ORNL/LTR-2016/544

Oak Ridge National Laboratory Official LTR Cover: Progress Update on Iterative Reconstruction of Neutron Tomographic Images

Paul Hausladen
Jens Gregor

September 15, 2016

Approved for public release.
Distribution is unlimited.

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE US DEPARTMENT OF ENERGY

Progress Update on Iterative Reconstruction of Neutron Tomographic Images

Paul Hausladen and Jens Gregor
September 15, 2016

Introduction

This report satisfies the fiscal year 2016 technical deliverable to report on progress in development of fast iterative reconstruction algorithms for project OR16-3DTomography-PD2Jb, "3D Tomography and Image Processing Using Fast Neutrons." This project has two overall goals. The first of these goals is to extend associated-particle fast neutron transmission and, particularly, induced-reaction tomographic imaging algorithms to three dimensions. The second of these goals is to automatically segment the resultant tomographic images into constituent parts, and then extract information about the parts, such as the class of shape and potentially shape parameters. This report addresses of the component of the project concerned with three-dimensional (3D) image reconstruction.

Most readers are familiar with the notion, if not the details, of computed tomography from medical physics. There, projection data from a number of angles are used to reconstruct cross-sectional images of patients for diagnostic purposes. The crucial elements of this process are that measured data divide the object (patient) into lines of response, and multiple views through the object can be mathematically combined to estimate an image. Computed tomography (CT), and particularly x-ray CT, has been commonplace in medicine for some decades, but emission tomography such as positron emission tomography (PET) and single photon emission computed tomography (SPECT) also see wide diagnostic use. For emission tomography, iterative reconstruction is often used to obtain the best possible images from data whose quality may be limited by a number of factors, including counting statistics, resolution limitations of the detectors, and the number of projections.

While x-ray CT, PET, and SPECT have existed for some time, Oak Ridge National Laboratory (ORNL) recently developed a suite of novel fast neutron induced-reaction (induced fission and elastic scatter) imaging techniques to characterize objects of interest [1,2]. At the outset of the present work, transmission and induced-reaction image reconstruction codes used for this purpose existed as a set of Matlab scripts, based on the maximum likelihood expectation maximization (MLEM) formalism [3], developed to demonstrate proof of concept. While functional, these scripts have a variety of drawbacks. Foremost among these is the limitation that the scripts can only reconstruct two-dimensional images, and are sufficiently slow that extension to three dimensions would make them unusable in practice. This limitation is contrary to the inherently 3D nature of induced-reaction images whose data have contributions from points in the reconstructed plane to detectors that are out of plane. In addition, the scripts cannot correctly weight

errors and are not readily suited to the addition of penalties, constraints, or other priors to select for desirable image properties.

The intent of the present work is to combine modern, parallel iterative reconstruction code with novel associated-particle fast-neutron induced-reaction (induced fission and elastic scatter) and transmission imaging techniques recently developed by ORNL. In this way, the increased speed of reconstruction will make routine use of 3D reconstruction of fast neutron images practical. In particular, an increase of reconstruction speed of approximately two orders of magnitude will make 3D reconstructions approximately as time consuming as the present 2D reconstructions. The necessary increases in speed will be realized through the use of:

- Compiled, not interpreted code
- Sparse matrices in order not to perform operations on a large number of zero-value elements
- Multi-threaded architecture to take advantage of multi-core computers
- The use of ordered subsets to increase convergence speed and enable use of fewer iterations (ordered subsets enable updating the guessed solution based on a fraction of the data and generally speed convergence by the number of such subsets)

Moreover, once the iterative reconstruction machinery is in place, it can be used, along with modifications of the update equation, to accommodate other algorithms that correctly weight errors for data that is background subtracted or that select for favorable solutions by the use of regularization, penalties, or constraints.

The original plan for this project stipulated development of a fully three-dimensional transmission imaging code in the first year. However, after initiating work, it was considered more prudent to implement both transmission imaging and induced reaction imaging in 2D prior to implementation in 3D. This would enable identification of any aspects of the induced-reaction reconstruction that were not consistent with the implementation used for transmission imaging. As a result, at present, imaging reconstruction algorithms for transmission imaging, induced-fission imaging, and elastic scatter imaging have been incorporated in a common framework that is readily extensible to 3D. Update equations based on the MLEM algorithm as well as the Simultaneous Iterative Reconstruction Technique (SIRT) [4], a gradient descent method for solving least squares problems, have been implemented. The SIRT method enables weighting by appropriate factors. The new code reduces reconstruction times for transmission images by more than two orders of magnitude. For induced reaction images, reconstruction times are dominated by the time to calculate the induced-count gate efficiencies over the reconstruction volume. The duration of these calculations is also reduced by approximately two orders of magnitude.

Associated-Particle Imaging

In the associated-particle method, fast (14 MeV) neutrons are made in a D-T neutron generator via the $d + t \rightarrow \alpha + n$ reaction. Because the alpha particle and neutron are emitted simultaneously in (nearly) opposite directions, detection of the time and position of the alpha particle determine the time and direction of the associated neutron. A fast timing, position-sensitive alpha-particle detector inside the D-T neutron generator records the time and position of alpha-particle detections, as shown in Figure 1.

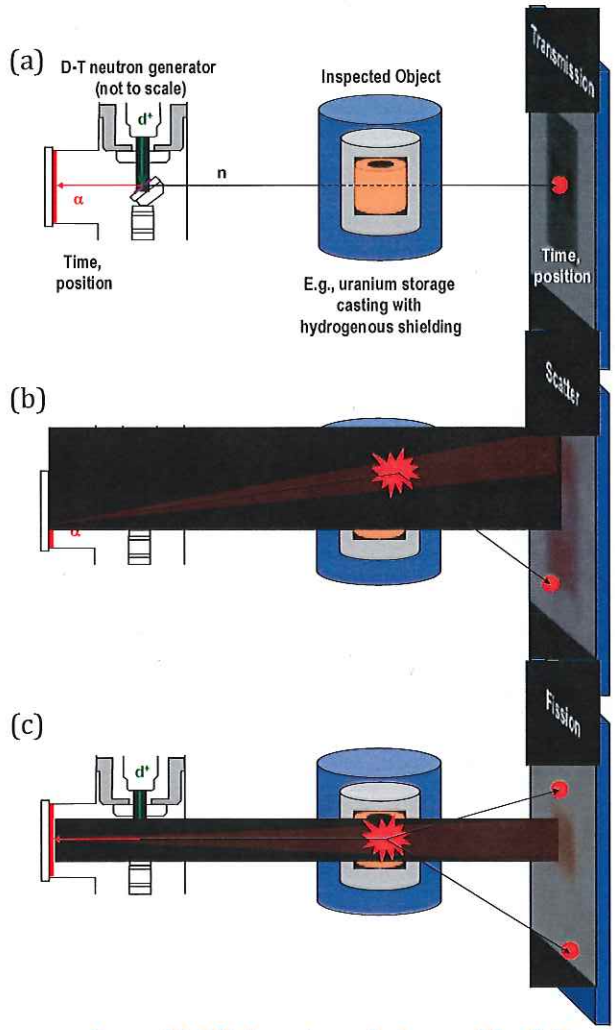

Figure 1: Schematic diagram of tagged 14 MeV neutrons being used for (a) transmission imaging, (b) elastic scatter imaging, and (c) induced-fission imaging.

Likewise, a position-sensitive fast neutron detector records the time and position of neutrons. In the neutron data, transmitted, scattered, and induced neutrons are identified via their arrival times, positions, and number.

Transmission imaging is performed when no interactions occur and the transmitted neutron is detected on the other side of the object. The transmitted neutrons

decrease with thickness of the object like $e^{-\mu x}$. A transmission radiograph gives an image of μx along each path through the object from the source to the detector pixel, and is extracted from measured data by taking the logarithm of the ratio of the number of neutrons with and without the object present. In instances where a number of views through an object can be combined, tomographic reconstruction of the transmission data gives the value of the attenuation coefficient μ for each pixel in a slice through the object. The interpretation of the value of the attenuation coefficient is the probability per unit length of the neutron interacting. Fast neutron images have contrast because different materials have different values of μ, originating either from different total nuclear cross sections or different number densities of nuclei. Neutron transmission imaging is similar to x-ray transmission imaging in principle, although neutrons are much better at penetrating high-atomic-number materials than even the highest energy x-rays and maintain contrast for low-atomic number materials such as hydrogenous materials.

Elastic-scatter imaging is performed when a single elastic scatter off of hydrogen takes place and the scattered-but-transmitted neutron is detected at an appropriate angle and time on the other side of the object. Elastic single scatters vary with the thickness of the object approximately like $\mu_s x \cdot e^{-\mu x}$, where μ_s is the attenuation coefficient due to elastic scatter on hydrogen. Similar to transmission imaging, the interpretation of the value μ_s is the probability per unit length that a neutron elastically scatters from hydrogen. An elastic-scatter radiograph gives an image of the number of detected scatter counts per emitted alpha along each direction from the alpha detector through the object. In instances where a number of views through an object can be combined, tomographic reconstruction of the transmission data gives the value of the attenuation coefficient μ_s for each pixel in a slice through the object. This imaging technique differs from transmission imaging in that each point along each ray from the source toward the detectors has a different efficiency for contributing a count to the measured scatter data. As a result, this efficiency must be explicitly calculated, and includes the attenuation from the source to the image pixel, the average attenuation from the image pixel to the detector (weighted by the solid angle of each detector pixel), the probability of scatter at a particular angle, the relative efficiency at the scattered energy, and the probability of being detected at the appropriate angle-time gate given the timing and direction resolution.

Induced-fission imaging is performed by counting neutrons at arrival times consistent with fission-energy neutrons. Similar to scatter imaging, fission neutrons are created with probability per unit length $\mu_f \bar{\nu}_{chain}$, where μ_f is the attenuation coefficient for fission (probability of fission per unit length) and $\bar{\nu}_{chain}$ is the average number of neutrons emitted in the subsequent fission chain. This property depends on the constituent materials as well as the overall geometry. Likewise, fission neutron pairs are created with probability per unit length $\mu_f \overline{\nu(\nu-1)}_{chain}$, where $\overline{\nu(\nu-1)}_{chain}$ is the average number of neutron pairs emitted in the fission chain. Induced fission images can be constructed based on induced neutrons or induced

neutron pairs. Induced-fission imaging is similar to elastic-scatter imaging in that each point along each ray from the source toward the detectors has a different efficiency for contributing a count to the measured induced neutron data. Again, this efficiency must be explicitly calculated. For neutron singles, it includes the attenuation from the source to the image pixel, the average attenuation from the image pixel to the detector (weighted by the solid angle of each detector pixel), and the detection efficiency for fission-spectrum neutrons to fall within the bounds of the fission time gate. For neutron pairs, an additional calculation is performed to see whether the relative position and time of the two neutrons satisfies a cut to eliminate inter-detector scattering.

Iterative Reconstruction

Suppose an image x with pixel values x_j and the corresponding measured projection data (or sinogram) $\mathbf{y}$ with values y_i are connected by a system matrix $\boldsymbol{A}$ such that $\boldsymbol{A}x = y$, or equivalently, $y_i = \sum_j a_{ij}x_j$. Then, image reconstruction is the process of solving for x given y and $\boldsymbol{A}$. Generally speaking, $\boldsymbol{A}$ will be ill conditioned, making the explicit solution a poor choice since it will compound errors in y. There are a number of methods for solving the problem iteratively of which several are implemented. However, all require the ability to "forward project" a guessed image to get a trial sinogram.

Forward projection of a trial image to get a calculated sinogram is illustrated in Figure 2, where the values of one row of the system matrix $\boldsymbol{A}$ correspond to a ray that originates at the source and terminates on a particular detector pixel. Forward projection in the case of transmission imaging then corresponds to summing the attenuation values along that ray.

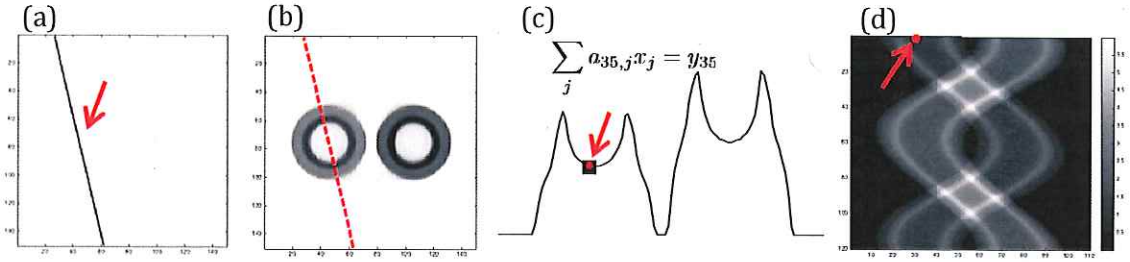

Figure 2: Forward projection of two cylinders to get a trial sinogram. The values of one row of the system matrix sums the attenuation values along a path through the object that terminates on a detector pixel.

For transmission imaging, the system matrix can be fairly simple. In contrast, for induced-reaction imaging, it can be quite complex. However, most of the complexity can be encoded in an efficiency ε_{ij}, the components of which were enumerated in the previous section. In addition, the ray connecting the source and the detectors widens as it approaches the latter. This is shown pictorially in Figure 3, where the

smeared ray is shown in (a) and the efficiency is shown using a logarithmic gray scale in (b).

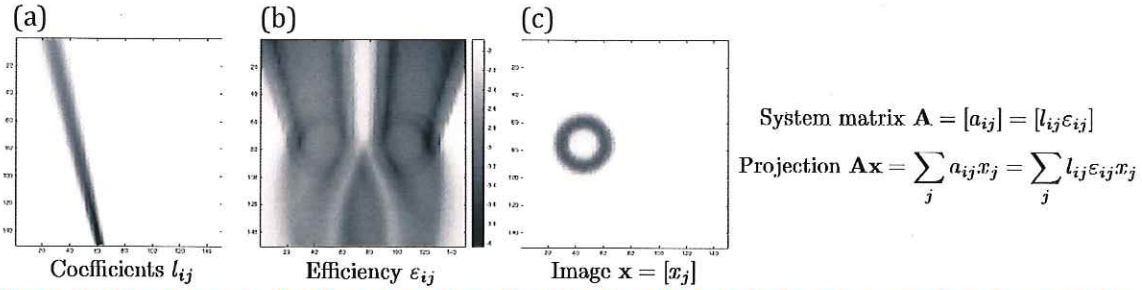

Figure 3: Forward projection for induced-reaction imaging which includes (a) a spread ray between the source and detector and (b) a complicated efficiency map that encodes the induced-reaction physics

The Matlab scripts for performing image reconstruction were based on the maximum likelihood expectation maximization algorithm. That algorithm is derived from formally maximizing the log likelihoods, assuming the values y to be Poisson distributed about the solution $\boldsymbol{A}x$. This maximization results in an update equation for the $k+1^{st}$ image values $x_j^{(k+1)}$ that is multiplicative.

$$x_j^{(k+1)} = x_j^{(k)} \frac{1}{\sum_i a_{ij}} \sum_i \frac{y_i}{\sum_h a_{ih} x_h^{(k)}} a_{ij}$$

In words, the k^{th} trial image is forward projected to get a trial sinogram. The ratio of the measured to calculated sinograms is then back projected to update the image. The derivation of the update equation is only valid if the y_i are Poisson distributed. This is not true of the transmission and induced-reaction values. In the former case, the values of the transmission sinogram are not counts, but the logarithm of a ratio of counts. In the latter case, the induced counts are background-subtracted. However, we have applied MLEM as a general multiplicative algorithm for image reconstruction.

In order to handle errors in the data consistently, a least squares formalism is more appropriate. In this case, image reconstruction corresponds to finding an image $\mathbf{x}$ that minimizes the weighted Euclidian norm:

$$\frac{1}{2}\|\boldsymbol{A}x - y\|_{\boldsymbol{W}}^2$$

When $\boldsymbol{W}$ is defined to scale the individual error terms to be per unit-length (thereby allowing longer rays to carry more error), minimization results in an additive update equation:

$$x_j^{(k+1)} = x_j^{(k)} + \alpha \frac{1}{\sum_i a_{ij}} \sum_i \frac{1}{\sum_h a_{ih}} \left(y_i - \sum_h a_{ih} x_h^{(k)} \right) a_{ij}$$

This is a gradient descent method known as SIRT (Simultaneous Iterative Reconstruction Technique). In words, the k^{th} trial image is forward projected to get a trial sinogram. The difference between the measured to calculated sinograms, scaled by matrix row and column sums as well as a scalar α chosen to both ensure and speed up convergence, is then back projected to update the image. In matrix notation, this can be written:

$$x^{(k+1)} = x^{(k)} + \alpha \boldsymbol{C}\boldsymbol{A}'\boldsymbol{R}\left(y - \boldsymbol{A}x^{(k)}\right)$$

Here, $\boldsymbol{A}'$ denotes the transpose of $\boldsymbol{A}$. Matrix $\boldsymbol{C}$ serves to precondition the computation (the role is somewhat similar to a catalyst in a chemical reaction).

A more general form of SIRT enables penalizing undesirable solutions using a regularizing matrix $\boldsymbol{Q}$. Also, it possible to replace $\boldsymbol{R}$ by a statistical weighting matrix $\boldsymbol{W}$ that represents the data variances, if the associated preconditioning matrix $\widetilde{\boldsymbol{C}}$ is chosen properly, namely,

$$x^{(k+1)} = \left(\boldsymbol{I} - \alpha\beta\widetilde{\boldsymbol{C}}\boldsymbol{Q}'\boldsymbol{Q}\right)x^{(k)} + \alpha\widetilde{\boldsymbol{C}}\boldsymbol{A}'\boldsymbol{W}\left(y - \boldsymbol{A}x^{(k)}\right)$$

where

$$\widetilde{\boldsymbol{C}} \equiv \mathrm{diag}\left\{1/\sum_i w_i\left(\sum_h a_{ih}\right)a_{ij}\right\}$$

Parameter β is defined by the user to control the degree to which minimization of the data fidelity term drives the computation versus the desire to regularize the solution.

Status and Future Plans

At the end of the first year of effort, we have implemented iterative reconstruction code for transmission, hydrogen scatter, and induced fission (singles and doubles) as compiled C code. The code shares a common framework for forward projection, back projection, and calculating and incorporating efficiencies; this framework is used to implement a number of algorithms, including MLEM and a number of variations of the SIRT. All algorithms are multi-threaded and implemented to make use of ordered subsets acceleration techniques.

The SIRT algorithm can be used with or without weighting. At present, the weighting is calculated automatically and is only appropriate for transmission imaging. Future work will implement weights of user-supplied variances.

We have implemented three different forms of regularization $\boldsymbol{Q}$:

1. No regularization, $\boldsymbol{Q} = 0$
2. Minimum norm regularization, $\boldsymbol{Q} = I$ for which $\|\boldsymbol{Q}x\|^2 = \sum_j x_j^2$
3. Finite differences regularization, $\|\boldsymbol{Q}x\|^2 = \sum_j \sum_{k \in N_j}\left(x_j - x_k\right)^2$, where N_j denotes the set of lexicographical predecessor neighbors (that is, neighboring image pixels)

Minimum norm regularization serves to provide numerical stability when the least squares problem solved is ill conditioned. Finite differences regularization serves to penalize roughness when the data is noisy which leads to a smoother appearance of the image.

For the induced-reaction reconstruction codes, at present the forward projector does not include a point spread function (smearing on the ray from the source to detectors). That is, it uses the ray shown in Figure 2 (a) rather than that shown in

Figure 3 (a). We anticipate that a point spread function can be added to the system model thru use of FFTs thereby keeping the computational cost down.

The speed performance of the compiled reconstruction code has increased more than two orders of magnitude over the Matlab script. As a measure of reconstruction time, the example sinogram shown in Figure 2 (d) was reconstructed using the Matlab script and the compiled C code. The sinogram consisted of 110 rays by 120 projections. For the reconstruction, the sinogram was oversampled so that 440 rays sampled the reconstruction volume.

Table 1: Speed improvement of the new compiled code

	Matlab script, MLEM, 40 iterations	C code, MLEM, 40 iterations, 1 core	C code, MLEM, 40 iterations, 4 cores	C code, MLEM, 10 iterations, 4 ordered subsets, 4 cores
Time (s)	286	13.6	4.2	1.1
Improvement	N/A	21.0	68.4	267

In addition to a factor of ~20 from efficient, compiled code, there are additional factors roughly proportional to the number of cores and the number of ordered subsets. In total, this test indicates that future 3D reconstructions will take about the same amount of time as 2D reconstructions previously took.

Comparison images derived using the (a) Matlab MLEM reconstruction code for 40 iterations and the (b) compiled C code MLEM reconstruction for 10 iterations and 4 ordered subsets are shown in Figure 4. Note that the conversion between image pixels and centimeters is slightly different in each case, but visual inspection indicates that the result is essentially identical.

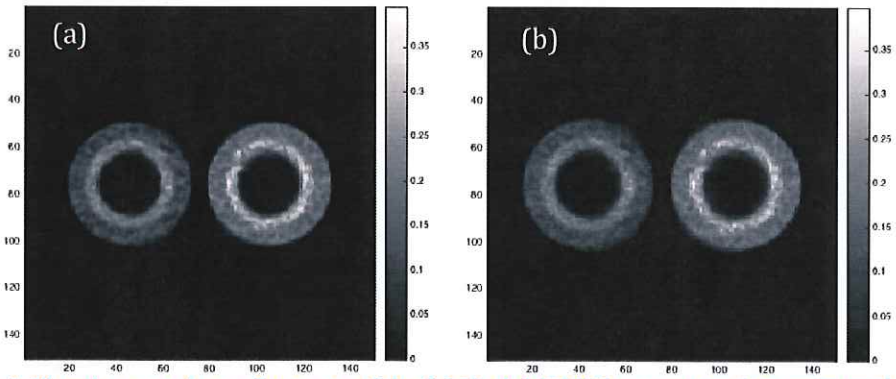

Figure 4: Visual comparison of outputs of the (a) Matlab MLEM reconstruction script and the (b) compiled C code MLEM reconstruction.

In the upcoming year, a number of improvements are anticipated for the reconstruction software. These improvements include:

- The ability to specify the variance of each data point (given that the data points are background-subtracted differences or log ratios)
- Implementation of a point spread function (PSF) in the forward projection and back projection operators for induced-reaction imaging
- Implementation of a total variation constraint to achieve smoother images while preserving sharp edges [5]
- Extension of the reconstruction code to 3D

References

[1] P. A. Hausladen, et al., "Induced-fission imaging of nuclear material," 51st Annual Meeting of the Institute of Nuclear Materials Management, Baltimore, Maryland, USA, July 2010.

[2] Matthew A. Blackston and Paul A. Hausladen, "Fast-Neutron Elastic-Scatter Imaging for Material Characterization", Proceedings of the 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference, San Diego, CA.

[3] David S. Lalush and Miles N. Wernick, "Chapter 21 – Iterative Image Reconstruction" *Emission Tomography, the Fundamentals of PET and SPECT*, Ed. Miles N. Wernick and John N. Aarsvold. Academic Press, 2004. 443-472.

[4] Jens Gregor and Jeffrey A. Fessler, "Comparison of SIRT and SQS for Regularized Weighted Least Squares Image Reconstruction", IEEE Transactions on Computational Imaging vol. 1, no. 1 (2015) 44-55.

[5] Jens Gregor, Philip Bingham, and Lloyd F. Arrowood, "Total Variation Constrained Weighted Least Squares Using SIRT and Proximal Mappings", 4th International Conference on Image Formation in X-Ray Computed Tomography, Bamberg, Germany, July 2016.