

# Fast Variance Prediction for Iteratively Reconstructed CT with Arbitrary Geometries

Stephen M. Schmitt and Jeffrey A. Fessler

**Abstract**—Fast variance prediction for iteratively reconstructed CT images is useful for the analysis of reconstruction algorithms and potentially for automatic tube current modulation. Prior methods are either computationally intractable or require impractical computation times to produce a map of the reconstructed image variance. In this paper we present the extension of prior work for fast variance prediction, which was specific to limited classes of CT geometries, to arbitrary CT geometries. We compare the results of our method to an empirical variance map produced from repeated axial CT scans of a chest phantom.

## I. INTRODUCTION

Iterative reconstruction methods for CT offer improved resolution and noise properties compared to FBP-like reconstruction methods [1]. However, the statistical properties of iteratively reconstructed images are more difficult to analyze than those of FBP-like images.

Prior work has provided closed form but computationally intractable expressions for the covariance of an iteratively reconstructed CT image [2]. Other work has made evaluating this closed form more tractable by using frequency-domain approximations [3]. These methods require computing a projection and back-projection of each voxel of interest and are impractical for producing a map of the variance for an entire 3D volume.

In this paper, we apply further approximations to the frequency-domain approximation that significantly accelerate these methods, allowing us to produce a variance map in less time than methods that require a projection and back-projection. We compare the resulting prediction to an empirical variance map produced from repeated CT scans of the same object.

## II. METHODS

### A. Problem Domain

In this work, we consider statistically reconstructed images of the form

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} L(\mathbf{Y}; \mathbf{x}) + \alpha R(\mathbf{x}) \quad (1)$$

Here,  $L$  is the negative log-likelihood of the vectorized observations  $\mathbf{Y}$  given an image vector  $\mathbf{x}$ . The function  $R(\mathbf{x})$  is a regularization penalty. We assume:

- 1) The covariance of  $\mathbf{Y}$  is diagonal, and can be estimated from the data and knowledge of the instrumentation.

This work was supported in part by NIH grant R01 HL-098686 and by equipment donations from Intel.

Stephen 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)

- 2) Given an image  $\mathbf{x}$ , the elements of  $\mathbf{Y}$  are statistically independent, and the likelihood of a particular observation  $Y_i$  is modeled in terms of the projection  $[\mathbf{A}\mathbf{x}]_i \triangleq \sum_j a_{ij} x_j$ , such that

$$L(\mathbf{Y}; \mathbf{x}) = \sum_{i=1}^{N_d} L_i(Y_i; [\mathbf{A}\mathbf{x}]_i); \quad (2)$$

the matrix  $\mathbf{A}$  is a projection matrix with elements  $a_{ij}$  representing the projection of voxel  $j$  onto observation  $i$ . We denote the number of observations  $N_d$ .

- 3) The regularizer takes the general form

$$R(\mathbf{x}) = \sum_{d=1}^{N_c} r_d \sum_k \psi([\mathbf{C}_d \mathbf{x}]_k). \quad (3)$$

In the common case of a regularizer that penalizes first differences between neighboring voxels,  $d$  indexes the directions over which we take the differences,  $\mathbf{C}_d$  is a first differencing matrix between voxels in that direction, and  $r_d$  is the relative strength of the regularizer in that direction. We assume the regularizer penalty  $\psi$  is twice-differentiable at 0, and scaled such that  $\psi''(0) = 1$ .

- 4) Let  $\tilde{\mathbf{x}}$  denote the reconstruction using noise-free data  $\tilde{\mathbf{Y}}$ . We assume that the Hessian of the regularizer, evaluated at  $\tilde{\mathbf{x}}$ , can be approximated by  $\mathbf{P}$ :

$$\nabla^2 R(\tilde{\mathbf{x}}) \approx \mathbf{P} \triangleq \sum_{d=1}^{N_c} r_d \mathbf{C}_d^T \mathbf{C}_d. \quad (4)$$

This approximation is accurate except near edges.

### B. Methods

Previous work has computed variance predictions using local frequency domain expressions for  $\mathbf{A}^T \mathbf{W} \mathbf{A}$ , where  $\mathbf{W}$  is a diagonal statistical weighting matrix, and for  $\mathbf{P}$ , using an approximation of local shift-invariance.

The local impulse response (LIR) of  $\mathbf{A}^T \mathbf{W} \mathbf{A}$  for the voxel  $j$  is defined by

$$\mathbf{h}_j^W \triangleq \mathbf{A}^T \mathbf{W} \mathbf{A} \mathbf{e}_j, \quad (5)$$

where  $\mathbf{e}_j$  is defined as the unit vector with a single 1 at position  $j$ . This LIR can be written exactly as the impulse  $\mathbf{e}_j$  operated on by a local frequency-domain filter  $H_j^W(\vec{v})$ :

$$\mathbf{h}_j^W = \mathcal{F}^* \mathcal{D} \left\{ H_j^W \right\} \mathcal{F} \mathbf{e}_j, \quad (6)$$

where  $\mathcal{D}$  is a “diagonalization” operator:  $(\mathcal{D}\{H\}X)(\vec{v}) = H(\vec{v})X(\vec{v})$ , and  $\mathcal{F}$  is the DSFT with the spatial extent limited by the image support. We refer to  $H_j^W$  as a *local* frequency response (LFR). In the region near voxel  $j$ ,  $\mathbf{A}^T \mathbf{W} \mathbf{A}$

is typically *approximately* spatially shift-invariant, leading to an approximation

$$[\mathbf{A}^T \mathbf{W} \mathbf{A}]_{kj} \approx \mathbf{e}_k^T \mathcal{F}^* \mathcal{D} \{H_j^W\} \mathcal{F} \mathbf{e}_j, \quad (7)$$

for voxel  $k$  near voxel  $j$ , which is suggested by (5) and (6). Except at the edges of the reconstructed image,  $\mathbf{P}$  can be represented in terms of its frequency response  $R(\vec{v})$ :

$$\mathbf{P} = \mathcal{F}^* \mathcal{D} \{R\} \mathcal{F}. \quad (8)$$

In [4], we develop a separable approximation to  $H_j^W$ :

$$H_j^W(\vec{v}) \approx J(\vec{v}) E_j^W(\vec{\Theta}), \quad (9)$$

where  $\vec{\Theta} \triangleq \vec{v}/\|\vec{v}\|$  is the angle of  $\vec{v}$ . The utility of this factorization is that  $E_j^W$  is the only term dependent on the weighting  $\mathbf{W}$  and voxel location  $j$ , but as a function of  $\vec{\Theta}$  rather than  $\vec{v}$ ,  $E_j^W$  is a function of one fewer dimension than  $H_j^W$ . The  $J(\vec{v})$  term does not depend on the weighting and voxel location.

We describe applying this factorization to accelerate computing variance maps and predicting numerical observer SNRs using frequency-domain methods.

### C. Variance Prediction

For reconstructions made under the assumptions described in Section II-A, the covariance of the reconstruction, denoted  $\Sigma_{\hat{\mathbf{x}}}$ , is given approximately as [2]:

$$\Sigma_{\hat{\mathbf{x}}} \approx (\mathbf{A}^T \mathbf{W} \mathbf{A} + \alpha \nabla^2 R(\vec{\mathbf{x}}))^{-1} \mathbf{A}^T \hat{\mathbf{W}} \mathbf{A}, \quad (10)$$

$$(\mathbf{A}^T \mathbf{W} \mathbf{A} + \alpha \nabla^2 R(\vec{\mathbf{x}}))^{-1},$$

where the diagonal matrices  $\mathbf{W}$  and  $\hat{\mathbf{W}}$  are defined as:

$$[\mathbf{W}]_{ii} \triangleq \left. \frac{\partial^2}{\partial y^2} L_i(Y_i; y) \right|_{y=[\mathbf{A}\vec{\mathbf{x}}]_i} \quad (11)$$

$$[\hat{\mathbf{W}}]_{ii} \triangleq \text{var}(Y_i) \cdot \left. \frac{\partial^2}{\partial y \partial Y_i} L_i(Y_i; y) \right|_{y=[\mathbf{A}\vec{\mathbf{x}}]_i}. \quad (12)$$

Expression (10) is impractical in CT due to the inversion of a large matrix. Approximating the Hessian using (4) and using the frequency-domain approximations of (7) and (8), finding one element of (10) simplifies to

$$\text{var}(\hat{x}_j) \approx \int_{[-\frac{1}{2}, \frac{1}{2}]^n} \frac{H_j^{\hat{\mathbf{W}}}(\vec{v})}{(H_j^W(\vec{v}) + \alpha R(\vec{v}))^2} d\vec{v}. \quad (13)$$

Using the factorization of (9) in (13), we can reduce this integral by a dimension:

$$\text{var}(\hat{x}_j) \approx \alpha^{-1} \int_{\mathbb{S}^n} \frac{E_j^{\hat{\mathbf{W}}}(\vec{\Theta})}{E_j^W(\vec{\Theta})} G(\alpha^{-1} E_j^W(\vec{\Theta}), \vec{\Theta}) d\vec{\Theta}, \quad (14)$$

where  $G(\gamma, \vec{\Theta})$  is a function defined as

$$G(\gamma, \vec{\Theta}) \triangleq \int_0^{\rho_{\max}(\vec{\Theta})} \frac{\gamma J(\rho, \vec{\Theta})}{(\gamma J(\rho, \vec{\Theta}) + R(\rho, \vec{\Theta}))^2} \rho^{n-1} d\rho, \quad (15)$$

and where  $\rho_{\max}(\vec{\Theta}) = 1/(2\|\vec{\Theta}\|_{\infty})$  is the maximum extent of  $\rho$  in  $[-1/2, 1/2]^n$ . In general,  $G$  cannot be computed in a

closed form, but it is well-behaved and depends only on voxel shape (which determines  $J(\vec{v})$ ) and regularizer (which determines  $R(\vec{v})$ ). We precompute a single table of values of  $G$  and use that table to predict variance maps via (14) for multiple voxels, any regularization parameter  $\alpha$ , any weighting  $\mathbf{W}$ , any voxel spacing or scan geometry.

Specific to small cone angle 3DCT geometries, in [5] we proposed another factorization like (9). This factorization is separable not in spherical coordinates  $(\rho, \vec{\Theta})$  but in cylindrical coordinates  $(\rho, \Phi, \nu_3)$ :

$$H_j^W(\vec{v}) \approx J_{\text{cyl}}(\vec{v}) E_{j, \text{cyl}}^W(\Phi). \quad (16)$$

Like (9),  $J_{\text{cyl}}$  does not depend on the voxel location or weighting;  $E_{j, \text{cyl}}^W$  does, but is a function of only the cylinder angle  $\Phi$ . Like (14), we simplify (13) using (16):

$$\text{var}(\hat{x}_j) \approx \alpha^{-1} \int_0^{2\pi} \frac{E_{j, \text{cyl}}^{\hat{\mathbf{W}}}(\Phi)}{E_{j, \text{cyl}}^W(\Phi)} G_{\text{cyl}}(\Phi, \alpha^{-1} E_{j, \text{cyl}}^W(\Phi)) d\Phi, \quad (17)$$

where we define another object-independent function  $G_{\text{cyl}}$ :

$$G_{\text{cyl}}(\Phi, \gamma) \triangleq \int_0^{\rho_{\max}(\Phi)} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\gamma J_{\text{cyl}}(\vec{v})}{(\gamma J_{\text{cyl}}(\vec{v}) + R(\vec{v}))^2} \rho d\nu_3 d\rho. \quad (18)$$

In this case,  $\rho_{\max} = 1/(2 \max\{|\cos \Phi|, |\sin \Phi|\})$ . Again,  $G_{\text{cyl}}$  has no closed form but is a well-behaved function of only two parameters that we precompute and tabulate. We compute this table only once for a given regularizer and voxel shape. Using the table, variance prediction via (17) simply requires looking up values of  $G_{\text{cyl}}$  and numerically integrating them in 1D. This integration can use a coarse discretization of  $\Phi$  with reasonably accurate predicted variance. While the derivation differs, (17) is the form for fast variance prediction given in [4], [5], which also reduces to the form given in [6] for quadratic regularization and an axial geometry.

## III. RESULTS

To evaluate our fast variance prediction approach (17), we compared it to an empirical variance map. We scanned a thorax phantom with added spherical nodules 10 times with a GE Discovery CT750 HD scanner and reconstructed each of the 10 sinograms separately to produce the empirical variance map of the reconstruction. Each scan was a one-rotation axial scan—since we could not ensure that each scan began at the same starting angle, using multiple realizations of the same *helical* scan to produce an empirical variance was not possible with our physical CT scanner. With the axial scans, we used a projection matrix  $\mathbf{A}$  that was correctly aligned to the starting angle of each scan so that each reconstruction was aligned to the same voxel grid. We used a 40mA tube current and 120 kVp tube voltage. The scan time was 0.5 seconds.

We reconstructed each of the 10 sinograms using statistical reconstruction methods. The size of the reconstruction was  $512 \times 512 \times 32$  voxels with voxel size  $\Delta_x \times \Delta_z = 0.9764 \times 0.625\text{mm}$ . Each reconstruction used 100 iterations of an ordered-subset method [7] using 64 subsets. We performed

the reconstructions using two different regularizers. In the first case, the regularization used a quadratic penalty and was spatially varying strength using the method of [8] to produce uniform spatial resolution. In the second case, the penalty function used the Huber potential with a threshold  $\delta$  of 10 Hounsfield units but was not spatially varying in regularization strength. In both cases, the elements of the weighting matrix  $\mathbf{W}$  were chosen to correspond to the CT scanner’s estimate of the inverse of the variance of each ray given the scanner-specific corrections used [9].

Figures 1(a) (with spatially varying, quadratic regularization) and 2(a) (uniform, Huber-penalized regularization) show axial, sagittal, and coronal slices of the 3D map of the empirical standard deviation from our simulated reconstructions. As in the simulated empirical standard deviation maps, the empirical maps were noisy, so we blurred the empirical variance maps with a 2D gaussian kernel with a FWHM of 5 voxels each in each direction. Figures 1(b) and 2(b) show the corresponding slices through the 3D predicted standard deviation map from (17). We computed the standard deviation once per  $4 \times 4 \times 1$  block and used nearest-neighbor interpolation to fill in the rest. Figures 1(c) and 2(c) show the absolute magnitude of the error of our approximated standard deviation compared to the empirical results. Figures 1(d) and 2(d) show the empirical and predicted standard deviation along a one-dimensional coronal profile through the center of the image, along with the standard deviation as computed from (13) using the DSFT of  $\mathbf{h}_j^W$  as the LFR (labelled ‘DFT-based’).

Table I compares the computation time required to find the empirical variance with the computation time required to predict the variance for the entire volume using the DFT-based method and using our methods. We used the DFT-based method only to produce the one-dimensional profiles shown in Figure 1(d), and 2(d); since the computation time is large, we extrapolate to predict the DFT computation time for the entire volume for Table I.

	<b>Empirical</b>	<b>DFT-based</b>	<b>Proposed</b>
	$3.63 \cdot 10^5$	$1.07 \cdot 10^8$	$6.73 \cdot 10^2$
(10 realizations)			

TABLE I: Computation time of variance prediction methods (CPU seconds)

#### IV. DISCUSSION

We demonstrated a method that is fairly accurate in the case of quadratic regularization, and accurate away from image edges in the case of edge-preserving regularization. From the profiles in Figures 1(d) and 2(d), we can see that the majority of the error in our method is incurred in the step of approximating (10) with (13), and not in approximating (13) with (17). However, we can compute (17) faster than (13) by a factor of over  $10^5$ .

#### V. FUTURE WORK

One main area of future work is using the separable approximation (9) in fast prediction of the performance of a

linear image observer for binary classification of iteratively reconstructed images. For example, the squared SNR of the ideal non-prewhitened image observer for detecting whether a feature  $\mathbf{f}$  has been added to a background image  $\mathbf{x}$  is given by:

$$\text{SNR}^2 = \frac{(\hat{\mathbf{f}}^T \hat{\mathbf{f}})^2}{\hat{\mathbf{f}}^T \Sigma_{\hat{\mathbf{x}}} \hat{\mathbf{f}}}, \quad (19)$$

where  $\hat{\mathbf{f}}$  is the difference in the mean reconstructions with and without the feature present. While this is not easily computable due to the presence of  $\Sigma_{\hat{\mathbf{x}}}$  and  $\hat{\mathbf{f}}$ , we can make frequency-domain approximations specific to the numerator and denominator and then accelerate them by using our factorization.

	<b>Empirical</b>	<b>DFT-based</b>	<b>Proposed</b>
<b>SNR<sup>2</sup></b>	4.0	5.0	5.3
<b>Time (CPU sec.)</b>	$3.00 \cdot 10^6$	16.4	$4.32 \cdot 10^{-4}$

TABLE II: Comparison of SNR prediction methods.

Table II shows preliminary results of this observer performance prediction for a simulated 2D problem. These results are similar to those for variance prediction, namely, that there is a notable speed up using our method at the cost of decreased accuracy.

#### REFERENCES

- [1] 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.*, vol. 34, pp. 4526–44, Nov. 2007.
- [2] J. A. Fessler, “Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography,” *IEEE Trans. Im. Proc.*, vol. 5, pp. 493–506, Mar. 1996.
- [3] 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.*, vol. 18, pp. 293–305, Apr. 1999.
- [4] S. M. Schmitt and J. A. Fessler, “Fast variance prediction for iteratively reconstructed CT images,” *IEEE Trans. Med. Imag.*, 2015. In revision.
- [5] S. M. Schmitt and J. A. Fessler, “Fast variance computation for iterative reconstruction of 3D helical CT images,” in *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med.*, pp. 162–5, 2013.
- [6] S. Schmitt and J. A. Fessler, “Fast variance computation for quadratically penalized iterative reconstruction of 3D axial CT images,” in *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pp. 3287–92, 2012.
- [7] H. Erdogan and J. A. Fessler, “Ordered subsets algorithms for transmission tomography,” *Phys. Med. Biol.*, vol. 44, pp. 2835–51, Nov. 1999.
- [8] J. A. Fessler and W. L. Rogers, “Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs,” *IEEE Trans. Im. Proc.*, vol. 5, pp. 1346–58, Sept. 1996.
- [9] Z. Chang, J.-B. Thibault, K. Sauer, and C. Bouman, “Statistical X-ray computed tomography from photon-starved measurements,” in *Proc. SPIE 9020 Computational Imaging XII*, p. 90200G, 2014.

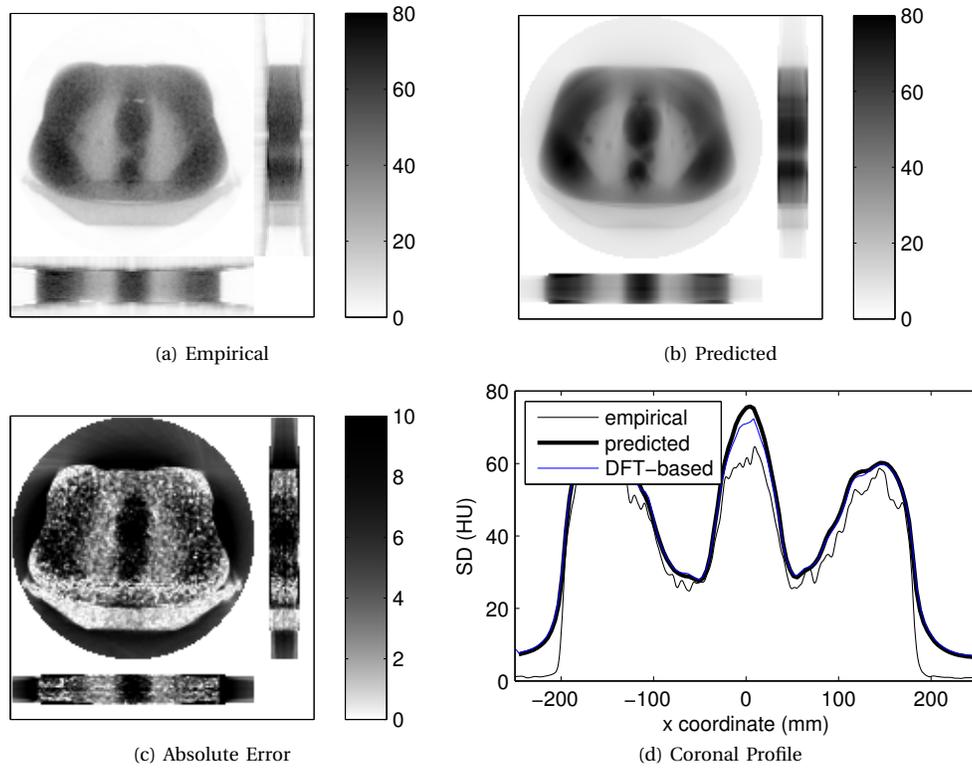


Fig. 1: Three slices of standard deviation maps using spatially varying, quadratic regularization (Hounsfield units). Coronal and sagittal slices were stretched in the trans-axial direction by a factor of two for visualization.

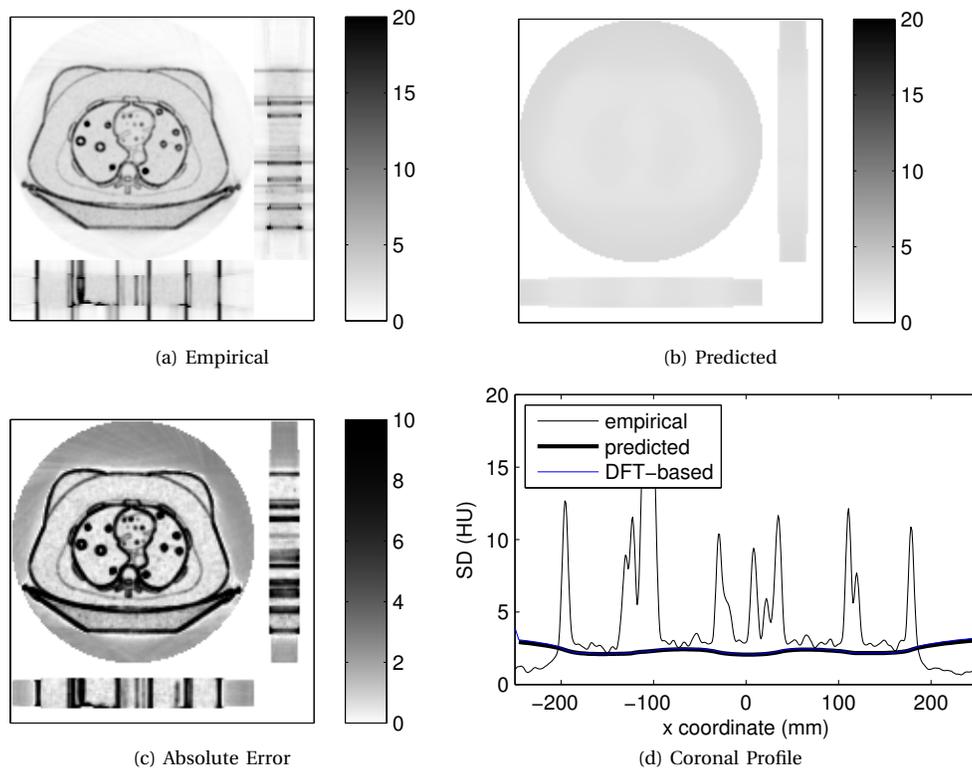


Fig. 2: Three slices of standard deviation maps using spatially uniform, Huber-penalized regularization (Hounsfield units). Coronal and sagittal slices were stretched in the trans-axial direction by a factor of two for visualization.