# Combining Non-Diagonal Preconditioning and Ordered-Subsets for Fully 3D Iterative CT Reconstruction

Lin Fu, Jeffrey A. Fessler, Paul E. Kinahan, and Bruno De Man

*Abstract*– **Preconditioning methods can accelerate the convergence rate of iterative algorithms. Classic iterative tomographic reconstruction algorithms are typically based on relatively simple diagonal preconditioners that ignore the spatial correlation between voxels thus providing only limited acceleration. Non-diagonal preconditioners have the potential to offer more substantial acceleration, but they are more complicated to design and implement for space-variant problems, especially in three dimensions (3D). This paper describes a new non-diagonal preconditioning method for space-variant reconstruction problems in x-ray computed tomography (CT). The proposed preconditioner is based on an approximation of the local spectral response of the Hessian matrix that serves as an approximate Fourier-domain majorizer of the cost function. Based on a multi-channel formulation, the preconditioner consists of independent frequency bands to allow flexible adjustment of location-dependent responses. The preconditioner also allows simple closed-form iterations that can be easily combined with other acceleration techniques such as ordered subsets, line search, conjugate gradient, and Nesterov's optimal gradient methods. Its computational overhead is roughly a fast Fourier transform and its inverse per iteration. Applications to clinical CT data illustrate that the proposed method provides more substantial acceleration to space-variant 3D CT reconstruction than classic diagonal preconditioning.**

*Index Terms*— **computed tomography, iterative image reconstruction, preconditioner.**

## I. Introduction

Within the past decade, model-based iterative reconstruction (MBIR) [1] has become available on clinical x-ray computed tomography (CT) scanners. Based on the principles of maximum *a posteriori* (MAP) estimation, MBIR incorporates models of three-dimensional (3D) system optics, data noise statistics, and image prior information via the optimization of a cost function [2]. Reduction of radiation dose of up to 80% relative to the classical filtered backprojection (FBP) reconstruction has been reported in several clinical studies [3]–[5].

Despite the improved image quality, the expensive computational cost of MBIR remains an impediment to its widespread use in clinical environments. It is well known that large-scale tomographic problems are ill-conditioned, thus conventional gradient-based iterations suffer from slow convergence rates. The computational cost of MBIR is also exacerbated by the complicated geometrical, physical, and statistical models incorporated in the cost function; and by the increasing size of data acquired by today's high-resolution, wide-coverage and multi-energy CT scanners.

Preconditioning methods can accelerate the convergence rate of iterative algorithms without altering their ultimate solutions. In a preconditioned MBIR algorithm, a transformation of variables, called the preconditioner, is applied to reduce the condition number of the Hessian matrix, i.e., the matrix of second-order partial derivatives, of the MBIR cost function. Several widely-used reconstruction algorithms in emission and transmission tomography can be viewed as diagonally preconditioned gradient descent, e.g. SIRT [6][7][8][9], EM [10][11][12], ML-TR [13] [14], and SQS [15]. However, they are typically derived by separable surrogate functions that consider only the largest eigenvalue of the Hessian matrix. Without considering the spectral characteristics of all eigenvalues, these classic diagonal preconditioners provide only limited acceleration.

To further accelerate convergence, it is appealing to develop non-diagonal preconditioners that incorporate the off-diagonal structure of the Hessian matrix, such as the $1/|\nu_r|$ spectral characteristic of tomographic problems. The simplest form of non-diagonal preconditioning are "Fourier" preconditioners that assume the Hessian matrix is approximately a space-invariant convolution operator and the preconditioner is the corresponding deconvolution operator. Clinthorne presented an apodized ramp filter as a preconditioner in unweighted-least-squares reconstruction and observed impressive acceleration [16]. Nuyts designed a similar frequency amplification filter to boost convergence of high spatial-frequency image features in Poisson maximum likelihood reconstruction [17]. Unlike Iterative FBP algorithms [18][19] or its variants [20], the preconditioning methods theoretically converge to the solution defined by the original MBIR cost function. However, such Fourier preconditioners in general are ineffective for highly space-variant tomographic problems due to non-uniform statistical noise modeling and location-dependent regularization [21][22].

To better address space-variant reconstruction, Booth and Fessler [21] suggested that the Hessian matrix in positron emission tomography (PET) problems can be *locally* approximated as a convolution operator and proposed a "combined

The work was supported by National Institutes of Health under grants R01-CA-160253, R01-HL-098686, and U01 EB018753. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Lin Fu and Bruno De Man are with GE Global Research, Niskayuna, NY 12309 (emails: fulin@ge.com, deman@ge.com).

J. A. Fessler is with the Electrical Engineering and Computer Science department, University of Michigan, Ann Arbor, MI 48109.

Paul E. Kinahan is with Department of Radiology, University of Washington, Seattle, WA 98195.

circulant-diagonal" preconditioner where a Fourier preconditioner is modulated by a location-dependent scaling factor. This method yields substantially faster convergence speed than either the Fourier or diagonal preconditioning alone and has been later applied to 3D CT MBIR [23]. However, the degree of freedom of the preconditioner remains limited to account for space-variant regularization. In a subsequent work, Fessler and Booth [22] improved the preconditioner by location-dependent interpolation among multiple Fourier kernels to provide more flexibility in addressing location- and edge-dependent regularization, although higher computational overhead is caused by multiple fast Fourier transforms (FFTs) per iteration. This multi-FFT method has shown promising behavior for non-quadratically regularized MBIR in 2D PET, but to our knowledge, has not been applied to MBIR in 3D CT, where it remains unknown whether interpolation among a small number of Fourier kernels is sufficient to address the complicated anisotropic behaviors. For example, in both [21] and [22], the geometric response of the PET system is approximately isotropic, whereas the geometric sampling pattern in helical CT is anisotropic when the helical pitch is not close to 0.5 or 1.0.

This paper extends the multi-channel preconditioner in [24] and proposes a new space-variant non-diagonal preconditioner for 3D CT reconstruction. Instead of directly modeling the location-dependent, anisotropic local responses of the Hessian matrix, the new preconditioner is based on an approximate upper bound of the local response of the Hessian matrix, which further leads to an approximate Fourier-domain majorizer of the cost function. Similar to [24], the new preconditioner is composed of multiple channels representing different frequency sub-bands and/or orientations to provide flexibility in controlling its local spectral response. Certain preconditioning channels are implemented as image-domain filters with a very small footprint to reduce the computational cost relative to FFTs. Finally, unlike previous non-diagonal preconditioning methods that typically rely on a line search to ensure convergence, the proposed method is empirically implemented as closed-form iterations with a fixed step size and thus can be easily combined with other acceleration techniques such as ordered subsets (OS) and Nesterov's optimal gradient methods. The new preconditioning method can be used to accelerate the convergence rate of a wide range of gradient-based simultaneous-update algorithms that are highly parallelizable and suitable for implementation on many-core computing devices.

## II. Background

### A. Cost function

Consider MBIR that minimizes a regularized negative log-likelihood function:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \Phi(\boldsymbol{x}), \qquad \Phi(\boldsymbol{x}) \triangleq -L(\boldsymbol{y}; \mathbf{A}\boldsymbol{x}) + R(\boldsymbol{x}).$$

where $\boldsymbol{x} \triangleq \{x_1, \dots, x_N\}$ denotes an image of the object, e.g. attenuation coefficients at each voxel location; $\boldsymbol{y} \triangleq \{y_1, \dots, y_M\}$ represents projection measurements; $\mathbf{A} = \{a_{ij}\}$ with $a_{ij} \geq 0$ is an $M \times N$ "system" matrix representing a discrete-discrete model of the Radon or x-ray transform; $[\mathbf{A}\boldsymbol{x}]_i \triangleq \hat{p}_i \triangleq \sum_{j=1}^{N} a_{ij} x_j$ is an estimated line integral value in measurement $i$; $-L(\cdot,\cdot)$ is a negative log-likelihood function (i.e., a data-fit term); $R(\cdot)$ is an image-domain regularization function.

We further assume $y_i$'s are statistically independent, thus their joint log-likelihood is the sum of individual log-likelihoods:

$$L(\boldsymbol{y}; \mathbf{A}\boldsymbol{x}) \triangleq \sum_{i=1}^{M} L_i(y_i; \hat{p}_i).$$

A widely-used choice of $L_i(\cdot,\cdot)$ for non-photon-starved CT data is the weighted-least-squares (WLS)

$$L_i^{\mathrm{WLS}}(y; \hat{p}) \triangleq -\frac{1}{2} w_i \left(\hat{p} - \log \frac{I_i}{y_i}\right)^2,$$

where $w_i \geq 0$ is a statistical weighting factor that typically reflects the variance of the statistical noise in $y_i$ [25][26]; $I_i > 0$ is a blank measurement of un-attenuated x-ray intensity. The WLS formulation assumes any non-positive measured projection values and physical factors such as beam hardening have been pre-corrected so that $y_i > 0$ [27][28]. For more general non-quadratic log-likelihood functions, the statistical weight $w_i$ is defined as the second derivative of $L_i(\cdot,\cdot)$ with respect to $\hat{y}_i$:

$$\mathbf{W} \triangleq -\mathrm{diag}\left\{\frac{\partial^2}{\partial \hat{y}_i^2} L_i(y_i; \hat{y}_i)\right\} = \mathrm{diag}\{w_i\}.$$

The regularization function $R(\boldsymbol{x})$ in this paper corresponds to the log prior of a Markov random field (MRF)

$$R(\boldsymbol{x}) = \sum_{j=1}^{N} \sum_{k>j}^{N} b_{jk} q(\Delta_{jk}).$$

with $b_{jk} = b_{kj} \geq 0$ representing the penalty strength between pixels $j$ and $k$, $\Delta_{jk} \triangleq x_j - x_k$ denoting the difference in pixel values, and $q(\Delta) = q(|\Delta|)$ denoting a prior potential function. We assume $q(\cdot)$ is convex and twice differentiable. We also assume and $q''(\cdot)$ is bounded. Without loss of generality, $q(\cdot)$ is scaled such that

$$0 \leq q''(\cdot) \leq 1. \tag{1}$$

### B. Hessian matrix and local spectral analysis

The $N \times N$ Hessian matrix of the cost function is the sum of the data-fit and regularization components

$$\mathbf{H} \triangleq \nabla^2 \Phi(\boldsymbol{x}) = \mathbf{A}^T \mathbf{W} \mathbf{A} + \mathbf{R}, \tag{2}$$

with its entries denoted by $h_{jk} = g_{jk} + r_{jk}$, $g_{jk} = \sum_{i=1}^{M} a_{ij} w_i a_{ik}$, and $r_{jk} = -b_{jk} q''(\Delta_{jk})$ for $k \neq j$ and $r_{jj} = -\sum_{k \neq j} r_{jk}$.

The Hessian matrix is space-variant but can be assumed *locally* space-invariant [29], [30]. For a local neighborhood around voxel $j$, one can approximate $\mathbf{H}$ by a convolution operator whose impulse response is the sum of the impulse responses of the data-fit and regularization terms:

$$\boldsymbol{h}^j = \boldsymbol{g}^j + \boldsymbol{r}^j, \tag{3}$$

where $\boldsymbol{h}^j \triangleq \mathbf{H}\boldsymbol{e}^j$, $\boldsymbol{g}^j \triangleq \mathbf{A}^T\mathbf{W}\mathbf{A}\boldsymbol{e}^j$, $\boldsymbol{r}^j \triangleq \mathbf{R}\boldsymbol{e}^j$, and $\boldsymbol{e}^j$ denotes the $j$th unit vector. Under this assumption, $\mathbf{H}$ can be locally diagonalized by the Fourier transform. The local spectral representation of $\boldsymbol{h}^j$ is

$$\boldsymbol{h}^j = \mathbf{Q}^{-1}\mathrm{diag}\{\boldsymbol{\xi}^j\}\mathbf{Q}\boldsymbol{e}^j = \mathbf{Q}^{-1}\mathrm{diag}\{\boldsymbol{\lambda}^j + \boldsymbol{\mu}^j\}\mathbf{Q}\boldsymbol{e}^j,$$

where $\boldsymbol{\xi}^j \triangleq \mathbf{Q}\boldsymbol{h}^j$, $\boldsymbol{\lambda}^j \triangleq \mathbf{Q}\boldsymbol{g}^j$, $\boldsymbol{\mu}^j \triangleq \mathbf{Q}\boldsymbol{r}^j$ are discrete local spectra (or eigenvalues) of $\mathbf{H}$, $\mathbf{A}^T\mathbf{W}\mathbf{A}$, and $\mathbf{R}$, respectively, and $\mathbf{Q}$ represents the discrete 3D Fourier transform. The continuous-frequency representation of $\boldsymbol{\xi}^j$, $\boldsymbol{\lambda}^j$, and $\boldsymbol{\mu}^j$ (i.e., discrete-space Fourier transform, DSFT, of the discrete signals $\boldsymbol{h}^j$, $\boldsymbol{g}^j$, and $\boldsymbol{\mu}^j$) are denoted by $\xi^j(\vec{\nu})$, $\lambda^j(\vec{\nu})$, and $\mu^j(\vec{\nu})$, respectively, where $\vec{\nu} \triangleq (\nu_\mathrm{x}, \nu_\mathrm{y}, \nu_\mathrm{z})^T$ denotes spatial frequencies in 3D and each element of $\vec{\nu}$ has units of cycles per pixel. The $\lambda^j(\vec{\nu})$ spectrum resembles the well-known $1/|\nu_\mathrm{r}|$ low-pass spectral characteristics of tomographic problems, with $|\nu_\mathrm{r}| \triangleq \sqrt{\nu_\mathrm{x}^2 + \nu_\mathrm{y}^2}$ denoting the magnitude of the radial component of $\vec{\nu}$. Conversely, the $\mu^j(\vec{\nu})$ spectrum has high-pass characteristics. These local impulse responses and spectra are location-dependent and anisotropic due to the high dynamic range of the statistical weight $w_i$ and, in helical scans, also due to the space-variant geometric responses.

### C. Preconditioner

We first consider the preconditioned gradient descent algorithm with a unit step size as an example:

$$\begin{aligned}\hat{\boldsymbol{x}}^{(n+1)} &= \hat{\boldsymbol{x}}^{(n)} - \mathbf{M}\left[\nabla\Phi\left(\hat{\boldsymbol{x}}^{(n)}\right)\right] \\ &= \hat{\boldsymbol{x}}^{(n)} + \mathbf{M}\left\{\mathbf{A}^T\left[\nabla L\left(\boldsymbol{y}; \mathbf{A}\hat{\boldsymbol{x}}^{(n)}\right)\right] - \nabla R\left(\hat{\boldsymbol{x}}^{(n)}\right)\right\},\end{aligned} \tag{4}$$

where $\hat{\boldsymbol{x}}^{(n)}$ denotes the reconstructed image at the $n$th iteration; the preconditioner $\mathbf{M}$ is a positive-definite $N \times N$ matrix; $\nabla L(\boldsymbol{y}; \hat{\boldsymbol{p}}) \triangleq \left[\frac{\partial L}{\partial \hat{p}_1}, \frac{\partial L}{\partial \hat{p}_2}, \ldots, \frac{\partial L}{\partial \hat{p}_M}\right]^T$ denotes the gradient vector of $L(\boldsymbol{y}; \hat{\boldsymbol{p}})$ with respect to $\hat{\boldsymbol{p}}$; and $\nabla R(\boldsymbol{x}) \triangleq \left[\frac{\partial R}{\partial x_1}, \frac{\partial R}{\partial x_2}, \ldots, \frac{\partial R}{\partial x_N}\right]^T$ denotes the gradient vector of the regularizer.

The choice of the preconditioner $\mathbf{M}$ does not alter the fixed point(s) of (4), but can dramatically impact its convergence properties. For quadratic cost functions, (4) is a contraction if the spectral radius $\rho(\mathbf{MH}) < 2$. The asymptotic rate of convergence of (4) is governed by the condition number of $\mathbf{MH}$. Ideally $\mathbf{M}_\mathrm{ideal}\mathbf{H} = \mathbf{H}\mathbf{M}_\mathrm{ideal} = \mathbf{I}$, which implies

$$\mathbf{H}\boldsymbol{m}_\mathrm{ideal}^j = \boldsymbol{e}^j, \quad j = 1,2,\ldots,N. \tag{5}$$

where $\boldsymbol{m}_\mathrm{ideal}^j$ is the $j$th column of $\mathbf{M}_\mathrm{ideal}^{-1}$. Because $\mathbf{H}$ is locally a convolution operator, (5) can be written as

$$\boldsymbol{h}^j \otimes \boldsymbol{m}_\mathrm{ideal}^j \approx \boldsymbol{e}^j, \tag{6}$$

where $\otimes$ denotes 3D convolution. To proceed, $\mathbf{M}$ is also assumed *locally* a convolution operator. The local impulse response of $\mathbf{M}$ at voxel $j$ is denoted by

$$\boldsymbol{m}^j \triangleq \mathbf{M}\boldsymbol{e}^j = \mathbf{Q}^{-1}\mathrm{diag}\{\boldsymbol{\zeta}^j\}\mathbf{Q}\boldsymbol{e}^j, \tag{7}$$

where $\boldsymbol{\zeta}^j \triangleq \mathbf{Q}\boldsymbol{m}^j$ is the discrete local spectrum (or eigenvalues) of $\mathbf{M}$. The continuous form of $\boldsymbol{\zeta}^j$ (i.e., DSFT of the discrete signal $\boldsymbol{m}^j$) is denoted by $\zeta^j(\vec{\nu})$. Under the local space-invariance assumption, a Fourier-domain representation of (6) is

$$\zeta_\mathrm{ideal}^j(\vec{\nu}) \approx \frac{1}{\xi^j(\vec{\nu})}. \tag{8}$$

However, directly implementing (6) or (8) is impractical because of the high computational cost associated with generating, storing, and applying such a preconditioner. The central topic of this paper is designing preconditioner that is both effective and efficient.

## III. Fourier Domain Upper Bound of hessian matrix

This section derives approximate upper bounds of the local spectra $\lambda^j(\vec{\nu})$, $\mu^j(\vec{\nu})$, and $\xi^j(\vec{\nu})$.

### A. Upper bound of data-fit term $\breve{\lambda}^j(\vec{\nu})$

Analytically modeling the anisotropic and space-variant behavior in $\lambda^j(\vec{\nu})$ could be possible [31][32], but likely to be overly complicated for the purpose of preconditioning. For simplification, an upper bound of $|\lambda_\theta^j(\vec{\nu})|$ is considered. After a series of approximations, which we will describe in more details in a following publication, we obtain an approximate upper bound of $\left|\lambda^j(\vec{\nu})\right|$

$$\left|\lambda^j(\vec{\nu})\right| \leq \breve{\lambda}^j(\vec{\nu}) \triangleq \mathrm{median}\left\{\breve{\lambda}_\mathrm{dc}^j, \frac{S\breve{\lambda}_\mathrm{ac}^j}{|\nu_\mathrm{r}|}, \breve{\lambda}_\mathrm{ac}^j\right\}, \tag{9}$$

where $S$ is a constant scale factor. Typically $\breve{\lambda}_\mathrm{dc}^j \gg \breve{\lambda}_\mathrm{ac}^j$. The $\breve{\lambda}_\mathrm{dc}^j$ parameter governs the DC response, the $\breve{\lambda}_\mathrm{ac}^j$ parameter governs the high frequency AC response, and the $1/|\nu_\mathrm{r}|$ term governs the intermediate frequency response.

### B. Upper bound of regularization term $\breve{\mu}^j(\vec{\nu})$

We make a similar spectral analysis for the local impulse response of the Hessian matrix of the regularization term, defined as $\boldsymbol{r}^j$ in (3). The DSFT of $\boldsymbol{r}^j$, denoted by $\mu^j(\vec{\nu})$, has high-pass characteristics: the smallest value in $|\mu^j(\vec{\nu})|$ is zero, which corresponds to DC; and the largest value in $|\mu^j(\vec{\nu})|$ corresponds to the Nyquist frequency. With edge preserving regularization, $\mu^j(\vec{\nu})$ is object-dependent and anisotropic, making preconditioning difficult. This paper proposes an object-independent upper bound of $|\mu^j(\vec{\nu})|$ to simplify the design of the preconditioner. The dependency on the object $\boldsymbol{x}$ is removed based on (1). After a series of approximations, which we will describe in more details in a following publication, we obtain an approximate upper bound of $|\mu^j(\vec{\nu})|$ in 3D

$$\begin{aligned} |\mu^j(\vec{\nu})| &\le \breve{\mu}^j(\vec{\nu}) \\ &\triangleq \breve{\mu}_{\mathrm{r}}^j \sin^2 \pi\nu_{\mathrm{r}} \cos^2 \pi\nu_{\mathrm{z}} + \breve{\mu}_{\mathrm{rz}}^j \sin^2 \pi\nu_{\mathrm{r}} \sin^2 \pi\nu_{\mathrm{z}}, \end{aligned} \quad (10)$$

where $\breve{\mu}_{\mathrm{r}}^j$ parameter represents the highest frequency response in-plane, and the $\breve{\mu}_{\mathrm{rz}}^j$ parameter represents the highest frequency response in 3D. The $\breve{\mu}^j(\vec{\nu})$ spectrum is approximately circular symmetric in-plane and depends on only two parameters $\breve{\mu}_{\mathrm{r}}^j$ and $\breve{\mu}_{\mathrm{rz}}^j$. The $\breve{\mu}_{\mathrm{r}}^j$ term corresponds to high frequency in the radial direction but low frequency in the z direction, whereas the $\breve{\mu}_{\mathrm{rz}}^j$ term corresponds to high frequency in both the radial and z directions. Not considered in (10) is the low frequency response in the radial direction, because the magnitude of such a component is very small when compared to the low frequency response of the data-fit term $\lambda^j(\vec{\nu})$.

### C. *Upper bound of overall Hessian $\breve{\xi}^j(\vec{\nu})$*

The local spectrum of the overall Hessian matrix is bounded by the sum of the data-fit term (9) and the regularization term (10)

$$\left|\xi^j(\vec{\nu})\right| \le \left|\lambda^j(\vec{\nu})\right| + \left|\mu^j(\vec{\nu})\right| \le \breve{\lambda}^j(\vec{\nu}) + \breve{\mu}^j(\vec{\nu}) \triangleq \breve{\xi}^j(\vec{\nu}),$$

where

$$\begin{aligned} \breve{\xi}^j(\vec{\nu}) = \mathrm{median}&\left(\breve{\lambda}_{\mathrm{dc}}^j, \frac{S\breve{\lambda}_{\mathrm{ac}}^j}{|\nu_{\mathrm{r}}|}, \breve{\lambda}_{\mathrm{ac}}^j\right) \\ &+ \breve{\mu}_{\mathrm{r}}^j \sin^2 \pi\nu_{\mathrm{r}} \cos^2 \pi\nu_{\mathrm{z}} \\ &+ \breve{\mu}_{\mathrm{rz}}^j \sin^2 \pi\nu_{\mathrm{r}} \sin^2 \pi\nu_{\mathrm{z}}. \end{aligned} \quad (11)$$

is determined by four coefficients: $\breve{\lambda}_{\mathrm{dc}}^j$, $\breve{\lambda}_{\mathrm{ac}}^j$, $\breve{\mu}_{\mathrm{r}}^j$, and $\breve{\mu}_{\mathrm{rz}}^j$. The low-frequency end of the $\breve{\xi}^j(\vec{\nu})$ spectrum is dominated the data term $\lambda^j(\vec{\nu})$, whereas the high-frequency end of the spectrum is more influenced by $\mu^j(\vec{\nu})$. Overall, the $\breve{\xi}^j(\vec{\nu})$ spectrum is an approximate majorizer of $\left|\xi^j(\vec{\nu})\right|$. Unlike conventional surrogate functions that are constructed in the space-domain, here $\breve{\xi}^j(\vec{\nu})$ is constructed as an approximate surrogate function in the Fourier domain.

## IV. Multi-channel preconditioner

The local spectrum of the proposed preconditioner, denoted by $\zeta^j(\vec{\nu})$, is intended to be approximately the reciprocal of $\breve{\xi}^j(\vec{\nu})$:

$$\zeta^j(\vec{\nu}) \approx \frac{1}{\breve{\xi}^j(\vec{\nu})} \le \frac{1}{|\xi^j(\vec{\nu})|} \approx |\zeta_{\mathrm{ideal}}^j(\vec{\nu})|. \quad (12)$$

Compared to the ideal preconditioner (6), the $\zeta^j(\vec{\nu})$ spectrum is based on the analytical expression (11) which depends on only four parameters per voxel location $j$. Equation (12) also suggests that the proposed preconditioner has lower frequency amplification compared to the ideal preconditioner. Thus, the spectral radius of the preconditioned Hessian matrix is less than one:

$$\rho(\mathbf{MH}) \approx \max\left|\zeta^j(\vec{\nu})\xi^j(\vec{\nu})\right| \le \max\left|\zeta^j(\vec{\nu})\right|\left|\breve{\xi}^j(\vec{\nu})\right| \approx 1.$$

This condition suggests that the convergence of (4) may be achieved by a unit step size without the need of a line search.

However, (11) and (12) do not immediately provide a practical implementation because $\zeta^j(\vec{\nu})$ still depends on the voxel location $j$, and at each location (12) corresponds to an impractical infinite impulse response filter typically. To further simplify, recall that $\breve{\xi}^j(\vec{\nu})$ is dominated by different factors in different frequency bands, suggesting $\zeta^j(\vec{\nu})$ may be empirically approximated by a channelized or filter-bank formulation where each channel represents a frequency sub-band that is much simpler to implement than (14). This paper proposes to approximate $\frac{1}{\breve{\xi}^j(\vec{\nu})}$ by four ($K = 4$) frequency channels:

$$\frac{1}{\breve{\xi}^j(\vec{\nu})} \approx \zeta^j(\vec{\nu}) \triangleq \sum_{k=1}^{K} \zeta_k^j(\vec{\nu}) \triangleq \sum_{k=1}^{K} \eta_k^j \zeta_k(\vec{\nu}), \quad (13)$$

where the spectrum of the $k$th channel, denoted by $\zeta_k^j(\vec{\nu})$, is factored as $\eta_k^j\zeta_k(\vec{\nu})$ so that its dependences on $\vec{\nu}$ and $j$ are decoupled: the "kernel" of the $k$th channel, denoted by $\zeta_k(\vec{\nu})$, is space-invariant (independent of $j$), but a modulation factor $\eta_k^j$ depends on the voxel location $j$ [22]. The proposed $\zeta_k(\vec{\nu})$ and $\eta_k^j$ are detailed below.

1) The first channel ($k = 1$) in (13) accounts for the DC and very low-frequencies.

$$\eta_1^j \triangleq \frac{1}{\breve{\lambda}_{\mathrm{dc}}^j}, \qquad \zeta_1(\vec{\nu}) \triangleq 1.$$

This channel can be simply implemented in the space domain as a voxel-wise scaling operation like the SQS preconditioner. It has a minimal computational overhead.

2) The second channel ($k = 2$) accounts for the low-to-medium frequencies.

$$\eta_2^j \triangleq \frac{1}{S\breve{\lambda}_{\mathrm{ac}}^j}, \qquad \zeta_2(\vec{\nu}) \triangleq \frac{\cos^2 \pi\nu_{\mathrm{r}}}{\frac{1}{|\nu_{\mathrm{r}}|} + k_0 \sin^2 \pi\nu_{\mathrm{r}}},$$

where $\zeta_2(\vec{\nu})$ is an apodized ramp-filter, with its gain (or slope) modulated by $\eta_2^j$. The Hann window in the numerator and the $k_0 \sin^2 \pi|\nu_{\mathrm{r}}|$ term in the denominator suppress its high frequency response. We empirically choose $k_0$=50, a unitless scale factor. Further optimization of the apodization terms has not been performed. This channel is implemented in the Fourier domain by FFTs like the diagonal-circulant preconditioner in [21], but the frequency response $\zeta_2(\vec{\nu})$ and the spatial weighting factor $\eta_2^j$ proposed here are derived differently.

3) The third channel ($k = 3$) accounts for the high frequency in-plane (x-y) and low frequency across-plane (z).

$$\eta_3^j \triangleq \frac{1}{\breve{\lambda}_{\mathrm{ac}}^j + \breve{\mu}_{\mathrm{r}}^j}, \qquad \zeta_3(\vec{\nu}) \triangleq \sin^2 \pi\nu_{\mathrm{r}} \cos^2 \pi\nu_z.$$

Instead of implementing this channel in the Fourier domain by FFTs, we more efficiently approximate it as an image-domain filter with a very small $3\times3\times3$ kernel:

$$\mathrm{ker}_3 = \frac{1}{64}\begin{bmatrix}-1 & -2 & -1\\ -2 & 12 & -2\\ -1 & -2 & -1\end{bmatrix} \otimes [1,2,1]_z$$

where $[\ldots]_z$ denotes a vector in the z dimension, and $\otimes$ denotes 3D convolution. The $1/64$ factor normalizes the frequency response to unity at the Nyquist frequency.

4) The fourth channel ($k = 4$) accounts for the high frequency both in in-plane (x-y) and across-plane (z).

$$\eta_4^j \triangleq \frac{1}{\breve{\lambda}_{\mathrm{ac}}^j + \breve{\mu}_{\mathrm{rz}}^j}, \qquad \zeta_4(\vec{v}) \triangleq \sin^2 \pi\nu_{\mathrm{r}} \sin^2 \pi\nu_{\mathrm{z}}.$$

Similar to $\zeta_3(\vec{v})$, we also approximated this channel as an image-domain filter with a $3\times3\times3$ kernel

$$\mathrm{ker}_4 = \frac{1}{64}\begin{bmatrix}-1 & -2 & -1\\ -2 & 12 & -2\\ -1 & -2 & -1\end{bmatrix} \otimes [-1,2,-1]_z.$$

The channelized approximation (13) leads to the proposed multi-channel preconditioner:

$$\mathbf{M} = \sum_{k=1}^{K} \mathrm{diag}\{\boldsymbol{\eta}_k\}^{1/2} \mathbf{M}_k \,\mathrm{diag}\{\boldsymbol{\eta}_k\}^{1/2}\ , \tag{14}$$

where $\boldsymbol{\eta}_k \triangleq (\eta_k^1, \ldots, \eta_k^N)^T$ is the vector form of the spatial gain factor $\eta_k^j$; $\mathbf{M}_k = \mathbf{Q}^{-1}\mathrm{diag}\{\boldsymbol{\zeta}_k\}\mathbf{Q}$ is a positive-definite convolution operator; $\boldsymbol{\zeta}_k$ defines the spectral kernel of the $k$th channel; and $\boldsymbol{\eta}_k$ is a spatial weighting factor for the channel at different spatial locations. We obtain $\boldsymbol{\zeta}_k$ by discrete sampling of the continuous spectrum $\zeta_k(\vec{v})$. By splitting the preconditioner into different channels, the gains of individual channels can be independently and space-variantly controlled, giving certain flexibility to incorporate space-variant effects. If only the first channel is enabled (other channels gains set to zero), the multichannel preconditioner reduces to a diagonal practitioner. Similarly, if only the second channel is enabled, it reduces to a “combined circulant-diagonal” preconditioner [21][23]. As described earlier, we implement the individual preconditioning channels $\mathbf{M}_k$ in space-domain or Fourier-domain for the best computational efficiency.

### A. Combination with other acceleration techniques

In addition to applying the proposed preconditioner (14) directly to gradient descent iterations (4), its property as a Fourier domain surrogate function allows relatively straightforward combinations with other independent acceleration techniques. In addition to the widely used combination of preconditioning with CG and line search [12][34][22], the proposed preconditioner can be combined with ordered subsets (OS) [35][36][15] and Nesterov’s optimal gradient method [37], [38], and their combinations [39][40][41][42][43]. These combined algorithms will be described in more details in a following publication.

## V. RESULTS

### A. CT data and reconstruction settings

Evaluation of the proposed preconditioner was performed with a retrospective adult patient dataset who underwent chest CT as part of clinical work-up, with institutional review board approval and written informed consent. The dataset was acquired on a GE Discovery CT750 HD scanner (GE Healthcare, Waukesha, WI) with 64-row collimation, helical pitch of 33/64, 120 kVp, 20 mA, and gantry speed of 0.5 s per rotation. Images were reconstructed on a grid of 512×512×25 over a field-of-view of 50 cm and with a slice thickness of 0.625 mm. All MBIR reconstructions were based on the same cost function formed by a post-log WLS data-fit term and a $q$-GGMRF edge-preserving regularizer [25]. The distance-driven model was used in forward and back projectors [44]. All MBIR reconstructions were initialized with FBP images with standard kernel. No non-negativity constraint of image voxel values was enforced during MBIR. All parameters of the proposed preconditioner were determined at the beginning of the iterations.

### B. Speed of convergence

The convergence speeds of 11 gradient-based simultaneous-update MBIR algorithms with various preconditioning and acceleration techniques were compared (Tab. 1).

Tab. 1. MBIR algorithms implemented.

| | |
|---|---|
| SQS-GD | Gradient descent (4) with the SQS diagonal preconditioner [15] |
| SQS-Nes | SQS with Nesterov momentum acceleration [42] |
| SQS-OS12 | SQS using OS with 12 subsets in bit reversal order |
| SQS-OS6-Nes | SQS using OS with 6 subsets and Nesterov momentum acceleration [42] |
| SQS-OS12-Nes | Same as SQS-OS6-Nes, but with 12 subsets |
| SQS-CG | SQS with conjugate gradient and line search [22] |
| MM-GD | Gradient descent (4) with the proposed multi-channel preconditioner (14). |
| MM-Nes | MM with Nesterov momentum acceleration [42] |
| MM-OS12 | MM using OS with 12 subsets in bit reversal order |
| MM-OS6-Nes | MM using OS with 6 subsets and Nesterov momentum acceleration |
| MM-CG | MM with conjugate gradient and line search [22] |

Fig. 1 compares the convergence speeds of various MBIR algorithms. The level of convergence was quantified by root-mean-square difference (RMSD) relative to the reference reconstruction obtained by 4,000 SQS-GD iterations. MM-GD achieved substantially faster reduction of the RMSD than the classic diagonal SQS preconditioner (SQS-GD). When combined with independent acceleration techniques such as Nesterov momentum (MM-Nes), OS (MM-OS12), OS+Nesterov (MM-OS6-Nes), and conjugate gradient with line search (MM-CG), the proposed preconditioner achieved further acceleration relative to the standard preconditioned gradient descent (MM-GD). These combined algorithms based on the proposed preconditioner were also faster than the counterpart algorithms based on the SQS preconditioner (SQS-Nes, SQS-OS12, SQS-OS6-Nes, and SQS-CG. Overall, among all the convergent, non-OS algorithms, “MM-CG” achieved the fastest convergence speed. Among all the semi-convergent OS algorithms, both MM-OS6-Nes and SQS-OS12-Nes achieved the fastest initial convergence speeds, although SQS-OS12-Nes resulted in instability and diverged after 20 iterations.

A small residual RMSD of about 3 HU remained after 100 iterations of the proposed algorithms as shown in Fig. 1. This suggests that the iterative process may have a slow-converging component regardless of the various acceleration techniques used. The residual error may be related to the boundary effect of the reconstruction field of view, where the local shift invariance approximation may be less accurate.

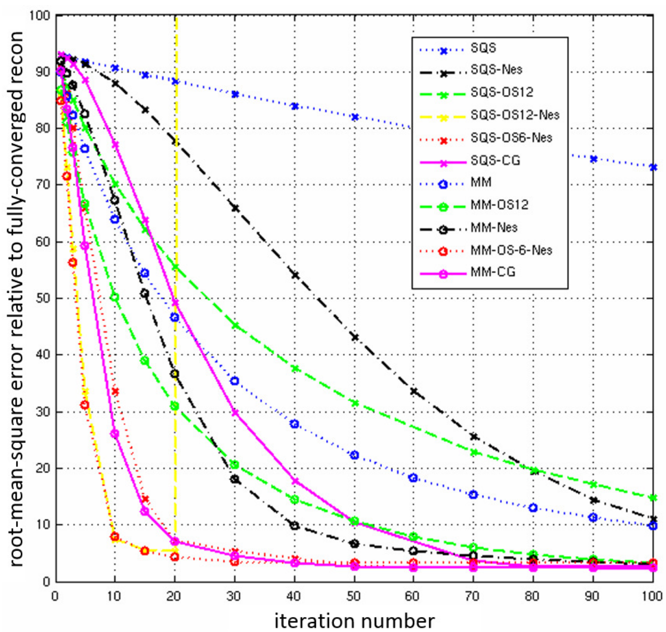


Fig. 1. Convergence curves of different algorithms.

## VI. Conclusions

A new non-diagonal preconditioner is proposed for acceleration of space-variant 3D reconstruction problems in x-ray CT. The proposed preconditioner is based on an approximate majorizer of the cost function. Leveraging a multi-channel formulation, the proposed preconditioner results in closed-form iterations with a computational overhead of one Fourier transform and its inverse. The proposed method can be further accelerated by independent techniques including ordered subsets, Nesterov momentum, conjugate gradient, and their combinations. Image reconstruction from clinical CT data shows that the proposed preconditioner, with and without combination with other acceleration techniques, provides substantially faster convergence than the counterpart algorithms based on the classic diagonal preconditioner.

## References

[1] B. De Man and J. A. Fessler, “Statistical iterative reconstruction for X-ray computed tomography,” in *Biomedical Mathematics*, G. Wang, Y. Censor, and J. Ming, Eds. Springer, 2009.

[2] J. Nuyts, B. De Man, J. A. Fessler, W. Zbijewski, and F. J. Beekman, “Modelling the physics in the iterative reconstruction for transmission computed tomography.,” *Phys. Med. Biol.*, vol. 58, no. 12, pp. R63--96, Jun. 2013.

[3] P. J. Pickhardt, M. G. Lubner, D. H. Kim, J. Tang, J. A. Ruma, A. M. del Rio, and G.-H. Chen, “Abdominal CT with model-based iterative reconstruction (MBIR): initial results of a prospective Trial comparing ultralow-dose with standard-dose imaging,” *Am. J. Roentgenol.*, vol. 199, no. 6, pp. 1266–1274, 2012.

[4] F. A. Miéville, L. Berteloot, A. Grandjean, P. Ayestaran, F. Gudinchet, S. Schmidt, F. Brunelle, F. O. Bochud, and F. R. Verdun, “Model-based iterative reconstruction in pediatric chest CT: assessment of image quality in a prospective study of children with cystic fibrosis.,” *Pediatr. Radiol.*, vol. 43, no. 5, pp. 558–67, Mar. 2013.

[5] A. Neroladaki, D. Botsikas, S. Boudabbous, C. D. Becker, and X. Montet, “Computed tomography of the chest with model-based iterative reconstruction using a radiation exposure similar to chest X-ray examination: preliminary observations,” *Eur. Radiol.*, vol. 23, no. 2, pp. 360–366, 2013.

[6] P. Gilbert, “Iterative methods for the three-dimensional reconstruction of an object from projections,” *J. Theor. Biol.*, vol. 36, no. 1, pp. 105–117, 1972.

[7] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1999.

[8] M. Jiang and G. Wang, “Convergence of the Simultaneous Algebraic Reconstruction Technique (SART),” *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 957–961, 2003.

[9] J. Gregor and J. A. Fessler, “Comparison of SIRT and SQS for Regularized Weighted Least Squares Image Reconstruction,” *IEEE Trans. Comput. Imaging*, vol. 1, no. 1, pp. 44–55, Mar. 2015.

[10] L. A. Shepp and Y. Vardi, “Maximum Likelihood Reconstruction for Emission Tomography,” *IEEE Trans Med Imaging1*, vol. M1-1, no. 2, pp. 113–122, 1982.

[11] K. Lange and R. Carson, “EM reconstruction algorithms for emission and transmission tomography.,” *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–16, Apr. 1984.

[12] L. Kaufman, “Implementing and accelerating the EM algorithm for positron emission tomography,” *IEEE Trans. Med. Imaging*, vol. 6, no. 1, pp. 37–51, 1987.

[13] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, “Iterative reconstruction for helical CT: a simulation study,” *Phys. Med. Biol.*, vol. 43, no. 4, pp. 729–737, Apr. 1998.

[14] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, “An iterative maximum-likelihood polychromatic algorithm for CT,” *IEEE Trans. Med. Imag.*, vol. 20, no. 10, pp. 999–1008, 2001.

[15] H. Erdoğan and J. A. Fessler, “Ordered subsets algorithms for transmission tomography.,” *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–2851, 1999.

[16] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos, “Preconditioning methods for improved convergence rates in iterative reconstructions.,” *IEEE Trans. Med. Imag.*, vol. 12, no. 1, pp. 78–83, 1993.

[17] J. Nuyts, P. Suetens, and L. Mortelmans, “Acceleration of maximum likelihood reconstruction, using frequency amplification and attenuation compensation.,” *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 643–652, 1993.

[18] L.-T. Chang, “A method for attenuation correction in radionuclide computed tomography,” *IEEE Trans. Nucl. Sci.*, vol. 25, no. 1, pp. 638–643, 1978.

[19] J. Sunnegårdh, “Iterative Filtered Backprojection Methods for Helical Cone-Beam CT,” Linköping University, 2009.

[20] D. S. Lalush and B. M. W. Tsui, “Improving the convergence of iterative filtered backprojection algorithms,” *Med. Phys.*, vol. 21, no. 8, pp. 1283–1286, Aug. 1994.

[21] S. D. Booth and J. A. Fessler, “Combined diagonal/Fourier preconditioning methods for image reconstruction in emission tomography,” in *Proceedings International Conference on Image Processing*, 1995, vol. 2, pp. 441–444.

[22] J. A. Fessler and S. D. Booth, “Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction,” *IEEE Trans. Imag. Proc.*, vol. 8, no. 5, pp. 688–699, 1999.

[23] L. Fu, B. De Man, K. Zeng, T. M. Benson, Z. Yu, G. Cao, and J.-B. Thibault, “A preliminary investigation of 3D preconditioned conjugate gradient reconstruction for cone-beam CT,” in *Proc. SPIE. 8313*, 2012, p. 83133O.

[24] L. Fu, Z. Yu, J.-B. Thibault, B. De Man, M. G. McGaffin, and J. A. Fessler, “Space-Variant Channelized Preconditioner Design for 3D Iterative CT Reconstruction,” in *The 12th International meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2013, pp. 205–208.

[25] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, “A three-dimensional statistical approach to improved image quality for multislice helical CT,” *Med. Phys.*, vol. 34, no. 11, pp. 4526–4544, 2007.

[26] K. Sauer and C. A. Bouman, “A local update strategy for iterative reconstruction from projections,” *IEEE Trans. Signal Proc.*, vol. 41, no. 2, pp. 534–548, 1993.

[27] J. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, “A Recursive Filter for Noise Reduction in Statistical Iterative Tomographic Imaging,” in *Proc SPIE/IS&T Symp Elec Im Sci Tech - Comp Im, Vol 6065, San Jose, CA*, 2006, vol. 6065, pp. 15–19.

[28] Z. Chang, R. Zhang, J.-B. Thibault, D. Pal, L. Fu, K. Sauer, and C. Bouman, “Modeling and pre-treatment of photon-starved CT data for iterative reconstruction,” *IEEE Trans Med Imaging*, p. (ahead of print. available online), 2016.

[29] J. A. Fessler and W. L. Rogers, “Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs,” *IEEE Trans. Imag. Proc.*, vol. 5, no. 9, pp. 1346–1358, 1996.

[30] J. Qi and R. M. Leahy, “Resolution and noise properties of MAP reconstruction for fully 3-D PET,” *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 493–506, 2000.

[31] A. H. Delaney and Y. Bresler, “A fast and accurate Fourier algorithm for iterative parallel-beam tomography,” *IEEE Trans. Image Process.*, vol. 5, no. 5, pp. 740–753, May 1996.

[32] S. Schmitt, M. Goodsitt, and J. Fessler, “Fast Variance Prediction for Iteratively Reconstructed CT Images with Locally Quadratic Regularization,” *IEEE Trans. Med. Imaging*, no. 1, pp. 17–26, 2017.

[33] D. S. Lalush and B. M. W. Tsui, “MAP-EM And WLS-MAP-CG Reconstruction Methods For Transmission Imaging In Cardiac SPECT,” in *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, 1993, pp. 1174–1178.

[34] H. M. Hudson and R. S. Larkin, “Accelerated Image Reconstruction Using Ordered Subsets of Projection Data,” *IEEE Trans Med Imaging*, vol. 13, no. 4, pp. 601–609, 1994.

[35] C. Kamphuis and F. J. Beekman, “Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm,” *IEEE Trans Med Imaging*, vol. 17, no. 6, pp. 1101–1105, 1998.

[36] D. Kim, D. Pal, J.-B. Thibault, and J. A. Fessler, “Accelerating ordered subsets image reconstruction for X-ray CT using spatially nonuniform optimization transfer.,” *IEEE Trans. Med. Imaging*, vol. 32, no. 11, pp. 1965–78, Nov. 2013.

[37] Y. Nesterov, “A method for unconstrained convex minimization problem with the rate of convergence o(1/k^2),” *Dokl. AN SSSR Transl. as Sov. Math Dokl.*, vol. 269, pp. 543–547, 1983.

[38] Y. Nesterov, “Smooth minimization of non-smooth functions,” *Math. Program.*, vol. 103, no. 1, pp. 127–152, 2005.

[39] D. Kim, S. Ramani, and J. A. Fessler, “Ordered subsets with momentum for accelerated X-ray CT image reconstruction,” in *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, 2013, pp. 920–923.

[40] D. Kim, S. Ramani, and J. A. Fessler, “Accelerating X-ray CT ordered subsets image reconstruction with Nesterov’s first-order methods,” in *fully three-dimensional image reconstruction in radiology and nuclear medicine*, 2013, pp. 22–25.

[41] D. Kim and J. A. Fessler, “Ordered subsets acceleration using relaxed momentum for x-ray CT image reconstruction,” in *IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2013.

[42] D. Kim, S. Ramani, and J. A. Fessler, “Combining ordered subsets and momentum for accelerated x-ray CT image reconstruction,” *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 167–178, 2015.

[43] A. S. Wang, J. W. Stayman, Y. Otake, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, “Accelerated statistical reconstruction for C-arm cone-beam CT using Nesterov’s method,” *Med. Phys.*, vol. 42, no. 5, pp. 2699–2708, 2015.

[44] B. De Man and S. Basu, “Distance-driven projection and backprojection in three dimensions,” *Phys. Med. Biol.*, vol. 49, no. 11, pp. 2463–2475, 2004.