate an image with four emitters that are randomly distributed in a 3D space of size 0.8 μm × 0.8 μm × 0.8 μm, with intensity of 300 photons each, as shown in Fig. 1 (a). For CSSTORM, an up-sampling factor of 8 in lateral direction and 9 in axial direction is used in the discretization. The output from CSSTORM typically requires post-processing such as de-biasing [7] in order to mitigate the gridding error, while TVSTORM does not include post-processing steps. Fig. 1 (b) and (c) show the image reconstruction from CSSTORM before and after de-biasing, where we use an ellipsoid to represent the spatial locations of an emitter with the center representing its lateral position and the shape representing its axial position. The reconstruction is shown in red, while the ground truth is shown in white. As seen from Fig. 1 (b), the reconstruction from CSSTORM before de-biasing contains many false positives. After de-biasing, nearby output emitters are clustered together but one emitter is missing, as shown in Fig. 1 (c). Contrarily, the reconstruction from TVSTORM, as shown in Fig. 1 (d), identifies all emitters correctly with high precision.

Next, to evaluate the average performance of TVSTORM, we generate a series of STORM images under different densities (0.75 emitter/μm^3 to 11.25 emitters/μm^3). The emitters are randomly distributed in a volume of 0.8 μm × 0.8 μm × 0.8 μm with the intensity set as 500. The images are then corrupted with Poisson noise. Fig. 2 compares the performance of TVSTORM with CSSTORM in terms of identified density, false discovery rate, precision and execution time with respect to the emitter density. Indeed, TVSTORM is able to detect more emitters with an improved precision while maintaining a lower false discovery rate than CSSTORM. Additionally, the execution time of TVSTORM is much faster than that of CSSTORM due to the elimination of a fine-grained grid during optimization.



(a) Original image (b) CSSTORM before de-biasing

(c) CSSTORM after de-biasing (d) TVSTORM

Fig. 1. Emitter localization using CSSTORM and TVSTORM. (a) Original image; (b) CSSTORM before debiasing; (c) CSSTORM after debiasing; and (d) TVSTORM.

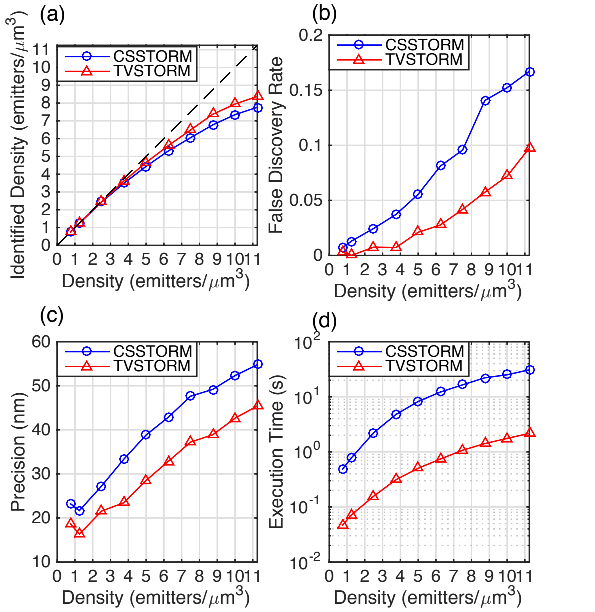


Fig. 2. Performance comparisons between TVSTORM and CSSTORM: (a) identified density, (b) false discovery rate, (c) precision and (d) execution time with respect to the emitter density.

5. CONCLUSION

In this paper, TVSTORM is proposed for 3D super-resolution image reconstruction, which is a penalized maximum likelihood estimator under the Poisson noise by the total variation norm of the activated emitters. TVSTORM avoids the intrinsic bias of CSSTORM due to gridding, and is computationally more efficient, with better detection rate, false discovery rate, and precision. Furthermore, TVSTORM can be easily adapted to 2D super-resolution image reconstruction or other single-molecule microscopy with different PSF configurations.

6. REFERENCES

[1] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *J. Microsc.*, vol. 198, no. 2, pp. 82–87, 2000.

[2] J. Fölling, M. Bossi, H. Bock, R. Medda, C. A. Wurm, B. Hein, S. Jakobs, C. Eggeling, and S. W. Hell, "Fluorescence nanoscopy by ground-state depletion and single-molecule return," *Nat. Methods*, vol. 5, no. 11, pp. 943–945, 2008.

[3] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods*, vol. 3, no. 10, pp. 793–796, 2006.

[4] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[5] B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," *Science*, vol. 319, no. 5864, pp. 810–813, 2008.

[6] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," *Nat. Methods*, vol. 9, no. 7, pp. 721–723, 2012.

[7] J. Huang, M. Sun, K. Gumpper, Y. Chi, and J. Ma, "3D multifocus astigmatism and compressed sensing (3D MACS) based superresolution reconstruction," *Biomed. Opt. Express*, vol. 6, no. 3, pp. 902–917, 2015.

[8] J. Min, S. J. Holden, L. Carlini, M. Unser, S. Manley, and J. C. Ye, "3D high-density localization microscopy using hybrid astigmatic/biplane imaging and sparse image reconstruction," *Biomed. Opt. Express*, vol. 5, no. 11, pp. 3935–3948, 2014.

[9] Y. Chi, L. L. Scharf, A. Pezeshki, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. on Signal Process.*, vol. 59, no. 5, pp. 2182–2195, 2011.

[10] J. Huang, K. Gumpper, Y. Chi, M. Sun, and J. Ma, "Fast two-dimensional super-resolution image reconstruction algorithm for ultra-high emitter density," *Opt. Lett.*, vol. 40, no. 13, pp. 2989–2992, 2015.

[11] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2014.

[12] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *arXiv preprint arXiv:1507.01562*, 2015.

[13] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: sparse poisson intensity reconstruction algorithms–theory and practice," *IEEE Trans. on Image Process.*, vol. 21, no. 3, pp. 1084–1096, 2012.