Published in final edited form as:

IEEE Trans Med Imaging. 2010 May ; 29(5): 1097–1113. doi:10.1109/TMI.2009.2034961.

Band-Restricted Estimation of Noise Variance in Filtered Backprojection Reconstructions Using Repeated Scans

Adam Wunderlich and Frédéric Noo [Member, IEEE]

The authors are with the Utah Center for Advanced Imaging Research, Department of Radiology, University of Utah, Salt Lake City, Utah, USA.

Adam Wunderlich: awunder@uclair.med.utah.edu; Frédéric Noo: noo@uclair.med.utah.edu

Abstract

We introduce a new estimator for noise variance in tomographic images reconstructed using algorithms of the filtered backprojection type. The new estimator operates on data acquired from repeated scans of the object under examination, is unbiased, and is shown to have significantly lower variance than the conventional unbiased estimator for many scenarios of practical interest. We provide an extensive theoretical analysis of this estimator, highlighting the circumstances under which it is most effective. This analysis includes both general and specific data correlation patterns. Moreover, we have applied our estimator to real x-ray computed tomography data and present preliminary results that support the theory and provide experimental evidence of the new estimator's efficacy.

Index Terms

noise; variance; estimation; CT; FBP

I. INTRODUCTION

In x-ray computed tomography (CT) and other medical imaging modalities, changes in the image formation process generally impact image quality. The changes can vary from the data acquisition specifics to the image reconstruction algorithm. To establish whether or not an improvement has been achieved, a careful and thorough assessment of image quality is required. Such assessment generally involves estimation of image noise, because statistical effects are a prominent aspect of image quality. This important issue was already acknowledged in the early years of CT; see e.g., [1]–[8]. Recently, we proposed a new variance estimator and also a new covariance estimator for the noise in images reconstructed with filtered backprojection (FBP) algorithms [9], which are currently used by all medical CT scanners. Here, we present a thorough, detailed analysis of the new noise variance estimator.

A simple approach to image variance estimation involves the substitution of spatial averages for ensemble averages. That is, the noise variance in a particular region of interest in an image may be estimated from a single noisy image realization by computing the sample variance of all pixel values in the region. Such an approach makes the strong assumption that the pixel values are taken from an ergodic, and hence stationary, random process [10]. Moreover, it requires that there are enough pixels in the region of interest so that statistical variations in the variance estimate are small. These constraints are not very satisfactory for CT imaging, which is our topic of interest here. For investigations of image noise in CT, it is preferable to use ensemble averages instead of spatial averages because CT image noise is highly non-stationary. (The non-stationarity of CT noise is primarily caused by the object-dependent nature of statistical variations in the data.)

Often, the initial evaluation of new CT reconstruction algorithms is performed from computer-simulated data of mathematically-defined phantoms. In this context, two techniques can be used to obtain an ensemble estimate of the image variance. The first technique involves computing multiple independent, identically distributed (iid) realizations of the noisy CT data set, reconstructing one image from each of these data sets, and then computing the variance of an individual pixel value from the sample of images; see, e.g., [11]–[14]. Since the data is simulated, it is usually possible to generate a large number of image realizations. However, in some situations, such as for 3-D volumes reconstructed from cone-beam data, it may not be practical to compute the large number of reconstructions required in order to get an accurate estimate of the image variance. In the second technique, the image variance is computed by analytically modeling each step of the reconstruction process [13], [15]–[20]. Such analytical models typically assume that the CT data is uncorrelated and that the variance of the CT data is known, which is the case when using simple noise models and mathematically-defined phantoms. Unfortunately, analytical propagation of noise through the reconstruction algorithm can sometimes be challenging and the computational cost can be large.

Other procedures, such as evaluation of refinements in the data acquisition or late-stage characterization of new reconstruction algorithms, require using real data. In these circumstances, the two techniques described in the last paragraph may be implemented from repeated CT scans of a phantom, but they have limitations. The first approach is straightforward, but demands a large number of scans to be accurate, and there are practical limits to the number of scans that can be reasonably acquired. The second approach, which relies on analytical modeling, can be implemented by estimating the CT data variance from the repeated scans and assuming that the data is uncorrelated. Unfortunately, since real CT data is typically correlated due to effects such as detector crosstalk and afterglow, and because this correlation is difficult to characterize accurately, variance estimators based on analytical models are bound to yield biased results, with a bias that is difficult to predict.

Compared to analytical models, the image variance estimator that we present in this paper is always unbiased and makes relatively weak assumptions on the correlations present in the CT data. Namely, our estimator only requires knowledge of one parameter, A , which specifies that each view of filtered data is correlated with at most A previous views and A subsequent views. In other words, the filtered CT projection data are allowed to be correlated within each view, and also between some views as described by A . Moreover, knowledge of the correlation magnitudes is not required. If each view of filtered data is known to be correlated with at most A previous views and A subsequent views, then we show that compared to the conventional approach of estimating image variance directly from a sample of images, our estimator has lower variance by a factor of as much as $N/(2A + 1)$ (approximately), where N is the number of views; see the end of section V-B. As a specific example, if $N = 1000$ and $A = 2$, then our estimator may have as much as 200 times lower variance than the conventional estimator, and it thus requires many fewer scans to achieve comparable accuracy. Hence, we believe that our new image variance estimator presents strong advantages over the other estimation strategies mentioned above, especially for application to CT data sets where a small value for A can be safely assumed.

This paper is organized as follows. First, section II outlines our notation and reviews key concepts. Then, section III introduces our noise variance estimator in a very general form, applicable to any FBP reconstruction algorithm. We have performed a theoretical analysis of this estimator including a comparison to a conventional estimator (the first technique described in the third paragraph of this introduction). This analysis is given in sections IV and V. Specifically, section IV considers the special case in which filtered data from different views are uncorrelated, and section V turns to the general case in which each view of filtered data is correlated with a known number of adjacent views. Our analysis is not only supported by

mathematical arguments, but also by preliminary experimental results using real x-ray CT data given in section VI. These results are in excellent agreement with the theory. Finally, section VII summarizes our results and conclusions.

II. PRELIMINARIES

In this section, we set our notation and review relevant material. Throughout the paper, we write vectors in boldface and denote the transpose of vectors and matrices with the superscript T , e.g., $\mathbf{x}^T$ is the transpose of the vector $\mathbf{x}$. Also, indicial notation is used to refer to individual entries of vectors and matrices. For example, x_j denotes the j th entry of vector $\mathbf{x}$ and M_{jk} denotes the element of matrix M that is in the j th row and the k th column.

We assume that the reader is familiar with the multivariate normal distribution [21], [22]. A $p \times 1$ random vector $\mathbf{X} \in \mathbb{R}^p$ following a multivariate normal distribution with mean $\boldsymbol{\mu}$ and nonsingular covariance matrix Σ will be denoted $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$.

A. The Sample Covariance Matrix

Consider a $p \times 1$ random vector $\mathbf{X} \in \mathbb{R}^p$ with mean $\boldsymbol{\mu} := E[\mathbf{X}]$ and covariance matrix $\Sigma := E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$, where $E[\cdot]$ denotes the expectation operator. Suppose that we have $n > 1$ independent, identically distributed (iid) measurements of $\mathbf{X}$, given by $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(n)}$ and that we wish to estimate Σ . The conventional unbiased estimator for Σ is the sample covariance matrix [21], [22],

$$S := (n - 1)^{-1} \sum_{i=1}^n (\mathbf{X}^{(i)} - \bar{\mathbf{X}})(\mathbf{X}^{(i)} - \bar{\mathbf{X}})^T, \quad (1)$$

where $\bar{\mathbf{X}} := (1/n) \sum_{i=1}^n \mathbf{X}^{(i)}$ is the sample mean. Throughout this paper, we will use an overbar to denote the sample mean.

The following useful lemma characterizes the covariance between two entries of the sample covariance matrix, S , when the samples follow a multivariate normal distribution.

Lemma 1—If S is computed from independent samples $\mathbf{X}^{(i)} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ for $i = 1, \dots, n$, then for any $j, k, l, m \in \{1, \dots, p\}$, $\text{Cov}[S_{jk}, S_{lm}] = (\Sigma_{jl}\Sigma_{km} + \Sigma_{jm}\Sigma_{kl})/(n - 1)$.

Proof: The proposition follows from Theorems 3.3.6 and 3.3.15(i) of [23].

B. FBP Reconstruction Methods

We consider image reconstruction from x-ray CT projections collected at N view positions, $\mathbf{a}(\lambda_j)$, on a given source trajectory, with $j = 1, 2, \dots, N$. A filtered backprojection (FBP) reconstruction algorithm [24] consists of two basic steps. First, there is a filtering step, in which the CT projection data is modified through some combination of rebinning, weighting, and filtering. Second, there is a backprojection step in which the filtered data is interpolated, possibly weighted, and then summed over λ for each point $\mathbf{x}$. The interpolated, filtered data from view j used for reconstruction at $\mathbf{x}$ is written as $g_j(\mathbf{x})$, and the backprojection weight for view j and $\mathbf{x}$ is denoted by $b_j(\mathbf{x})$, which is assumed to be strictly positive. Typically, the backprojection weight is voxel-dependent and is inversely proportional to a power of $\|\mathbf{x} - \mathbf{a}(\lambda_j)\|$.

Since the steps outlined above may be implemented in various ways, and since we wish to keep our discussion general, the development in this paper will be based on a generic expression for the final backprojection step, given by

$$f(\mathbf{x}) = \sum_{j=1}^N b_j(\mathbf{x}) g_j(\mathbf{x}). \quad (2)$$

Above, $f(\mathbf{x})$ is the FBP reconstruction at the point $\mathbf{x}$. Equivalently, we may express (2) as the vector inner product

$$f(\mathbf{x}) = \mathbf{b}_\mathbf{x}^T \mathbf{g}_\mathbf{x}, \quad (3)$$

$$\mathbf{b}_\mathbf{x} := [b_1(\mathbf{x}), b_2(\mathbf{x}), \dots, b_N(\mathbf{x})]^T, \quad (4)$$

$$\mathbf{g}_\mathbf{x} := [g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_N(\mathbf{x})]^T. \quad (5)$$

C. Noise in FBP Reconstructions

We wish to calculate the variance of the FBP reconstruction at a given location $\mathbf{x}$, as defined by (2), from multiple CT data sets obtained through repeated scans of the object of interest. For this purpose, we view these multiple data sets as repeated measurements of the interpolated, filtered data vector $\mathbf{g}_\mathbf{x}$. Statistically speaking, $\mathbf{g}_\mathbf{x}$ is a random vector. The variance estimators which we consider in this paper apply without any specific assumptions on the distribution of $\mathbf{g}_\mathbf{x}$, except that its mean and covariance matrix are well-defined. However, for all theoretical analyses we assume that $\mathbf{g}_\mathbf{x}$ is multivariate normal. For most CT data sets, we believe that this assumption is valid as long as the number of photons reaching the detector for each ray is large enough. This belief is in agreement with experimental observations made in [25], [26].

For an image reconstructed with the FBP formula of (3), the variance of a pixel value may be expressed in terms of covariances of the interpolated, filtered data. Let the covariance matrix of $\mathbf{g}_\mathbf{x}$ be denoted as

$$\Gamma(\mathbf{x}) := E[(\mathbf{g}_\mathbf{x} - E[\mathbf{g}_\mathbf{x}])(\mathbf{g}_\mathbf{x} - E[\mathbf{g}_\mathbf{x}])^T]. \quad (6)$$

Using (3), the variance of the image pixel at $\mathbf{x}$ is seen to be

$$\text{Var}[f(\mathbf{x})] = \mathbf{b}_\mathbf{x}^T \Gamma(\mathbf{x}) \mathbf{b}_\mathbf{x}. \quad (7)$$

Alternatively, the above equation may be written as

$$\text{Var}[f(\mathbf{x})] = \sum_{j=1}^N \sum_{k=1}^N b_j(\mathbf{x}) b_k(\mathbf{x}) \Gamma_{jk}(\mathbf{x}). \quad (8)$$

D. A Conventional Image Variance Estimator

Suppose that we perform $n > 1$ repeated CT scans of an object, and that we wish to estimate the variance of an image pixel, assuming that the reconstruction is performed with an FBP method. A conventional approach to noise estimation is to first reconstruct the n iid CT data sets to obtain n iid image realizations, and then apply the standard unbiased estimator for variance [27, p. 212]. Specifically, if we denote the value of the i th image realization at pixel location $\mathbf{x}$ as $f^{(i)}(\mathbf{x})$, then the image variance may be estimated as

$$\widehat{\text{Var}}_c[f(\mathbf{x})] := (n-1)^{-1} \sum_{i=1}^n \left(f^{(i)}(\mathbf{x}) - \overline{f(\mathbf{x})} \right)^2, \quad (9)$$

where $\overline{f(\mathbf{x})} = (1/n) \sum_{i=1}^n f^{(i)}(\mathbf{x})$ is the sample mean. We call the image noise estimator given by (9) the conventional variance estimator (CVE). In our notation, the hat emphasizes that the quantity defined by (9) is an estimator, and the subscript c on the estimator stands for *conventional*.

The CVE may also be written in terms of the interpolated, filtered data $\mathbf{g}_{\mathbf{x}}$. Suppose that each of the n CT data sets is filtered and interpolated to obtain n realizations of $\mathbf{g}_{\mathbf{x}}$. Denote the i th realization as $\mathbf{g}_{\mathbf{x}}^{(i)}$, and define the sample covariance matrix for $\mathbf{g}_{\mathbf{x}}$ as

$$G(\mathbf{x}) := (n-1)^{-1} \sum_{i=1}^n (\mathbf{g}_{\mathbf{x}}^{(i)} - \bar{\mathbf{g}}_{\mathbf{x}})(\mathbf{g}_{\mathbf{x}}^{(i)} - \bar{\mathbf{g}}_{\mathbf{x}})^T, \quad (10)$$

where $\bar{\mathbf{g}}_{\mathbf{x}} := (1/n) \sum_{i=1}^n \mathbf{g}_{\mathbf{x}}^{(i)}$. Utilizing (3), the value of the i th image realization at pixel location $\mathbf{x}$ is $f^{(i)}(\mathbf{x}) = \mathbf{b}_{\mathbf{x}}^T \mathbf{g}_{\mathbf{x}}^{(i)}$. Inserting this relation into (9), the CVE may be written as

$$\widehat{\text{Var}}_c[f(\mathbf{x})] = \mathbf{b}_{\mathbf{x}}^T G(\mathbf{x}) \mathbf{b}_{\mathbf{x}}. \quad (11)$$

In summation form, the above equation becomes

$$\widehat{\text{Var}}_c[f(\mathbf{x})] = \sum_{j=1}^N \sum_{k=1}^N b_j(\mathbf{x}) b_k(\mathbf{x}) G_{jk}(\mathbf{x}). \quad (12)$$

Observe that (11) and (12) are simply straightforward sample estimates of (7) and (8), respectively.

III. THE BAND-RESTRICTED VARIANCE ESTIMATOR

For data collected with a typical CT scanner, the x-ray measurements within each projection are correlated, and each projection may be correlated with projections collected at adjacent views. However, a given projection is not typically correlated with projections from distant views. Moreover, the filtering and interpolation steps required to compute $\mathbf{g}_{\mathbf{x}}$ in an FBP algorithm typically induce correlations only within a view and possibly between adjacent views. We will use this a priori knowledge to obtain a better estimator for image variance.

Suppose that the interpolated, filtered data within any view j is correlated with the interpolated, filtered data from at most A previous views and A subsequent views. That is, A is the smallest non-negative integer which ensures that, for any j and k , $\Gamma_{jk}(\mathbf{x}) = 0$ if $|j - k| > A$. Using this assumption together with (8), it follows that the variance of an image pixel is

$$\text{Var}[f(\mathbf{x})] = \sum_{(j,k) \in \Delta} b_j(\mathbf{x}) b_k(\mathbf{x}) \Gamma_{jk}(\mathbf{x}), \quad (13)$$

where

$$\Delta := \{(j, k) \mid 1 \leq j \leq N, 1 \leq k \leq N, |j - k| \leq A\}. \quad (14)$$

Now, we use (13) to motivate an image variance estimator. Suppose we have a sample of $n > 1$ data sets resulting from n repeated CT scans. From these data sets, we compute n realizations of the filtered, interpolated data, $\mathbf{g}_\mathbf{x}$. Observe that the covariances $\Gamma_{jk}(\mathbf{x})$ in (13) may be estimated with the corresponding entries $G_{jk}(\mathbf{x})$ of the sample covariance matrix, defined by (10). Doing this, we obtain the following estimator for image variance:

$$\widehat{\text{Var}}_r[f(\mathbf{x})] := \sum_{(j,k) \in \Delta} b_j(\mathbf{x}) b_k(\mathbf{x}) G_{jk}(\mathbf{x}). \quad (15)$$

Equation (15) may also be expressed in the vector-matrix form

$$\widehat{\text{Var}}_r[f(\mathbf{x})] = \mathbf{b}_\mathbf{x}^T G^{\mathcal{R}}(\mathbf{x}) \mathbf{b}_\mathbf{x}, \quad (16)$$

where the entries of the band-restricted sample covariance matrix, $G^{\mathcal{R}}(\mathbf{x})$, are defined as

$$G_{jk}^{\mathcal{R}}(\mathbf{x}) := \begin{cases} G_{jk}(\mathbf{x}) & \text{if } |j - k| \leq A \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

We call the image variance estimator defined by (15) the band-restricted variance estimator (BRVE). The subscript r on the estimator stands for *restricted*. The name for this estimator emphasizes that it depends on entries of the sample covariance matrix, $G(\mathbf{x})$, only within a specified band of width $2A + 1$. On the other hand, the CVE does not have this restriction, since it uses all entries of $G(\mathbf{x})$; see (11).

By construction, the BRVE is unbiased and is consistent (i.e., converges to the true image variance as $n \rightarrow \infty$). However, there is no guarantee that the new estimator is non-negative, because $G^{\mathcal{R}}(\mathbf{x})$ is not necessarily a positive definite matrix, unless $A = 0$. Nonetheless, we will show that this disadvantage is largely compensated by attractive convergence properties compared to the CVE.

Our definition of the BRVE falls within the topic of estimating covariance matrices with a known pattern, which has received attention in the statistics literature [28]–[37]. Optimal estimation of such matrices is a challenging problem, especially as far as obtaining analytical results is concerned. To our knowledge, none of the theoretical results that we present here have been published. As an alternative to explicit analytical estimators, many investigators have proposed algorithms to compute the maximum likelihood estimator of patterned covariance matrices, e.g., [29]–[32], [37]. These algorithms are iterative, and they are also computationally demanding relative to the BRVE. Moreover, little is known concerning the magnitude of the bias and variance of these estimates (and functions of them) for finite n as compared to the CVE.

IV. THEORETICAL ANALYSIS OF THE BRVE: $A = 0$

Here, we consider the performance of the BRVE in the $A = 0$ case, i.e., when adjacent projections of the interpolated, filtered data are uncorrelated. We start by assessing the optimality of the BRVE and subsequently compare the BRVE to the CVE. This $A = 0$ case applies to evaluation of many FBP reconstruction algorithms from computer-simulated data. It may also be useful for evaluations from real data sets provided that the effects which create correlations between views, such as detector afterglow, are well-mitigated.

As discussed in section II-C, the interpolated, filtered data vector, $\mathbf{g}_\mathbf{x}$, is assumed to follow a multivariate normal distribution, with mean denoted by $\boldsymbol{\eta}(\mathbf{x})$, i.e., $\mathbf{g}_\mathbf{x} \sim \mathcal{N}(\boldsymbol{\eta}(\mathbf{x}), \Gamma(\mathbf{x}))$. In the

development below, functional dependencies on $\mathbf{x}$ are frequently suppressed for notational convenience.

A. Optimality Evaluation

Since the BRVE is an unbiased estimator of image variance, it is important to ask whether or not it is the uniformly minimum variance unbiased (UMVU) estimator [38], [39]. The following theorem answers this question for $A = 0$:

Theorem 1—*Suppose that $A = 0$ and that the BRVE is computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. Then the BRVE is the unique UMVU estimator for the image variance.*

Proof: The joint probability distribution function (pdf) of the measurements, $\mathbf{g}_x^{(i)}$, is

$$p(\mathbf{g}_x^{(1)}, \dots, \mathbf{g}_x^{(n)}) = (2\pi)^{-nN/2} |\Gamma|^{-n/2} \exp \left[-\frac{\omega}{2} \right], \quad (18)$$

where $|\Gamma|$ denotes the determinant of Γ , and

$$\omega := \sum_{i=1}^n (\mathbf{g}_x^{(i)} - \eta)^T \Gamma^{-1} (\mathbf{g}_x^{(i)} - \eta). \quad (19)$$

With the aid of the additive and cyclic properties of the trace, written $\text{tr}(\cdot)$, we may express ω as

$$\omega = \text{tr} \left(\Gamma^{-1} \left[\sum_{i=1}^n (\mathbf{g}_x^{(i)} - \eta)(\mathbf{g}_x^{(i)} - \eta)^T \right] \right). \quad (20)$$

Rearranging the above summation in brackets yields

$$\sum_{i=1}^n (\mathbf{g}_x^{(i)} - \eta)(\mathbf{g}_x^{(i)} - \eta)^T = \sum_{i=1}^n (\mathbf{g}_x^{(i)} - \bar{\mathbf{g}}_x)(\mathbf{g}_x^{(i)} - \bar{\mathbf{g}}_x)^T + n(\bar{\mathbf{g}}_x - \eta)(\bar{\mathbf{g}}_x - \eta)^T \quad (21)$$

$$= (n-1)G + n\bar{\mathbf{g}}_x\bar{\mathbf{g}}_x^T + n\eta\eta^T - n\eta\bar{\mathbf{g}}_x^T - n\bar{\mathbf{g}}_x\eta^T, \quad (22)$$

where G is defined by (10). Next, we use the assumption that Γ is diagonal together with (20) and (22) to find that ω takes the form

$$\omega = \sum_{k=1}^N \Gamma_{kk}^{-1} \left((n-1)G_{kk} + n\bar{g}_k^2 + n\eta_k^2 - 2n\eta_k\bar{g}_k \right), \quad (23)$$

where $\bar{g}_k$ denotes the k th component of $\bar{\mathbf{g}}_x$. Combining (18) and (23), the joint pdf for the sample may be rewritten as

$$p(\mathbf{g}_x^{(1)}, \dots, \mathbf{g}_x^{(n)}) = (2\pi)^{-nN/2} |\Gamma|^{-n/2} \exp \left[-\frac{n}{2} \eta^T \Gamma^{-1} \eta \right] \times \exp \left[-(1/2) \sum_{k=1}^N \Gamma_{kk}^{-1} ((n-1)G_{kk} + n\bar{g}_k^2) \right] \times \exp \left[n \sum_{k=1}^N \eta_k \Gamma_{kk}^{-1} \bar{g}_k \right]. \quad (24)$$

By the Fisher-Neyman factorization theorem [38, Thm. 6.5, p. 35] [39, Prop. IV.C.1, p. 159], the statistic

$$T := [(n-1)G_{11} + n\bar{g}_1^2, (n-1)G_{22} + n\bar{g}_2^2, \dots, (n-1)G_{NN} + n\bar{g}_N^2, \bar{g}_1, \bar{g}_2, \dots, \bar{g}_N] \quad (25)$$

is sufficient. Moreover, because the expression in (24) has the form of a full rank exponential family [38, p. 23–24], T is a complete statistic [38, Thm. 6.22, p. 42]. Since (i) T is a complete sufficient statistic, (ii) the BRVE is an unbiased estimator of image variance, and (iii)

$\widehat{\text{Var}}_r[f(\mathbf{x})] = E[\widehat{\text{Var}}_r[f(\mathbf{x})] | T]$, i.e., the BRVE is a function of T only, the Lehmann-Scheffé Theorem [38, Thm. 1.11, p. 88] [39, p. 164] implies that the BRVE is the unique UMVU estimator of image variance.

In addition, the BRVE for the $A = 0$ case is closely related to the maximum likelihood estimator (MLE) [38], [39] for image variance. Specifically, the MLE asymptotically approaches the BRVE with increasing n , while at any given n , the MLE has slightly lower variance than the BRVE but is biased. The next theorem clarifies this relationship.

Theorem 2—*Suppose that $A = 0$ and that the BRVE is computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. Then the MLE for image variance is equal to the BRVE multiplied by $(n - 1)/n$.*

Proof: We start by finding the MLE for the diagonal covariance matrix, Γ . For this purpose, we use a technique employed by Muirhead [21, Thm. 3.1.5, p. 83–84].

The joint pdf of the sample is given by (18) and (19). Using (20) and (21) together with the cyclic property of the trace, we see that the likelihood function is

$$L(\eta, \Gamma) = (2\pi)^{-nN/2} |\Gamma|^{-n/2} \exp[-0.5(n-1) \text{tr}(\Gamma^{-1}G)] \times \exp[-(n/2)(\bar{\mathbf{g}}_x - \eta)^T \Gamma^{-1}(\bar{\mathbf{g}}_x - \eta)]. \quad (26)$$

Clearly,

$$L(\eta, \Gamma) \leq (2\pi)^{-nN/2} |\Gamma|^{-n/2} \exp[-0.5(n-1) \text{tr}(\Gamma^{-1}G)] \quad (27)$$

with equality if and only if $\eta = \bar{\mathbf{g}}_x$, since Γ^{-1} is positive definite. Hence, $L(\eta, \Gamma)$ is maximized only if $\eta = \bar{\mathbf{g}}_x$.

Next, we find the Γ which maximizes the function $\mathcal{L}(\Gamma) := L(\bar{\mathbf{g}}_x, \Gamma)$, or equivalently, any logarithm of this function. Because Γ is diagonal, $\mathcal{L}(\Gamma)$ takes the form

$$\mathcal{L}(\Gamma) = (2\pi)^{-nN/2} \left(\prod_{k=1}^N G_{kk}^{-1} (\Gamma_{kk}^{-1} G_{kk}) \right)^{n/2} \times \exp \left[-0.5(n-1) \sum_{k=1}^N \Gamma_{kk}^{-1} G_{kk} \right]. \quad (28)$$

Taking the natural logarithm of both sides yields

$$\log \mathcal{L}(\Gamma) = -(nN/2) \log(2\pi) - (n/2) \sum_{k=1}^N \log(G_{kk}) + \frac{1}{2} \sum_{k=1}^N [n \log(\Gamma_{kk}^{-1} G_{kk}) - (n-1) \Gamma_{kk}^{-1} G_{kk}]. \quad (29)$$

Let $\lambda_k := \Gamma_{kk}^{-1} G_{kk}$ and consider the function $f(\lambda_k) = n \log \lambda_k - (n-1) \lambda_k$. Because this function is maximized at $\lambda_k = n/(n-1)$, it follows that

$$\log \mathcal{L}(\Gamma) \leq -(nN/2) \log(2\pi) - (n/2) \sum_{k=1}^N \log(G_{kk}) + (nN/2) [\log(n/(n-1)) - 1], \quad (30)$$

where equality is achieved if and only if $\Gamma_{kk}^{-1} G_{kk} = n/(n-1)$ for every k . In other words, $\log \mathcal{L}(\Gamma)$ is maximized if and only if $\Gamma_{kk} = [(n-1)/n] G_{kk}$ for all k . Therefore, in this case, we conclude that the MLE for Γ is $[(n-1)/n] G^{\mathcal{R}}$, where $G^{\mathcal{R}}$ is defined by (17) with $A = 0$.

By the invariance property of maximum likelihood estimators [40, Thm. 7.4, p. 185] [41, Lem. 5.1.3, p. 76–77], (7) implies that the MLE for image variance is equal to $[(n-1)/n]\mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x$. Comparing this expression with (16), we see that the MLE is equal to the BRVE multiplied by a factor of $(n-1)/n$.

B. Comparison with the CVE

From Theorem 1, we know that when $A = 0$, the BRVE has smaller variance than the CVE. It is instructive to further examine both the origin and the magnitude of this difference. First, we state a useful result which shows that the diagonal entries of the sample covariance matrix, G , are uncorrelated with the off-diagonal entries.

Lemma 2—*Suppose $A = 0$ and G is computed from independent samples*

$\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, \dots, n$. Then for any $k, l, m \in \{1, \dots, N\}$ with $l \neq m$, $\text{Cov}[G_{kk}, G_{lm}] = 0$.

Proof: Lemma 1 with $j = k$ yields $\text{Cov}[G_{kk}, G_{lm}] = [2/(n-1)]\Gamma_{kl}\Gamma_{km}$. Since $l \neq m$, either $k = l$ or $k = m$. Consequently, since Γ is diagonal, either $\Gamma_{kl} = 0$ or $\Gamma_{km} = 0$. The preceding formula thus implies that $\text{Cov}[G_{kk}, G_{lm}] = 0$.

Observe from (11) and (16) that the BRVE for the $A = 0$ case uses only the diagonal entries of the sample covariance matrix, G , whereas the CVE incorporates all entries of this matrix. Moreover, (11) and (16) imply that the CVE may be written as the sum of the BRVE plus a linear combination of the off-diagonal entries of G , i.e.,

$$\widehat{\text{Var}}_c[f(\mathbf{x})] = \widehat{\text{Var}}_r[f(\mathbf{x})] + \mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x. \quad (31)$$

From Lemma 2, it follows that when $A = 0$, the BRVE is uncorrelated with the second term above. Therefore, the variance of the CVE consists of the variance of the BRVE plus the variance of the second term. Specifically,

$$\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = \text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] + \text{Var}[\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x]. \quad (32)$$

The second term on the right side of (32) may be written as

$$\text{Var}[\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] = \sum_{(j,k) \in \nabla} \sum_{(l,m) \in \nabla} b_j b_k b_l b_m \text{Cov}[G_{jk}, G_{lm}], \quad (33)$$

where ∇ is defined as

$$\nabla := \{(l, m) \mid 1 \leq l \leq N, 1 \leq m \leq N, |l - m| > A\}. \quad (34)$$

Using Lemma 1, (33) becomes

$$\text{Var}[\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] = \sum_{(j,k) \in \nabla} \sum_{(l,m) \in \nabla} \frac{b_j b_k b_l b_m}{n-1} (\Gamma_{jl}\Gamma_{km} + \Gamma_{jm}\Gamma_{kl}). \quad (35)$$

As discussed in section II-B, all of the backprojection weights are strictly positive. Therefore, given that the diagonal entries of Γ are strictly positive due to Γ being positive definite, all of the summation terms in (35) are either zero or strictly positive. Moreover, the summation includes nonzero terms; such terms appear, for example, when $(j, k) = (l, m) \in \nabla$, in which case we get contributions of the form $b_j b_k b_l b_m \Gamma_{jl}\Gamma_{km} = b_l^2 b_m^2 \Gamma_{ll}\Gamma_{mm} > 0$. Hence,

$\text{Var}[\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] > 0$, demonstrating that the variance of the CVE is strictly larger than the variance of the BRVE because the CVE includes an additional term that depends only on the off-diagonal entries of G .

We may gain further insight by recalling from the proof of Theorem 1 that the BRVE is a function of the complete sufficient statistic T , defined by (25). Since any complete sufficient statistic is minimal sufficient [38, p. 42], the BRVE is therefore a function of a minimal sufficient statistic. Intuitively, this means that the BRVE is based on a statistic that optimally reduces the available data while still retaining all of the relevant information. By contrast, the CVE cannot be obtained as a function of the minimal sufficient statistic T , and hence, the CVE makes an inefficient use of the data.

The next proposition characterizes the magnitude of the variance improvement offered by the BRVE when $A = 0$.

Theorem 3—*Suppose that $A = 0$ and that the CVE and the BRVE are computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. Then*

$$\frac{\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]]}{\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]]} = \frac{(\sum_{j=1}^N b_j^2 \Gamma_{jj})^2}{\sum_{j=1}^N b_j^4 \Gamma_{jj}^2}.$$

Proof: Utilizing (3), the value of the i th image realization at pixel location $\mathbf{x}$ is $f^{(i)}(\mathbf{x}) = \mathbf{b}_x^T \mathbf{g}_x^{(i)}$. Since any linear combination of a multivariate normal vector is also normal [21, Thm. 1.2.6, p. 6], it follows that the iid realizations of $f(\mathbf{x})$ are each normal. Hence, Lemma 1 with $p = 1$ implies that $\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = [2/(n-1)](\text{Var}[f(\mathbf{x})])^2$. Using the structure of Γ , (8) reduces to $\text{Var}[f(\mathbf{x})] = \sum_{j=1}^N b_j^2 \Gamma_{jj}$. Thus, the variance of the CVE is

$$\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = 2(n-1)^{-1} \left(\sum_{j=1}^N b_j^2 \Gamma_{jj} \right)^2. \quad (36)$$

In the $A = 0$ case, (15) for the BRVE reduces to $\widehat{\text{Var}}_r[f(\mathbf{x})] = \sum_{j=1}^N b_j^2 G_{jj}$. Since all of the G_{jj} are independent, $\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] = \sum_{j=1}^N b_j^4 \text{Var}(G_{jj})$. Moreover, Lemma 1 with $j = k = l = m$ implies that $\text{Var}[G_{jj}] = 2\Gamma_{jj}^2/(n-1)$. Hence, the variance of the BRVE is

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] = 2(n-1)^{-1} \sum_{j=1}^N b_j^4 \Gamma_{jj}^2. \quad (37)$$

The ratio of (36) and (37) yields the stated formula.

Observe that the variance ratio in Theorem 3 above is independent of n . If we further assume that there is a positive constant d such that $b_j^2 \Gamma_{jj} = d$ for all j , then Theorem 3 yields Corollary 1 below. The assumption that $b_j^2 \Gamma_{jj} = d$ for all j applies, for example, in good approximation when the classical parallel-beam FBP algorithm is used for reconstruction (that is, $b_j = 1$ for all j) and data is acquired with a perfectly compensating bowtie filter (i.e., with a shaped piece

of material that is placed between the x-ray source and the object so that all rays hitting the detector have the same intensity).

Corollary 1: Suppose that $A = 0$ and that the CVE and the BRVE are computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. If there exists a positive constant d such that $b_j^2 \Gamma_{jj} = d$ for each $j \in \{1, 2, \dots, N\}$, then

$$\frac{\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]]}{\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]]} = N.$$

Since a typical commercially-available CT scanner uses $N \approx 1000$, the variance improvement offered by the BRVE can thus be quite substantial in some circumstances.

In the other extreme, the energy of $b_j^2 \Gamma_{jj}$ could be very concentrated around a given j value. In such a case, Theorem 3 implies that the BRVE only offers a slight advantage over the CVE.

For illustration, assume that there exists some index $J \in \{1, 2, \dots, N\}$ such that $b_j^2 \Gamma_{jj} = d$ for $j = J$ and $b_j^2 \Gamma_{jj} = \varepsilon$ for $j \neq J$ where d and ε are positive constants with ε much smaller than d . Under this assumption, Theorem 3 shows that the ratio of the CVE variance to the BRVE variance is close to one. Situations where the energy of $b_j^2 \Gamma_{jj}$ is concentrated around a given j value appear when data collected from one view have much higher noise than data from other views, or when $\mathbf{x}$ is close to the source trajectory and the backprojection weights are inversely proportional to a positive power of $\|\mathbf{x} - \mathbf{a}(\lambda)\|$.

V. THEORETICAL ANALYSIS OF THE BRVE: $A > 0$

Next, we examine the performance of the BRVE in the general case when $A > 0$. After investigating the optimality of the BRVE, we contrast the BRVE with the CVE. As before, we assume that the interpolated, filtered data vector, $\mathbf{g}_x$, follows a multivariate normal distribution, and we frequently suppress functional dependencies on $\mathbf{x}$ for notational convenience.

For the following discussion, it is informative to note that in practice, A is usually small compared to the number of views, N . In particular, on a typical commercial CT scanner collecting $N \approx 1000$ projections, we have observed that the amount of between-view correlation is commonly well-approximated by values of A less than or equal to two.

A. Optimality Evaluation

We start with a heuristic argument supporting the effectiveness of the BRVE. For this argument, we decompose the covariance matrix for $\mathbf{g}_x$ as $\Gamma = \Gamma^d + \Gamma^o$, where Γ^d and Γ^o are respectively composed of the diagonal and off-diagonal entries of Γ , i.e.,

$\Gamma_{jk}^d = \Gamma_{jj}$ if $j=k$, $\Gamma_{jk}^d = 0$ if $j \neq k$, and $\Gamma_{jk}^o = 0$ if $j=k$, $\Gamma_{jk}^o = \Gamma_{jk}$ if $j \neq k$. By definition, Γ^d is diagonal and Γ^o is a band matrix of bandwidth $2A + 1$.

Denote the $N \times N$ identity matrix by I and let

$$\Omega := (\Gamma^d)^{-1/2} \Gamma^o (\Gamma^d)^{-1/2} \quad (38)$$

be the matrix of off-diagonal correlation coefficients. For a typical CT scanner, the between-view correlation can be expected to decay quickly. Therefore, it is plausible that $\|\Omega\| < 1$ for a

suitable matrix norm $\|\cdot\|$. If this is this case, then the inverse of $I + \Omega$ exists, and the following Neumann series converges [42, p. 301] [43, p. 254]:

$$(I + \Omega)^{-1} = (I - (-\Omega))^{-1} = \sum_{k=0}^{\infty} (-1)^k \Omega^k. \quad (39)$$

Now, consider the decomposition

$$\Gamma = (\Gamma^d)^{1/2} (I + (\Gamma^d)^{-1/2} \Gamma^o (\Gamma^d)^{-1/2}) (\Gamma^d)^{1/2} \quad (40)$$

$$= (\Gamma^d)^{1/2} (I + \Omega) (\Gamma^d)^{1/2}. \quad (41)$$

Since Γ is positive definite, it follows that $(\Gamma^d)^{1/2}$ and $I + \Omega$ are invertible. Therefore, the inverse of Γ may be written as

$$\Gamma^{-1} = (\Gamma^d)^{-1/2} (I + \Omega)^{-1} (\Gamma^d)^{-1/2}. \quad (42)$$

Under the assumption that $\|\Omega\| < 1$, the expression given by (39) may be substituted into (42) to find that

$$\Gamma^{-1} = (\Gamma^d)^{-1/2} \left[\sum_{k=0}^{\infty} (-1)^k \Omega^k \right] (\Gamma^d)^{-1/2}. \quad (43)$$

For any integer $L \geq 0$, we define the L th order approximation of Γ^{-1} as

$$\Theta := \sum_{k=0}^L (-1)^k (\Gamma^d)^{-1/2} \Omega^k (\Gamma^d)^{-1/2}. \quad (44)$$

Because Ω is a symmetric band matrix of bandwidth $2A + 1$, the k th power of Ω is a symmetric band matrix with bandwidth $2kA + 1$. Consequently, it follows from (44) that Θ is a symmetric band matrix of bandwidth $2LA + 1$.

Suppose that the BRVE is computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. By combining (18), (20), and (22), the joint pdf of the sample can be seen to take the form

$$p(\mathbf{g}_x^{(1)}, \mathbf{g}_x^{(2)}, \dots, \mathbf{g}_x^{(n)}) = (2\pi)^{-nN/2} |\Gamma|^{-n/2} \exp \left[-(n/2) \eta^T \Gamma^{-1} \eta \right] \times \exp \left[-(1/2) \text{tr} \left(\Gamma^{-1} ((n-1)G + n \bar{\mathbf{g}}_x \bar{\mathbf{g}}_x^T) \right) \right] \times \exp \left[n \text{tr} \left(\Gamma^{-1} (\eta \bar{\mathbf{g}}_x^T) \right) \right]. \quad (45)$$

If the L th-order approximation given in (44) is accurate, then the joint pdf of the sample is well-approximated as

$$p(\mathbf{g}_x^{(1)}, \mathbf{g}_x^{(2)}, \dots, \mathbf{g}_x^{(n)}) \approx (2\pi)^{-nN/2} |\Gamma|^{-n/2} \exp \left[-(n/2) \eta^T \Theta \eta \right] \times \exp \left[-(1/2) \text{tr} \left(\Theta ((n-1)G + n \bar{\mathbf{g}}_x \bar{\mathbf{g}}_x^T) \right) \right] \times \exp \left[n \text{tr} \left(\Theta (\eta \bar{\mathbf{g}}_x^T) \right) \right]. \quad (46)$$

The expression on the right of (46) is not necessarily a valid pdf since Θ may be singular and since $|\Gamma| \neq |\Theta|$. Nonetheless, in the heuristic discussion below, we treat this expression as though it actually is a pdf.

Let B be the $N \times N$ matrix defined as

$$B := (n-1)G + n \bar{\mathbf{g}}_x \bar{\mathbf{g}}_x^T. \quad (47)$$

Since B is symmetric, it is easy to see that

$$\text{tr}(\Theta B) = \sum_{j=1}^N \sum_{k=1}^N \Theta_{jk} B_{jk}. \quad (48)$$

So (46) depends on the entry B_{jk} if and only if Θ_{jk} is nonzero. Define the matrix $\tilde{G}$ to contain the entries of G which correspond to nonzero entries of Θ , i.e.,

$$\tilde{G}_{jk} := \begin{cases} G_{jk} & \text{if } \Theta_{jk} \neq 0 \\ 0 & \text{otherwise} \end{cases}. \quad (49)$$

Now, consider the statistic $\tilde{T} = [\mathbf{g}_{\mathbf{x}}, \tilde{G}]$. The Fisher-Neyman factorization theorem [38, Thm. 6.5, p. 35] [39, Prop. IV.C.1, p. 159] suggests that $\tilde{T}$ is sufficient in the sense of (46).

The matrix Θ is, by definition, a symmetric band matrix of bandwidth $2LA + 1$. By contrast, Γ has a bandwidth of $2A + 1$. If approximation (46) with $L = 1$ is accurate, then the completeness theorem for exponential families [38, Theorem 6.22, p.42] [39, Prop. IV.C.3, p. 165] suggests that $\tilde{T}$ is complete. Thus, if the joint sample pdf is well-approximated by (46) with $L = 1$, then $\tilde{T}$ is a complete sufficient statistic in the sense of this approximation, and the Lehmann-Scheffé Theorem [38, Thm. 1.11, p. 88] [39, p. 164] implies that the BRVE is the unique UMVU estimator for the image variance. Hence, the BRVE is the UMVU estimator for the image variance as a first-order approximation in the above sense.

On the other hand, if approximation (46) is only accurate for some $L > 1$, then because the dimension of $\tilde{T}$ is larger than the dimension of the parameter space, the completeness theorem for exponential families [38, Theorem 6.22, p.42] [39, Prop. IV.C.3, p. 165] does not apply. Hence, we cannot easily determine whether or not $\tilde{T}$ is complete (and hence minimal sufficient) in this approximate setting.

In light of the above observations, it is natural to wonder if the BRVE is generally the UMVU estimator. The following theorem shows that the answer is negative.

Theorem 4—*Suppose that the BRVE is computed from independent samples*

$\mathbf{g}_{\mathbf{x}}^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. For any integers N and A with $0 < A < N - 1$, there exists a ball of nonzero volume in the parameter space over which the BRVE is not the minimum variance unbiased estimator.

Proof: See Appendix A.

Likewise, it is interesting to ask how the BRVE relates to the MLE for image variance in the general case when $0 < A < N - 1$. If one can show that the BRVE is a function of the MLE for Γ , then the relationship between the BRVE and the MLE for image variance follows from the invariance property of maximum likelihood estimators [40, Thm. 7.4, p. 185] [41, Lem. 5.1.3, p. 76–77]. (We followed this approach in the proof of Theorem 2 in section IV-A.)

Unfortunately, finding the MLE for Γ in the case when Γ is a non-diagonal, band matrix is not straightforward and is still a subject of active research [33], [34]. Moreover, we have verified to our satisfaction that $[(n-1)/n]G^{\mathcal{R}}$ is generally not equal to the MLE for Γ when $0 < A < N - 1$. On these grounds, we therefore conjecture that the BRVE is not simply linked to the MLE for image variance. Because explicit construction of the MLE for band matrices is difficult, proving this assertion is likely onerous.

B. Comparison with the CVE

We have learned that the BRVE is the UMVU estimator as a first-order approximation but is not the UMVU estimator in general. Below, we contrast the BRVE with the CVE.

Recall from the last section that the statistic $\tilde{T} = [\tilde{\mathbf{g}}_{\mathbf{x}}, \tilde{G}]$ is sufficient in the sense of approximation (46). Also, (45) and the Fisher-Neyman factorization theorem [38, Thm. 6.5, p. 35] [39, Prop. IV.C.1, p. 159] imply that the statistic $\hat{T} := [\mathbf{g}_{\mathbf{x}}, G]$ is sufficient. Thus, in an approximate sense, both $\tilde{T}$ and $\hat{T}$ contain all information in the sample. However, $\tilde{T}$ generally may be obtained as a function of $\hat{T}$, but not vice-versa.

In CT imaging, the truncated Neumann series in (44) is likely to be a good approximation to Γ^{-1} with L being very small, e.g., $L \approx 10$. Moreover, A is much smaller than N . Therefore, $\tilde{G}$ is a band matrix with a bandwidth which is much smaller than N , i.e., $\tilde{T}$ has much smaller dimension than $\hat{T}$. As a consequence, $\tilde{T}$ extracts the sample information much more efficiently than $\hat{T}$. Because the CVE is a function of $\hat{T}$ but not of $\tilde{T}$, it appears that the CVE makes a relatively inefficient use of the sample. Hence, since the BRVE is a function of $\tilde{T}$, we expect it to have better performance than the CVE. The following theorem shows that the BRVE outperforms the CVE when the interpolated, filtered data from different views all have a non-negative correlation.

Theorem 5—*Suppose that the CVE and the BRVE are computed from independent samples $\mathbf{g}_{\mathbf{x}}^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. If $\Gamma_{jk} = 0$ for all $j, k \in \{1, 2, \dots, N\}$, then*

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] < \text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]].$$

Proof: Using (11) and (17), the CVE may be rewritten as

$$\widehat{\text{Var}}_c[f(\mathbf{x})] = \mathbf{b}_{\mathbf{x}}^T G^{\mathcal{R}} \mathbf{b}_{\mathbf{x}} + \mathbf{b}_{\mathbf{x}}^T (G - G^{\mathcal{R}}) \mathbf{b}_{\mathbf{x}}. \quad (50)$$

Thus, the variance of the CVE is given by

$$\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = \text{Var}[\mathbf{b}_{\mathbf{x}}^T G^{\mathcal{R}} \mathbf{b}_{\mathbf{x}}] + \text{Var}[\mathbf{b}_{\mathbf{x}}^T (G - G^{\mathcal{R}}) \mathbf{b}_{\mathbf{x}}] + 2 \text{Cov}[\mathbf{b}_{\mathbf{x}}^T G^{\mathcal{R}} \mathbf{b}_{\mathbf{x}}, \mathbf{b}_{\mathbf{x}}^T (G - G^{\mathcal{R}}) \mathbf{b}_{\mathbf{x}}]. \quad (51)$$

Equations (16) and (51) imply that the difference in the variances of the CVE and the BRVE takes the form

$$\begin{aligned} \delta V &:= \text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] - \text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] \\ &= \text{Var}[\mathbf{b}_{\mathbf{x}}^T (G - G^{\mathcal{R}}) \mathbf{b}_{\mathbf{x}}] + 2 \text{Cov}[\mathbf{b}_{\mathbf{x}}^T G^{\mathcal{R}} \mathbf{b}_{\mathbf{x}}, \mathbf{b}_{\mathbf{x}}^T (G - G^{\mathcal{R}}) \mathbf{b}_{\mathbf{x}}]. \end{aligned} \quad (52)$$

The first term on the right side of (52) is expressed in summation form by (35). As discussed in section II-B, all of the backprojection weights are strictly positive. Because the entries of Γ are assumed to be non-negative, it then follows that each of the terms in (35) is non-negative. Because Γ is positive definite, its diagonal entries are strictly positive. Hence, if $(j, k) = (l, m) \in \nabla$, then $b_j b_k b_l b_m \Gamma_{jl} \Gamma_{km} = b_l^2 b_m^2 \Gamma_{ll} \Gamma_{mm} > 0$. Thus, there are some strictly positive terms in (35), and we see that $\text{Var}[\mathbf{b}_{\mathbf{x}}^T (G - G^{\mathcal{R}}) \mathbf{b}_{\mathbf{x}}] > 0$.

The covariance in the second term on the right side of (52) may be written in summation form as

$$\text{Cov}[\mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x, \mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] = \sum_{(j,k) \in \Delta} \sum_{(l,m) \in \nabla} b_j b_k b_l b_m \text{Cov}[G_{jk}, G_{lm}], \quad (53)$$

where the set Δ was defined by (14).

Using Lemma 1, (53) becomes

$$\text{Cov}[\mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x, \mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] = \sum_{(j,k) \in \Delta} \sum_{(l,m) \in \nabla} \frac{b_j b_k b_l b_m}{n-1} (\Gamma_{jl} \Gamma_{km} + \Gamma_{jm} \Gamma_{kl}). \quad (54)$$

Since the entries of Γ are non-negative and since the backprojection weights are strictly positive, it follows that $\text{Cov}[\mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x, \mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] \geq 0$. Thus, (52) and the above observations imply that $\delta V > 0$.

The next theorem shows that the BRVE can also outperform the CVE when some correlations between views are negative. For the statement of the theorem, we define the maximum modulus of the off-diagonal correlation coefficients as

$$\rho_{\max} := \max_{\substack{1 \leq j, k \leq N \\ j \neq k}} |\Gamma_{jk} / \sqrt{\Gamma_{jj} \Gamma_{kk}}|. \quad (55)$$

Theorem 6—*Suppose that the CVE and the BRVE are computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. For any $\mathbf{b}_x$ and Γ , there exists a strictly positive constant C such that if $\rho_{\max} < C$, then*

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] < \text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]].$$

Proof: From (52), (35), and (54), note that δV consists of only three kinds of terms:

$b_j^2 b_l^2 \Gamma_{jj} \Gamma_{ll}$, $b_j^2 b_l b_m \Gamma_{jj} \Gamma_{lm}$ with $l \neq m$, and $b_j b_k b_l b_m \Gamma_{jk} \Gamma_{lm}$ with $j \neq k$ and $l \neq m$. Let N_1 be the number of terms of the first kind, and notice that each such term is bounded below as follows:

$b_j^2 b_l^2 \Gamma_{jj} \Gamma_{ll} \geq b_{\min}^4 \Gamma_{\min}^2$, where $b_{\min}$ and $\Gamma_{\min}$ are defined as $b_{\min} := \min_{1 \leq j \leq N} b_j$ and $\Gamma_{\min} := \min_{1 \leq j \leq N} \Gamma_{jj}$. Let N_2 be the number of terms of the second kind; each of these terms satisfy

$b_j^2 b_l b_m \Gamma_{jj} \Gamma_{lm} \geq -b_{\max}^4 \rho_{\max}^2 \Gamma_{\max}^2$, where $b_{\max} := \max_{1 \leq j \leq N} b_j$ and $\Gamma_{\max} := \max_{1 \leq j \leq N} \Gamma_{jj}$. Finally, let N_3 be the number of terms of the third kind, which are such that

$b_j b_k b_l b_m \Gamma_{jk} \Gamma_{lm} \geq -b_{\max}^4 \rho_{\max}^2 \Gamma_{\max}^2$. This decomposition of δV and the lower bounds above yield

$$\delta V \geq b_{\max}^4 \Gamma_{\max}^2 (n-1)^{-1} [N_1 (b_{\min}^2 \Gamma_{\min} / (b_{\max}^2 \Gamma_{\max}))^2 - N_2 \rho_{\max} - N_3 \rho_{\max}^2]. \quad (56)$$

Considered as a function of $\rho_{\max}$, the quantity on the right side of (56) is a concave down parabola that attains its maximum at a (fictitious) negative value of $\rho_{\max}$ and is strictly positive at $\rho_{\max} = 0$. Therefore, there exists a constant $C > 0$ such that if $\rho_{\max} < C$, then $\delta V > 0$.

The previous theorem is very general, as it applies to any value of A , but it comes with no explicit indication of how small $\rho_{\max}$ must be for the BRVE to outperform the CVE. Close examination of the proof shows that the constant C can potentially be small; unfavorable settings occur when A is large, because then N_2 and N_3 can be large, and when there are strong

variations in the diagonal entries of Γ or in the backprojection weights. Theorem 7 below provides an explicit expression for the constant C from Theorem 6 for the particular case when $A = 1$. Notice that Theorem 7 involves some of the notation introduced in the proof of Theorem 6. Nevertheless, we repeat for the reader's convenience all required symbols so that Theorem 7 can be understood without reference to the proof of Theorem 6.

Theorem 7—Suppose that $A = 1$ and that the CVE and the BRVE are computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. For any $\mathbf{b}_x$ and Γ , define the constant

$$\beta := \frac{2}{(N-2)^2} \sum_{j=1}^{N-2} \frac{b_j^2 \Gamma_{jj}}{b_{\max}^2 \Gamma_{\max}} \sum_{k=j+2}^N \frac{b_k^2 \Gamma_{kk}}{b_{\max}^2 \Gamma_{\max}},$$

where

$$b_{\max} := \max_{1 \leq j \leq N} b_j \quad \text{and} \quad \Gamma_{\max} := \max_{1 \leq j \leq N} \Gamma_{jj}.$$

If

$$\rho_{\max} < \frac{N-1}{2(N-2)} \left[-1 + \sqrt{1 + \left(\frac{N-2}{N-1} \right)^2 \beta} \right],$$

then

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] < \text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]].$$

Proof: See Appendix B.

To better understand how small $\rho_{\max}$ must be to satisfy the hypothesis in Theorem 7, suppose that $b_j^2 \Gamma_{jj} = b_{\max}^2 \Gamma_{\max}$ for all $j \in \{1, 2, \dots, N\}$. This assumption applies, for example, in good approximation when the classical parallel-beam FBP algorithm is used for reconstruction and the data is acquired with a perfectly compensating bowtie filter. In this case, it is easy to see

that $\beta = 2/[(N-2)^2] \sum_{j=1}^{N-2} (N-j-1) = (N-1)/(N-2)$. Thus, for large N , $\beta \approx 1$, and the right side of the $\rho_{\max}$ inequality in Theorem 7 is roughly equal to $(1/2)[-1 + \sqrt{2}] \approx 0.2$. Since it may be reasonable to assume that $\rho_{\max} < 0.2$ in experimental settings, this result suggests that the BRVE offers a potential advantage over the CVE for many situations of practical interest.

On the other hand, if the energy of $b_j^2 \Gamma_{jj} / (b_{\max}^2 \Gamma_{\max})$ is very concentrated, then $\rho_{\max}$ needs to be very small for Theorem 7 to be satisfied, indicating that the BRVE might then not easily outperform the CVE. For example, suppose that there exists some index $J \in \{1, 2, \dots, N\}$ such that $b_J^2 \Gamma_{JJ} = b_{\max}^2 \Gamma_{\max}$ for $j = J$ and $b_j^2 \Gamma_{jj} = \epsilon b_{\max}^2 \Gamma_{\max}$ for $j \neq J$ where ϵ is a positive constant much smaller than one. In this situation, β is nearly 0, and hence, the upper bound on $\rho_{\max}$ is very small.

So far, we have not investigated the magnitude of the estimator variance ratio for the CVE and the BRVE in the general case when $A > 0$. First, note that this ratio is always independent of n for normally distributed measurements, as was observed for the $A = 0$ case (see Theorem 3); this is easily seen from (12) and (15) together with Lemma 1, as n appears only through a common multiplicative factor in the estimators' variance. The next two theorems characterize the magnitude of the estimator variance ratio for the cases $A = 1$ and $A = 2$ when the covariance matrix, Γ , is Toeplitz and the backprojection weights are constant.

Theorem 8—Let $\gamma_o > 0$, $\rho \in [-1, 1]$, and $d > 0$ be constants. Suppose that $A = 1$ and that the CVE and the BRVE are computed from independent samples $\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, \dots, n$. If $\Gamma_{jj} = \gamma_o$ and $b_j = d$ for each $j \in \{1, \dots, N\}$, and if $\Gamma_{j,j+1} = \Gamma_{j+1,j} = \rho\gamma_o$ for each $j \in \{1, \dots, N-1\}$ with ρ such that $\Gamma(\mathbf{x})$ is positive definite, then

$$\frac{\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]]}{\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]]} = \frac{[N(1+2\rho) - 2\rho]^2}{N(3+8\rho+8\rho^2) - 2 - 8\rho - 12\rho^2}.$$

Proof: See Appendix C.

In the above theorem, ρ may be interpreted as a correlation coefficient. Asymptotic theory for the minimum eigenvalue of a Toeplitz matrix [44, Lemma 4.1, p. 41] may be used to show that the domain of ρ where Γ is positive definite quickly converges to the region $|\rho| < 0.5$ as N increases. Also, when N is much larger than 1, the formula in Theorem 8 may be approximated as

$$\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]]/\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] \approx NR_1(\rho), \quad (57)$$

where

$$R_1(\rho) := (1+2\rho)^2/(3+8\rho+8\rho^2). \quad (58)$$

A plot of R_1 versus $\rho \in (-0.5, 0.5)$ is given in Fig. 1 (left). When N is large, we observe that, except for a small interval bounded below by $\rho = -0.5$, the BRVE has significantly smaller variance than the CVE. It is important to note, however, that the variance gain offered by the BRVE for the $A = 1$ case is not as great as for the $A = 0$ case; see Corollary 1 in section IV-B. The relative size of the interval where the CVE has smaller variance than the BRVE is plotted as a function of N in Fig. 1 (right) with a solid line. For $N = 1000$ views, the CVE has lower variance than the BRVE over only 1.58% of the ρ values in the interval $(-0.5, 0.5)$.

Theorem 9—Let $\gamma_o > 0$, $\rho_1 \in [-1, 1]$, $\rho_2 \in [-1, 1]$, and $d > 0$ be constants. Suppose that $A = 2$ and that the CVE and the BRVE are computed from independent samples

$\mathbf{g}_x^{(i)} \sim \mathcal{N}_N(\eta(\mathbf{x}), \Gamma(\mathbf{x}))$ for $i = 1, 2, \dots, n$. If $\Gamma_{jj} = \gamma_o$ and $b_j = d$ for each $j \in \{1, 2, \dots, N\}$, if $\Gamma_{j,j+1} = \Gamma_{j+1,j} = \rho_1\gamma_o$ for each $j \in \{1, 2, \dots, N-1\}$, and if $\Gamma_{j,j+2} = \Gamma_{j+2,j} = \rho_2\gamma_o$ for each $j \in \{1, 2, \dots, N-2\}$, where ρ_1 and ρ_2 are such that $\Gamma(\mathbf{x})$ is positive definite, then

$$\frac{\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]]}{\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]]} = \frac{C(\rho_1, \rho_2, N)}{D(\rho_1, \rho_2, N)}.$$

where

$$C(\rho_1, \rho_2, N) := [N(1+2\rho_1+2\rho_2) - 2\rho_1 - 4\rho_2]^2,$$

$$D(\rho_1, \rho_2, N) := N(5+16\rho_1+12\rho_2+24\rho_1\rho_2+16\rho_1^2+12\rho_2^2) - 6 - 24\rho_1 - 24\rho_2 - 56\rho_1\rho_2 - 32\rho_1^2 - 36\rho_2^2.$$

Proof: See Appendix D.

Similar to the previous theorem, ρ_1 and ρ_2 may be interpreted as correlation coefficients. If N is much larger than 1, then the formula in Theorem 9 may be approximated as

$$\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]]/\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] \approx NR_2(\rho_1, \rho_2), \quad (59)$$

where

$$R_2(\rho_1, \rho_2) := \frac{(1+2\rho_1+2\rho_2)^2}{16\rho_1+12\rho_2+24\rho_1\rho_2+16\rho_1^2+12\rho_2^2+5}. \quad (60)$$

Asymptotic theory for the minimum eigenvalue of a Toeplitz matrix [44, Lemma 4.1, p. 41] may again be used to find the domain of (ρ_1, ρ_2) where Γ is positive definite for arbitrarily large N . Fig. 1 (center) contains a contour plot of $R_2(\rho_1, \rho_2)$ over this domain. Examining this figure, we see that for (ρ_1, ρ_2) values outside a region bounded by the line $\rho_2 = -\rho_1 - 1/2$, the BRVE has much smaller variance than the CVE. However, the maximum variance gain afforded by the BRVE for the $A = 2$ case is not as great as for the $A = 1$ case. The relative size of the region near the line $\rho_2 = -\rho_1 - 1/2$ where the CVE has smaller variance than the BRVE is plotted as a function of N in Fig. 1 (right) with a dashed line. Interestingly, we find that the plot for the $A = 2$ case is very similar to the plot obtained in the $A = 1$ case. For the specific case of $N = 1000$ views, the CVE has lower variance than the BRVE over only 1.46% of the (ρ_1, ρ_2) values in the region over which Γ is positive definite.

For both the $A = 1$ and $A = 2$ cases discussed above, the regions where the CVE has lower variance than the BRVE correspond to situations when the variance of the CVE is nearly zero. From Lemma 1 with $p = 1$ it follows that $\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = [2/(n-1)](\text{Var}[f(\mathbf{x})])^2$, and hence, the variance of the CVE is nearly zero only when the variance of the image pixel is nearly zero. Therefore, the CVE outperforms the BRVE in the two above cases only when the correlations in the filtered data are such that the variance of the reconstructed image pixel is very close to zero, which is impractical. In other words, the likelihood that the CVE outperforms the BRVE is small in practice.

Finally, note that $R_1(\rho)$ and $R_2(\rho_1, \rho_2)$ both behave as $1/(2A+1)$ when the correlations are small. In this case, the variance of the BRVE is roughly $N/(2A+1)$ times smaller than the variance of the CVE, as mentioned in the introduction.

VI. EXPERIMENTAL RESULTS

To give further evidence for the effectiveness of the BRVE, we now present some preliminary results regarding the application of the BRVE to real x-ray CT data. In particular, we compare variance maps obtained with the CVE and the BRVE corresponding to images reconstructed with the classical FBP algorithm for direct reconstruction of fan-beam data acquired over a full circular scan; see [19] for a description of this algorithm as we used it. The reconstructions assume a rectangular apodization window for the ramp filter kernel, a data redundancy weight

of $1/2$ for all rays, and are computed on a 640×640 grid of pixels, where each pixel has dimensions of $0.075 \text{ cm} \times 0.075 \text{ cm}$ and the grid covers a field-of-view with radius 24 cm.

For the experimental data acquisition, we used a Siemens *SOMATOM Sensation 64* CT scanner to repeatedly scan a thorax phantom 186 times over a circular source trajectory. The phantom consisted of a torso constructed by QRM (Möhrendorf, Germany) [45] together with two different water bottles attached to the sides to simulate arms. A reconstruction of the phantom is shown in Fig. 2. The scans were executed in a thorax scan mode using a 2-slice acquisition with a slice thickness of 1 mm; the x-ray tube settings were 25 mAs and 120 kVp, and the rotational speed was 3 revolutions per second, which implies that no flying-focal-spot was involved. Also, the *Care Dose* feature (x-ray tube current modulation) was turned off. Over the 186 repeated scans, the measurements for the first of the two slices amounted to 186 fan-beam data sets that constituted the repeated fan-beam measurements for our experiment.

The CT data that we used was the data that was saved by the scanner at scanning-time. This data differs from purely raw measurements due to proprietary pre-processing steps performed by the scanner, but was deemed appropriate for our experiment, especially since it is the data that the scanner uses for reconstruction. We read the data using software supplied by Siemens, and we fed it directly into our own implementation of the classical direct fan-beam FBP algorithm, without performing any additional pre-processing prior to reconstruction.

Before applying the BRVE to the acquired set of CT data realizations, it was necessary to choose a value for A , the amount of between-view correlation. Such a selection could be based on inspection of sample covariance matrices for various $\mathbf{g}_x$. Unfortunately, this approach was not practical here as the limited number of repeated scans available to us resulted in too much statistical uncertainty in the estimated covariance matrices. For this reason, we resorted to the indirect approach of mutually comparing the BRVE estimates obtained with a number of values of A . Still, statistical uncertainty needed to be mitigated. To control this uncertainty, we used averages in x and y . Fig. 3 and Fig. 4 show a comparison of such averages obtained with $A = 0, 1, 2$, and 10 using all $n = 186$ repeated scans. The comparison is based on relative differences computed using $A = 1$ as a reference. This comparison shows that the correlations between views have only a small impact on the BRVE estimate, as the difference between using $A = 1$ and using $A = 0$ is small. The plots in the figures also show that using A larger than 1 does not bring any apparent structural information, although it does increase statistical variability in the estimates, as expected. Hence, we decided to proceed with $A = 1$. We note that the relative difference between $A = 1$ and $A = 10$ exhibits strong deviations from zero at large x and y values, but these deviations are most likely due to the reduction in the number of pixels involved in the averaging process, which is due to the estimated variance images being defined only in the circular field of view of the scanner.

Fig. 5 presents a comparison of variance images obtained with the CVE and the BRVE. The left side of the figure displays the CVE and BRVE variance estimates computed from $n = 186$ CT data realizations. Examining these images, we see that the BRVE estimate appears to be similar to the CVE estimate, but with much lower variance. On the right side of Fig. 5, the CVE and BRVE estimates computed from only $n = 5$ data realizations are shown. While the CVE estimate exhibits strong statistical uncertainties in this case, the BRVE estimate based on only $n = 5$ data realizations is comparable to the CVE estimate obtained from $n = 186$ data realizations.

Fig. 6 shows the profiles along the x -axis for the BRVE image variance estimates with $n = 186$ and $n = 5$, which are displayed in the bottom row of Fig. 5. As a reference, the CVE estimate computed from all $n = 186$ CT data sets is shown in light gray in both plots. On the left side, the BRVE estimate based on all $n = 186$ data sets is plotted in black. Here, the BRVE result

with $n = 186$ appears close to the $n = 186$ CVE result, but with a much lower variance. On the right, the BRVE estimate based on only $n = 5$ data sets is plotted in black. Inspecting this plot, we observe that the BRVE estimate with $n = 5$ closely follows the trends of the $n = 186$ CVE result, with slightly lower variance.

Together, Fig. 5 and Fig. 6 demonstrate good agreement with our theoretical analysis of the BRVE and show its potential for improved image variance estimation in x-ray CT. Of course, a more thorough experimental investigation is desirable because the simple experiment reported here does not shed light on all aspects of the BRVE and since the presented results suffer from statistical variability.

VII. DISCUSSION AND CONCLUSIONS

We introduced a new unbiased estimator, named the band-restricted variance estimator (BRVE), for noise variance in images reconstructed with FBP algorithms. The BRVE operates on an iid sample of CT data sets obtained from repeated CT scans of an object, and requires no statistical information for the data other than one parameter, A , which describes the maximum amount of between-view correlation in the interpolated, filtered data, $\mathbf{g}_x$. Assuming that $\mathbf{g}_x$ has a multivariate normal distribution, we showed that the BRVE is the uniformly minimum variance unbiased (UMVU) estimator for image variance when $A = 0$. In the general case when $A > 0$, we proved that the BRVE is not the UMVU estimator. Nonetheless, we found that the BRVE has lower variance than the conventional variance estimator (CVE) if either all between-view correlations are positive, or if the maximum modulus of the between-view correlations is small enough. Furthermore, we found precise formulae describing the magnitude of the CVE to BRVE variance ratio in the cases $A = 0$, $A = 1$, and $A = 2$, when the backprojection weights are constant and the covariance matrix of $\mathbf{g}_x$ is Toeplitz. These formulae demonstrate that the variance reduction offered by the BRVE can be quite large when A is small relative to N . Finally, we presented preliminary experimental results establishing that the BRVE can be effectively applied to real x-ray CT data sets, and showing that in the case of real fan-beam CT data, the BRVE may require many fewer repeated scans to achieve comparable accuracy to the CVE. Although our experimental results involved data with no x-ray tube current modulation, the BRVE is of course not limited to such data. In other words, tube current modulation does not hinder use of the BRVE, but the value of A may change.

A comprehensive experimental assessment of the BRVE was beyond the scope of this paper. In a forthcoming publication, we will present a thorough evaluation using both computer-simulated and real CT data; see [9] for preliminary results. In this evaluation, we will examine the effectiveness of the BRVE for different data acquisition geometries, e.g., parallel-beam, fan-beam and cone-beam, and different FBP reconstruction algorithms, e.g., classical direct reconstruction from fan-beam data [24], classical fan-to-parallel beam rebinning reconstruction [24], and Katsevich's algorithm for the helical cone-beam geometry [46]. We will also experimentally investigate how to best select A and how the performance of the BRVE is impacted by the assumption of different A values as well as by variations in the backprojection weights; Theorems 3 and 7 give an indication of how the backprojection weights play a role. Furthermore, we will present a formula for the variance of the BRVE in the general case, and illustrate how this formula may be applied to estimate the BRVE variance effectively; see [9] for a statement of the BRVE variance formula.

We did not discuss the computational complexity of the BRVE here. It is not difficult to see that for a fixed number of scans n , the BRVE has computational requirements similar to those of the CVE. However, since the BRVE can achieve accuracy comparable to the CVE with many fewer data realizations, a large computational savings will often result from implementation of the BRVE.

Although the discussion was focused on filtered backprojection methods, the BRVE also applies directly to calculation of the variance in a differentiated backprojection (see e.g., [47], [48]). We note that, in such cases, some of the backprojection weights may be negative, particularly for formulations involving M -lines [48]. This difficulty is overcome by incorporating the sign of the negative backprojection weights within the definition of the interpolated, filtered data. Extension of the BRVE methodology to iterative reconstruction algorithms is also of high interest, but does not appear to be straightforward. Initial investigations could possibly build on the work in [49].

In a future work, we will explore the generalization of the BRVE to the estimation of covariance between image pixels. We first described this estimator in [9] and we call it the band-restricted covariance estimator (BRCE). The BRCE is potentially very exciting, because it may enable accurate estimation of the full image covariance matrix from a relatively small number of CT scans. One possible application is to task-based image quality evaluation using model observers, such as the channelized Hotelling observer [50].

Acknowledgments

This work was partially supported by NIH grants R01 EB007236, R21 EB009168, and by a generous grant from the Ben B. and Iris M. Margolis Foundation. Its contents are solely the responsibility of the authors.

APPENDIX A

We now prove Theorem 4, which shows that the BRVE is generally not the UMVU estimator of image variance. In the proof, we construct another image variance estimator which has smaller variance than the BRVE over a ball of nonzero volume of the parameter space. For notational convenience, we denote the BRVE as $\hat{\theta}_1$ and the alternative image variance estimator as $\hat{\theta}_2$. Also, the functional dependence of quantities on $\mathbf{x}$ is suppressed below.

Recall the definition of the restricted sample covariance matrix, $G^{\mathcal{R}}$, given by (17). Define a modified version of this matrix as

$$\hat{G}_{jk}^{\mathcal{R}} := \begin{cases} G_{11} + G_{1,A+2} & \text{if } (j, k) = (1, 1) \\ G_{jk}^{\mathcal{R}} & \text{otherwise.} \end{cases} \quad (61)$$

From (16), we have $\hat{\theta}_1 = \mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x$. The definition for the other image variance estimator is $\hat{\theta}_2 = \mathbf{b}_x^T \hat{G}^{\mathcal{R}} \mathbf{b}_x$. Observe that $\hat{\theta}_2 = \hat{\theta}_1 + b_1^2 G_{1,A+2}$ and that $\hat{\theta}_2$ is unbiased.

The difference of the variances for $\hat{\theta}_2$ and $\hat{\theta}_1$ is

$$\delta(\Gamma) := \text{Var}(\hat{\theta}_2) - \text{Var}(\hat{\theta}_1) \quad (62)$$

$$= b_1^4 \text{Var}(G_{1,A+2}) + 2b_1^2 \text{Cov}(\hat{\theta}_1, G_{1,A+2}). \quad (63)$$

The functional dependence of δ on Γ is clarified next.

Using the fact that $\Gamma_{1,A+2} = 0$ together with Lemma 1 shows that the variance of $G_{1,A+2}$ is

$$\text{Var}(G_{1,A+2}) = (n-1)^{-1} \Gamma_{11} \Gamma_{A+2,A+2}. \quad (64)$$

Since $G^{\mathcal{R}}$ is a symmetric matrix, $\hat{\theta}_1$ may be expressed as

$$\widehat{\theta}_1 = \sum_{j=1}^N b_j^2 G_{jj} + 2 \sum_{j=1}^{N-1} \sum_{k=j+1}^{k_o(j)} b_j b_k G_{jk}, \quad (65)$$

where $k_o(j) := \min(j + A, N)$. This expression may be used to see that the covariance of $\widehat{\theta}_1$ and $G_{1,A+2}$ is

$$\text{Cov}(\widehat{\theta}_1, G_{1,A+2}) = \sum_{j=1}^N b_j^2 \text{Cov}(G_{jj}, G_{1,A+2}) + 2 \sum_{j=1}^{N-1} \sum_{k=j+1}^{k_o(j)} b_j b_k \text{Cov}(G_{jk}, G_{1,A+2}). \quad (66)$$

Lemma 1 yields the relations

$$\text{Cov}(G_{jj}, G_{1,A+2}) = 2(n-1)^{-1} \Gamma_{j1} \Gamma_{j,A+2}, \quad (67)$$

$$\text{Cov}(G_{jk}, G_{1,A+2}) = \frac{\Gamma_{j1} \Gamma_{k,A+2} + \Gamma_{j,A+2} \Gamma_{k1}}{n-1}. \quad (68)$$

Substituting the formulae from (64), (66), (67), and (68) into (63), we find that the difference of estimator variances is

$$\delta(\Gamma) = (n-1)^{-1} \left[b_1^4 \Gamma_{11} \Gamma_{A+2,A+2} + 4b_1^2 \sum_{j=1}^N b_j^2 \Gamma_{j1} \Gamma_{j,A+2} + 4b_1^2 \sum_{j=1}^{N-1} \sum_{k=j+1}^{k_o(j)} b_j b_k (\Gamma_{j1} \Gamma_{k,A+2} + \Gamma_{j,A+2} \Gamma_{k1}) \right]. \quad (69)$$

The estimator $\widehat{\theta}_2$ has smaller variance than $\widehat{\theta}_1$ if $\delta(\Gamma) < 0$. This places one condition on the entries of Γ . Moreover, since Γ is positive definite, it must satisfy Sylvester's criterion [51], i.e., the N leading principal minors must be positive. These $N+1$ inequalities define a region of the covariance matrix parameter space which is the intersection of $N+1$ open sets. We call this region Λ . Since the intersection of a finite number of open sets is open, Λ is an open set, and it follows that for any $\Gamma^* \in \Lambda$, there is a ball of nonzero volume in the parameter space centered on Γ^* for which $\text{Var}(\widehat{\theta}_2) < \text{Var}(\widehat{\theta}_1)$. Thus, it remains to show that Λ is nonempty.

Let α be a constant which satisfies

$$\alpha > b_1^4 / (b_{A+1}^2 b_{A+2}^2), \quad (70)$$

and consider the matrix

$$\Gamma_{jk}^* := \begin{cases} 1 & \text{if } j=k \neq A+1 \\ \alpha & \text{if } j=k=A+1 \\ -(b_1^2)/(b_{A+1} b_{A+2}) & \text{if } (j,k)=(1,A+1) \\ -(b_1^2)/(b_{A+1} b_{A+2}) & \text{if } (j,k)=(A+1,1) \\ 0 & \text{otherwise.} \end{cases} \quad (71)$$

Clearly, the first A leading principal minors of Γ^* are one. The $(A+1)$ th through the N th leading principal minors are all equal to

$$\alpha + (-1)^{A+2} \left(\frac{-b_1^2}{b_{A+1} b_{A+2}} \right) (-1)^{A+1} \left(\frac{-b_1^2}{b_{A+1} b_{A+2}} \right) = \alpha - b_1^4 / (b_{A+1}^2 b_{A+2}^2). \quad (72)$$

From (70), it is clear that the value in (72) is positive. Thus, Γ^* is positive definite by Sylvester's criterion [51].

We still need to verify that $\delta(\Gamma^*) < 0$. To do this, we examine each term in (69). Since

$\Gamma_{11}^* = \Gamma_{A+2,A+2}^* = 1$, the first term is just $b_1^4/(n-1)$. The second term is zero since Γ_{j1}^* is nonzero only if either $j = 1$ or $j = A+1$, but in both cases, $\Gamma_{j,A+2}^* = 0$. In the first part of the third term, only the $(j, k) = (A+1, A+2)$ term contributes to the sum. Therefore, since $\Gamma_{A+1,1}^* = -b_1^2/(b_{A+1}b_{A+2})$ and $\Gamma_{A+2,A+2}^* = 1$, the first part of the third term is equal to $-4b_1^4/(n-1)$. The second part of the third term is zero since the only way for $\Gamma_{j,A+2}\Gamma_{k1}$ to be nonzero is to have $(j, k) = (A+2, 1)$ or $(j, k) = (A+2, A+1)$, which are impossible cases due to the summation limits. Collecting these observations gives $\delta(\Gamma^*) = -3b_1^4/(n-1) < 0$. Hence, $\Gamma^* \in \Lambda$, i.e., Λ is nonempty.

APPENDIX B

Next, we prove Theorem 7, which gives an explicit condition under which the BRVE has smaller variance than the CVE when $A = 1$. This condition specifies a restriction on the maximum modulus of the between-view correlations, $\rho_{\max}$. We start by finding a lower bound on δV , the difference between the variance of the CVE and the variance of the BRVE. To do this, we derive a lower bound on each term in (52). Subsequently, we show that the resulting lower bound on δV is positive if $\rho_{\max}$ satisfies the inequality stated in Theorem 7.

First, observe that when $A = 1$, the BRVE takes the form

$$\mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x = \sum_{j=1}^N b_j^2 G_{jj} + 2 \sum_{j=1}^{N-1} b_j b_{j+1} G_{j,j+1}. \quad (73)$$

Similarly, the difference of the CVE and the BRVE is

$$\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x = 2 \sum_{l=1}^{N-2} \sum_{m=l+2}^N b_l b_m G_{lm}. \quad (74)$$

Thus, the covariance of the quantities in (73) and (74) is

$$\text{Cov}[\mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x, \mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] = 2 \sum_{j=1}^N \sum_{l=1}^{N-2} \sum_{m=l+2}^N b_j^2 b_l b_m \text{Cov}[G_{jj}, G_{lm}] + 4 \sum_{j=1}^{N-1} \sum_{l=1}^{N-2} \sum_{m=l+2}^N b_j b_{j+1} b_l b_m \text{Cov}[G_{j,j+1}, G_{lm}]. \quad (75)$$

Lemma 1 implies that $\text{Cov}[G_{jj}, G_{lm}] = 2\Gamma_{jl}\Gamma_{jm}/(n-1)$. Assuming that $2 \leq j \leq N-1$, $1 \leq l \leq N-2$, and $l+2 \leq m \leq N$, we may rewrite this covariance as

$$\text{Cov}[G_{jj}, G_{lm}] = 2(n-1)^{-1} \Gamma_{j,j-1} \Gamma_{j,j+1} \delta_{l,j-1} \delta_{m,j+1}, \quad (76)$$

where δ_{jk} is the usual Kronecker delta function. As a consequence, we have the inequality

$$\text{Cov}[G_{jj}, G_{lm}] \geq -2(n-1)^{-1} \Gamma_{\max}^2 \rho_{\max}^2 \delta_{l,j-1} \delta_{m,j+1}. \quad (77)$$

Also, observe that if $j = 1$ and $m \leq l+2$, then $\text{Cov}[G_{jj}, G_{lm}] = 2\Gamma_{1l}\Gamma_{1m}/(n-1) = 0$ since $A = 1$. Likewise, if $j = N$ and $l \leq N-2$, then $\text{Cov}[G_{jj}, G_{lm}] = 2\Gamma_{Nl}\Gamma_{Nm}/(n-1) = 0$. Therefore, (77) holds for $1 \leq j \leq N$, $1 \leq l \leq N-2$, and $l+2 \leq m \leq N$.

Next, Lemma 1 implies that $\text{Cov}[G_{j,j+1}, G_{lm}] = (\Gamma_{jl}\Gamma_{j+1,m} + \Gamma_{jm}\Gamma_{j+1,l})/(n-1)$. Suppose that $m \neq l+2$ and consider the following implications. Because $A = 1$, $\Gamma_{j+1,l} = 0$ only if $l = j, j+1, j+2$, which is true only if $l = j$. Since $m \neq l+2$, it follows that $\Gamma_{j+1,l} = 0$ only if $m = j+2$, i.e., when $\Gamma_{jm} = 0$. Therefore, we must have $\Gamma_{jm}\Gamma_{j+1,l} = 0$, i.e., the second term in the expression for $\text{Cov}[G_{j,j+1}, G_{lm}]$ above is zero when $m \neq l+2$. Furthermore, the first term in the above expression may be rewritten to show that

$$\text{Cov}[G_{j,j+1}, G_{lm}] = (n-1)^{-1}(\Gamma_{j,j-1}\Gamma_{j+1,j+1}\delta_{l,j-1}\delta_{m,j+1} + \Gamma_{j,j-1}\Gamma_{j+1,j+2}\delta_{l,j-1}\delta_{m,j+2} + \Gamma_{jj}\Gamma_{j+1,j+2}\delta_{lj}\delta_{m,j+2}). \quad (78)$$

Namely, we use the fact that $\Gamma_{jl} = 0$ only if $l = j-1, j, j+1$. Then, we observe that if $l = j-1$, the assumption that $m \neq l+2$ implies that $\Gamma_{j+1,m} = 0$ only if $m = j+1$ or $m = j+2$. Similarly, if $l = j$, then $\Gamma_{j+1,m} = 0$ only if $m = j+2$, and if $l = j+1$, then $\Gamma_{j+1,m} = 0$. Thus, these observations imply that (78) is valid if $2 \leq j \leq N-2$, $1 \leq l \leq N-2$, and $l+2 \leq m \leq N$. Consequently, we have the lower bound

$$\text{Cov}[G_{j,j+1}, G_{lm}] \geq -(n-1)^{-1}\Gamma_{\max}^2\rho_{\max}[\delta_{l,j-1}\delta_{m,j+1} + \delta_{lj}\delta_{m,j+2} + \rho_{\max}\delta_{l,j-1}\delta_{m,j+2}]. \quad (79)$$

A larger range of indices for the lower bound (79) may be justified. Note that if $j = 1$ and $m \neq l+2$, then $\text{Cov}[G_{j,j+1}, G_{lm}] = \Gamma_{1l}\Gamma_{2m}/(n-1)$ is nonzero only if $l = 1$ and $m = 3$, so that $\text{Cov}[G_{j,j+1}, G_{lm}] = \Gamma_{jj}\Gamma_{j+1,j+2}\delta_{lj}\delta_{m,j+2}/(n-1)$, i.e.,

$\text{Cov}[G_{j,j+1}, G_{lm}] \geq -\Gamma_{\max}^2\rho_{\max}\delta_{lj}\delta_{m,j+2}/(n-1)$ in this case. Likewise, if $j = N-1$, $l \leq N-2$, and $l+2 \leq m \leq N$, then $\text{Cov}[G_{j,j+1}, G_{lm}] = \Gamma_{N-1,l}\Gamma_{N,m}/(n-1)$ is nonzero only if $l = N-2$ and $m = N$ so that $\text{Cov}[G_{j,j+1}, G_{lm}] = \Gamma_{j,j-1}\Gamma_{j+1,j+1}\delta_{l,j-1}\delta_{m,j+1} \geq -\Gamma_{\max}^2\rho_{\max}\delta_{l,j-1}\delta_{m,j+1}/(n-1)$ in this case. Hence, (79) is valid for $1 \leq j \leq N-1$, $1 \leq l \leq N-2$, and $l+2 \leq m \leq N$.

Substituting (77) and (79) into (75) and using the definition of $b_{\max}$ yields the lower bound

$$\text{Cov}[\mathbf{b}_x^T G^{\mathcal{R}} \mathbf{b}_x, \mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] \geq -4(n-1)^{-1}\Gamma_{\max}^2 b_{\max}^4 [\rho_{\max} \alpha_1 + \rho_{\max}^2 (\alpha_2 + \alpha_3)], \quad (80)$$

where

$$\alpha_1 := \sum_{j=1}^{N-1} \sum_{l=1}^{N-2} \sum_{m=l+2}^N (\delta_{l,j-1}\delta_{m,j+1} + \delta_{lj}\delta_{m,j+2}) \quad (81)$$

$$\alpha_2 := \sum_{j=1}^N \sum_{l=1}^{N-2} \sum_{m=l+2}^N \delta_{l,j-1}\delta_{m,j+1} \quad (82)$$

$$\alpha_3 := \sum_{j=1}^{N-1} \sum_{l=1}^{N-2} \sum_{m=l+2}^N \delta_{l,j-1}\delta_{m,j+2}. \quad (83)$$

Through straightforward calculations, one may show that

$$\alpha_1 = 2(N-2), \quad \alpha_2 = N-2, \quad \text{and} \quad \alpha_3 = N-3. \quad (84)$$

Now, we turn to the variance of the quantity in (74):

$$\text{Var}[\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] = 4 \sum_{j=1}^{N-2} \sum_{k=j+2}^N \sum_{l=1}^{N-2} \sum_{m=l+2}^N b_j b_k b_l b_m \text{Cov}[G_{jk}, G_{lm}] \quad (85)$$

$$= \sum_{j=1}^{N-2} \sum_{k=j+2}^N \sum_{l=1}^{N-2} \sum_{m=l+2}^N \frac{4b_j b_k b_l b_m}{n-1} (\Gamma_{jl} \Gamma_{km} + \Gamma_{jm} \Gamma_{kl}), \quad (86)$$

where we used Lemma 1 to find the expression in (86). Consider the second term in (86). Since $A = 1$, $\Gamma_{kl} = 0$ only if $l = k - 1, k, k + 1$, i.e., only if $l = k - 1$. If $k = j + 2$ and $m = l + 2$, then it follows that $\Gamma_{kl} = 0$ only if $m = j + 3$, but this condition implies that $\Gamma_{jm} = 0$. Therefore, the second term in (86) contributes nothing to the sum and we have

$$\text{Var}[\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] = 4(n-1)^{-1} \sum_{j=1}^{N-2} \sum_{k=j+2}^N \sum_{l=1}^{N-2} \sum_{m=l+2}^N b_j b_k b_l b_m \Gamma_{jl} \Gamma_{km}. \quad (87)$$

Next, by using similar methods to those used to find (80), a lower bound for $\Gamma_{jl} \Gamma_{km}$ may be derived which then leads to:

$$\text{Var}[\mathbf{b}_x^T (G - G^{\mathcal{R}}) \mathbf{b}_x] \geq \frac{4}{n-1} \left[\sum_{j=1}^{N-2} \sum_{k=j+2}^N b_j^2 b_k^2 \Gamma_{jj} \Gamma_{kk} - \Gamma_{\max}^2 b_{\max}^4 (\rho_{\max} \alpha_4 + \rho_{\max}^2 \alpha_5) \right], \quad (88)$$

where

$$\alpha_4 := \sum_{j=1}^{N-2} \sum_{k=j+2}^N \sum_{l=1}^{N-2} \sum_{m=l+2}^N (\delta_{l,j-1} \delta_{mk} + \delta_{lj} \delta_{m,k-1} + \delta_{lj} \delta_{m,k+1} + \delta_{l,j+1} \delta_{mk}) \quad (89)$$

$$\alpha_5 := \sum_{j=1}^{N-2} \sum_{k=j+2}^N \sum_{l=1}^{N-2} \sum_{m=l+2}^N (\delta_{l,j-1} \delta_{m,k-1} + \delta_{l,j-1} \delta_{m,k+1} + \delta_{l,j+1} \delta_{m,k-1} + \delta_{l,j+1} \delta_{m,k+1}). \quad (90)$$

Tedious, but straightforward calculations, show that

$$\alpha_4 = 2(N-2)(N-3) \quad \text{and} \quad \alpha_5 = 2(N-3)^2. \quad (91)$$

Substituting (80) and (88) into (52), we find that

$$\delta V \geq (4b_{\max}^4 \Gamma_{\max}^2)(n-1)^{-1} B, \quad (92)$$

with

$$B := (N-2)^2 \beta / 2 - (2\alpha_1 + \alpha_4) \rho_{\max} - (2\alpha_2 + 2\alpha_3 + \alpha_5) \rho_{\max}^2, \quad (93)$$

where β was defined in the statement of Theorem 7. Equations (84) and (91) together with (93) then yield the expression

$$B := (N-2)^2 [\beta / 2 - 2 \frac{N-1}{N-2} \rho_{\max} - 2 \rho_{\max}^2]. \quad (94)$$

Considered as a function of $\rho_{\max}$, B is a concave down parabola that attains its maximum at a (fictitious) negative value of $\rho_{\max}$, is strictly positive at $\rho_{\max} = 0$, and has a positive zero at $\rho_{\max} = \rho^*$, where

$$\rho^* = \frac{N-1}{2(N-2)} \left[-1 + \sqrt{1 + \left(\frac{N-2}{N-1} \right)^2 \beta} \right]. \quad (95)$$

Therefore, if $\rho_{\max} < \rho^*$, then $B > 0$, and thus, $\delta V > 0$.

APPENDIX C

In this appendix, we prove Theorem 8, which describes the magnitude of the variance improvement offered by the BRVE compared to the CVE when Γ is tri-diagonal Toeplitz and the backprojection weights are constant.

As discussed in the proof of Theorem 3, the iid realizations of $f(\mathbf{x})$ are each normal. Hence, Lemma 1 with $p = 1$ implies that $\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = [2/(n-1)](\text{Var}[f(\mathbf{x})])^2$. Using the structure of Γ , (7) reduces to $\text{Var}[f(\mathbf{x})] = d^2 \gamma_o [N + 2(N-1)\rho]$. Thus, the variance of the CVE is

$$\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = 2d^4 \gamma_o^2 (n-1)^{-1} [N + 2(N-1)\rho]^2. \quad (96)$$

Since $A = 1$ and $\mathbf{b}_x = [d, \dots, d]$, the BRVE takes the form

$$\widehat{\text{Var}}_r[f(\mathbf{x})] = T_1 + T_2, \quad (97)$$

where

$$T_1 := d^2 \sum_{j=1}^N G_{jj} \quad (98)$$

$$T_2 := 2d^2 \sum_{j=1}^{N-1} G_{j,j+1}. \quad (99)$$

Hence, the variance of the BRVE is

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] = \text{Var}(T_1) + \text{Var}(T_2) + 2 \text{Cov}(T_1, T_2). \quad (100)$$

Applying Lemma 1, the first term in (100) becomes

$$\text{Var}(T_1) = d^4 \sum_{j=1}^N \sum_{k=1}^N \text{Cov}[G_{jj}, G_{kk}] \quad (101)$$

$$= 2d^4 (n-1)^{-1} \sum_{j=1}^N \sum_{k=1}^N \Gamma_{jk}^2. \quad (102)$$

Using the Toeplitz structure of Γ , we then get

$$\text{Var}(T_1) = 2d^4(n-1)^{-1} [\Gamma_{11}^2 + \Gamma_{12}^2 + \sum_{j=2}^{N-1} (\Gamma_{j,j-1}^2 + \Gamma_{j,j}^2 + \Gamma_{j,j+1}^2) + \Gamma_{N,N-1}^2 + \Gamma_{N,N}^2] \quad (103)$$

$$= 2d^4(n-1)^{-1} [2(\gamma_o^2 + \gamma_o^2 \rho^2) + (N-2)(\gamma_o^2 + 2\gamma_o^2 \rho^2)] \quad (104)$$

$$= 2d^4 \gamma_o^2 (n-1)^{-1} (N+2(N-1)\rho^2). \quad (105)$$

Similar, but tedious calculations also show that

$$\text{Var}(T_2) = 4d^4 \gamma_o^2 (n-1)^{-1} [(N-1) + (3N-5)\rho^2] \quad (106)$$

$$\text{Cov}(T_1, T_2) = 8d^4 \gamma_o^2 (n-1)^{-1} (N-1)\rho. \quad (107)$$

Substituting (105)–(107) into (100) and simplifying yields

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] = 2d^4 \gamma_o^2 (n-1)^{-1} [3N - 2 + 8(N-1)\rho + 4(2N-3)\rho^2]. \quad (108)$$

The desired result follows from the ratio of (96) and (108).

APPENDIX D

Here, we prove Theorem 9, which characterizes the magnitude of the variance improvement offered by the BRVE compared to the CVE when Γ is penta-diagonal Toeplitz and the backprojection weights are constant.

As discussed in the proof of Theorem 3, the iid realizations of $f(\mathbf{x})$ are each normal. Hence, Lemma 1 with $p = 1$ implies that $\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = [2/(n-1)](\text{Var}[f(\mathbf{x})])^2$. Using the structure of Γ , (7) reduces to $\text{Var}[f(\mathbf{x})] = d^2 \gamma_o [N + 2(N-1)\rho_1 + 2(N-2)\rho_2]$. Thus, the variance of the CVE is

$$\text{Var}[\widehat{\text{Var}}_c[f(\mathbf{x})]] = 2d^4 \gamma_o^2 (n-1)^{-1} [N + 2(N-1)\rho_1 + 2(N-2)\rho_2]^2. \quad (109)$$

Since $A = 2$ and $\mathbf{b}_x = [d, \dots, d]$, the BRVE takes the form

$$\widehat{\text{Var}}_r[f(\mathbf{x})] = U_1 + U_2 + U_3, \quad (110)$$

where

$$U_1 := d^2 \sum_{j=1}^N G_{jj}, \quad (111)$$

$$U_2 := 2d^2 \sum_{j=1}^{N-1} G_{j,j+1}, \quad (112)$$

$$U_3 := 2d^2 \sum_{j=1}^{N-2} G_{j,j+2}. \quad (113)$$

Hence, the variance of the BRVE is

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] = \text{Var}(U_1) + \text{Var}(U_2) + \text{Var}(U_3) + 2[\text{Cov}(U_1, U_2) + \text{Cov}(U_1, U_3) + \text{Cov}(U_2, U_3)]. \quad (114)$$

By applying Lemma 1 and then using the Toeplitz structure of Γ , tedious calculations similar to that of $\text{Var}(T_1)$ in Appendix C yield the following expressions:

$$\text{Var}(U_1) = \frac{2d^4\gamma_o^2}{n-1} [N + 2(N-1)\rho_1^2 + 2(N-2)\rho_2^2] \quad (115)$$

$$\text{Var}(U_2) = \frac{4d^4\gamma_o^2}{n-1} [(N-1) + 2(N-2)\rho_2 + (3N-5)\rho_1^2 + 2(N-3)\rho_2^2] \quad (116)$$

$$\text{Var}(U_3) = \frac{4d^4\gamma_o^2}{n-1} [(N-2) + 2(N-3)\rho_1^2 + (3N-10)\rho_2^2] \quad (117)$$

$$\text{Cov}(U_1, U_2) = \frac{4d^4\gamma_o^2}{n-1} [2(N-1)\rho_1 + 2(N-2)\rho_1\rho_2] \quad (118)$$

$$\text{Cov}(U_1, U_3) = \frac{4d^4\gamma_o^2}{n-1} [2(N-2)\rho_2 + (N-2)\rho_1^2] \quad (119)$$

$$\text{Cov}(U_2, U_3) = \frac{8d^4\gamma_o^2}{n-1} [(N-2)\rho_1 + (2N-5)\rho_1\rho_2]. \quad (120)$$

Substituting (115)–(120) into (114) and simplifying yields

$$\text{Var}[\widehat{\text{Var}}_r[f(\mathbf{x})]] = \frac{2d^4\gamma_o^2}{n-1} [(5N-6) + 8(2N-3)\rho_1 + 12(N-2)\rho_2 + 8(3N-7)\rho_1\rho_2 + 16(N-2)\rho_1^2 + 12(N-3)\rho_2^2]. \quad (121)$$

The desired result follows from the ratio of (109) and (121).

REFERENCES

1. Tanaka E, Iinuma TA. Correction functions and statistical noises in transverse section picture reconstruction. *Comput. Biol. Med.* 1976; vol. 6(no. 4):295–306. [PubMed: 1000956]
2. Brooks RA, Di Chiro G. Statistical limitations in x-ray reconstructive tomography. *Med. Phys.* 1976 Jul-Aug; vol. 3(no. 4):237–240. [PubMed: 785201]
3. Chesler DA, Riederer SJ, Pelc NJ. Noise due to photon counting statistics in computed x-ray tomography. *J. Comput. Assist. Tomo.* 1977; vol. 1(no. 1):64–74.
4. Gore J, Tofts P. Statistical limitations in computed tomography. *Phys. Med. Biol.* 1978; vol. 23(no. 6):1176–1182. [PubMed: 733907]
5. Riederer SJ, Pelc NJ, Chesler DA. The noise power spectrum in computed x-ray tomography. *Phys. Med. Biol.* 1978; vol. 23(no. 3):446–454. [PubMed: 674361]
6. Alpert N, Chesler D, Correia J, Ackerman R, Chang J, Finklestein S, Davis S, Brownell G, Taveras J. Estimation of the local statistical noise in emission computed tomography. *IEEE Trans. on Med. Imag.* 1982 Oct. vol. 1(no. 2):142–146.

7. Huesman R. A new fast algorithm for the evaluation of regions of interest and statistical uncertainty in computed tomography. *Phys. Med. Biol.* 1984; vol. 29(no. 5):543–552. [PubMed: 6610883]
8. Kijewski MF, Judy PF. The noise power spectrum of CT images. *Phys. Med. Biol.* 1987 May; vol. 32 (no. 5):565–575. [PubMed: 3588670]
9. Wunderlich, A.; Noo, F. Rapidly converging image covariance estimation for FBP reconstruction methods; *IEEE Nucl. Sci. Symp. Conf. Rec*; 2008 Oct.. p. 5187-5190.
10. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill; 1965.
11. Pan X. Optimal noise control in and fast reconstruction of fan-beam computed tomography image. *Med. Phys.* 1999 May; vol. 26(no. 5):689–697. [PubMed: 10360528]
12. La Rivière PJ, Pan X. Favorable noise uniformity properties of Fourier-based interpolation and reconstruction approaches in single-slice helical computed tomography. *Med. Phys.* 2002 Jun; vol. 29(no. 6):943–951.
13. Zeng GL. Nonuniform noise propagation by using the ramp filter in fan-beam computed tomography. *IEEE Trans. Med. Imag.* 2004 Jun; vol. 23(no. 6):690–695.
14. Xia, D.; Cho, S.; Pan, X. Image noise properties in circular sinusoid cone-beam CT; *IEEE Nucl. Sci. Symp. Conf. Rec*; 2007 Oct.. p. 3082-3084.
15. Pan X, Yu L. Image reconstruction with shift-variant filtration and its implication for noise and resolution properties in fan-beam computed tomography. *Med. Phys.* 2003 Apr; vol. 30(no. 4):590–600. [PubMed: 12722811]
16. Yu L, Pan X. Half-scan fan-beam computed tomography with improved noise and resolution properties. *Med. Phys.* 2003 Oct.vol. 30(no. 10):2629–2637. [PubMed: 14596299]
17. Wang J, Lu H, Li T, Liang Z. An alternative solution to the nonuniform noise propagation problem in fan-beam FBP image reconstruction. *Med. Phys.* 2005 Nov.vol. 32(no. 11):3389–3394. [PubMed: 16372414]
18. Zhu L, Starlack J. A practical reconstruction algorithm for CT noise variance maps using FBP reconstruction. *Proc. SPIE: Med. Imag.* 2007 Mar.vol. 6510:651023.
19. Wunderlich A, Noo F. Image covariance and lesion detectability in direct fan-beam x-ray computed tomography. *Phys. Med. Biol.* 2008 May; vol. 53(no. 10):2471–2493. [PubMed: 18424878]
20. Borsdorf, A.; Kappler, S.; Raupach, R.; Hornegger, J. Analytic noise-propagation in indirect fan-beam FBP reconstruction; *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc*; 2008 Aug.. p. 2701-2704.
21. Muirhead, RJ. *Aspects of Multivariate Statistical Theory*. Hoboken, NJ: Wiley; 1982.
22. Anderson, TW. *An Introduction to Multivariate Statistical Analysis*. 3rd ed.. New York: Wiley; 2003.
23. Gupta, AK.; Nagar, DK. *Matrix Variate Distributions*. Boca Raton, FL: Chapman & Hall/CRC; 2000.
24. Kak, AC.; Slaney, M. *Principles of Computerized Tomographic Imaging*, ser. *Classics in Applied Mathematics* #33. Philadelphia, PA: SIAM; 2001.
25. Wang J, Lu H, Liang Z, Eremina D, Zhang G, Wang S, Chen J, Manzione J. An experimental study on the noise properties of x-ray CT sinogram data in Radon space. *Phys. Med. Biol.* 2008 Jun; vol. 53(no. 12):3327–3341. [PubMed: 18523346]
26. Massoumzadeh P, Don S, Hildebolt CF, Bae KT, Whiting BR. Validation of CT dose-reduction simulation. *Med. Phys.* 2009 Jan.vol. 36(no. 1):174–189. [PubMed: 19235386]
27. Casella, G.; Berger, RL. *Statistical Inference*. 2nd ed.. Belmont, CA: Duxbury Press; 2001.
28. Szatrowski TH. Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Ann. Stat.* 1980; vol. 8(no. 4):802–810.
29. Rubin DB, Szatrowski TH. Finding maximum likelihood estimates of patterned covariance matrices by the EM algorithm. *Biometrika.* 1982 Dec.vol. 69(no. 3):657–660.
30. Calvin JA, Dykstra RL. REML estimation of covariance matrices with restricted parameter spaces. *J. Am. Stat. Assoc.* 1995 Mar; vol. 90(no. 429):321–329.
31. Chaudhuri S, Drton M, Richardson TS. Estimation of a covariance matrix with zeros. *Biometrika.* 2007; vol. 94(no. 1):199–216.
32. Christensen, LP. An EM-algorithm for band-Toeplitz covariance matrix estimation; *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process*; 2007 Apr. p. III–1021-III–1024.

33. Andrushchenko Z, Ohlson M, von Rosen D. Estimation of banded covariance matrices in a multivariate normal distribution. Swedish Univ. of Ag. Sci.: Centre of Biostochastics, Tech. Rep. 2008. 2008:2.
34. Ohlson M, Andrushchenko Z, von Rosen D. Explicit estimators under m -dependence for a multivariate normal distribution. Ann. Inst. Stat. Math. 2009 To Appear.
35. Ohlson, M. Ph.D. dissertation. Linköping, Sweden: Linköping University; 2009. Studies in estimation of patterned covariance matrices.
36. Bickel PJ, Levina E. Regularized estimation of large covariance matrices. Ann. Stat. 2008; vol. 36 (no. 1):199–227.
37. Chafaï D, Concordet D. A new method for the estimation of variance matrix with prescribed zeros in nonlinear mixed effects models. Stat. Comput. 2009 Jun; vol. 19(no. 2):129–138.
38. Lehmann, EL.; Casella, G. Theory of Point Estimation. 2nd ed.. New York: Springer; 1998.
39. Poor, HV. An Introduction to Signal Detection and Estimation. 2nd ed.. New York: Springer; 1994.
40. Kay, SM. Fundamentals of Statistical Signal Processing. Vol. vol. 1. Upper Saddle River, NJ: Prentice Hall; 1993. Estimation Theory
41. Giri, NC. Multivariate Statistical Inference. New York: Academic Press; 1977.
42. Horn, RA.; Johnson, CR. Matrix Analysis. Cambridge, UK: Cambridge Univ. Press; 1985.
43. Naylor, AW.; Sell, GR. Linear Operator Theory in Engineering and Science. New York: Springer; 1982.
44. Gray, RM. Toeplitz and Circulant Matrices: A Review. Hanover, MA: Now Publishers; 2006.
45. QRM GmbH. <http://www.qrm.de/>
46. Katsevich A. An improved exact filtered backprojection algorithm for spiral computed tomography. Adv. Appl. Math. 2004; vol. 32:681–697.
47. Zou Y, Pan X. Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT. Phys. Med. Biol. 2004 Mar; vol. 49(no. 6):941–959. [PubMed: 15104318]
48. Pack JD, Noo F, Clackdoyle R. Cone-beam reconstruction using the backprojection of locally filtered projections. IEEE Trans. Med. Imag. 2005 Jan.vol. 24(no. 1):70–85.
49. Fessler JA. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. IEEE Trans. Image Process. 1996 Mar; vol. 5(no. 3):493–506. [PubMed: 18285134]
50. Barrett, HH.; Myers, KJ. Foundations of Image Science. New York: Wiley; 2004.
51. Gilbert GT. Positive definite matrices and Sylvester's criterion. Am. Math. Mon. 1991 Jan.vol. 98 (no. 1):44–46.

Fig. 1.

Comparison of the variances for the BRVE and CVE when Γ is Toeplitz with either $A = 1$ or $A = 2$ over the range of correlations for which Γ is positive definite: (left) plot of $R_1(\rho)$; (center) contour plot of $R_2(\rho_1, \rho_2)$ with the non-relevant region shown in white (this region corresponds to correlations for which Γ is not positive definite); (right) proportion of correlation coefficient values (in %) for which the CVE has lower variance than the BRVE plotted as a function of the number of views.

Fig. 2.
A reconstruction of the phantom. The grayscale window is $[-400, 400]$ HU.

Fig. 3.

Relative differences between the average in x of the BRVE estimate obtained with $A = 1$ and the averages in x of the BRVE estimates obtained with $A = 0$, $A = 2$, $A = 3$, and $A = 10$. These differences are expressed in % relative to the average obtained for $A = 1$ and are plotted as functions of y , measured in cm.

Fig. 4. Relative differences between the average in y of the BRVE estimate obtained with $A = 1$ and the averages in y of the BRVE estimates obtained with $A = 0$, $A = 2$, $A = 3$, and $A = 10$. These differences are expressed in % relative to the average obtained for $A = 1$, and are plotted as functions of x , measured in cm.

Fig. 5.

Estimated image variance maps computed from n data realizations: (top left) the CVE result with $n = 186$; (bottom left) the BRVE result with $n = 186$ and $A = 1$; (top right) the CVE result with $n = 5$; (bottom right) the BRVE result with $n = 5$ and $A = 1$. The grayscale window is $[0, 6000] \text{ HU}^2$.

Fig. 6.

Image variance estimates along the x -axis: In both plots, the CVE estimate computed from $n = 186$ data realizations is shown in light gray and the BRVE estimate is plotted in black. (left) the BRVE estimate with $n = 186$ and $A = 1$; (right) the BRVE estimate with $n = 5$ and $A = 1$.