

Accelerating error correction in tomographic reconstruction

Sajid Ali ^{1,4}✉, Matthew Otten^{2,5} & Z. W. Di³

Spurred by recent advances in detector technology and X-ray optics, upgrades to scanning-probe-based tomographic imaging have led to an exponential growth in the amount and complexity of experimental data and have created a clear opportunity for tomographic imaging to approach single-atom sensitivity. The improved spatial resolution, however, is highly susceptible to systematic and random experimental errors, such as center of rotation drifts, which may lead to imaging artifacts and prevent reliable data extraction. Here, we present a model-based approach that simultaneously optimizes the reconstructed specimen and sinogram alignment as a single optimization problem for tomographic reconstruction with center of rotation error correction. Our algorithm utilizes an adaptive regularizer that is dynamically adjusted at each alternating iteration step. Furthermore, we describe its implementation in a software package targeting high-throughput workflows for execution on distributed-memory clusters. We demonstrate the performance of our solver on large-scale synthetic problems and show that it is robust to a wide range of noise and experimental drifts with near-ideal throughput.

¹Applied Physics, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA. ²Center for Nanoscale Materials, Argonne National Laboratory, 9700 Cass Ave., Lemont, USA. ³Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL 60439, USA. ⁴Present address: Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA. ⁵Present address: HRL Laboratories, LLC, 3011 Malibu Canyon Road, Malibu, CA 90265, USA. ✉email: sasyed@fnal.gov

Tomography is an imaging technique based on combining 2D projections from multiple rotation angles to recover a 3D sample. It is a versatile technique that is used in the context of medical imaging with x-rays where radiation doses are kept to a minimum^{1,2}, in (cryo) electron microscopy^{3–6}, and in x-ray microscopy at synchrotrons^{7,8}, among many others.

A big challenge in achieving high-resolution images is sample/beam position drifts (caused by experimental imperfections), which corrupt the captured sinograms. One consequence of such sample/beam drifts is that the center of rotation is no longer at the center of the rotation (CoR) plane, as is assumed by standard tomographic reconstruction methods. Moreover, these errors can be different per projection instead of being a single offset for the entire acquisition range. While some imaging apparatus incorporates advanced metrology systems to correct for these errors^{9–12}, their usage is limited. Briefly speaking, two types of approaches account for the per-projection center of rotation drifts when performing reconstructions. The first type treats error correction as a preprocessing step by aligning the projections first, followed by a standard reconstruction. Alignment strategies include cross-correlation between adjacent projections¹³, an inspection of opposite projections and tracking the movement of reference (or marker) objects, feature-tracking-based alignment systems^{14–16}, and marker-free alignment by utilizing a metric such as the center of mass^{17,18} or geometric moments¹⁹. The main drawback of such two-step approaches is that they neglect the correlation between experimental configuration and reconstruction results and risk the accumulation and propagation of subsequent drift across the whole experiment process.

Recent development has focused on a second approach: the simultaneous error correction and reconstruction for more coherent and robust performance. For example, reprojection methods²⁰ seek to reproject the sinogram either as part of an iterative reconstruction process^{21–23} or jointly alongside the reconstruction^{24,25}. In some cases, multiscale methods are employed that downsample the projections and align them at the lower resolution prior to working with the full-scale data²⁶. Optimization-based methods seek to minimize some cost function that accounts for the drifts and reconstruction simultaneously, in which gradient-based methods are often employed as the underlying solver^{23,27–30}.

Owing to ongoing (and planned) upgrades at synchrotron light sources to fourth-generation storage rings^{31,32}, alongside advances in detector technology and x-ray optics, the amount of available photon fluence is expected to grow by 2–3 orders of magnitude. Improved resolution will reveal sample structures with an unprecedented clarity; but the results are highly susceptible to uncertainty, scale, non-stationarity, noise, and heterogeneity, which are fundamental impediments to progress at all phases of the pipeline for creating knowledge from data. The increases in data rates necessitate the need for high-throughput software that can process the experimental data, while being robust to experimental drifts and noise. Specifically, robust methods and libraries are needed that deliver good error correction capability and can deal with large-scale datasets through high-performance computing, either by increasing efficiency on a single compute node or by harnessing multiple compute nodes. A large number of distributed-memory parallel solvers for standard tomographic reconstructions exist^{33–39}; however, they lack the capabilities of robust error correction.

In this article, we present “parallel iterative reconstruction for tomography,” hereafter referred to as “PIRT”. While we intend to have PIRT as a general-purpose tomographic reconstruction library given its good scalability, we emphasize its algorithmic capability enabled by the underlying model-based approach, enables simultaneous optimization of the sample reconstruction and sinogram alignment as a single problem while leveraging L1 regularization. The objective of PIRT is to address the need for robust error

correction while also being suitable for distributed-memory clusters and high-performance computing, specific to parallel-beam x-ray tomography. We begin with quantitative examinations of our solver, followed by examinations of scaling and throughput in “Results and discussion”. We then detail our mathematical model and present implementation details in “Methods”.

Results and discussion

Model overview. We formulate the center of rotation drifts based on the model proposed in ref. ²⁷ (see “Methods”). Briefly, we note that each center of rotation drift is parameterized by a scalar parameter (thus, the center of rotation drifts for a complete scan is represented by a vector of these parameters). The drift at each projection is modeled by convolving the drift-free sinogram with a Gaussian function whose parameters are determined by the magnitude of the drift. One way to find the optimal values for both the drift parameter and sample parameter, as proposed in ref. ²⁷, is to combine the drifts and sample into a single unknown parameter set and optimize for both simultaneously; we refer to this as “joint”. Alternatively, we propose to perform the optimization following an alternating fashion between the drifts and the sample. We use a bound-constrained quasi-Newton solver with line search as the underlying solver for the corresponding optimization problem, where the only constraint we employ is the non-negativity on the sample as a natural physics-aware constraint. We run the optimization solver for each subproblem (drifts and sample) for a fixed number of iterations. Our proposed PIRT approach is outlined in Algorithm 1.

Algorithm 1. Alternating tomography solver

```

Input: Experimental data  $\mathbf{D}$ 
Result: Reconstructed sample & corrected sinogram
/* parse inputs, create vectors & distribution
   layouts *
 $\mathcal{W} \leftarrow 0$ 
/* Execute warm-starts *
for  $n_{\text{warm-start-idx}}=1, n_{\text{total-warm-starts}}$  do
  /* corrected sinogram from previous warm-start
   as input *
  if  $n_{\text{warm-start-idx}}=1$  then
    |  $\tilde{\mathbf{D}} \leftarrow \mathbf{D}$ 
  end

   $\mathbf{P} \leftarrow 0$ 
  /* set regularization weight *
   $\lambda = \lambda_0 * (1 - \gamma * \frac{n_{\text{warm-start-idx}}}{n_{\text{total-warm-starts}}})$ 
  AltSolve( $\mathcal{W}, \tilde{\mathbf{D}}, \mathbf{P}$ )
end

Procedure AltSolve( $\mathcal{W}, \tilde{\mathbf{D}}, \mathbf{P}$ )
  for  $n_{\text{alt-outer-idx}}=1, n_{\text{alt-outer-its}}$  do
    |  $\tilde{\mathbf{D}} \leftarrow \text{driftsSolve}(\mathcal{W}, \tilde{\mathbf{D}})$ 
    |  $\mathcal{W} \leftarrow \text{SampleSolve}(\tilde{\mathbf{D}})$ 
  end
  return

Procedure SampleSolve( $\tilde{\mathbf{D}}$ )
  /* fix  $\mathbf{P}$ , minimize Eq. (8) wrt to  $\mathcal{W}$  *
  L1-regularized BQNLS solver for sample for
   $n_{\text{alt-sample-its}}$ ;
  return  $\mathcal{W}$ 

Procedure driftsSolve( $\mathcal{W}, \tilde{\mathbf{D}}$ )
  /* fix  $\mathcal{W}$ , minimize Eq. (8) wrt to  $\mathbf{P}$  *
  BQNLS solver for drifts for  $n_{\text{alt-drifts-its}}$ ;
  return  $\tilde{\mathbf{D}}$ 

```

We use XDesign⁴⁰ to generate phantoms of varying sizes and illustrate them alongside the results of our computational

experiments. We simulate sinograms with the center of rotation drifts following a normal distribution, where the percentage of drifts generated is with respect to the sample domain. The projection angles for all cases are uniformly generated between 1 and 2π .

We add varying amounts of Gaussian noise to the generated sinograms to test the robustness of our solver to the presence of noise. The Gaussian noise generated is centered on 0 with the standard deviation proportional to the highest beam intensity in the generated sinogram. Although photon shot noise present at detectors is more accurately described by a Poisson distribution, given the large photon count in hard x-ray sources, it can be well approximated by a Gaussian distribution. The added negative noise may cause nonphysical negative values in the sinogram as measured intensity by the detector; therefore, during the random drawing process, we keep only the positive intensities.

The projection approximation⁴¹ states that diffraction within the sample can be neglected, thereby allowing the projections from each 2D slice of a 3D sample to be completely independent of other slices. We assume its validity for this study. First, we extensively characterize the solvers using 2D images before moving on to experiments on throughput for solving 3D samples.

Ill-posedness. In theory, the number of projection angles N_θ required to fill the Fourier space of projection data is given by the Crowther criterion⁴², which sets the required number for a square sample with N^2 pixels as

$$N_{\text{Crowther}} = \frac{\pi}{2} N. \quad (1)$$

In general, however, it is impossible to satisfy the Crowther criterion, that is, $N_\theta < N_{\text{Crowther}}$, which makes that problem an ill-posed inverse problem; in other words, the given data is not sufficient for a unique reconstruction of the sample. Consequently, the corresponding optimization problem has nonunique local minima. In addition, the added drift parameters further complicate and expand the degree of freedom and result in an even more ill-posed problem. For example, there exists more than one set of drifts that give rise to the same sinogram from a given phantom depending on where the center of the field of view is. Therefore, an error metric that is invariant to translation is necessary. We use the normalized translation-invariant root mean square error (hereafter referred to as nTIE)⁴³, as implemented in the `scikit-image` package⁴⁴, as our error metric for all results in this article. Since the dimensions of the reference phantom and the reconstructed object are fixed even when the number of projection angles change, we always report the nTIE of the reconstructed object measured against the reference phantom unless otherwise noted. One could choose to measure the nTIE of the corrected sinogram against a drift-free and noise-free reference sinogram, but when the number of projection angles changes, the dimensions of the reference sinogram are also affected, thereby preventing the use of a single reference sinogram.

A critical question for the proposed alternating mode concerns the number of optimization iterations for each subproblem (drifts and sample). Let us denote the number of inner iterations for the drifts and sample solver by $n_{\text{alt-drifts-its}}$ and $n_{\text{alt-sample-its}}$, respectively. Let $n_{\text{alt-outer-its}}$ denote the total number of outer iterations; that is, one outer iteration consists of $n_{\text{alt-drifts-its}}$ iterations of drifts optimization and $n_{\text{alt-sample-its}}$ iterations of sample optimization. We can expect that performing too many iterations for either the drifts or the sample can cause the optimization to become stuck in an undesired local minimum. To investigate the role of the subproblem iteration numbers and deduce the (approximately) optimal set of parameters for the

alternating mode of reconstruction, we perform a parameter sweep over all possible combinations of optimization iterations for the solver and drifts for a given number of total outer iterations, as shown in Fig. 1a. As a result, we choose the final parameters for the alternating mode of the reconstruction as $n_{\text{alt-drifts-its}} = 2$ and $n_{\text{alt-sample-its}} = 5$ throughout the rest of the tests. Increasing the number of outer iterations for the alternating method does not necessarily lead to a better solution even as the cost function decreases. The reason is partially due to the ill-posed nature of tomography but also due to the well-known “discrepancy principle” that states that the solution from an inverse problem is always bounded by the noise present in data⁴⁵.

We avoid getting stuck in a local minimum by performing warm start of the optimization procedure where we checkpoint the corrected sinogram and restart the optimization process with the checkpointed sinogram instead of the original sinogram. To be more specific, our warm start is implemented in the way that after each complete alternating solve procedure, the current corrected sinogram will be used as the input experimental data and restart the alternating solve using the default initial guess for the drifts and sample. As seen in Fig. 2, both the joint and alternating solvers get to a better solution with an increasing number of warm starts. To further stabilize the performance and account for the discrepancy principle, we introduce an adaptive L1 regularizer for the alternating solver, where the regularizer parameter is dynamically adjusted along each warm start.

We first test the validity of our algorithm by comparing the results using a null initialization with the exact drifts (that were used to generate the sinogram) as initial guesses. For this test, we use a 1024^2 softwood sample generated by XDesign⁴⁰, and we generate the sinogram while inducing 10% center of rotation drifts, with no added noise. The results of this experiment are illustrated in Fig. 3. With a null initialization, we see that both of the solvers are able to recover the sample; however, the alternating solver outperforms the joint solver by recovering a normalized translation-invariant RMS error of 0.190 as compared with 0.572 from the joint solver. As a confidence check, when the drifts are initialized by the drifts that were used to generate the sinogram (hereafter referred to as “exact drifts”), we see that both solvers are able to incorporate this knowledge and remain within the neighborhood of this optimal solution with comparable results.

The next-generation light sources will improve the spatial resolution to reveal sample structures with an unprecedented clarity; however, they are also highly susceptible to systematic and random error. Next, we investigate the performance of the methods for various levels of ill-posedness, namely, with a reduced number of projection angles at various noise levels. We use the same softwood phantom as before but generate sinograms by varying the number of projection angles at three different noise levels, while inducing 10% center of rotation drifts. The results of this experiment are illustrated in Fig. 4. As expected, the solution quality improves as we increase the number of projection angles (as indicated by the nTIE metric). Moreover, with an extremely limited number of projection angles, both the joint and alternating solvers are able to outperform the standard reconstruction without error correction, while the alternating solver consistently performs better than the joint solver.

Our next experiment involves studying the effect of noise on the reconstruction quality. Given the same 1024^2 softwood sample with a fixed 200 projection angles, we add varying amounts of the center of rotation drifts and added Gaussian noise to its ideal sinogram. The Gaussian noise is generated as a distribution centered on 0 and width given as a percentage of the maximum value in the sinogram; any negative values that arise after the addition of the noise are clipped to ensure non-

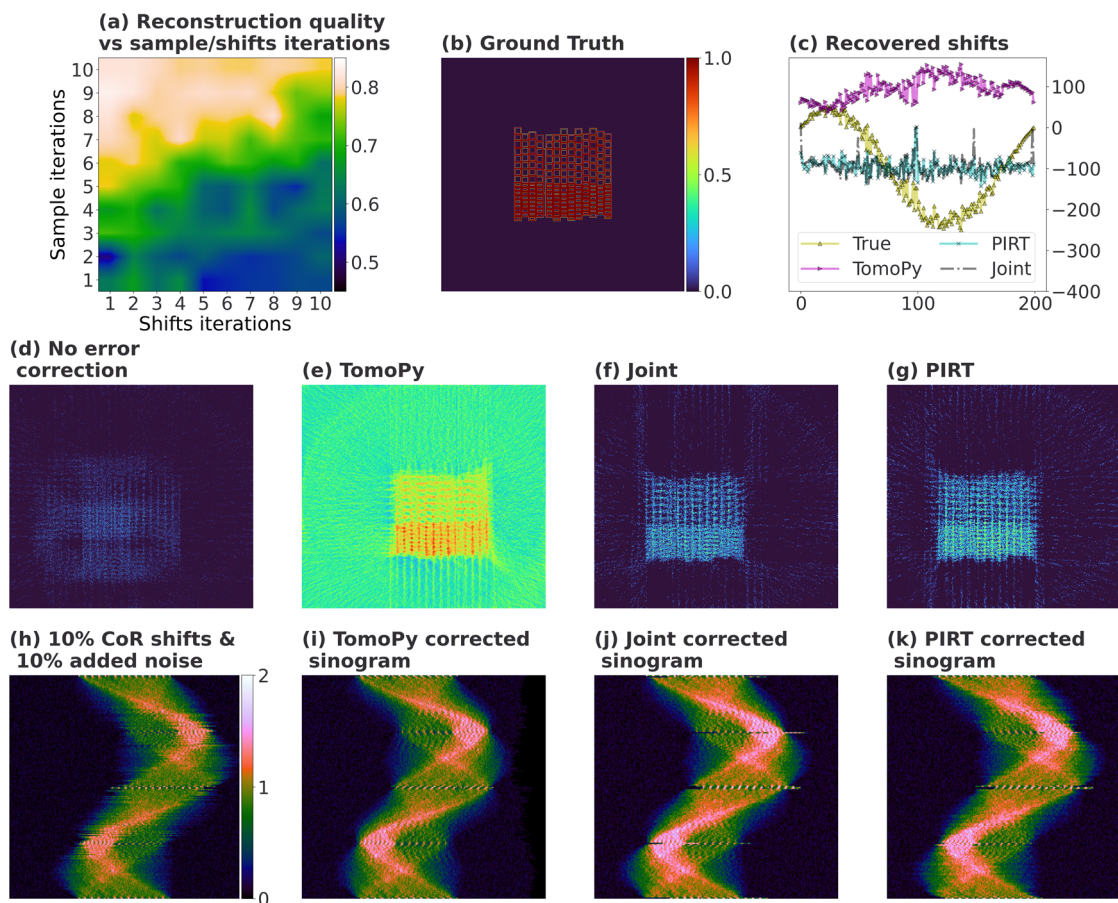


Fig. 1 Reconstruction comparison of our proposed method PIRT and the standard existing approaches. **a** Solution quality (as measured by the normalized, translation-invariant root mean square error (nTIE)⁴³) of our proposed alternating reconstruction mode as a function of iteration budget distribution. The color bar alongside the figure indicates the color scale of the nTIE metric (lower is better). **b** The ground truth phantom generated by XDesign of size 1024 × 1024. All reconstructed samples use the same color scale. **c** Illustration of the true and reconstructed center of rotation drifts returned by PIRT, joint, and TomoPy, respectively. **d** Reconstruction without error correction. **e, i** Reconstruction by the reprojection method implemented in TomoPy²⁴, yielding an nTIE of 0.79 and its corresponding corrected sinogram. **h** Simulated sinogram with 200 projections (well below the Crowther criterion), 10% CoR drifts, and 10% added Gaussian noise. All sinograms use the same color scale. **f, j** Reconstruction by the joint method proposed in refs. 27,28 yielding an nTIE of 0.81 and its corrected sinogram. **g, k** Reconstruction by our proposed PIRT yielding an nTIE of 0.77 and its corresponding corrected sinogram. The fact that seems very different sets of drifts all lead to reasonable reconstructions implies that the solution to the reconstruction with error correction is not unique. This test was conducted on a dual-socket system with Intel Xeon Silver 4110 CPUs.

negativity of the resulting sinogram. The results from this experiment are shown in Fig. 5. We clearly see that the two error-correcting modes of reconstruction perform far better than the non-error-correcting mode, with the alternating mode of reconstruction being more robust to induced drifts and added noise than is the joint mode of reconstruction.

Throughput results. Another fundamental impediment to today’s tomographic reconstruction is its increasing data collection rate, which puts a tremendous computational burden on data processing. We first demonstrate the scalability of PIRT when solving a single large 2D slice. We again use the XDesign⁴⁰ package to generate phantoms of varying sizes to reconstruct via PIRT. For the strong-scaling study, we set the sample size to be 2896² and the number of projection angles to be 800. For the weak-scaling study, we set the sample size and number of angles per node to be 1024² and 100, respectively. The naive data distribution over MPI ranks, described in the section “Distributed-memory parallel implementation”, causes poor scalability. While we could gain a limited amount of scalability by reducing the load imbalance (whereby each MPI rank gets approximately the same amount of work assigned to it for the sparse matrix-vector

multiplication routines that dominate the overall runtime), we do not explore the space of tomography-specific decomposition schemes^{36,46,47}, since one of our goals is to have a flexible error-correcting capability. However, instead of focusing on achieving good scalability for solving a single 2D slice, we instead focus on achieving better throughput for solving a set of 2D slices that make up a 3D sample.

Owing to the poor scalability of the 2D solver, we avoid solving for a 3D sample as a series of sequential 2D solves, each of which is executed in parallel (i.e., where all the nodes are used to solve for one 2D slice at a given time and the set of 2D slices are solved sequentially). Instead, we propose a two-level hierarchical parallelism in our solver, whereby we split the MPI communicator into smaller MPI subcommunicators. We then launch independent instances of the solver on each “task” subcommunicator and execute all instances concurrently. This approach limits each solver instance to a regime of high performance (by not running it in its strong-scaling regime) and gives us high overall throughput, as measured by the number of slices solved per unit of time, illustrated in Fig. 6. Compared with the naive concurrent parallelism afforded by launching independent jobs (via, say, SLURM’s job array), our approach creates the

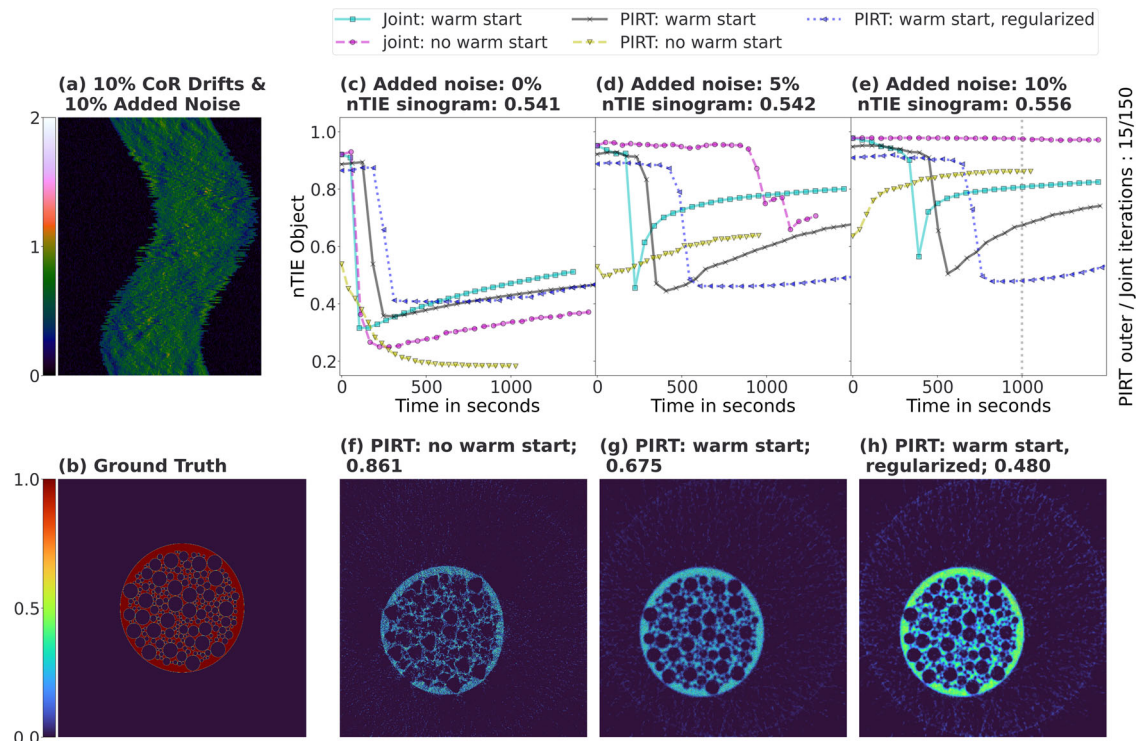


Fig. 2 Performance of warm start and regularization. **a** The input sinogram, generated by using 200 projection angles and with 10% drifts and 10% added noise. **b** The ground truth phantom generated by XDesign of size 1024×1024 . All reconstructed samples use the same color scale. **c–e** The result (as indicated by the nTIE metric) of each variation of the PIRT method for varying amounts of added noise, as a function of runtime. **f–h** The results from each variation of the PIRT method after a runtime of 1,000 seconds (as indicated by the dotted gray line on the corresponding plot of error metric) for 10% added noise, where the no-warm-start method yields an nTIE of 0.861, the warm start method yields an nTIE of 0.675, and the regularized warm start method yields an nTIE of 0.480.

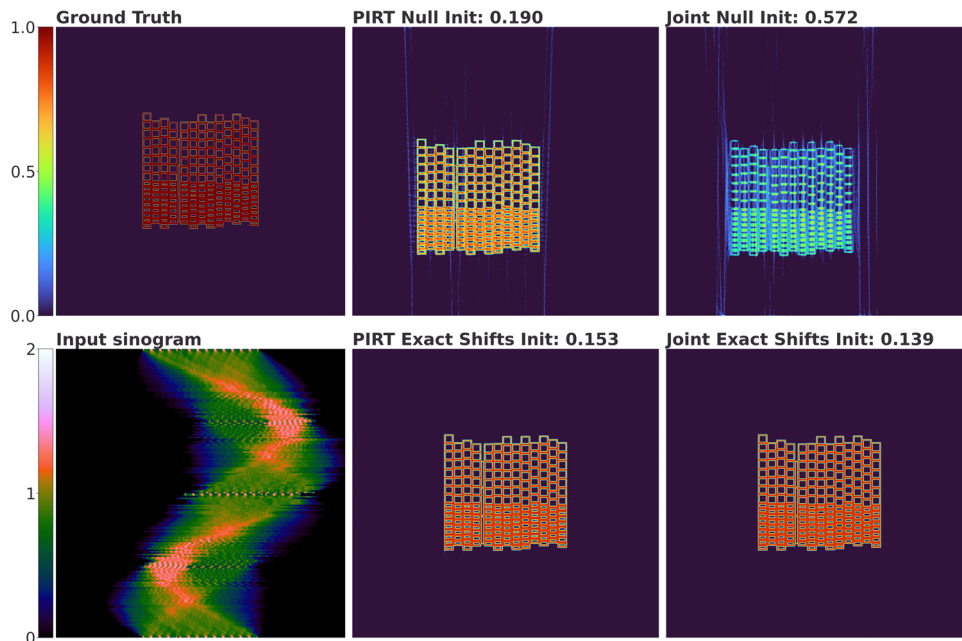


Fig. 3 Sensitivity of PIRT to different initial guesses. Comparing the reconstruction qualities between a null initialization and exact drifts as initialization. Compared with the joint solver, PIRT is much more robust to different initial guess qualities and is able to produce good reconstruction even with an initial guess far from the true solution. All the object images use the same color scale.

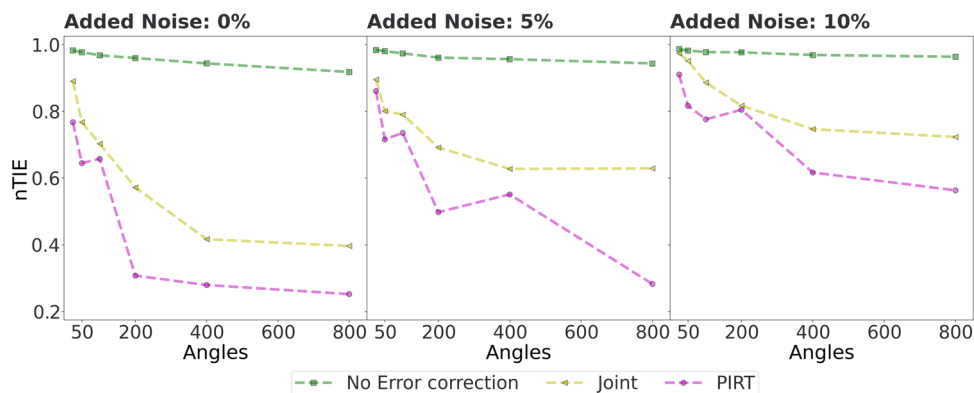


Fig. 4 Sensitivity of different methods to the amount of available data (in terms of the number of projection angles). While for all the methods the solution quality improves as the number of projection angles increases, the error correction feature provided by the joint solver and by PIRT improves the reconstruction quality dramatically. Moreover, PIRT outperforms the joint solver even further, especially for the severely ill-posed cases with limited angles.

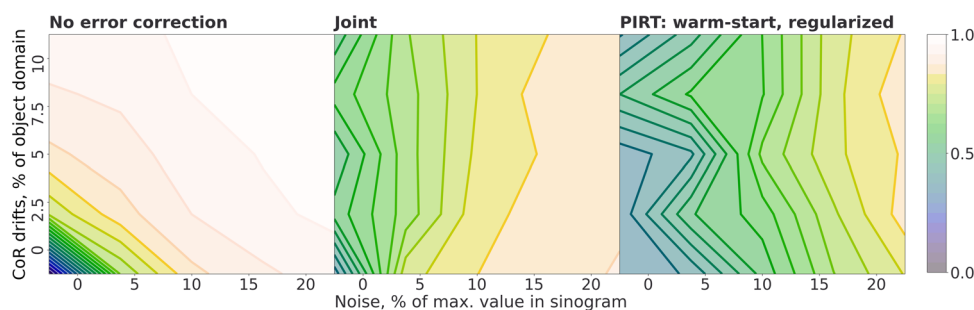


Fig. 5 Performance of the three modes of reconstruction evaluated by the nTIE metric as a function of the center of rotation drifts and noise. The alternating mode of reconstruction had an adaptive regularizer term as described in Eq. (11). The joint mode of reconstruction had a fixed number of warm-starts (=15). Note that the plots have been generated by smoothing over 25 data points (five values of noise and five values of CoR drifts were used).

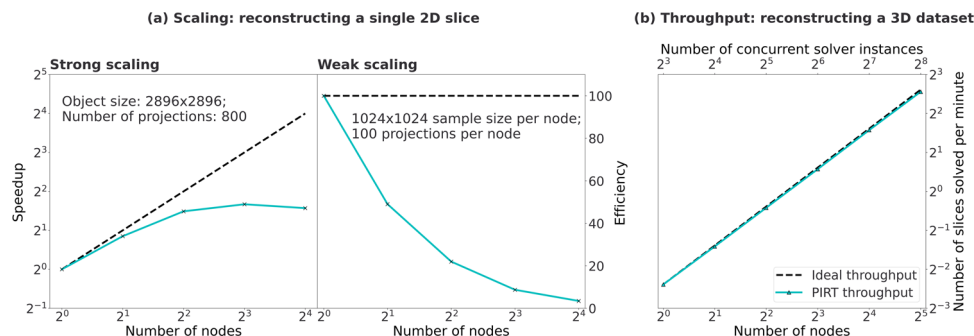


Fig. 6 Strong and weak scaling results for reconstructing a single slice and throughput for reconstructing a batch of slices. **a** Strong-scaling results for solving a single 2D slice of dimension 2896² and 800 projections over a varying number of nodes using the regularized alternating solver, with warm starts. **a** Weak-scaling results for solving a single 2D slice, with sample size and projection angles per node set to 1024² and 100, respectively, using the regularized alternating solver with warm starts. The naive distribution of data over MPI ranks hinders scalability. **b** Throughput for the regularized alternating solver, with warm starts. With two-level hierarchical MPI parallelism, PIRT achieves high throughput. These experiments were conducted on the Bebop cluster at the Argonne Laboratory Computing Resource Center⁶⁰, where each node contains a dual-socket Intel Broadwell Xeon, which are connected together by an Intel OmniPath fabric interconnect.

tomography projection matrix once per solver instance and reads the projection angles only once (for the entire job).

Conclusions. We propose an algorithm for error-correcting tomography, namely, the regularized alternating solver with warm starts, which outperforms the existing drift correction methods including the joint solver and the reprojection approach (as illustrated in Fig. 1). We further improve the scalability of the proposed algorithm by implementing a distributed-memory

parallel solver PIRT using the PETSc/TAO framework^{48–51}. PIRT’s capability to execute multiple instances concurrently in parallel allows for high-throughput, error-correcting reconstruction of very large datasets as expected from the imminent upgrades to synchrotron light sources.

Methods

Mathematical model. The Radon transform^{52,53} allows one to transform an image from real space to “sinogram” space. Its action upon a function, $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, is

given by

$$\mathbf{R}f(\tau, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\tau - x \cos(\theta) - y \sin(\theta)) dx dy, \quad (2)$$

where δ is the Dirac delta function, the (x, y) coordinates represent the real space, and the (τ, θ) coordinates represent the sinogram space, with τ referring to the position along the detector and θ referring to the projection angle.

Denote the center of rotation for the projection angle θ as $(x_{\theta}^*, y_{\theta}^*)$. The corresponding projection is given by

$$\begin{aligned} \mathbf{R}f(\tau, \theta, x_{\theta}^*, y_{\theta}^*) \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\tau - P_{\theta} - x \cos(\theta) - y \sin(\theta)) dx dy, \end{aligned} \quad (3)$$

where

$$P_{\theta} = x_{\theta}^* (1 - \cos(\theta)) + y_{\theta}^* \sin(\theta). \quad (4)$$

Therefore, we establish the relationship between the true projection (with COR $(x_{\theta}^*, y_{\theta}^*)$) and the theoretical projection (with COR $(0, 0)$) as

$$\mathbf{R}f(\tau, \theta, 0, 0) = \mathbf{R}f(\tau - P_{\theta}, \theta, x_{\theta}^*, y_{\theta}^*). \quad (5)$$

In practice, because of the nature of the physical experiment, we discretize the space (containing the compact sample) being imaged into $N \times N$ pixels and vectorize it into an array \mathcal{V} . Let \mathcal{W}_v be the value of the sample property we intend to recover on the pixel $v \in \mathcal{V}$. Then we have $\mathcal{W} = \{\mathcal{W}_v : v \in \mathcal{V}\}$ denoting the discretized image. Given parameters τ and θ , we calculate a discrete Radon transform of \mathcal{W} via

$$\mathbf{R}_{\tau, \theta}(\mathcal{W}) = \sum_{v \in \mathcal{V}} L_{v, \tau, \theta}^{\tau, \theta} \mathcal{W}_v, \quad (6)$$

where $L_{v, \tau, \theta}^{\tau, \theta}$ denotes the proportional amount of contribution from pixel v to the corresponding projection, which is approximated by the length of the intersection of the beam described by τ and θ with the pixel v .

One can explicitly recover the experimental configuration (e.g., sample and scanning position) to address the misalignment, which can be embedded in the forward Radon transform operator. Doing so, however, is computationally expensive given the iterative construction of individual $L_{v, \tau, \theta}^{\tau, \theta}$ with the dependency on the unknown experiment parameters. Instead, we adapt the numerical approach proposed in ref. 27, which seeks the optimal drifting parameters to translate the misaligned projection (i.e., with drift error) to drift-free projection in order to optimally match the nominal $\mathbf{L} = [L_{v, \tau, \theta}^{\tau, \theta}] \in \mathbb{R}^{N_{\theta} N_r \times N^2}$. Let $\mathbf{D} = [\mathbf{D}_{\theta}] \in \mathbb{R}^{N_{\theta} \times N_r}$ be the misaligned (i.e., observed) sinogram, and $\tilde{\mathbf{D}} = [\tilde{\mathbf{D}}_{\theta}] \in \mathbb{R}^{N_{\theta} \times N_r}$ be the drift-free sinogram. For each angle θ , we approximate $\tilde{\mathbf{D}}_{\theta}$ by convolving \mathbf{D}_{θ} with a P_{θ} -mean Gaussian function:

$$\tilde{\mathbf{D}}_{\theta} = \mathbf{D}_{\theta} * \left(\frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ \frac{-(\tau - P_{\theta})^2}{2\sigma^2} \right\} \right), \quad (7)$$

where $*$ denotes convolution and σ determines the amount of smoothing introduced by the Gaussian to \mathbf{D}_{θ} . In practice, we choose $\sigma = 2.0/2.355 \approx 0.84$, where the full width at half maximum (FWHM) of the Gaussian is 1 unit of the beam width (2.355 is the FWHM of the standard Gaussian), so that smoothing artifacts are smaller than the actual beam width. Therefore, we have the final optimization problem to recover both $\mathbf{P} = [P_{\theta}]_{\theta=1}^{N_{\theta}}$ and \mathcal{W} :

$$\min_{\mathcal{W} \geq 0, \mathbf{P}} \phi(\mathcal{W}, \mathbf{P}) = l(\mathcal{W}, \mathbf{P}) + \lambda R(\mathcal{W}), \quad (8)$$

where $l(\mathcal{W}, \mathbf{P}) = \frac{1}{2} \|\mathbf{L}\mathcal{W} - \mathbf{g}(\mathbf{D}, \mathbf{P})\|^2$ is the misfit term, $R(\mathcal{W})$ is a regularization term to incorporate prior knowledge on \mathcal{W} , and λ is the regularization weight. We can analytically compute the gradient of l as

$$\nabla l(\mathcal{W}, \mathbf{P}) = [\mathbf{L}^T, \nabla_{\mathbf{P}} \mathbf{g}(\mathbf{D}, \mathbf{P})]^T (\mathbf{L}\mathcal{W} - \mathbf{g}(\mathbf{D}, \mathbf{P})), \quad (9)$$

where

$$\nabla_{\mathbf{P}} \mathbf{g} = \left[\mathbf{D}_{\theta, \cdot} * \left(\frac{1}{\sigma \sqrt{2\pi}} \exp \left(\frac{-(\tau - \mathbf{P}_{\theta})^2}{2\sigma^2} \right) \right) \circ \frac{\tau - \mathbf{P}_{\theta}}{\sigma^2} \right]_{\theta=1}^{N_{\theta}}. \quad (10)$$

L1 regularization is known to promote sparsity in inverse problems⁵⁴, which we adapt in this work as $R(\mathcal{W})$. Having a regularizer weight that is too large prevents the solver from obtaining the optimum solution, and having a regularizer weight that is too small fails to prevent the instability we observe. Thus, we choose an adaptive L1 regularizer where the weight of the regularizer decreases as the number of warm starts increases⁵⁵. With the regularizer we see that the solution quality is stabilized and that, except for the noise-free case, the solution quality is superior to the nonregularized version. Given that the noise-free case is not of practical importance, henceforth we use the adaptive regularizer for the alternating solver.

We thus have

$$\lambda = \lambda_0 * \left(1 - \gamma \frac{n_{\text{warm-start-idx}}}{n_{\text{total-warm-starts}}} \right), \quad (11)$$

where λ is the regularization weight for the current warm start, λ_0 is the initial

regularization weight, γ represents the rate of decrease, and $n_{\text{warm-start-idx}}/n_{\text{total-warm-starts}}$ refer to the index of the current cycle and total number of warm starts, respectively. In this study, we choose $\lambda_0 = 5 \times 10^{-3}$ and $\gamma = 10^{-2}$ for the case of $n_{\text{total-warm-starts}} = 15$.

Optimization. Given the analytical derivatives, we employ a gradient-based optimization solver and focus on a quasi-Newton method to efficiently exploit the curvature information for robust convergence. We observe that the overall unknown parameters $[\mathbf{P}, \mathcal{W}]$ in Eq. (8) present a partial separability property, that is, the property that the Jacobian exhibits a structured sparsity pattern or, equivalently, a sparse Hessian. It was shown in ref. 56 that a quasi-Newton method without a specific treatment of the partially separable structure tends to poorly approximate the full Hessian and causes slow and inefficient convergence. Therefore, in contrast to joint minimization developed in ref. 27, we propose to optimize \mathbf{P} and \mathcal{W} separately and perform their update in an alternating fashion.

Distributed-memory parallel implementation. We implement PIRT using the PETSc/TAO framework^{48–51}, which provides the tools for developing large-scale applications using distributed-memory parallelism (via the use of the Message Passing Interface (MPI), as well as the use of graphical processing units via a range of frameworks). The Toolkit for Advanced Optimization (TAO)⁵¹ builds on the PETSc data structures and provides distributed-memory optimization routines. In addition to PETSc/TAO, we use the HDF5 library⁵⁷ to enable parallel file I/O (via its PETSc interface), Boost geometry⁵⁸ for intersection routines needed to generate the discretized Radon transform, and FFTW⁵⁹ for Fourier space convolutions.

We distribute the experimental data such that each MPI rank is assigned a set of projections from contiguous projection indices. This keeps the Fourier transforms local; that is, all the inputs for each transform are local to the MPI rank, thereby simplifying the process of convolving the data with a Gaussian kernel to account for the center of rotation drifts (as described by Eq. (7)). A simple formula that tries to keep the number of local projection angles assigned to different MPI ranks the same is used to determine the distribution of projection angles over MPI ranks.

To execute multiple reconstructions concurrently, PIRT uses PETSc subcommunicators, which are based on MPI subcommunicators but include some utilities to simplify usage. The global MPI communicator is split into task subcommunicators, and each task subcommunicator works on one tomography slice at any given time. In order to manage the distribution of slices over task subcommunicators (in other words to decide which task subcommunicator works on which slice at any given time), an organizer subcommunicator is also created. Each task subcommunicator is given a unique identifier: the highest MPI rank in the global MPI communicator of all the constituent ranks.

The organizer subcommunicator is created by combining all MPI ranks whose rank in the task subcommunicator is 0 (though its rank on global rank may not be 0), and a PETSc index set is then defined on this subcommunicator. The indices that are assigned to each rank (on this organizer subcommunicator) are taken to be the indices (or slice indices) that the associated task subcommunicator reconstructs.

Data availability

The data that support the findings of this study are available from the corresponding author on request.

Code availability

The code described in this manuscript is available at <https://gitlab.com/pirt/pirt>.

Received: 29 March 2022; Accepted: 17 June 2022;

Published online: 11 July 2022

References

1. Padole, A., Ali Khawaja, R. D., Kalra, M. K. & Singh, S. CT radiation dose and iterative reconstruction techniques. *Am. J. Roentgenol.* **204**, W384–W392 (2015).
2. Seeram, E. Computed tomography: a technical review. *Radiol. Technol.* **89**, 279CT–305CT (2018).
3. Saggi, Z. & Midgley, P. A. Electron tomography in the (s)tem: From nanoscale morphological analysis to 3D atomic imaging. *Annu. Rev. Mater. Res.* **42**, 59–79 (2012).
4. Hayashida, M. & Malac, M. Practical electron tomography guide: Recent progress and future opportunities. *Micron* **91**, 49–74 (2016).
5. Turk, M. & Baumeister, W. The promise and the challenges of cryo-electron tomography. *FEBS Lett.* **594**, 3243–3261 (2020).

6. Benjin, X. & Ling, L. Developments, applications, and prospects of cryo-electron microscopy. *Protein Sci.* **29**, 872–882 (2020).
7. Attwood, D. & Sakdinawat, A. *X-rays and Extreme Ultraviolet Radiation*, 2nd edn. (Cambridge University Press, Cambridge, 2017).
8. Jacobsen, C. *X-ray Microscopy* (Cambridge University Press, Cambridge, 2020).
9. Holler, M. et al. An instrument for 3D x-ray nano-imaging. *Rev. Sci. Instruments* **83**, 073703 (2012).
10. Wang, J. et al. Automated markerless full field hard x-ray microscopic tomography at sub-50-nm 3-dimension spatial resolution. *Appl. Phys. Lett.* **100**, 143107 (2012).
11. Kim, J., Lauer, K., Yan, H., Chu, Y. S. & Nazaretski, E. Compact prototype apparatus for reducing the circle of confusion down to 40 nm for x-ray nanotomography. *Rev. Sci. Instruments* **84**, 035006 (2013).
12. Villar, F. et al. in *MEDSI Mechanical Engineering Design of Synchrotron Radiation Equipment and Instrumentation* (Australian Synchrotron, 2014).
13. Brandt, S. S. *Markerless Alignment in Electron Tomography*, 187–215 (Springer New York, 2006).
14. Brandt, S., Heikkonen, J. & Engelhardt, P. Automatic alignment of transmission electron microscope tilt series without fiducial markers. *J. Struct. Biol.* **136**, 201–213 (2001).
15. Sorzano, C. O. S. et al. Marker-free image registration of electron tomography tilt-series. *BMC Bioinformatics* **10**, 124 (2009).
16. Castaño-Díez, D., Scheffer, M., Al-Amoudi, A. & Frangakis, A. S. Alignator: A GPU powered software package for robust fiducial-less alignment of cryo tilt-series. *J. Struct. Biol.* **170**, 117–126 (2010).
17. Azevedo, S., Schneberk, D., Fitch, J. & Martz, H. Calculation of the rotational centers in computed tomography sinograms. *IEEE Trans. Nucl. Sci.* **37**, 1525–1540 (1990).
18. Liu, Y., Penczek, P. A., McEwen, B. F. & Frank, J. A marker-free alignment method for electron tomography. *Ultramicroscopy* **58**, 393–402 (1995).
19. Wang, S. et al. Jitter correction for transmission X-ray microscopy via measurement of geometric moments. *J. Synchrotron Radiat.* **26**, 1808–1814 (2019).
20. Dengler, J. A multi-resolution approach to the 3D reconstruction from an electron microscope tilt series solving the alignment problem without gold particles. *Ultramicroscopy* **30**, 337–348 (1989).
21. Mayo, S., Miller, P., Gao, D. & Sheffield-Parker, J. Software image alignment for x-ray microtomography with submicrometre resolution using a sem-based x-ray microscope. *J. Microsc.* **228**, 257–263 (2007).
22. Parkinson, D. Y., Knoechel, C., Yang, C., Larabell, C. A. & Le Gros, M. A. Automatic alignment and reconstruction of images for soft x-ray tomography. *J. Struct. Biol.* **177**, 259–266 (2012).
23. Cheng, C.-C., Ching, Y.-T., Ko, P.-H. & Hwu, Y. Correction of center of rotation and projection angle in synchrotron x-ray computed tomography. *Sci. Rep.* **8**, 9884 (2018).
24. Gürsoy, D. et al. Rapid alignment of nanotomography data using joint iterative reconstruction and reprojection. *Sci. Rep.* **7**, 11818 (2017).
25. Wang, C.-C. Joint iterative fast projection matching for fully automatic marker-free alignment of nano-tomography reconstructions. *Sci. Rep.* **10**, 7330 (2020).
26. Latham, S. J., Kingston, A. M., Recur, B., Myers, G. R. & Sheppard, A. P. in *Developments in X-Ray Tomography X* (eds Stock, S. R., Müller, B. & Wang, G.) 50–61 (International Society for Optics and Photonics, SPIE, 2016).
27. Austin, A. P., Di, Z., Leyffer, S. & Wild, S. M. Simultaneous sensing error recovery and tomographic inversion using an optimization-based approach. *SIAM J. Sci. Comput.* **41**, B497–B521 (2019).
28. Di, Z. W. et al. Optimization-based simultaneous alignment and reconstruction in multi-element tomography. *Optics Lett.* **44**, 4331–4334 (2019).
29. Yang, C., Ng, E. G. & Penczek, P. A. Unified 3-d structure and projection orientation refinement using quasi-Newton algorithm. *J. Struct. Biol.* **149**, 53–64 (2005).
30. Bleichrodt, F. & Batenburg, K. J. in *Image Analysis* (eds Kämäräinen, J.-K. & Koskela, M.) 489–500 (Springer Berlin Heidelberg, 2013).
31. *J. Synchrotron Radiat.* (eds Eriksson, M. & van der Veen, J. F.) <https://journals.iucr.org/s/issues/2014/05/00/> (2014).
32. Eriksson, M., van der Veen, J. F. & Quitmann, C. Diffraction-limited storage rings – a window to the science of tomorrow. *J. Synchrotron Radiat.* **21**, 837–842 (2014).
33. Palenstijn, W. J., Bédorf, J., Sijbers, J. & Batenburg, K. J. A distributed ASTRA toolbox. *Adv. Struct. Chem. Imaging* **2**, 19 (2016).
34. Bicer, T. et al. Trace: a high-throughput tomographic reconstruction engine for large-scale datasets. *Adv. Struct. Chem. Imaging* **3**, 6 (2017).
35. Wang, X. et al. Massively parallel 3D image reconstruction. in *Proc. International Conference for High Performance Computing, Networking, Storage and Analysis, SC '17* (Association for Computing Machinery, New York, NY, USA, 2017).
36. Hidayetoglu, M. et al. MemXCT: Memory-Centric X-Ray CT Reconstruction with Massive Parallelization. In *Proc. International Conference for High Performance Computing, Networking, Storage and Analysis, SC '19* (Association for Computing Machinery, New York, 2019).
37. Chen, P., Wahib, M., Takizawa, S., Takano, R. & Matsuoka, S. IFDK: A scalable framework for instant high-resolution image reconstruction. In *Proc. International Conference for High Performance Computing, Networking, Storage and Analysis, SC '19* (Association for Computing Machinery, New York, 2019).
38. Marchesini, S., Trivedi, A., Enfedaque, P., Perciano, T. & Parkinson, D. in *Computational Science – ICCS 2020* (eds Krzhizhanovskaya, V. V. et al.) 248–261 (Springer International Publishing, Cham, 2020).
39. Hidayetoglu, M. et al. Petascale XCT: 3D image reconstruction with hierarchical communications on multi-GPU nodes. In *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*, 1–13 (2020).
40. Ching, D. J. & Gürsoy, D. XDesign: an open-source software package for designing x-ray imaging phantoms and experiments. *Journal of Synchrotron Radiation* **24**, 537–544 (2017).
41. Paganin, D. *Coherent X-ray Optics* (Oxford University Press, Oxford; New York, 2006).
42. Crowther, R. A., DeRosier, D. J. & Klug, A. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc. R. Soc. London. A. Math. Phys. Sci.* **317**, 319–340 (1970).
43. Fienup, J. R. Invariant error metrics for image reconstruction. *Appl. Opt.* **36**, 8352–8357 (1997).
44. van der Walt, S. et al. scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
45. Morozov, V. A. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.* **7**, 414–417 (1966).
46. Parashar, M. & Browne, J. On partitioning dynamic adaptive grid hierarchies. *Proc. HICSS-29: 29th Hawaii Int. Conf. Syst. Sci.* **1**, 604–613 (1996).
47. Campbell, P. M., Devine, K. D., Flaherty, J. E., Gervasio, L. G. & Teresco, J. D. Dynamic octree load balancing using space-filling curves. *Tech. Rep.* (2003).
48. Balay, S. et al. PETSc Web page. <https://www.mcs.anl.gov/petsc> (2019).
49. Balay, S. et al. PETSc users manual. ANL-95/11 Revision 3.15. *Tech. Rep.* <https://www.mcs.anl.gov/petsc> (2020).
50. Balay, S., Gropp, W. D., McInnes, L. C. & Smith, B. F. in *Modern Software Tools in Scientific Computing* (eds Arge, E., Bruaset, A. M. & Langtangen, H. P.) 163–202 (Birkhäuser Press, 1997).
51. Dener, A. et al. Tao users manual. ANL/MCS-TM-322 Rev. 3.15. *Tech. Rep.* <https://www.mcs.anl.gov/petsc> (2020).
52. Radon, J. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Akad. Wiss.* **69**, 262–277 (1917).
53. Radon, J. On the determination of functions from their integral values along certain manifolds. *IEEE Transact. Med. Imaging* **5**, 170–176 (1986).
54. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. Series B (Methodological)* **58**, 267–288 (1996).
55. Huang, X., Wild, S. M. & Di, Z. W. Calibrating sensing drift in tomographic inversion. In *2019 IEEE International Conference on Image Processing (ICIP)*, 1267–1271 (2019).
56. Nocedal, J. & Wright, S. *Numerical Optimization* (Springer Science & Business Media, 2006).
57. Folk, M., Heber, G., Koziol, Q., Pourmal, E. & Robinson, D. An overview of the HDF5 technology suite and its applications. In *Proc. EDBT/ICDT 2011 Workshop on Array Databases, AD '11*, 36–47 (Association for Computing Machinery, New York, NY, USA, 2011).
58. Schling, B. *The Boost C++ Libraries* (XML Press, 2011).
59. Frigo, M. & Johnson, S. G. The design and implementation of FFTW3. *Proc. IEEE* **93**, 216–231 (2005).
60. Laboratory, A. N. Bebob, laboratory computing resource center. <https://www.lrcr.anl.gov/systems/resources/bebob/> (2021).

Acknowledgements

We thank PETSc developers on the petsc-users mailing list for general guidance and bug fixes. In particular, we thank Alp Dener for helpful discussions and for implemented bug fixes in the TAO library. We gratefully acknowledge the computing resources provided on Bebob (and/or Blues), a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory. This material was based upon work supported by the U.S. Department of Energy under contract DE-AC02-06CH11357.

Author contributions

Z.D. developed the initial mathematical model that provided the basis for the software described in this article and provided the mathematical intuition and guidance for the development of various solver configurations. M.O. developed the first version of the software described here. S.A. refactored the software, implemented the hierarchical parallelism, and carried out all the experiments.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Sajid Ali.

Peer review information *Communications Materials* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Aldo Isidori.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022