

# Fast Variance Computation for Quadratically Penalized Iterative Reconstruction of 3D Axial CT Images

Stephen M. Schmitt, *Student Member, IEEE*, and Jeffrey A. Fessler, *Fellow, IEEE*

**Abstract**—Finding the variance of iteratively-reconstructed 3D axial CT images is useful for statistical analysis of these images or could be useful for dynamic, on-line beam intensity adjustment. Previous methods for finding the variance are either computationally intractable or approximations that take minutes per voxel to compute. In this paper, we propose a method that can generate a variance image for an entire volume in seconds by making approximations specific to third-generation 3D axial CT geometries. We compare the computation time and error of the resulting approximate image to the empirical image formed from many simulated realizations of a reconstruction.

## I. INTRODUCTION

Iterative reconstruction (IR) methods in computed tomography have received attention for their resolution and noise properties that improve on FBP [5]. However, the statistical properties of IR methods are more difficult to analyze than FBP. When using quadratic regularization with IR, there is a closed-form expression for the covariance matrix of the reconstruction for a fixed statistical weighting matrix  $\mathbf{W}$  [2]. Knowledge of the variance could be used for further statistical analysis, or potentially in a clinical setting to inform a radiologist of the relative accuracy of a region, or to dynamically adjust the X-ray beam intensity of a CT scan to reduce dose while providing a reconstructed image with more uniform variance.

Calculating and storing the covariance is intractable, motivating methods to find the variance of each reconstructed voxel quickly. Prior work has given expressions for the variance in terms of the forward and back-projection of each voxel, under an approximation of local shift-invariance [4]. There are methods that greatly speed up these approximations by removing the need to forward and back-project a voxel for specific CT geometries, such as 2D fan-beam [7] and 3D step-and-shoot CT [8], but to our knowledge, there is no such method for 3D axial CT.

In this paper, we use methods similar to those in [7], [8] to develop an approximation to the variance for 3D axial CT that reduces the computational cost of finding this variance by several orders of magnitude compared to existing methods.

This work was supported in part by NIH grant R01 HL-098686.

Steve M. Schmitt and Jeffrey A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (smschm@umich.edu, fessler@umich.edu)

## II. METHODS

In our approximation, we assume that we are only interested in the variance of each voxel, i.e., the diagonal of the covariance matrix; that the observed sinogram elements are statistically independent given a scanned object; and that the projection of a single voxel is, to a local approximation, shift-invariant.

For a quadratically-penalized least squares reconstruction problem with a data-fit term and a regularization term

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \frac{\alpha}{2} \|\mathbf{R}\mathbf{x}\|_2^2, \quad (1)$$

the closed-form (but generally uncomputable) solution is given by

$$\hat{\mathbf{x}} = (\mathbf{H} + \alpha \mathbf{R}^T \mathbf{R})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y},$$

where  $\mathbf{H} = \mathbf{A}^T \mathbf{W} \mathbf{A}$  is the Hessian of the data-fit term. If the weighting  $\mathbf{W}$  is chosen so that  $\text{cov}(\mathbf{y}) = \mathbf{W}^{-1}$ , then the reconstruction covariance is [2]

$$\text{cov}(\hat{\mathbf{x}}) = \mathbf{K} = (\mathbf{H} + \alpha \mathbf{R}^T \mathbf{R})^{-1} \mathbf{H} (\mathbf{H} + \alpha \mathbf{R}^T \mathbf{R})^{-1}.$$

Evaluating the elements of this matrix is computationally intractable because of the need to invert a large matrix.

### A. Fourier-based methods

Previous work ([7], [4]) has demonstrated that an element on the diagonal of this matrix, representing the variance of the reconstruction of a single voxel  $j$ , can be approximated as

$$\mathbf{K}_{jj} \approx \frac{1}{N_x N_y N_z} \sum_{\vec{k} \in \mathcal{K}} \frac{\Gamma_j[\vec{k}]}{(\Gamma_j[\vec{k}] + \alpha \Omega_j[\vec{k}])^2}, \quad (2)$$

where  $\mathcal{K} = \{0, \dots, N_x - 1\} \times \{0, \dots, N_y - 1\} \times \{0, \dots, N_z - 1\}$  is a set of discrete 3D frequencies;  $\Gamma_j = \mathcal{F}_j^3\{[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{*j}\}$  and  $\Omega_j = \mathcal{F}_j^3\{[\mathbf{R}^T \mathbf{R}]_{*j}\}$  represent 3D frequency responses of  $\mathbf{A}^T \mathbf{W} \mathbf{A}$  (the Hessian of the data-fit term, a weighted project-and-backproject matrix) and  $\mathbf{R}^T \mathbf{R}$  (the Hessian of the regularization term) *local* to the voxel  $j$ . This local 3DFT centered at  $\vec{n}_j$  is defined as

$$(\mathcal{F}_j^3 \{\mathbf{x}\})[\vec{k}] = \sum_{\ell=1}^{|\vec{N}|} \mathbf{x}_\ell \exp(-i2\pi(\vec{k} \oslash \vec{N}) \cdot (\vec{n}_\ell - \vec{n}_j)), \quad (3)$$

where  $\oslash$  represents element-wise division,  $\vec{n}_j$  represents the spatial position of voxel  $j$  in units of voxels, and  $\vec{N} =$

$(N_x, N_y, N_z)$ . We define a continuous-frequency analog to (3);

$$\left(\mathcal{F}_{j,\text{cont}}^3\{\mathbf{x}\}\right)(\vec{v}) = \sum_{\ell=1}^{|\vec{N}|} \mathbf{x}_\ell \exp(-i2\pi\vec{v} \cdot (\vec{n}_\ell - \vec{n}_j)), \quad (4)$$

where  $\vec{v}$  has units of cycles per sample; we then approximate (2) as

$$\mathbf{K}_{jj} \approx \int_{[-\frac{1}{2}, \frac{1}{2}]^3} \frac{H_j(\vec{v})}{(H_j(\vec{v}) + \alpha R(\vec{v}))^2} d\vec{v}, \quad (5)$$

where  $H_j \triangleq \mathcal{F}_{j,\text{cont}}^3\{[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{*j}\}$  and  $R \triangleq \mathcal{F}_{j,\text{cont}}^3\{[\mathbf{R}^T \mathbf{R}]_{*j}\}$ . The closed-form expression [8] for  $R$  is independent of voxel position (so long as the voxel is not along an image border), and so, to evaluate (5), we focus on developing a closed-form approximation to  $H_j(\vec{v})$ , the local frequency response of the Hessian of the data-fit term. In a parallel-ray geometry with noise that is stationary within one view angle, as in [1],  $H_j(\vec{v})$  would also be independent of voxel, and  $\mathbf{K}_{jj}$  would be uniform.

### B. Local frequency response for axial CT

Let  $a_{\beta,j}(\vec{s})$  be the continuous footprint of the  $j$ th voxel when the source is at an angle  $\beta$ , as a function of 2D detector position  $\vec{s}$ . With this notation,  $\mathbf{A}_{ij} = a_{\beta,i}(\vec{s}_i)$ , where  $\vec{s}_i$  and  $\beta_i$  are the detector position and source angle, respectively, of the  $i$ th projection element.

Similarly, we can define  $w_\beta(\vec{s})$  to be a continuous extension of the elements of the weighting matrix  $\mathbf{W}$ , defined such that  $w_{\beta,i}(\vec{s}_i) = \mathbf{W}_{ii}$ . We assume that the weighting changes slowly as a function of  $\beta$  and  $\vec{s}$ , such that the continuous interpolation from the set of defined points is reasonably accurate. We also specify  $w_\beta(\vec{s}) = 0$  when  $\vec{s}$  is not located on the detector.

Writing the matrix multiplication defining  $\mathbf{A}^T \mathbf{W} \mathbf{A}$ , one particular value of  $[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{kj}$  is:

$$[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{kj} = \sum_{i=1}^{N_s N_\beta} w_{\beta,i}(\vec{s}_i) a_{\beta,i,k}(\vec{s}_i) a_{\beta,i,j}(\vec{s}_i). \quad (6)$$

Approximating this sum as an integral:

$$[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{kj} \approx \Pi \int_{\mathbb{R}^2} \int_0^{2\pi} w_\beta(\vec{s}) a_{\beta,k}(\vec{s}) a_{\beta,j}(\vec{s}) d\beta d\vec{s}, \quad (7)$$

where  $\Pi = 1/\Delta_s \Delta_t \Delta_\beta$ ;  $\Delta_s$  and  $\Delta_t$  are the spacings on the detector between pixels in the  $s$  and  $t$  directions, and  $\Delta_\beta$  is the spacing, in radians, between source positions. We define

$$\xi_{k,j\beta} \triangleq \int_{\mathbb{R}^2} a_{\beta,k}(\vec{s}) a_{\beta,j}(\vec{s}) d\vec{s} \quad (8)$$

$$\check{w}_{k,j\beta} \triangleq \frac{\int_{\mathbb{R}^2} w_\beta(\vec{s}) a_{\beta,k}(\vec{s}) a_{\beta,j}(\vec{s}) d\vec{s}}{\int_{\mathbb{R}^2} a_{\beta,k}(\vec{s}) a_{\beta,j}(\vec{s}) d\vec{s}} \quad (9)$$

so that (7) becomes

$$[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{kj} \approx \Pi \int_0^{2\pi} \xi_{k,j\beta} \check{w}_{k,j\beta} d\beta. \quad (10)$$

We now find approximations for  $\xi$  (an unweighted correlation between footprints) and  $\check{w}$  (the weighted correlation normal-

ized by the unweighted correlation) that allow us to further simplify (5).

We can interpret  $\xi_{k,j\beta}$  as an unweighted correlation between the footprints of voxels  $k$  and  $j$ , or as the  $k$ th voxel in the 3D backprojection (using only projection angle  $\beta$ ) of the projection of voxel  $j$ .

### C. Normalized weighted correlation approximation

Usually,  $w_\beta(\vec{s})$  varies slowly relative to  $a_{\beta,j}(\vec{s})$ , and so, approximately, the definition of  $\check{w}$  in (9) essentially sifts out one particular value of  $w_\beta$ . This value can be any suitable value of  $w_\beta(\vec{s})$  near the position  $\vec{s}_j^*$  on the detector that maximizes  $a_{\beta,j}(\vec{s}_j^*)$ . We choose the point on the detector where a ray from the source at angle  $\beta$  passing through voxel  $j$  intercepts. Since this is likely not exactly the exact parameters of an observation, we round to the nearest detector element and source angle. We define  $\hat{w}_{j\beta} \triangleq w_\beta(\vec{s}_j^*)$  where, for  $\vec{n}_j \approx \vec{n}_k$ ,  $w_\beta(\vec{s}_j^*) \approx \check{w}_{k,j\beta}$ .

With this approximation, (10) becomes

$$[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{kj} \approx \Pi \int_0^{2\pi} \xi_{k,j\beta} \hat{w}_{j\beta} d\beta.$$

Since  $\hat{w}$  no longer depends on  $k$ , we can propagate  $\mathcal{F}_{j,\text{cont}}^3$  inside the integral, giving:

$$H_j(\vec{v}) \approx \Pi \int_0^{2\pi} \Xi_{j\beta}(\vec{v}) \hat{w}_{j\beta} d\beta, \quad (11)$$

where  $\Xi_{j\beta}(\vec{v}) \triangleq \mathcal{F}_{j,\text{cont}}^3\{\xi_{*j\beta}\}$  is the local 3DFT of (8). We now focus on approximating (11) to simplify it.

### D. Unweighted footprint correlation approximation

Voxel  $j$  has detector footprint  $a_{\beta,j}(\vec{s})$ , and to an approximation of the detector curvature as a plane tangent to this footprint, a voxel  $k$  near  $j$  has footprint

$$a_{\beta,k}(\vec{s}) \approx a_{\beta,j}(\vec{s} - \mathbf{P}_{\beta,j}(\vec{n}_k - \vec{n}_j)),$$

for a matrix  $\mathbf{P}_{\beta,j} \in \mathbb{R}^{2 \times 3}$  that is determined by the CT geometry (see Appendix A). Using this approximation, we define the approximation  $\tilde{\xi}_{k,j\beta}$ :

$$\begin{aligned} \xi_{k,j\beta} \approx \tilde{\xi}_{k,j\beta} &\triangleq \int_{\mathbb{R}^2} a_{\beta,j}(\vec{s} - \mathbf{P}_{\beta,j}(\vec{n}_k - \vec{n}_j)) a_{\beta,j}(\vec{s}) d\vec{s} \\ &= r_{a,\beta,j}(\mathbf{P}_{\beta,j}(\vec{n}_k - \vec{n}_j)), \end{aligned} \quad (12)$$

where  $r_{a,\beta,j}(\vec{s})$  is the 2D auto-correlation of  $a_{\beta,j}$ . Defining  $\tilde{\Xi}_{j\beta}(\vec{v}) \triangleq \mathcal{F}_{j,\text{cont}}^3\{\tilde{\xi}_{*j\beta}\}$ , it follows from (12) that:

$$\tilde{\Xi}_{j\beta}(\vec{v}) = |\mathbf{P}_{\beta,j} \mathbf{P}_{\beta,j}^T|^{-1/2} |A_{\beta,j}(\mathbf{P}_{\beta,j}^T \vec{v})|^2 \delta(\vec{p}_{\beta,j}^\perp \cdot \vec{v}), \quad (13)$$

where  $A_{\beta,j}(\vec{v})$  is the 2DFT of the footprint  $a_{\beta,j}(\vec{s})$  and  $\vec{p}_{\beta,j}^\perp$  is the unit vector whose span is the kernel of  $\mathbf{P}_{\beta,j}$  (see Appendix B).

Using the approximation (13) in (11), with the sifting property of the Dirac impulse, the local frequency response simplifies

to

$$H_j(\vec{v}) \approx \Pi \sum_{\beta \in \mathcal{B}_j(\vec{v})} \frac{|A_{\beta,j}(\mathbf{P}_{\beta,j}^T \vec{v})|^2 \hat{w}_{j\beta}}{|\mathbf{P}_{\beta,j} \mathbf{P}_{\beta,j}^T|^{1/2} \zeta_{\beta,j}(\vec{v})}, \quad (14)$$

where  $\mathcal{B}_j(\vec{v})$  is the set of values of  $\beta$  that are solutions to  $\vec{p}_{\beta,j}^\perp \cdot \vec{v} = 0$ ; for a 360-degree rotation axial scan,  $|\mathcal{B}_j(\vec{v})| = 2$  when  $\vec{v} \neq \vec{0}$ . The scaling factor

$$\zeta_{\beta,j} \triangleq \left| \frac{\partial}{\partial \beta} (\vec{p}_{\beta,j}^\perp \cdot \vec{v}) \right|,$$

arises from the sifting of the Dirac impulse in (13).

### E. Reduction to single integral

We define  $(\rho, \Phi, v_3)$  to be  $\vec{v}$  in cylindrical coordinates and  $D_{s0} \cdot (r, \phi, \tilde{z})$  to be the location of voxel  $j$  in space in cylindrical coordinates, normalized to  $D_{s0}$ , so that  $r = 0$  at the isocenter and  $r = 1$  at the detector. Appendix A shows that (14) simplifies to

$$H_j(\vec{v}) \approx \frac{\Pi \Delta_x^3 \Delta_z D_{sd}^2}{D_{s0}^2} \sum_{\beta \in \mathcal{B}} \frac{\hat{w}_{\beta,j} \text{sinc}(v_1)^2 \text{sinc}(v_2)^2 \text{sinc}(v_3)^2}{\rho d_{\beta,j} \sqrt{1 - r^2 \cos^2(\phi - \Phi)}}, \quad (15)$$

where  $d_{\beta,j}$  is the distance from the source at angle  $\beta$  to the projection of voxel  $j$  into the  $xy$ -plane, divided by  $D_{s0}$ ; this is approximately equal to 1 near the center regardless of  $\beta$ . For ease of notation, define:

$$\begin{aligned} C &= \Pi \Delta_x^3 \Delta_z D_{sd}^2 / D_{s0}^2 \\ D(\rho, \Phi, v_3) &= \text{sinc}(\rho \cos \Phi)^2 \text{sinc}(\rho \sin \Phi)^2 \text{sinc}(v_3)^2 \\ E_j(\Phi) &= \sum_{\beta \in \mathcal{B}(\Phi)} \frac{\hat{w}_{\beta,j}}{d_{\beta,j} \sqrt{1 - r^2 \cos^2(\phi - \Phi)}}, \end{aligned}$$

so that  $H_j(\vec{v}) \approx C \cdot D(\vec{v}) \cdot E_j(\Phi) / \rho$ , and  $E_j(\Phi)$  is the only one of these terms that depends on the voxel location and the weights, and hence the object being scanned.

We define the following expression, which we tabulate,

$$F(\Phi, \gamma) \triangleq \int_0^{\rho_{\max}(\Phi)} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\gamma \cdot D(\rho, \Phi, v_3) \cdot \rho^2}{(\gamma \cdot D(\rho, \Phi, v_3) + \rho R(\vec{v}))^2} dv_3 d\rho, \quad (16)$$

which allows us to rewrite (5) as

$$\mathbf{K}_{jj} \approx \alpha^{-1} \int_0^{2\pi} F(\Phi, \alpha^{-1} C E_j(\Phi)) d\Phi. \quad (17)$$

This expression isolates all voxel location and object dependence into evaluation of  $E_j(\Phi)$ , so we can precompute  $F$  for a range of  $\Phi$  and  $k$  by numerical integration. Approximating the variance via (17), then, becomes a simple 1D numerical integration. This integration can be evaluated using a coarse discretization of  $\Phi$  with reasonably accurate predicted variance, especially given that the integrand is periodic and integrated over one period, a case in which numerical integration converges quickly [6].

## III. RESULTS

We evaluated the variance estimate by comparing it to the empirical variance computed by simulating 100 reconstructions of a  $512 \times 512 \times 48$  voxel XCAT phantom (pictured in Fig. 1) with voxel size  $\Delta_x \times \Delta_z = 0.9764 \times 0.625$ mm. The simulated scanner was a GE third-generation system geometry with a  $888 \times 32$  quarter-offset pixel detector with pixel size  $\Delta_x \times \Delta_t = 1.0239 \times 1.0964$ mm, 984 view angles,  $D_{sd} = 949.075$ mm source-to-detector distance, and  $D_{s0} = 408.075$ mm source-to-isocenter distance. Each reconstruction used 40 iterations of an ordered-subset method with 41 subsets.

The resulting empirical variance maps were noisy, so for the purposes of comparison and under the assumption that the variance image changes slowly (which is confirmed in [7]), we blurred the empirical variance image with a Gaussian kernel with a FWHM of 3 voxels. Fig. 2 shows the resulting standard deviation image. Fig. 3 shows the results of our method. Fig. 4 shows the percentage error of the fast method relative to the empirical method..

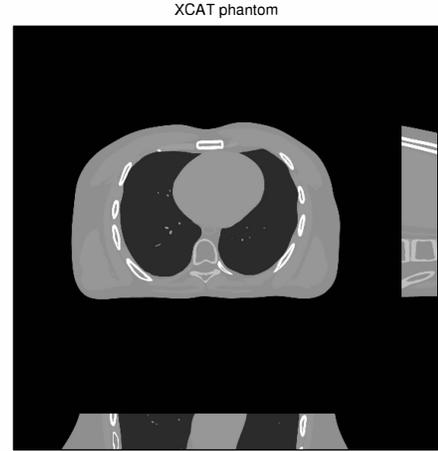


Fig. 1. XCAT phantom (center is axial slice through the center of the volume; bottom is coronal slice; right side is sagittal slice)

Fig. 5 shows horizontal profiles of the estimates through the center of the center slice with the empirical method in black and our method in red. Fig. 6 shows percentage errors of this profile.

The empirical reconstructions used to make Figure 2 took a mean of 14846 CPU-seconds (realizations ranged from 11995 CPU-sec. to 17934 CPU-sec. with a standard deviation of 1658 sec.) The FFT method takes 470 CPU-seconds *per voxel*. Our method takes less than  $10^{-5}$  CPU-seconds per voxel when evaluating (17) by numerically integrating using 24 sample points. Estimating the variance for the entire XCAT volume by evaluating (17) once for each  $4 \times 4 \times 1$  voxel block and interpolating only requires 4 CPU-seconds. Tabulating  $F$  in (16) takes approximately 15 minutes using an unaccelerated method. Since this is image-independent and fixed only to a

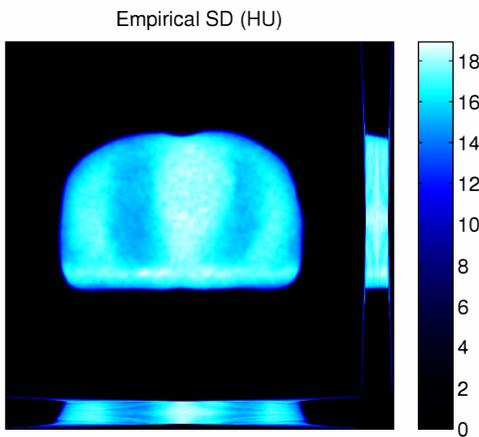


Fig. 2. Empirical standard deviation estimate

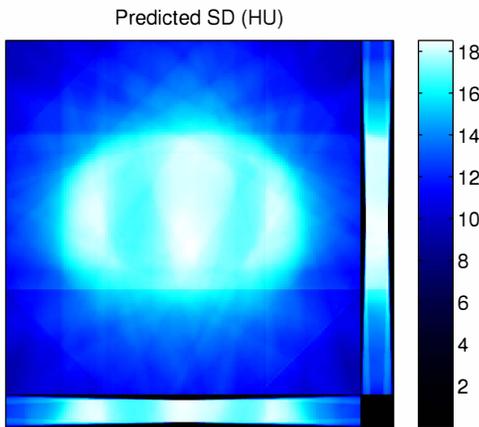


Fig. 3. Fast standard deviation estimate

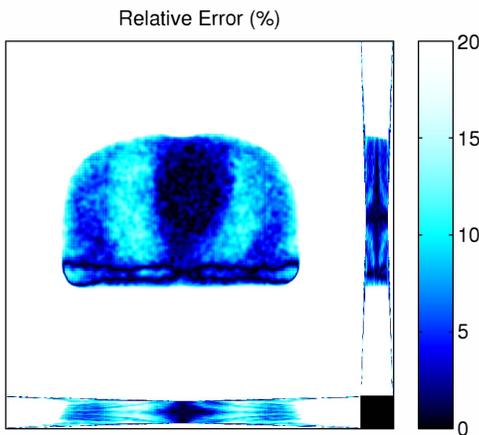


Fig. 4. Standard deviation error percentage

given regularization scheme and CT geometry, we consider this cost to be amortized.

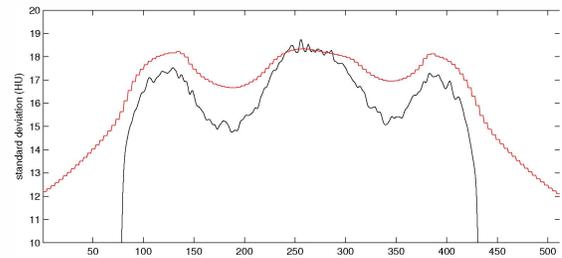


Fig. 5. Profile comparison of empirical and estimate

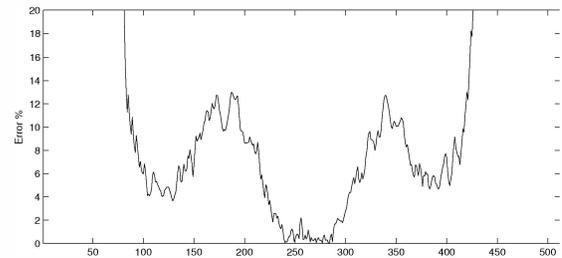


Fig. 6. Profile of error percentage

#### IV. DISCUSSION

The computational time of our method is much lower than empirical methods, using less than  $10^{-5}$  times the time of 100 reconstructions. FFT methods are slower for an entire volume than empirical methods, but they have the advantage of being able to find the variance of a specific voxel of interest. Our method is able to find the variance of one specific voxel in less than  $10^{-7}$  times the time of FFT methods. Our method is less exact, but the error relative to empirical standard deviation is within 20% within the support of the object; whether this error-time tradeoff is worthwhile depends on the application.

Our method is currently unable to accurately find variance outside the support of the object; non-negativity constraints in the reconstruction algorithm significantly reduce the variance in air regions outside the object, and our method does not take this into account. As a result, the relative error grows toward infinity as the empirical variance approaches zero. However, accurate variance prediction outside the object is probably not important for many applications. Alternatively, external air regions could be identified from a rough pilot reconstruction and accounted for separately.

#### V. CONCLUSIONS

We have presented a method for approximating the variance image of iteratively-reconstructed 3D axial CT images that saves significant computational time to existing methods while keeping the approximation error within reasonable bounds within the support of the image. Additionally, our method can be used to find the variance of any subset of voxels within the image, as opposed to only the entire image simultaneously for empirical methods. Future work will include extending our method to other CT geometries, particularly helical CT, and non-quadratic regularization.

For axial cone-beam CT, the position of the X-ray source, as a function of angle  $\beta$ , is given by:

$$\vec{p}_s(\beta) = D_{s0}(\cos\beta, \sin\beta, 0).$$

Let  $\mathbf{P}_2 = \text{diag}(1, 1, 0)$  be a projection matrix into the  $xy$ -plane. The vector  $\mathbf{P}_2(\vec{x}_j - \vec{p}_s(\beta))$  is then in the direction of a ray from the source at angle  $\beta$  passing through the projection of voxel  $j$  onto the  $xy$ -plane. A unit vector in this direction is given by

$$\vec{e}_{u,j}(\beta) = \mathbf{P}_2(\vec{x}_j - \vec{p}_s(\beta))/(D_{s0}d_{\beta,j}),$$

where  $d_{\beta,j} = \|\mathbf{P}_2(\vec{x}_j - \vec{p}_s(\beta))\|_2/D_{s0}$  is the distance from the source to the  $xy$ -projection of voxel  $j$ , normalized by  $D_{s0}$ . This vector is perpendicular to the direction of increasing  $s_1$  coordinate on the detector in the  $xy$ -plane, and so the vector

$$\vec{e}_{1,j}(\beta) = \mathbf{R}\mathbf{P}_2(\vec{x}_j - \vec{p}_s(\beta))/(D_{s0}d_{\beta,j}),$$

where  $\mathbf{R}$  is a matrix that rotates counterclockwise 90 degrees in the  $xy$ -plane, is a unit vector in the direction of increasing  $s_1$  coordinate. A shift of  $\Delta\vec{n}$  in voxel space in the  $xy$ -plane corresponds to a shift of  $\Delta_x\Delta\vec{n}$  in space, which is magnified by a factor of  $m_{\beta,j} = D_{sd}/D_{s0}d_{\beta,j}$ . Therefore this shift causes a shift in the  $s_1$  direction on the detector of  $\Delta_x m_{\beta,j}(\vec{e}_{1,j} \cdot \Delta\vec{n})$ .

A unit vector in the direction of increasing  $s_2$  coordinate is  $\vec{e}_2 = [0, 0, 1]^T$ , regardless of voxel and ray direction. Analogously to the  $s_1$  shift, a shift of  $\Delta\vec{n}$  in voxel space corresponds to a shift of  $\Delta_z m_{\beta,j}(\vec{e}_2 \cdot \Delta\vec{n})$  in the  $s_2$  direction on the detector.

These vectors are the rows of the  $2 \times 3$  matrix  $\mathbf{P}_{\beta,j}$  in Section II-D:

$$\mathbf{Q} \triangleq \begin{bmatrix} \frac{-\vec{x}_2 + D_{s0} \sin\beta}{D_{s0}d_{\beta,j}} & \frac{\vec{x}_1 - D_{s0} \cos\beta}{D_{s0}d_{\beta,j}} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{\Sigma} \triangleq \begin{bmatrix} \Delta_x m_{\beta,j} & 0 \\ 0 & \Delta_z m_{\beta,j} \end{bmatrix}$$

$$\mathbf{P}_{\beta,j} = \mathbf{\Sigma}\mathbf{Q}$$

$$\mathbf{P}_{\beta,j}^{+T} = \mathbf{\Sigma}^{-1}\mathbf{Q}$$

for which  $|\mathbf{P}_{\beta,j}\mathbf{P}_{\beta,j}^T|^{1/2} = |\mathbf{\Sigma}\mathbf{Q}\mathbf{Q}^T\mathbf{\Sigma}|^{1/2} = |\mathbf{\Sigma}| = \Delta_x\Delta_z m_{\beta,j}^2$ . Since  $\vec{e}_{u,j}(\beta)$  is a unit vector perpendicular to both  $\vec{e}_{1,j}$  and  $\vec{e}_2$ , it is in the kernel of  $\mathbf{P}_{\beta,j}$ , and so we can let it be  $\vec{p}_{\beta,j}^\perp$  in Section II-D.

The argument to the impulse in (13) is equal to:

$$a(\beta) \triangleq (\mathbf{P}_2(\vec{x}_j - \vec{p}_s(\beta)) \cdot \vec{v})/D_{s0}d_{\beta,j}.$$

With  $\vec{v}$  in cylindrical coordinates  $(\rho, \Phi, \nu_3)$  and  $\vec{x}_j$  in cylindrical coordinates  $(rD_{s0}, \phi, \vec{z}D_{s0})$ ,

$$a(\beta) = d_{\beta,j}^{-1}\rho(r \cos(\phi - \Phi) - \cos(\beta - \Phi)). \quad (18)$$

The values of  $\beta$  for which the argument to the impulse is zero, then, are

$$\mathcal{B} = \{\beta^+, \beta^-\} = \{\Phi \pm \arccos(r \cos(\phi - \Phi))\}.$$

At these values of  $\beta$ ,

$$\mathbf{P}_{\beta,j}^{+T}\vec{v} = \begin{bmatrix} (\Delta_x m_{\beta,j})^{-1}\rho \\ (\Delta_z m_{\beta,j})^{-1}\nu_3 \end{bmatrix}.$$

Using the separable footprint model for CT projection with a transaxial trapezoid and an axial rectangle [3], the 2DFT of voxel  $j$ 's footprint is

$$|A_{\beta,j}(\vec{u})| = \Delta_x^2 \Delta_z m_{\beta,j}^2 \text{sinc}(\Delta_x m_{\beta,j} u_1 \cos \theta_{\beta,j}) \cdot \text{sinc}(\Delta_x m_{\beta,j} u_1 \sin \theta_{\beta,j}) \text{sinc}(\Delta_z m_{\beta,j} u_2),$$

where  $\theta_{\beta,j}$  is the angle at which the ray passes through the voxel. When  $\beta \in \mathcal{B}$ , we have ensured that the angle of a ray passing through a voxel  $j$  is  $\Phi \pm \pi/2$ , since  $\vec{e}_{u,j}(\beta) \cdot \vec{v} = 0$ . A voxel's footprint is unchanged if  $\theta$  is rotated by an integer multiple of  $\pi/2$ , and so we can evaluate the footprint at  $\theta = \Phi$  for  $\beta \in \mathcal{B}$ . In this case,

$$|A_{\beta,j}(\mathbf{P}_{\beta,j}^{+T}\vec{v})| = \Delta_x^2 \Delta_z m_{\beta,j}^2 \text{sinc}(\nu_1) \text{sinc}(\nu_2) \text{sinc}(\nu_3).$$

Taking the derivative of (18) and evaluating at  $\beta \in \mathcal{B}$  gives

$$\zeta_{\beta,q} = d_{\beta,j}^{-1}\rho \sqrt{1 - r^2 \cos^2(\phi - \Phi)}.$$

## APPENDIX B 3DFT OF A 2D FUNCTION

This section derives (13). If  $g(\vec{x}_M)$  and  $G(\vec{v}_M)$  are  $M$ -dimensional Fourier transform pairs and  $h(\vec{x}_N)$  and  $H(\vec{v}_N)$  are  $N$ -dimensional Fourier transform pairs, then a separability property is that

$$gh(\vec{x}) \triangleq g(\mathbf{T}_M \vec{x})h(\mathbf{T}_N \vec{x}) \leftrightarrow G(\mathbf{T}_M \vec{v})H(\mathbf{T}_N \vec{v}) \triangleq GH(\vec{v})$$

are also  $(M+N)$ -dimensional Fourier transform pairs, where  $\mathbf{T}_M$  truncates a  $(M+N)$ -vector to its first  $M$  elements, and  $\mathbf{T}_N$  truncates it to its last  $N$  elements.

Let  $\mathbf{A} \in \mathbb{R}^{M \times (M+N)}$ ,  $\mathbf{B} \in \mathbb{R}^{N \times (M+N)}$  be two matrices that map  $(M+N)$ -vectors into fewer dimensions. Then

$$g(\mathbf{A}\vec{x})h(\mathbf{B}\vec{x}) \leftrightarrow |\mathbf{C}|^{-1}G(\mathbf{T}_M \mathbf{C}^{-T}\vec{v})H(\mathbf{T}_N \mathbf{C}^{-T}\vec{v}),$$

where  $\mathbf{C} \in \mathbb{R}^{(M+N) \times (M+N)}$  is the vertical concatenation of  $\mathbf{A}$  above  $\mathbf{B}$ . This follows naturally from defining  $\mathbf{C}$  and using it in the Fourier identity  $gh(\mathbf{C}\vec{x}) \leftrightarrow |\mathbf{C}|^{-1}GH(\mathbf{C}^{-T}\vec{v})$ .

A. With a constant as  $h$

If  $h: \mathbb{R}^N \rightarrow \mathbb{R}$  is a unit function,  $h(\cdot) = 1$ ; we can simplify this further since we can choose  $\mathbf{B}$  to be any  $N \times (M+N)$  matrix. We choose  $\mathbf{B} = \mathbf{A}_\perp$ , which is  $N$  orthonormal  $(M+N)$ -dimensional row vectors that are each orthogonal to each row in  $\mathbf{A}$ , that is:  $\mathbf{A}\mathbf{A}_\perp^T = \mathbf{0}_{M \times N}$ , and  $\mathbf{A}_\perp \mathbf{A}_\perp^T = \mathbf{I}_N$ . The inverse of  $\mathbf{C}$  is then given by:

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{A}^+ & \mathbf{A}_\perp^T \end{bmatrix},$$

from which  $\mathbf{T}_M \mathbf{C}^{-T} = \mathbf{A}^{+T}$  and  $\mathbf{T}_N \mathbf{C}^{-T} = \mathbf{A}_\perp$ .

From noting that, for this orthogonal  $\mathbf{A}_\perp$ ,

$$\mathbf{C}\mathbf{C}^T = \begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{A}\mathbf{A}_\perp^T \\ \mathbf{A}_\perp\mathbf{A}^T & \mathbf{A}_\perp\mathbf{A}_\perp^T \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix};$$

the determinant  $|\mathbf{C}|^{-1} = |\mathbf{C}\mathbf{C}^T|^{-1/2} = |\mathbf{A}\mathbf{A}^T|^{-1/2}$ . Therefore, the  $(M + N)$ -dimensional Fourier transform of  $g(\mathbf{A}\vec{x})$  is

$$\mathcal{F}^{M+N}\{g(\mathbf{A}\vec{x})\} = |\mathbf{A}\mathbf{A}^T|^{-1/2}G(\mathbf{A}^{+T}\vec{v})\delta^N(\mathbf{A}_\perp\vec{v}),$$

where  $\delta^N$  denotes the  $N$ -dimensional Dirac impulse.

### B. Application

In (12), we have  $M = 2$ ,  $N = 1$ ,  $g = r_a$ , the autocorrelation of a footprint,  $\mathbf{A} = \mathbf{P}_{\beta,j}$ , and  $\mathbf{A}_\perp = \vec{p}_{\beta,j}^\perp$ , a unit vector orthogonal to both rows of  $\mathbf{A}$ .

Directly applying the results of the previous section, while noting that the magnitude of  $R_a$ , the 2DFT of the autocorrelation  $r_a$ , is simply  $|A|^2$ , results in (13).

### REFERENCES

- [1] A. H. Delaney and Y. Bresler. A fast and accurate Fourier algorithm for iterative parallel-beam tomography. *IEEE Trans. Im. Proc.*, 5(5):740–53, May 1996.
- [2] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996.
- [3] Yong Long, J.A. Fessler, and J.M. Balter. 3D forward and back-projection for X-Ray CT using separable footprints. *Medical Imaging, IEEE Transactions on*, 29(11):1839–1850, nov. 2010.
- [4] J. Qi and R. M. Leahy. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *IEEE Trans. Med. Imag.*, 18(4):293–305, April 1999.
- [5] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med. Phys.*, 34(11):4526–44, November 2007.
- [6] J. A. C. Weideman. Numerical integration of periodic functions: A few examples. *The American Mathematical Monthly*, 109(1):pp. 21–36, 2002.
- [7] Y. Zhang-O’Connor and J. A. Fessler. Fast predictions of variance images for fan-beam transmission tomography with quadratic regularization. *IEEE Trans. Med. Imag.*, 26(3):335–46, March 2007.
- [8] Y. Zhang-O’Connor and J. A. Fessler. Fast variance predictions for 3D cone-beam CT with quadratic regularization. In *Proc. SPIE 6510, Medical Imaging 2007: Phys. Med. Im.*, pages 65105W:1–10, 2007.