This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2022.3194719

SparseAlign: A Grid-Free Algorithm for Automatic Marker Localization and Deformation Estimation in Cryo-Electron Tomography

Poulami Somanya Ganguly[1,2], Felix Lucka[1], Holger Kohr[3], Erik Franken[3], Hermen Jan Hupkes[2], and K. Joost Batenburg[1,4]

[1]Computational Imaging, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
[2]Mathematical Institute, Leiden University, Leiden, The Netherlands
[3]Thermo Fisher Scientific, Eindhoven, The Netherlands
[4]Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

***Abstract*—Tilt-series alignment is crucial to obtaining high-resolution reconstructions in cryo-electron tomography. Beam-induced local deformation of the sample is hard to estimate from the low-contrast sample alone, and often requires fiducial gold bead markers. The state-of-the-art approach for deformation estimation uses (semi-)manually labelled marker locations in projection data to fit the parameters of a polynomial deformation model. Manually-labelled marker locations are difficult to obtain when data are noisy or markers overlap in projection data. We propose an alternative mathematical approach for simultaneous marker localization and deformation estimation by extending a grid-free algorithm first proposed in the context of super-resolution single-molecule localization microscopy. Our approach does not require labelled marker locations; instead, we use an image-based loss where we compare the forward projection of markers with the observed data. We equip this marker localization scheme with an additional deformation estimation component and solve for a reduced number of deformation parameters. Using extensive numerical studies on marker-only samples, we show that our approach automatically finds markers and reliably estimates sample deformation without labelled marker data. We further demonstrate the applicability of our approach for a broad range of model mismatch scenarios, including experimental electron tomography data of gold markers on ice.**

***Index Terms*—Mathematical super-resolution, parallel-beam tomography, conditional gradient method, marker-based alignment.**

I. Introduction

Cryo-electron tomography (cryoET) is a powerful imaging technique to resolve the structures of biomolecules and cellular components *in situ* using an electron microscope [1]. In recent years, advancements in detector technology and image processing methods have greatly improved the resolution of structure determination routines using cryoET, down to near-atomic resolution [2].

A typical cryoET workflow consists of tilt-series acquisition, tilt-series alignment and reconstruction, followed by post-processing steps such as per-particle reconstruction refinement, segmentation and sub-tomogram averaging [3], [4].

The image formation process in cryoET is as follows. A frozen sample is inserted into a transmission electron microscope (TEM) where it is irradiated with an electron beam, and the resulting transmitted beam lands on the camera to form a TEM image. For biological samples, the observed image contrast is mainly phase contrast because such samples are made up of light materials and thus are weak scatterers [5]. In contrast, gold markers are strong scatterers and show clear image contrast even under low-dose acquisition conditions. In order to obtain a tomographic *tilt series* (i.e. a series of projection images for consecutive angles), images of the sample are acquired at different view angles by tilting the sample with respect to the electron beam.

Aspects of cryoET that distinguish it from other CT setups are as follows. Firstly, the geometry of the experimental system limits the extent to which the sample can be tilted. Moreover, the increase in apparent sample thickness with increasing tilt allows projection images to only be acquired for a limited angular range in cryoET, usually in $[-60°, 60°]$, resulting in a *missing wedge* of information that is not available during reconstruction [6]. Secondly, cryoET samples are dose-sensitive, which limits the total dose during acquisition and leads to very noisy projection images when a large number are acquired. Thirdly, the sample undergoes local and global movements during the acquisition procedure, making it difficult to reconstruct with a constant sample assumption. For a detailed discussion on the mathematics of electron tomography we refer the reader to [7].

The acquired tomographic tilt series must be corrected for global and local sample motion during tilt-series acquisition [8]. Types of global motion include rotations and shifts of the sample with respect to the field-of-view (FoV) captured by the camera. Local motion includes sample deformation induced by the electron beam. In addition, a build up of surface charges due to irradiation can lead to apparent sample motion due to a microlensing effect [9]. When not corrected, sample motion leads to blurred reconstructions and poor resolution of the biological structures extracted by further post-processing [10]. *Tilt-series alignment*, the process of figuring out geometric relationships between projections in the tilt series, provides a way to correct for these effects so that the highest possible resolution can be achieved in subsequent reconstructions.

Beam-induced local sample deformation is a crucial limiting factor in high-resolution cryoET studies [11]. In particular, as shown in Fig. 1(a), compensation of local motion during alignment leads to sharper reconstructions and thus more reliable structure determination. In [11], the authors propose a method to extend currently used alignment methods with a sample deformation term that takes into account local sample motion induced by the electron beam. It has previously been observed that cryoET samples undergo "doming" motion, where the sample exhibits an upward deformation perpendicular to the sample plane (Fig. 1(b)). The authors of [11] model this motion using polynomial surfaces with coefficients that can be estimated as part of a minimization scheme. In addition to global shifts and rotations, the parameters of the doming model are fitted by solving a non-linear least-squares problem.

One of the drawbacks of the doming model approach is that it requires labelled marker locations in the tilt series as input, where the same marker has to be identified in all tilt images such that its locations can be connected to a trace. Markers are usually identified and traced in tilt-series images by template matching, a procedure that is prone to errors when the signal-to-noise ratio in tilt images is low, when markers cluster together or when they overlap in projection while being separate in 3D [8]. Other, state-of-the-art approaches in local sample deformation correction such as emClarity [12] and M [13] rely on detecting features from reconstructed tomograms and using these as fiducials, and are computationally expensive.

An additional disadvantage of the doming model method is the large number of parameters that must be estimated because no additional prior information on the deformation field is incorporated. Without smoothness constraints on the time evolution of the deformation field, the model allows deformation parameters to vary freely over the tilt series and does not penalize unphysical deformations.

Though not always appropriate, smoothness constraints on local sample motion are reasonable in the context of continuous-tilt cryoET (CTT) data collection, where thousands of very noisy projection images are captured continuously while the stage is tilted with a constant rotation speed [14]. This allows for a reduction in the number of doming model parameters.

We propose extensions to the doming model approach that make it possible to align tilt-series images without labelling markers in the tilt series. Taking inspiration from algorithms proposed in the context of single-molecule localization microscopy [15], we use a continuous formulation of the marker localization problem, which enables us to formulate an image-based loss and identify marker locations with a localization precision greater than the pixel spacing of the acquired tilt-series data. We equip the localization scheme with an additional deformation estimation routine and solve for the parameters of the doming model.

In addition, we incorporate a polynomial time dependence of the deformation field, which assumes smoothness of the local sample motion after global motion correction. This assumption is motivated by the fact that local sample motion is the result of positive-charge accumulation on the sample due to irradiation with a high-energy electron beam [10], [16]. As charge accumulation happens continuously and smoothly over the acquisition time, we can assume that local sample motion is also smooth. This assumption helps us reduce the number of deformation parameters by orders of magnitude. An important aspect of our approach, however, is that it is independent of the choice of deformation field parametrization.

To validate our proposed method, we apply it to simulated data in 2D and 3D as well as experimental data containing gold markers on ice. As the main focus of our paper is on testing the properties and robustness of our proposed method, we focus on simulation studies with ground-truth marker locations and deformation fields. In experimental studies, we restrict ourselves to data of gold markers on ice to disentangle the marker localization and deformation estimation problem from the later image reconstruction problem. We study the robustness of our approach with respect to noise, forward model mismatch and deformation model mismatch. We show that we are able to estimate deformation fields and marker locations with similar accuracy as the doming model approach without the need for labelled marker data, and that our method estimates deformation parameters accurately despite model mismatch.

This paper is structured as follows. In Section II, we review the mathematical formulation of the alignment problem and discuss a unifying framework for solving it. We derive the doming model approach in [11] as *one* possible choice of alignment method. We also present the main contribution of our paper: a method that localizes markers and estimates deformation fields without marker labelling. In Section III, we give details of the optimization techniques used to solve our extended problem. In Section IV, we describe the numerical experiments performed, and discuss our results on 2D and 3D simulated data as well as experimental data in Section V. We end our paper with a critical discussion of our approach and point to possible extensions in Section VI.

II. Mathematical Formulation

We consider an initial sample $u_0(\rho)$, with $\rho \in \Omega \subset \mathbb{R}^d$ ($d = 2, 3$ for simulated data and $d = 3$ for experimental data), which consists of two distinct components with non-overlapping supports:

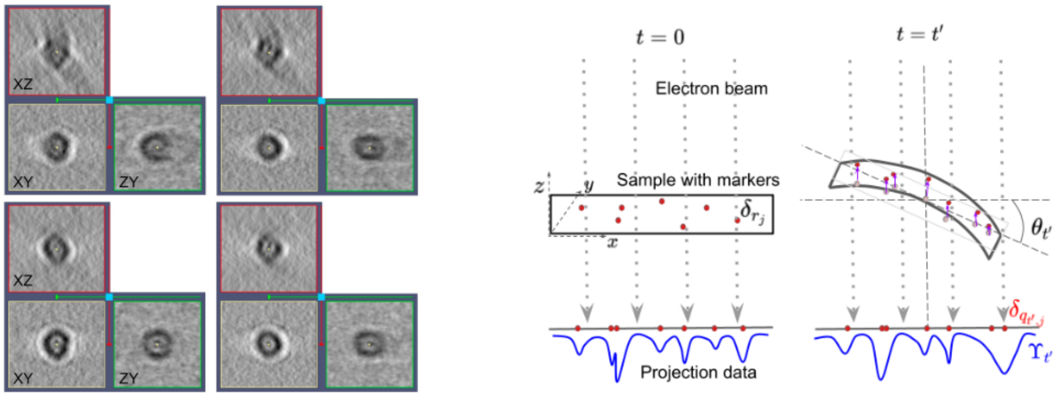
$$u_0(\rho) = u_0^m(\rho) + u_0^s(\rho), \tag{1}$$

where $u_0^m(\rho)$ represents markers and $u_0^s(\rho)$ represents the biological sample in the background.

This initial sample deforms over time, in the sense

$$u_t(\rho) = u_0(\rho - D_t(P)(\rho)) =: \mathcal{W}_{D_t(P)} u_0(\rho), \tag{2}$$

where $D_t(P, \rho) : \mathcal{P} \times \Omega \to \mathbb{R}^d$ is a time- and space-dependent deformation field parametrized by global parameters $P \in \mathcal{P}$. The action of this deformation field can be represented by a linear warping operator $\mathcal{W}_{D_t(P)}$. The global deformation parameters couple the reconstruction problems for individual markers. Later in this section we discuss appropriate parametrizations for the deformation field.



(a) Reconstructions of a gold marker

(b) Forward models

Fig. 1: (a) Reconstructions of a gold bead marker using (top two rows) standard alignment without sample deformation compensation and (bottom two rows) with sample deformation compensation. Images reproduced with permission from [11]. (b) Forward models used in SparseAlign and the doming model method. At $t = 0$ the sample with markers is not deformed. Projected marker locations (red dots) are convolved with a known shape function to yield projection data (blue line). As the sample is tilted, it undergoes doming deformation. At time $t = t'$, the change in marker locations caused by doming (purple upward arrows) leads to a change in the projection data.

Projection data Ψ_t of the deforming configuration are generated by applying the continuous Radon transform to $u_t(\rho)$:

$$\Psi_t = \mathcal{R}_{\theta_t} u_t(\rho) = \mathcal{R}_{\theta_t} \mathcal{W}_{D_t(P)} \left(u_0^m + u_0^s\right), \tag{3}$$

where θ_t is the projection angle and the Radon transform for $d = 2$ is defined as a line integral over rays:

$$\mathcal{R}_{\theta_t}[u](s) = \int_{l(s,\theta_t)} u(\rho)\, d\rho$$

$$l(s,\theta_t) = \{(x,y) \in \mathbb{R}^2 \,|\, x\cos\theta_t + y\sin\theta_t = s\}.$$

Projection in 3D for a parallel beam geometry, as in the case for cryoET, can be decomposed into a series of 2D projections [17].

The full tomographic data, obtained over discrete time points $t \in \{t_0, t_1, \ldots, t_T\}$ is a stack of individual projections:

$$\Psi := \begin{bmatrix} \Psi_0 \\ \Psi_1 \\ \cdots \\ \Psi_T \end{bmatrix} = \begin{bmatrix} \mathcal{R}_{\theta_1} \mathcal{W}_{D_0(P)} \\ \mathcal{R}_{\theta_2} \mathcal{W}_{D_1(P)} \\ \cdots \\ \mathcal{R}_{\theta_T} \mathcal{W}_{D_T(P)} \end{bmatrix} (u_0^m + u_0^s). \tag{4}$$

Solving the set of equations (4) when all the variables - u_0^m, u_0^s and D_t - are unknown amounts to solving a joint image reconstruction and alignment problem. Most approaches for solving the joint problem alternate between solving (4) for one of the three variables while keeping the others fixed. In such schemes, determining a good order for these updates is crucial.

As markers are designed to have a significantly higher contrast compared to the sample, we can often obtain reasonable first estimates for the marker configuration u_0^m and deformations D_t while ignoring the sample contribution. This corresponds to solving (4) by setting $u_0^s = 0$.

One way to parametrize the initial marker configuration u_0^m is to represent it using the continuous locations of markers at $t = 0$. Here we represent a single marker as a delta function at the location of its centre convolved with a fixed, known shape function; the marker configuration is then a sum of convolved delta functions in $\Omega \subset \mathbb{R}^d$:

$$u_0^m(x) = \sum_{j=1}^{M} \Big(G * \delta_{r_j}(\rho) \Big), \tag{5}$$

where r_j are the initial marker locations, M is the total number of markers and G is a known shape function, for instance a Gaussian.

For parallel beam projection, Theorem 1.2 in [17] states that:

$$\mathcal{R}_\theta(G * \delta_{r_j}(\rho)) = (\mathcal{R}_\theta G) * \left(\mathcal{R}_\theta \delta_{r_j}(\rho)\right) =: G_\theta^p * \left(\mathcal{R}_\theta \delta_{r_j}(\rho)\right). \tag{6}$$

Furthermore, the Radon transform of a delta function is a delta function in projection space:

$$\mathcal{R}_\theta \delta_{r_j}(\rho) = \delta_{A_\theta r_j}(s), \tag{7}$$

where $A_\theta \in \mathbb{R}^{(d-1)\times d}$ is a projection matrix that maps marker locations in configuration space to locations in projection space. We denote the resulting projected marker locations by $q_j := A_\theta r_j$.

We can assume that in contrast to the sample, markers are displaced over time, not deformed. Furthermore, when variations in the global deformation field D_t over the area covered by a marker are small, we can make the following approximation by commuting the deformation operator with convolution with the shape function:

$$\mathcal{W}_{D_t(P)}(G * \delta_r)(\rho) = (G * \delta_r)(\rho - D_t(P,\rho))$$
$$\approx G * \delta_r(\rho - D_t(P,\rho) = G * \delta_{r + D_t(P,\rho)}(\rho).$$

Thus, the deformed marker configuration is given by:

$$\mathcal{W}_{D_t} u_0^m(x) \approx \sum_{j=1}^{M} \Big(G * \delta_{r_j + D_t(P,r_j)}(\rho) \Big). \tag{8}$$

This assumption is accurate when the support of G is small and the deformation $D_t(P, \rho)$ is smooth over the support of $(G * \delta_{r_j})$. Setting $u_0^s = 0$ and inserting the ansatz above into (3) yields

$$\Psi_t = \mathcal{R}_{\theta_t}\mathcal{W}_{D_t(P)}u_0^m \approx \sum_{j=1}^{M}\left(G^p_{\theta_t} * \delta_{A_{\theta_t}(r_j+D_t(P,r_j))}\right) = \sum_{j=1}^{M}\left(G^p_{\theta_t} * \delta_{q_{t,j}}\right), \quad (9)$$

where

$$q_{t,j} = A_{\theta_t}(r_j + D_t(P, r_j)). \quad (10)$$

Using equation (9) amounts to localizing markers by matching their projection data $\Psi_t \in \mathbb{R}^{(N_\theta \times N_d)}$ (in 2D), where N_θ is the number of projection angles and N_d is the discretisation of the detector plane. A schematic of this forward model is shown in Fig. 1(b), where we indicate 1D projected data with blue lines.

In [11], the authors use projected marker locations over time as the input instead of image data (indicated with red dots in Fig. 1(b)) and use the following optimization problem for deformation estimation and marker localization:

$$\underset{r_j,P}{\text{minimize}} \quad \sum_{t=0}^{T}\sum_{j=1}^{M}\left\|\left(\tilde{q}_{t,j} - A_{\theta_t}(r_j + D_t(P, r_j))\right)\right\|_2^2. \quad (11)$$

Such an approach assumes that we can identify the projected marker locations $\tilde{q}_{t,j}$ directly, despite convolution with $G^p_{\theta_t}$. Here and elsewhere, we use symbols with a tilde (e.g. $\tilde{q}_{t,j}$) to denote measured data and symbols without a tilde (e.g. $q_{t,j}$) to denote model predictions.

Comparing equations (9) and (10), we find that for each t the dimensions of 2D data for (10) are $d \times M$ and those of the data for (9) are $N_\theta \times N_d$. Typical values for d, M, N_θ and N_d are $3, 20, 100$ and 4096, respectively, such that $d \times M = 3 \times 20$ and $N_\theta \times N_d = 100 \times 4096$, the latter being approximately 6000 times the former. Thus, (10) is a much lower-dimensional problem. Furthermore, the deformation field can be extracted from (10) in a more direct fashion as it directly describes the corresponding projected marker displacement, not the change in the projection image caused by it.

However, identifying markers robustly from data is not a trivial problem [8]. It involves solving an optimization problem of the form: $\text{minimize}_{q_{t,j}} \sum_t \|\tilde{\Psi}_t - \sum_j (G^p_{\theta_t} * \delta_{q_{t,j}})\|_2^2$. Marker labelling is generally performed using normalized cross-correlation-based schemes or template matching algorithms. Such methods are error-prone when projection data are noisy or when gold beads are occluded or cluster together in projection data. In such situations, users must manually annotate markers, or manually inspect and correct for incorrect and failed detection in one or more images in the tilt series. This manual intervention leads to time-consuming and subjective labelling.

To avoid solving the marker identification problem, we take a step back and start directly from (9). We solve for marker locations and the deformation field in a least-squares sense. In addition, we do not assume that we know the number of markers beforehand. The resulting optimization problem is as follows:

$$\underset{r_j,P,M}{\text{minimize}} \quad \sum_{t=0}^{T}\left\|\tilde{\Psi}_t - \sum_{j=1}^{M}\left(G^p_{\theta_t} * \delta_{A_{\theta_t}(r_j+D_t(P,r_j))}\right)\right\|_2^2. \quad (12)$$

The optimization problem above assumes a model for the markers, uses an image-based loss and does not need labelled marker locations like the problem in (11). In the following section, we discuss optimisation schemes for solving (12).

The deformation field D_t can be represented using different basis functions. If one uses localized basis functions, e.g. the B-spline basis functions often used in non-rigid image registration, one either needs a sufficiently dense sampling of the domain with markers or include suitable regularization constraints [18]. Global basis functions that are supported in the entire domain will only lead to a compact, low-dimensional description of the deformation field with sufficient accuracy if they are chosen based on *a priori* knowledge about the sample deformation.

In this paper, we use the global basis functions proposed in [11], where the beam-induced sample deformation is modeled with a set of polynomial surfaces. The parametrized sample deformation $D_t(P, r_j) := [D_{t,x}, D_{t,y}, D_{t,z}]$ is modelled with polynomials in (x, y, z) such that the deformation in each direction is given by

$$D_{t,k}(r, P) = \sum_{\substack{\alpha,\beta,\gamma \geq 0 \\ \alpha+\beta+\gamma \leq d_p}} \Big(P_{\alpha\beta\gamma}(t)\Big)_k x^\gamma y^\beta z^\alpha, \quad k \in \{x, y, z\}, \quad (13)$$

where $P_{\alpha\beta\gamma}$ are the coefficients of the polynomial and d_p is the degree of the polynomial. In [11], these polynomials are allowed to vary freely over the tilt series, resulting in a large number of free parameters. In 3D, we must estimate 18 parameters for each tilt for a quadratic deformation model, which amounts to thousands of parameters when the number of tilts is high. One way to reduce the number of parameters, used in [11], is by assuming that the deformation field is constant along the depth (z direction) of the sample. with $\frac{(d_p+2)(d_p+1)}{2}$ free parameters.

To further reduce the number of free parameters, we introduce a temporal dependence in (13), which reduces the number of parameters from 18 for each tilt to 18 for the entire tilt series, assuming a quadratic deformation model. Our time-dependent deformation field is given by:

$$D_{t,k}(r, P) = \sum_{\zeta=1}^{d_t}\sum_{\substack{\alpha,\beta,\gamma \geq 0 \\ \alpha+\beta+\gamma \leq d_p}} \Big(P_{\alpha\beta\gamma\zeta}\Big)_k x^\alpha y^\beta z^\gamma t^\zeta, t \in [0, 1]. \quad (14)$$

As we reconstruct the first image, there is no way to recover a zeroth order deformation in time. For simplicity, we consider linear time dependence in our experiments, which amounts to setting $d_t = 1$.

Our method is independent of the choice of parametrization of the deformation field. Other parametrizations, which take advantage of the possible symmetries of the deformation field or additional understanding of the physics underlying the sample behaviour, could also be suitable choices.

III. Optimization

In [15], [19], [20], convex approximations to the minimization problem (12) have been devised by mapping the problem onto the space of measures $\mathcal{M}(\Omega)$. We interpret the marker configuration as a measure $\mu := \sum_{j=1}^{M} w_j \delta_{r_j} \in \mathcal{M}(\Omega)$, where the weights w_j are introduced as a means of relaxing the optimization problem (12). The weights determine the relative "importance" of the markers and, as we show later, can be used to remove candidate markers that do not contribute significantly to the data. Mapping the problem to measure space enables us to express the forward operation shown in (9) in terms of a linear operator, $\Phi_t : \mathcal{M}(\Omega) \to \mathbb{R}^{N_d}$:

$$\Psi_t = \sum_{j=1}^{M} w_j \left(G^p_{\theta_t} * \delta_{q_{t,j}} \right) =: \Phi_t \mu, \qquad \Psi = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \dots \\ \Phi_T \end{bmatrix} \mu =: \Phi \mu \tag{15}$$

The minimization problem (12) can then be rewritten as the following problem in the space of measures, where the loss is convex in the measure μ:

$$\underset{\mu \in \mathcal{M}(\Omega)}{\text{minimize}} \quad \ell(\Phi\mu - \tilde{\Psi}), \qquad \ell(\cdot) := \|\cdot\|_2^2 \tag{16}$$

In [15], the authors devised an effective numerical scheme for solving infinite-dimensional convex problems of the type shown above by using a variant of the conditional gradient or Frank-Wolfe method [21]. They also showed that interleaving the convex Frank-Wolfe iterations with nonconvex local optimization steps improved the convergence of the algorithm. This algorithm, known as the alternating descent conditional gradient (ADCG) method, has been subsequently extended for and applied to a range of application areas [15], [19], [20].

In this paper, we adapt the ADCG algorithm to solve the marker localization and deformation estimation problems simultaneously. To do this, we perform the Frank-Wolfe iterations as-is but modify the block coordinate descent routine to include an additional deformation estimation step. At each iteration of the algorithm, we place a new marker at a candidate initial location by solving a linearized approximation of our optimisation problem. Then, we solve a linear optimisation problem to obtain estimates for the weights of all current markers. Local optimisation routines are used to solve for the parameters for the deformation field and to refine the marker support in a bounded region. Our modified ADCG routine, which we call SparseAlign, is shown in Algorithm 1. Below we describe each step in our method in detail.

a) Adding candidate marker locations: We use the conditional gradient method to obtain candidate marker locations in steps 2-3. The conditional gradient or Frank-Wolfe method [21] can be used to solve constrained optimization problems of the type $\text{minimize}_{x \in \mathcal{C}} f(x)$ iteratively, where C is a convex set. The first step in each iteration is to minimize a linearized version of the loss within a specified domain. The linear approximation to a function $f(x)$ at x_k is given by

$$f_{\text{lin}}(s) = f(x_k) + \langle \nabla f(x_k), s - x_k \rangle.$$

Minimizing $f_{\text{lin}}(s)$ over a domain $\mathcal{D}_s$ thus amounts to solving

$$\underset{s \in \mathcal{D}_s}{\text{minimize}} \quad \langle \nabla f(x_k), s - x_k \rangle.$$

Algorithm 1 SparseAlign

```
for $n = 1 : n_{\max}$ do
    1) Compute current residual: $\varrho_n \leftarrow \Phi\mu_n - \tilde{\Psi}$
    2) Find next marker: $r_n^* \leftarrow \arg\min_{r \in \text{grid}} \langle \nabla\ell(\varrho_n), \Psi(r) \rangle$

    3) Update support: $\boldsymbol{r}_{n+1} \leftarrow [\boldsymbol{r}_n, r_n^*]$
    4) Block coordinate descent:
      Repeat:
        (a) Compute weights:
            $w_{n+1} \leftarrow \arg\min_w \ell(\Phi\mu_{n+1} - \tilde{\Psi})$
        (b) Prune support:
            $(w_{n+1}, \boldsymbol{r}_{n+1}) \leftarrow \texttt{prune}(w_{n+1}, \boldsymbol{r}_{n+1})$
        (c) Fit deformation parameters:
            $P_{n+1} \leftarrow \arg\min_{P \in \mathcal{P}} \ell(\Phi\mu_{n+1} - \tilde{\Psi})$
        (d) Improve support:
            $\boldsymbol{r}_{n+1} \leftarrow \arg\min_{\boldsymbol{r} \in \mathcal{C}} \ell(\Phi\mu_{n+1} - \tilde{\Psi})$
end for
```

Using our forward model (15) and the loss function in (16), we can compute that the linear minimisation step at iteration n is the following optimisation problem over measures $s \in \mathcal{M}_s(\Omega) \subset \mathcal{M}(\Omega)$

$$\underset{s \in \mathcal{M}_s(\Omega)}{\text{minimize}} \quad \langle \nabla\ell(\varrho_n), \Phi s \rangle, \tag{17}$$

where $\varrho_n := \Phi\mu_n - \tilde{\Psi}$ is the residual at iteration n.

An optimal solution of the above problem is the addition a single new marker with positive weight to the current support of μ_n. This ensures that, at iteration n of the algorithm the measure μ is supported at n points. Adding only one location at a time has been shown to give the sparsest possible solution [15].

Practically, we solve (17) by gridding the domain of marker locations coarsely. The contribution of a single marker at each grid point, r_{grid}, is computed for a current guess of deformation parameters:

$$\psi(r_{\text{grid}}) = \begin{bmatrix} G^p_{\theta_t} * \delta_{A_{\theta_1}(r_{\text{grid}} + D_1(r_{\text{grid}}))} \\ G^p_{\theta_t} * \delta_{A_{\theta_2}(r_{\text{grid}} + D_2(r_{\text{grid}}))} \\ \dots \\ G^p_{\theta_t} * \delta_{A_{\theta_T}(r_{\text{grid}} + D_T(r_{\text{grid}}))} \end{bmatrix}$$

Then, the inner product of the current residual with the forward projection of a marker located at each grid location is calculated. The grid location r^*_{grid} with the smallest inner product with the residual is chosen as the next candidate location:

$$r^*_{\text{grid}} = \arg \min_{r \in \text{grid}} \quad \langle \nabla\ell(\varrho_n), \psi(r) \rangle. \tag{18}$$

b) Optimizing weights: Once we have optimized for marker locations, we can optimize the weights of each marker as shown in steps 4(a)-(b). Note that the model (15) depends linearly on the weights w_j, $j \in \{1, 2, \dots, M\}$. Thus, with the number of markers, marker locations and deformation parameters fixed, the weights w_j can be estimated by solving the following linear least-squares problem

$$\underset{w \in [0,1]^n}{\text{minimize}} \quad \|\ell(\Phi\mu_n - \tilde{\Psi})\|_2^2. \tag{19}$$

All weights w_j are constrained to lie in $[0,1]$ and represent the relative importance of marker contributions to the data. Markers with weights close to zero can be removed by an additional `prune` routine that removes all markers with a weight lower than a predefined threshold. In some cases an additional `prune` routine can be used to remove markers with small weights at the end of a full algorithm run. This further ensures that the solution obtained is the sparsest possible marker configuration required to explain the data $\tilde{\Psi}$.

c) Refining initial marker locations: At each iteration, we perform the nonconvex local optimization step shown in 4(d) to refine our estimates for the initial marker locations. This step was first proposed in [15] as a way to speed up convergence of the conditional gradient method.

Refining the support of the current measure μ_n without changing the number of markers ensures that markers are moved off the grid locations used in steps 2-3. It also imparts some of the rapid local convergence qualities of nonconvex optimisation [15]. In our implementation, we use the L-BFGS-B algorithm to perform local optimisation over initial marker locations.

d) Estimating deformation parameters: The optimization problem behind step 4(c) is given by

$$\underset{P\in\mathcal{P}}{\text{minimize}} \quad \sum_{t=0}^{T} \left\| \tilde{\Psi}_t - \sum_{j=1}^{M} w_j \left(G^p_{\theta_t} * \delta_{A_{\theta_t}(r_j + D_t(r_j,P))} \right) \right\|_2^2, \tag{20}$$

which is a difficult nonconvex problem that is often studied in the context of image correspondence problems such as image registration or optical flow estimation [22]. We use L-BFGS-B initialized at the current P_n to compute a local update P_{n+1} for the parameters of the deformation field.

e) Coarse-to-fine scheme for large data: One of the challenges of solving (20) is that the objective function is flat if the forward projection of the current marker configuration and the data do not share the same support, and gradient-based optimization schemes such as L-BFGS-B have a hard time locating a minima. This easily happens for small objects, such as markers, embedded in large projection images. The remedy is typically to smooth both images with a Gaussian, compute a deformation field on the smoothed problem, and use the solution of the smoothed problem to initialize the optimization of the original problem.

Gaussian smoothing followed by downsampling removes high image frequencies and one starts matching only the low frequencies. For noisy data, downsampling has the additional advantage of denoising the data. Furthermore, for large experimental data, where each tilt image has pixel dimensions 4096×4096, warm-starting the optimization at high resolutions with good initial values ensures that not many expensive iterations have to be performed.

For realistic simulation data and experimental data, we use a coarse-to-fine scheme where the marker localization and deformation estimation problem is solved at successively finer resolutions using the results at the coarser resolutions as initialization.

At full resolution, we generate the forward projection of a single marker using (6) followed by sampling on a spatial grid X_f with N_d grid points. Thus, the discretized forward projection of the full marker configuration can be written as

$$\Psi_t = \sum_j w_j S^f \mathcal{G}_{(q_{t,j},\tau_f)}, \tag{21}$$

where S^f is the sampling operator associated with the spatial grid X_f and $\mathcal{G}_{(q_{t,j},\tau_f)}$ is a Gaussian centred at $q_{t,j}$ with standard deviation τ_f.

For obtaining measured data at coarse resolutions, we downsampled the full-resolution measured data $\tilde{\Psi}_t$ at each time after Gaussian convolution to prevent aliasing artefacts [23]. Thus, the coarse-resolution data were given by $\tilde{\Psi}^c_t := \mathcal{H}^c(\mathcal{G}_{\tau_a} * \tilde{\Psi}_t)$, where $\mathcal{H}^c$ is a downsampling operator associated with a coarse grid X_c and $\mathcal{G}_{\tau_a}$ is an anti-aliasing Gaussian. For integer downsampling factors $\eta := |X_c|/|X_f|$, $\mathcal{H}^c$ only keeps pixels separated by η in the coarse-resolution image.

We approximated matching forward projection data Ψ^c_t directly from marker locations using our forward model (9) by sampling the Gaussian-convolved projected marker locations on the coarse grid X_c:

$$\Psi^c_t = \sum_j w_j S^c \mathcal{G}_{(q_{t,j},\tau_f)}, \tag{22}$$

where S^c is the sampling operator associated with the coarse grid X_c.

IV. Numerical Experiments

In this section we describe our experiments with simulated and real data. Implementation notes with details of software packages used are provided in Section S1 of the Supplementary Materials.

A. Illustrative 2D example

a) Ground truth: We used a simple simulated sample to elucidate properties of our algorithm in 2D. The FoV was taken to be $[-L/2, L/2]$ along both axes, with the canonical length scale $L = 1$. The ground truth sample consisted of 10 gold bead markers confined to a thin rectangular region: $x \in [-2L/5, 2L/5], z \in [-L/10, L/10]$. We chose this geometry for our 2D sample to mimic the geometry of experimental cryoET samples.

For simplicity, we considered deformation field components to be zero along the horizontal (x) direction. In the vertical (z) direction, we assumed the deformation to be given by a quadratic polynomial of x and z:

$$D_{t,z}(r,P) = (P_0 + P_1 x + P_2 z + P_3 x^2 + P_4 z^2 + P_5 xz)t =: D_{1,z}t, \tag{23}$$

with $P_0 = 0\ L$, $P_1 = P_2 = -1$, $P_3 = P_4 = P_5 = -1\ L^{-1}$, and t taking values in $[0,1]$

b) Projection data: We generated projection data using the forward model in (15) over a set of discrete projection angles $\theta \in [-70°, 70°)$, $N_\theta = 20$. Practically, we computed the continuous Radon transform of each marker, followed by a continuous 1D Gaussian convolution in projection space. The Gaussian-convolved projection was then discretized on a detector grid with $N_d = 64$. At each projection angle, the projection was then a 1D profile. All the projections were rearranged in a sinogram with dimensions $N_\theta \times N_d$.

For comparison, we also generated input data for the doming model method in [11]. These data were the projected locations of each marker over the same series of projection angles.

B. Simulated 3D examples

a) Ground truth: We used a 3D configuration of markers to test the robustness of our method to noise and to mismatches in the forward model. We used 20 randomly placed markers in a thin region in 3D with dimensions 819.2 nm $\times$ 819.2 nm $\times$ 100.0 nm. The sample used was the same as that described in IV-C.

We considered deformation field components to be non-zero only along the z direction; this component was then given by:

$$D_z(x, y, z, t) = (P_0 + P_1 x^2 + P_2 y^2)t, \qquad (24)$$

with $P_0 = 200$ nm, $P_1 = P_2 = -100$ nm^{-1}, and t taking values in $[0, 1]$.

b) Projection data: We generated projection data along 140 equispaced projection angles in $[-70^\circ, 70^\circ]$ using a Gaussian with standard deviation 15nm as the shape function of individual markers. Each projection image was discretized on a 64×64 pixel grid.

To convert the intensities in these generated images to meaningful electron counts, we used that the expected electron count in any pixel is given by $I = I_0 e^{-V_{\text{abs}} C \times \delta x}$, where I_0 is the incoming electron count, V_{abs} is the absorption potential of gold nanoparticles ($5.39V$ for a $300keV$ electron beam, treating the gold as amorphous), C is the interaction constant ($0.00653V^{-1}nm^{-1}$ at $300keV$) and δx is the path length travelled by electrons through a gold marker. This path length is equal to the product of the diameter of the gold bead, which we take to be 15nm, and the intensity in our generated images. For our experiments, we generated data with $I_0 = 2^n, n \in \{6, 7, 8, 10, 12, 14\}$.

c) Gaussian noise: To test the properties of our approach for noisy data, we performed experiments with data corrupted with additive Gaussian noise, such that

$$\Psi_{\text{noisy}} = \Psi_{\text{clean}} + \mathcal{N}(0, \sigma^2_{\text{noise}}),$$

where Ψ_{clean} are the data scaled to physical electron counts and σ^2_{noise} is the variance of the noise added.

We performed experiments using $\sigma^2 = 2^n$, $n \in \{7, 8, 10, 12, 14\}$. For each noise setting, multiple independent experiments were performed and the results were averaged to obtain mean values for the metrics. Each independent experiment was initialized with a with a different random seed.

d) Poisson noise: We also generated a series of Poisson noise-corrupted data by varying the electron count per pixel per frame, I_0. For $I_0 = 2^n$, $n \in \{6, 8, 10, 12, 13, 14\}$, we generated Poisson-distributed electron counts at each pixel using:

$$\Psi_{\text{noisy}} = \text{Poi}(\Psi_{\text{clean}}), \qquad (25)$$

where Ψ_{clean} are the data scaled to physical electron counts and $\text{Poi}(\cdot)$ denotes a Poisson random variable. The Poisson-noise data were generated to have comparable signal-to-noise ratios as those of the Gaussian-noise data. For each noise instance, we performed multiple independent experiments with different random seeds and averaged over the obtained metrics.

C. Realistic TEM simulations

We used the TEM-simulator software [24] to generate physically plausible simulations of TEM images from a specification of a 3D sample (see example projection images in Fig. 2(a) and (b)). To simplify matters, the sample consisted purely of gold particles in vacuum, thus disregarding the ice buffer and other sample structures. The purpose of this numerical experiment was to test our algorithm in situations where its forward model did not match the one used for data generation. In particular, the explicit assumption of Gaussian shape of gold particles and the implicit assumption of additive uncorrelated noise characteristics were violated.

The test sample consisted of 20 gold particles of 15nm diameter, randomly distributed in a slab of dimensions 819.2nm $\times$ 819.2nm $\times$ 100.0nm in x, y, z space. Over time, this sample was simulated to undergo a deformation described by the vector field

$$D_z(x, y, z, t) = (P_0 + P_1 x^2 + P_2 y^2)t, \quad D_x = D_y \equiv 0 \qquad (26)$$

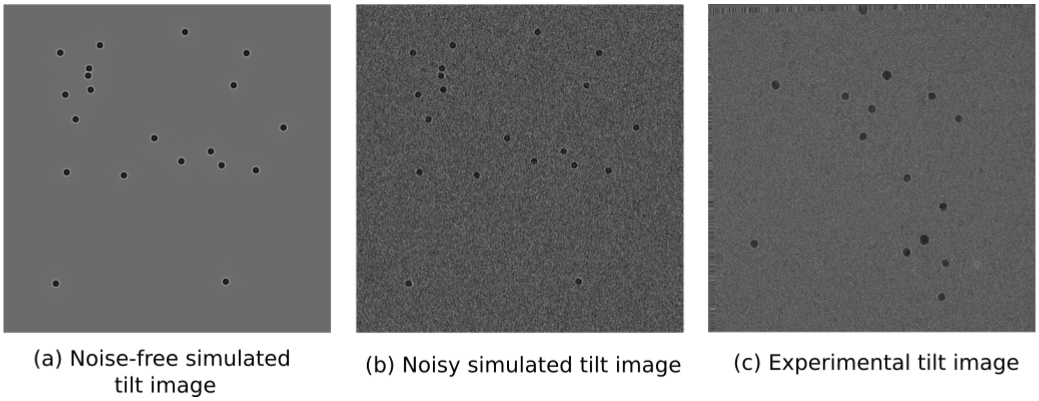
with $P_0 = 200$ nm, $P_1 = P_2 = -100$ nm^{-1}, and t taking values in $[0, 1]$. This amount of deformation (200 nm at $x = y = 0,\ t = 1$) is an exaggerated version of a doming motion observed in practice. The large amplitude was chosen to make the effects under investigation easier to observe.

Assuming constant tilt speed, the time t was mapped to a tilt angle θ according to $\theta_i = -70^\circ + t_i \cdot 140^\circ,\ t_i = \frac{i}{140}\,, i = 0, \ldots, 140$. At each tilt angle, a projection image was simulated according to the weak phase object approximation model [5], taking the contrast transfer function (CTF) of the optical system into account (see [24] for details). We used electrostatic potential values of $V = 0$ for vacuum and $V = (29.87 + \imath \cdot 5.39)$ Volt for (amorphous) gold. The CTF parameters were chosen as $\Delta z = 8$ μm (defocus), $C_C = 2.7$ mm (chromatic aberration) and $C_S = 2.7$ mm (spherical aberration).

The size of each projection image was chosen equal to the $x - y$ dimensions of the sample, subdivided into $(N_x, N_y) = (512, 512)$ pixels, each of size 1.6 nm. Simulated data were generated with 8x binning, with the full resolution pixel size equal to 0.19 nm. Binning was performed because of computational convenience.

a) Noiseless data: The noiseless images generated by TEM-Simulator correspond to probability densities of detecting an electron at a given location in the detector plane. Therefore, scaling with the average number of incoming electrons per pixel area results in each pixel value representing the expected number of electrons measured in that pixel, also referred to as “infinite dose” case.

b) Noise generation: In a real experiment, a finite number of electrons interacts with the sample and is detected at the camera. This process was modeled with a Poisson random variable $\text{Poi}(\lambda_k)$ per pixel, where the parameter $\lambda_k = I_0 \Psi_k$ equals the intensity of the $k-$th pixel in the scaled noiseless


(a) Noise-free simulated tilt image

(b) Noisy simulated tilt image

(c) Experimental tilt image

Fig. 2: Example tilt images generated using TEM-simulator (a) without noise and (b) with added correlated noise; (c) an experimental TEM image showing gold beads on vitrified ice.

image. This noise model applies to a perfect counting camera. However, cameras operating in integration mode have a non-trivial point spread function because charge from one incident electron can leak into neighboring pixels, triggering multiple detection events. Furthermore, signal and noise transfer vary with spatial frequency. These two effects are characterized by the MTF (modulation transfer function) and DQE (detective quantum efficiency) of the camera and lead to signal blur and noise correlation [5]. The noisy images in these numerical experiments made use of this model.

c) Pre-processing for noisy data: For data with correlated Poisson noise, we performed the following pre-processing steps. First, we used noiseless data to perform segmentation with Otsu's method [25]. We obtained a mask for the markers in the tilt series from this segmentation procedure, which we used to compute average background and marker intensities in the noisy tilt series. Second, we shifted the range of the noisy data by subtracting its minimum value and applied the Anscombe transform to our shifted data. Our forward model (15) assumes that the intensity in the background of a projection image is mean zero with constant variance and the intensity at gold beads is mean one with constant variance. The variance of data with Poisson noise varies with the mean, and thus differs from the assumption in our forward model. To reduce the discrepancy between our model assumptions and the simulated data, we used the Anscombe transform

$$\text{Anscombe}(\tilde{\Psi}) := 2\sqrt{\tilde{\Psi} + 3/8}$$

as a variance-stabilizing transformation to obtain data with an approximately constant variance and standard deviation [26]. Finally, we subtracted the average background intensity and divided by the average bead intensity in the data.

D. Experimental data

For our experimental data we used a sample with gold beads as the only prominent features. We deposited 20nm gold particles on a lacey carbon grid, which was plunge-frozen in liquid ethane using a Thermo Scientific Vitrobot. An example tilt image is shown in Fig. 2(c).

We acquired a tomographic tilt series using the Thermo Scientific Tomography 5.5 software package on a Thermo Scientific Titan Krios electron microscope equipped with a Thermo Scientific Falcon 3EC camera. An area in a hole with 15 gold beads was selected. A magnification of 37000x was chosen for a pixel size of $1.949\mathring{A}$ and a field of view of 800 nm. The sample was tilted from -60 to +60 degrees with a tilt step of 2 degrees. Each image in the tilt series had an electron dose of 0.198 $e^{-}/\mathring{A}^{2}$.

a) Cross-correlation-based global alignment: Projection images were globally shift-aligned using the cross-correlation-based routine in Thermo Scientific Inspect3D.

b) Data pre-processing: Not all projections were globally aligned correctly using the cross-correlation-based alignment routine. We inspected the tilt series visually for any misaligned projections and removed these. This resulted in a total of 27 projections that were then used for estimating local sample deformation. Next, we deleted 256 pixels from each of the four borders of the tilt series images to get rid of missing image data added by the cross-correlation-based alignment routine. Only one marker, near the top edge of the tilt series images, was discarded because of edge removal. As we expected correlated Poisson noise in these data, we applied the Anscombe transform to the raw tilt series to obtain data with approximately constant variance. After applying the Anscombe transform, we subtracted the mean of the tilt series; because most pixels were background pixels, this ensured that the average background intensity was close to 0. Finally, all tilt series pixels were divided by the average marker intensity to ensure that, in accordance with our forward model, the markers had an average intensity of approximately 1. To determine the average bead intensity in experimental data, we inspected the tilt series visually and used the average intensity in three small square regions around three beads.

E. Evaluation criteria

To quantify the accuracy of our estimated deformation fields with respect to the ground truth, where available, we used the following evaluation criteria. First, the estimated and ground truth deformation parameters were used to compute the deformation field at $t = 1$ on a gridded FoV of dimensions 1000×1000 (for 2D) and $1000 \times 1000 \times 1000$ (for 3D), using equation (23). Next, the vectorial difference between estimated and ground truth deformation fields at $t = 1$ was computed at each grid point:

$$E(r_{\text{grid}}) = \|D^{\text{gt}}_{1,z}(r_{\text{grid}}) - D^{\text{est}}_{1,z}(r_{\text{grid}})\|_2^2 \tag{27}$$

This deformation estimation error was averaged over the whole grid to obtain the global deformation estimation error and averaged only at the ground-truth marker locations to obtain the deformation estimation error at markers:

$$E_{\text{global}} = \frac{1}{N_{\text{grid}}} \sum_{\text{grid}} E(r_{\text{grid}}) \tag{28}$$

$$E_{\text{markers}} = \frac{1}{M} \sum_{j=1}^{M} E(r_j) \tag{29}$$

where $N_{\text{grid}} = 10^9$ for 3D and $N_{\text{grid}} = 10^6$ for 2D.

V. Results

a) SparseAlign adds markers with small displacements first: In Fig. 3(a) and (b), we show how SparseAlign localizes markers. At each iteration, markers are added by solving the linearized problem (18) on a coarse grid. We show the values of the objective function at each grid location in Fig. 3(a). The first marker added is a marker close to the centre of the field of view, where the displacement of markers is smallest. This corresponds with the fact that all deformation parameters are set to zero for the first iteration. After the first iteration, when we start optimizing for the deformation parameters, markers that show larger displacements are added. In Fig. 3(b), we show two examples of marker location refinement. The two plots on the left show marker addition and refinement at iteration 3; a new marker, indicated with a red star, is added at a grid location. Local optimization then allows us to move this marker as well as all currently placed markers (blue plus signs) off the grid and closer to the ground truth locations (green crosses). The two plots on the right show another step of marker addition and local optimization at iteration 7. In both cases, local optimization helps to improve the solution close to the region where the new marker is added. We indicate this region with a red rectangle in the plots.

b) SparseAlign's image-based loss is not convex with respect to deformation parameters: In Fig. 3(c), we plot the image-based loss in (12) as a function of each deformation parameter separately, while holding other parameters and marker locations fixed at their respective ground truth values. For comparison we also plot the marker-based loss in (11). Finally, each plot is normalized with a different normalization constant, equal to the maximum value of the loss for that parameter. For each parameter, the marker-based loss is a near-perfect quadratic function with a minimum at the ground truth parameter value. The image-based loss function shares the same minima but differs from the marker-based loss at higher parameter values. In general, the image-based loss function is only convex in a small region around the global minimum. As we move away from the minimum, the loss function increases for each parameter until, at large parameter values, markers move out of the field of view and the loss shows other minima (as in the plot for P_0) or flattens and dips (as in the plots for P_1 through P_3). Gradient-based schemes can thus get caught in local minima if parameter values are very far away from the true minimum at initialization.

c) SparseAlign estimates deformation parameters with an accuracy comparable to that of the doming model: In Fig. 4 we illustrate the differences between the doming model optimization used in [11] and our method. We use the simple 2D sample shown in Fig. 3 with a quadratic deformation field along the vertical (z) direction.

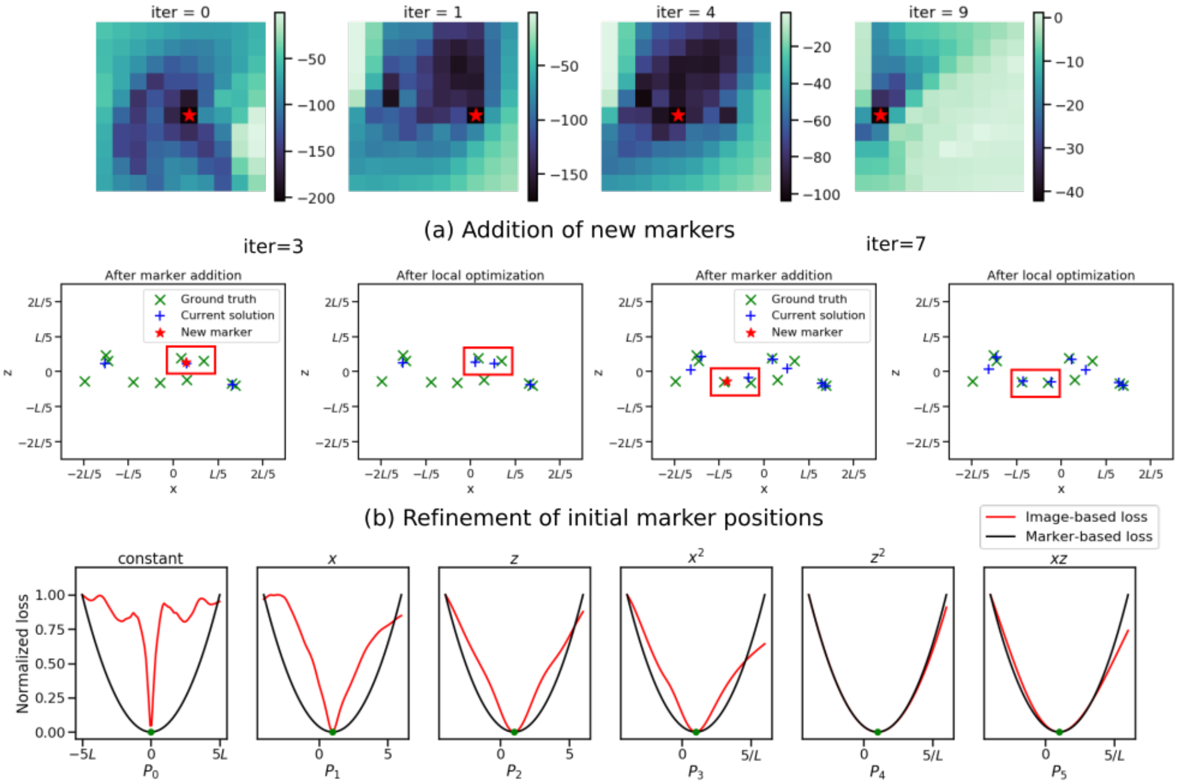
Input data for the doming model ('DM') optimization are indicated with red dots in Fig. 4(b); projection data for SparseAlign is a 1D profile indicated with a blue line. The set of line profiles can be rearranged to give a sinogram for the SparseAlign data.

In Fig. 4(c), we show the reconstructed deformation fields obtained using the two methods. In Fig. 4(d), we illustrate the vectorial deformation field error (27) in both cases. We observe that the error in the convex hull of the markers is comparable using both methods. This is true despite the fact that our method does not need labelled marker locations and minimizes a more complicated image-based loss function. In regions without markers, our method shows larger errors. This is an indication of the greater ill-posedness of our deformation estimation problem (20).

In Fig. 4(e-f), we compare mean deformation estimation errors (29) and (28) for both methods at the ground truth marker locations and in the entire FoV. Mean deformation estimation errors at marker locations are comparable for both methods although the global mean error is higher for SparseAlign. The larger global error, however, is not significant because the major contribution comes from boundaries where no sample is present. Marker localization using SparseAlign and DM gives comparable results, as illustrated in Fig. 4(g).

d) Deformation estimation accuracy reduces almost linearly for additive Gaussian noise: In Fig. 5, we perform a quantitative analysis of the robustness of our method with respect to noise in projection data. The ground truth marker configuration and deformation field are shown in Fig. 5(a). We used different noise settings to probe the properties of our method for data corrupted with Gaussian and Poisson noise, and for each noise level we performed 100 independent experiments by randomizing both the initial marker locations as well as using different noise realizations. The mean deformation estimation error plots for Gaussian noise show an almost linear decrease in deformation estimation accuracy for increasing signal-to-noise ratio (SNR, given by the standard deviation of the Gaussian noise). Moreover the spread of the distribution narrows for high SNRs, indicating that there are fewer catastrophic failure cases for deformation estimation.

The dependence of deformation estimation error on noise is more complicated in the case of Poisson noise. As shown in the plots in Fig. 5(c), we do not see a linear dependence as in the case of Gaussian noise. The difference in accuracy between deformation estimation results for low and high electron counts is also smaller. This suggests that the mismatch between Poisson noise data and data generated from our forward model is greater than the mismatch in the case of comparable Gaussian noise.



(a) Addition of new markers

(b) Refinement of initial marker positions

(c) Sensitivity of loss functions to changes in deformation parameters

Fig. 3: Three steps in SparseAlign. (a) Addition of new markers is performed on a coarse grid using the optimisation problem (18). The grid location with the smallest pixel intensity in the heatmap is chosen as the next candidate location, which is indicated with a red star. (b) Refinement of initial marker locations is performed using L-BFGS-B. The two leftmost plots show one step of marker addition followed by local optimization; the two rightmost plots show another step of marker addition and local optimization. In both cases, after addition of a new marker (red star), local optimization ensures that all current markers (blue plus signs) are brought closer to the ground truth locations (green crosses). We indicate the areas where this improvement is clearest with red rectangles. (c) Sensitivity of the marker-based loss (black line) used in the doming model approach and our image-based loss (red line) to changes in deformation parameter values. For each plot, the loss was normalized independently with respect to its maximum value.

e) Model mismatch does not affect deformation estimation significantly: We used physically plausible TEM simulations to generate data where the forward model of SparseAlign did not match the data generation model.

In these data, the shape function of a gold bead marker is not a Gaussian. In Fig. 6(a), we show the profile of a marker in projection data generated using the TEM-simulator package [24] and the profile of a marker using our forward model. We assumed that the size of gold bead markers and the pixel size of projection images are known, so that the width of the Gaussian can be computed.

We used binned simulated data, as detailed in IV-C, for these experiments. In Fig. 6(b), we show results on marker localization and deformation estimation using noiseless data. The ground truth marker configuration and deformation field are the same as those shown in Fig. 5(a). The results we show in Fig. 6(b) are those obtained at the final step of a coarse-to-fine scheme, where we solved for marker localizations and deformation parameters at increasing resolutions using downsampling factors $\eta = 1/16, 1/8, 1/4, 1/2$. The final result of such a scheme shows a good qualitative match between reconstructed and ground truth marker locations and deformation fields. We stopped at $\eta = 1/2$ because the effect of model mismatch, which we discuss in the next paragraph, is greatest at high resolutions. Moreover, our current implementation is unable to handle very large data sizes, an area we plan to improve in a future work. Nevertheless, our results indicate a good qualitative match between ground truth and estimated deformation fields, suggesting that the absence of higher-resolution data might not impact deformation estimation for the cases considered.

In Fig. 6(c), we show the effect of model mismatch at different resolutions using plots of the difference between our forward projected reconstructed markers and the observed data. We see that the effect of model mismatch is most pronounced at the finest resolutions. This indicates why using a coarse-to-fine scheme, where we obtain initial guesses for marker locations and deformation parameters by solving the problem in a coarse resolution first, leads to reasonable results despite the difference in forward models.

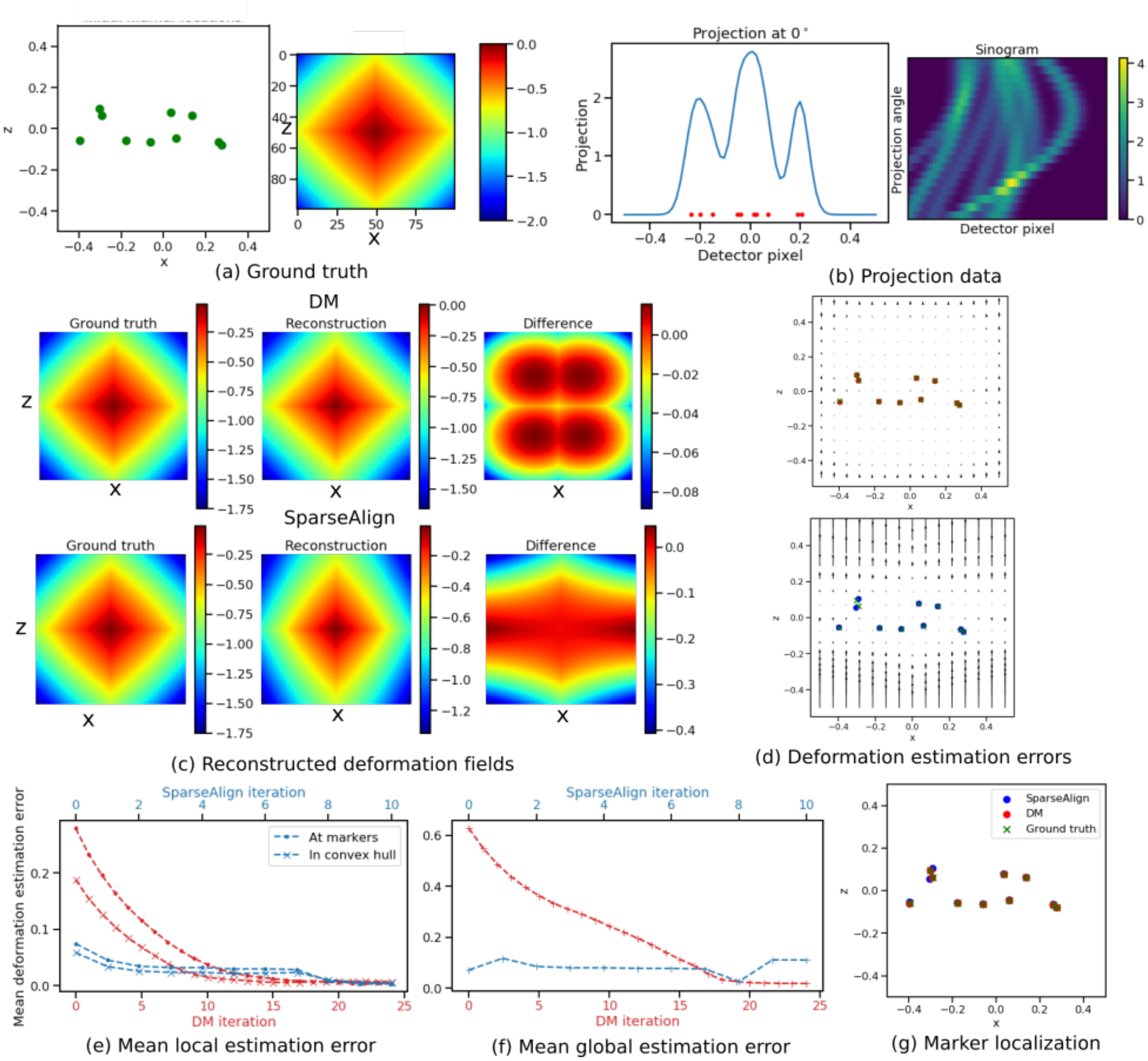


Fig. 4: Marker localization and deformation estimation using SpaseAlign and the doming model method (DM). (a) Ground truth initial marker locations and deformation field component along the z-axis at $t = t_1$, $D_{1,z}$. (b) Input data for DM are the projected marker locations indicated with red dots. Projection data for SparseAlign at 0° is a 1D profile that is a superposition of Gaussians; we indicate this data in blue. The full sinogram data is a stack of projections taken along tilt angles in [-60°, 60°). (c) Reconstructed deformation fields using DM and SparseAlign. In both cases, errors are largest at the boundaries of the field of view (FoV), where no markers are present. (d) Deformation estimation error (27) obtained using DM and SparseAlign. Errors are comparable in the convex hull of markers; errors outside the convex hull are larger when using SparseAlign. (e)-(f) Mean local and global deformation estimation errors (28)-(29) as a function of DM and SparseAlign iterations. (g) Localized initial marker locations using SparseAlign (blue circles) and DM (red circles) overlaid with the ground truth marker locations (green crosses).

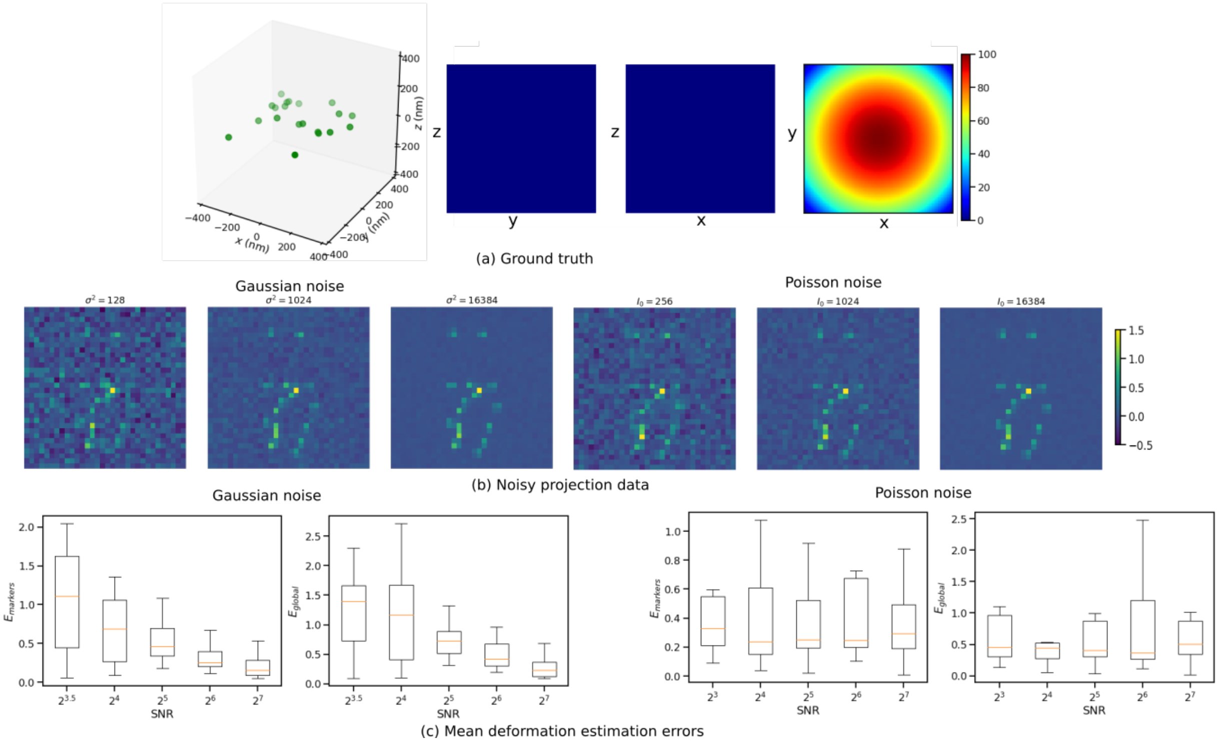
We plot mean deformation estimation errors (29) and (28) for each iteration in Fig. 6(d). Jumps in resolution are indicated with dotted lines. Here we observe that the maximum reduction in deformation estimation error is achieved at the coarsest resolution. The initial guesses obtained are then refined subsequently at each finer resolution. The stopping criterion we used to jump to a higher resolution was to check whether the absolute difference in loss at each new iteration was greater than a pre-set tolerance value (here, 10^{-6}).

Finally, in Fig. 6(e), we illustrate the deformation estimation error (27) at each resolution. Here we observe that, at the coarsest resolution, the error is already small near the centre of the FoV, where a majority of markers is present. At higher resolutions, the refinement in deformation parameters ensures smaller errors at the boundaries and indicates improvements in the values of estimated parameters.

f) Marker localization is poor for data with correlated Poisson noise: In Fig. S1, we show results of our method on data with realistic markers and realistic correlated noise using the ground truth marker configuration and deformation field in Fig. 5.

We observe that marker localization for correlated noise-

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2022.3194719



(a) Ground truth

(b) Noisy projection data

(c) Mean deformation estimation errors

Fig. 5: Deformation estimation in 3D with Gaussian and Poisson noise-corrupted data. (a) Ground truth configuration of markers (left) and ground truth deformation field in nm (right). (b) Projection image at 0° with different Gaussian noise and Poisson noise settings. The variance of Gaussian noise (σ^2) and the photon flux (I_0) were chosen to simulate comparable Gaussian noise and Poisson noise realizations. (c) Deformation field estimation errors as a function of iteration at markers (E_{markers}) and in the entire field-of-view (E_{global}) for various Gaussian and Poisson noise settings.

corrupted data is poorer than that for noiseless data (shown in Fig. 6). At the end of a coarse-to-fine scheme, two markers are not localized and a few markers with small weights are added to the reconstruction domain. These small weighted markers were removed with a further thresholding step, where markers with weights less than 0.1 were discarded. Improving marker localization might need changes to the forward model used, an aspect that needs further research; however, in our experiments, marker localization did not have a significant effect on deformation estimation accuracy, as seen from the reconstructed deformation field shown in Fig. S1(a).

In Fig. S1(b), we show plots of mean deformation estimation errors. Note that the same stopping criterion as that used for noiseless data ensured that more iterations were performed at finer resolutions for data with realistic noise.

In Fig. S1(c), we plot the deformation estimation error at different resolutions. Comparing these plots with those for noiseless data in Fig. 6, we see that the errors at the boundaries are higher for noisy data, which is most clearly observed at the coarse resolutions.

g) Deformation estimation is limited by the model basis: We performed experiments with realistic 3D simulated data where the ground truth deformation field along the z direction contained cubic terms in x and y in addition to the quadratic terms in (26). The ground truth deformation field used in these experiments was given by:

$$D_z(x, y, z, t) = (P_0 + P_1x^2 + P_2y^2 + P_3xy^2 + P_4x^2y)t \quad (30)$$

with $P_0 = 200$ nm, $P_1 = P_2 = -50$ nm^{-1}, $P_3 = P_4 = 25$ nm^{-2}. Although the ground truth contained cubic terms, we restricted the deformation terms used in our forward model to be quadratic in x and y. We performed experiments for both noiseless data and data corrupted with correlated Poisson noise. For both noiseless and noisy data, our algorithm was able to identify the quadratic terms in the deformation field (Fig. S2(a-b)). As there were no cubic terms in the forward model, the reconstructed deformation fields did not contain any cubic components. The effect of this mismatch is greatest at the two corners of the FoV where the contribution of cubic terms was the highest.

When we included cubic terms in the forward model, we found that both marker localization and deformation estimation improved as both quadratic and cubic terms were now estimated. The recovered deformation field in Fig. 6(c) is much closer to the ground truth. These results indicate that the accuracy of SparseAlign is limited by the basis used for deformation modelling.

h) SparseAlign locates markers reasonably in experimental data: We used an experimental dataset of gold beads embedded in ice to test the applicability of our method to experimental datasets. We used a coarse-to-fine scheme with downsampling factors $\eta = 1/128, 1/64, 1/32, 1/16, 1/8$ to localize gold bead markers and estimate the deformation field. We show an example tilt image in Fig. 7(a) and the same image at different downsampling factors in Fig. 7(b).

This article has been accepted for publication in IEEE Transactions on Computational Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCI.2022.3194719

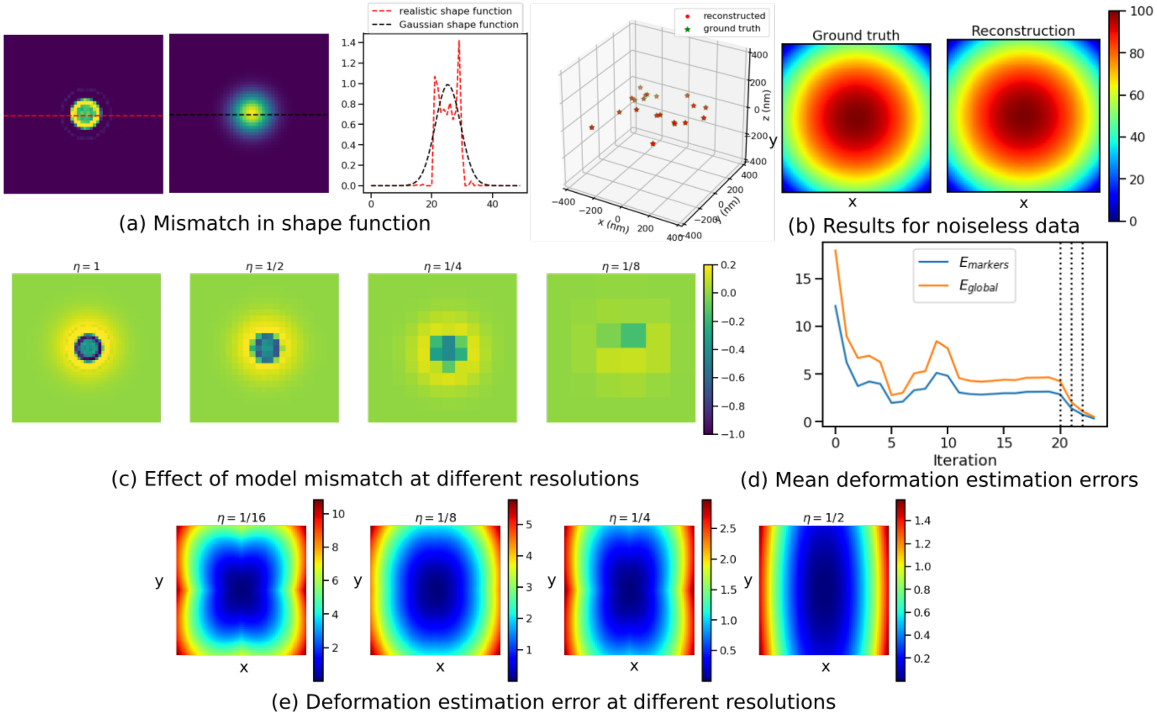


Fig. 6: (a) Mismatch in shape function. (left) 2D projection of a single marker generated using the TEM simulator. (centre) Projection of a Gaussian marker used in our forward model. (right) Profiles of both shape functions. (b) Marker localization results (left) and deformation estimation results in nm (right) for noiseless realistic data. (c) Difference between forward projected marker locations and observed data (a small region around a single marker is shown). The difference due to model mismatch is largest at the fine resolutions. (d) Mean deformation estimation error at ground truth marker locations and in the entire FoV for different iterations. Resolution changes in the coarse-to-fine scheme are indicated with black dotted lines. (d) Absolute error of estimated deformation field with respect to the ground truth at different values of the downsampling factor η.

Using a coarse-to-fine scheme we were able to localize several, but not all, markers. In Fig. 7(c), we show our marker localization results. We thresholded the localized markers according to their reconstructed weights. Here we show 15 markers with the highest weights. We estimated deformation along the z direction using a quadratic model:

$$D_{t,z}(r, P) = (P_0 + P_1 x + P_2 y + P_3 x^2 + P_4 y^2 + P_5 xy)t \quad (31)$$

Additionally, we set the x and y components of the deformation field to zero. It is probable that our assumed deformation field was insufficient to model sample deformation in the experimental data.

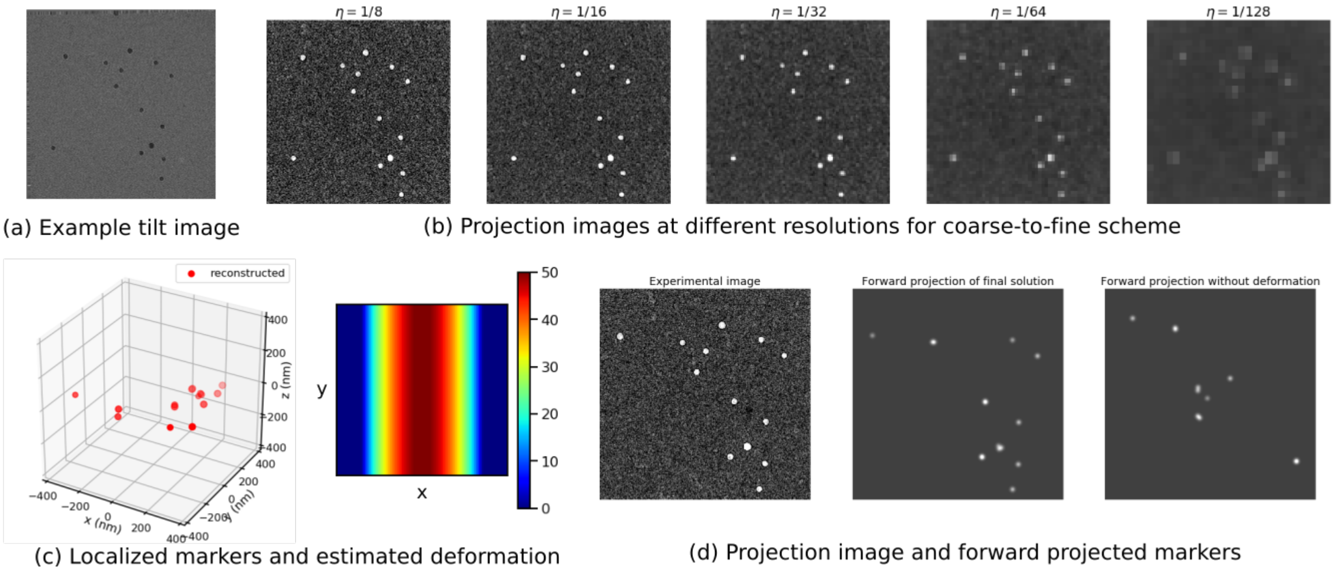
Our algorithm predicted a deformation field that is quadratic in x but constant in y, a model that could not be experimentally validated. Plugging the estimated deformation field and marker locations into our forward model, we computed the forward projection shown in Fig. 7(d). Comparing this image to the data, we see that not all markers have been localized correctly, but at least one marker was localized in each of location with a cluster of markers. Markers throughout the FoV were localized; this suggests that the deformation estimation routine did not do worse for certain spatial regions. Moreover, mismatch in the shapes of actual markers and the Gaussian used in our forward model did not hinder the localization of most markers. Using localized marker locations and setting deformation to zero leads to projection images that are qualitatively different from the experimental data (Fig. 7(d)).

VI. Conclusion and Discussion

Marker-based alignment is a important step for reconstruction improvement in cryoET. We have developed a mathematical approach called SparseAlign for modeling beam-induced local sample motion. In contrast to current approaches, our method does not need labelled marker locations, and directly uses projection data to localize markers and solve for the parameters of a polynomial deformation model. As a consequence, our method is more suited to data with low signal-to-noise ratios where markers cannot be reliably identified. The deformation fields estimated using our method can be used in a subsequent routine to compute a motion-compensated sample reconstruction.

Despite solving a more ill-posed problem for deformation estimation, SparseAlign localizes markers and estimates deformation parameters with an accuracy comparable to that of the doming model approach. Moreover, SparseAlign estimates deformation accurately even when the forward model for markers shows discrepancies with respect to marker projections in observed data.

The image-based loss (12) in this paper can be improved upon by using a more canonical loss as the objective function



(a) Example tilt image

(b) Projection images at different resolutions for coarse-to-fine scheme

(c) Localized markers and estimated deformation

(d) Projection image and forward projected markers

Fig. 7: Results on experimental data. (a) A raw projection image in the acquired tilt series. (b) One image from pre-processed data used for deformation estimation and marker localization with downsampling factor $\eta = 1/8, 1/16, 1/32, 1/64, 1/128$. (c) Localized markers (left) and estimated deformation along z (in nm). (d) One experimental projection image downsampled by $\eta = 1/8$ (left), forward projection of localized markers with estimated deformation field (centre) and forward projection of markers with deformation field set to zero (right).

for marker localization and deformation estimation. Unlike the ℓ^2 loss used in this paper, the Wasserstein loss measures distances between distributions and has non-zero gradients even when the supports of the ground truth and current solution do not overlap [27].

The application of our approach to experimental data is limited by the deformation model used. One way to choose the most suitable, sparse basis for deformation modelling is to optimize over a library of basis functions using the data-driven approach in [28].

In this paper, we have ignored the image contrast of the biological sample while estimating deformation parameters. Ideally, our approach would be the first step in an iterative scheme where we alternate between sample reconstruction and tilt-series alignment, taking both sample and marker contributions into account during deformation estimation.

Acknowledgments

This work was supported by the European Union's Marie Skłodowska-Curie Innovative Training Network MUMMERING (grant agreement no. 765604) and the Netherlands Organisation for Scientific Research (NWO) (project number 613.009.106). E.F. thanks Felix de Haas, Sebastian Unger, and Oliver Raschdorf for their help with sample preparation and experimental data collection. P.S.G. thanks Roberto Bondesan for fruitful discussions.

References

[1] M. Turk and W. Baumeister, "The promise and the challenges of cryo-electron tomography," *FEBS letters*, vol. 594, no. 20, pp. 3243–3261, 2020.

[2] R. I. Koning, A. J. Koster, and T. H. Sharp, "Advances in cryo-electron tomography for biology and medicine," *Annals of Anatomy - Anatomischer Anzeiger*, vol. 217, pp. 82–96, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0940960218300219

[3] M. Chen, J. M. Bell, X. Shi, S. Y. Sun, Z. Wang, and S. J. Ludtke, "A complete data processing workflow for cryo-et and subtomogram averaging," *Nature methods*, vol. 16, no. 11, pp. 1161–1168, 2019.

[4] E. Pyle and G. Zanetti, "Current data processing strategies for cryo-electron tomography and subtomogram averaging," *Biochemical Journal*, vol. 478, no. 10, pp. 1827–1845, 2021.

[5] M. Vulović, R. B. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lücken, H. Rullgård, O. Öktem, and B. Rieger, "Image formation modeling in cryo-electron microscopy," *Journal of structural biology*, vol. 183, no. 1, pp. 19–32, 2013.

[6] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," *IEEE signal processing magazine*, vol. 37, no. 2, pp. 58–76, 2020.

[7] O. Öktem, *Mathematics of electron tomography*, 2015, qC 20160218. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-181248

[8] F. Amat, D. Castaño-Díez, A. Lawrence, F. Moussavi, H. Winkler, and M. Horowitz, "Alignment of cryo-electron tomography datasets," *Methods in enzymology*, vol. 482, pp. 343–67, 12 2010.

[9] C. J. Russo and R. Henderson, "Charge accumulation in electron cryomicroscopy," *Ultramicroscopy*, vol. 187, pp. 43–49, 2018.

[10] A. F. Brilot, J. Z. Chen, A. Cheng, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher, R. Henderson, and N. Grigorieff, "Beam-induced motion of vitrified specimen on holey carbon film," *Journal of structural biology*, vol. 177, no. 3, pp. 630–637, 2012.

[11] J.-J. Fernandez, S. Li, T. A. Bharat, and D. A. Agard, "Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion," *Journal of structural biology*, vol. 202, no. 3, pp. 200–209, 2018.

[12] B. A. Himes and P. Zhang, "emclarity: software for high-resolution cryo-electron tomography and subtomogram averaging," *Nature methods*, vol. 15, no. 11, pp. 955–961, 2018.

[13] D. Tegunov, L. Xue, C. Dienemann, P. Cramer, and J. Mahamid, "Multi-particle cryo-em refinement with m visualizes ribosome-antibiotic complex at 3.5 å in cells," *Nature Methods*, vol. 18, no. 2, pp. 186–193, 2021.

[14] G. Chreifi, S. Chen, L. A. Metskas, M. Kaplan, and G. J. Jensen, "Rapid

tilt-series acquisition for electron cryotomography," *Journal of structural biology*, vol. 205, no. 2, pp. 163–169, 2019.

[15] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[16] S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, and D. A. Agard, "Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy," *Nature methods*, vol. 14, no. 4, pp. 331–332, 2017.

[17] F. Natterer, *The mathematics of computerized tomography*. SIAM, 2001.

[18] J. Modersitzki, *Numerical methods for image registration*. OUP Oxford, 2003.

[19] G. S. Alberti, H. Ammari, F. Romero, and T. Wintz, "Dynamic spike superresolution and applications to ultrafast ultrasound imaging," *SIAM Journal on Imaging Sciences*, vol. 12, no. 3, pp. 1501–1527, 2019.

[20] P. S. Ganguly, F. Lucka, H. J. Hupkes, and K. J. Batenburg, "Atomic super-resolution tomography," *arXiv preprint arXiv:2002.00710*, 2020.

[21] M. Frank, P. Wolfe *et al.*, "An algorithm for quadratic programming," *Naval research logistics quarterly*, vol. 3, no. 1-2, pp. 95–110, 1956.

[22] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1858–1865, 2008.

[23] F. J. Harris, *Multirate signal processing for communication systems*. River Publishers, 2021.

[24] H. Rullgård, L.-G. Öfverstedt, S. Masich, B. Daneholt, and O. Öktem, "Simulation of transmission electron microscope images of biological specimens," *Journal of microscopy*, vol. 243, no. 3, pp. 234–256, 2011.

[25] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[26] F. J. Anscombe, "The transformation of poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.

[27] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal mass transport: Signal processing and machine-learning applications," *IEEE signal processing magazine*, vol. 34, no. 4, pp. 43–59, 2017.

[28] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the national academy of sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.