A Three-Dimensional Statistical Approach to Improved Image Quality for Multi-Slice Helical CT

Jean-Baptiste Thibault*, Ken Sauer†, Charles Bouman‡ and Jiang Hsieh*

*Applied Science Laboratory, GE Healthcare
3000 N Grandview Bvd, W-1180, Waukesha, WI 53188
Email: jean-baptiste.thibault@med.ge.com, jiang.hsieh@med.ge.com
†Department of Electrical Engineering, 275 Fitzpatrick
University of Notre Dame, Notre Dame, IN 46556-5637
Email: sauer@nd.edu
‡School of Electrical Engineering
Purdue University, West Lafayette, IN 47907-0501
Email: bouman@ecn.purdue.edu

Abstract

Multi-slice helical Computed Tomography (CT) scanning offers the advantages of faster acquisition and wide organ coverage for routine clinical diagnostic purposes. However, image reconstruction is faced with the challenges of three-dimensional cone-beam geometry, data completeness issues, and low dosage. Of all available reconstruction methods, statistical iterative reconstruction (IR) techniques appear particularly promising since they provide the flexibility of accurate physical noise modeling and geometric system description. In this paper, we present the application of Bayesian iterative algorithms to real 3D multi-slice helical data to demonstrate significant image quality improvement over conventional techniques. We also introduce a novel prior distribution designed to provide flexibility in its parameters to fine-tune image quality. Specifically, enhanced image resolution and lower noise have been achieved, concurrently with the reduction of helical cone-beam artifacts, as demonstrated by phantom studies. Clinical results also illustrate the capabilities of the algorithm on real patient data. Although computational load remains a significant challenge for practical development, superior image quality combined with advancements in computing technology make IR techniques a legitimate candidate for future clinical applications.

I. INTRODUCTION

Multi-slice CT scanning is particularly attractive for clinical applications due to short acquisition times, thin slices, and large organ coverage. Those acquisition trajectories produce projection measurements that pass obliquely through the 2-D reconstructed image planes. As the pitch increases, deviation from conventional approximate two-dimensional planar data is further amplified. The accurate handling of this geometry is critical to the elimination of unwanted reconstruction artifacts and the enhancement of image quality. Recent developments in analytical inversion algorithms give reason to hope that for many applications, image quality may be adequate under single-pass, deterministic inversion culminating in data backprojection [1], [2]. Imaging applications arise, however, in which characteristics of the scanner hardware place a limit on the quality of reconstructions [3]. Helical “windmill” artifacts may originate from portions of patient anatomy, particularly in the case of abrupt edges in high-contrast materials, such as bones and prosthetics. Clinical diagnostics also require the resolution of detail beyond the capability of even recent scanners.

Traditionally, images have been reconstructed from CT data using so-called analytical reconstruction algorithms such as filtered backprojection (FBP) or convolution backprojection (CBP). Some algorithms completely neglect the cone-beam geometry of the measured data during backprojection, and rely instead on helical projection data interpolations to limit the extent of geometric artifacts [4]–[7]. Other algorithms consider the cone-beam geometry by calculating nutating image planes to minimize the error between the reconstruction plane and the projection ray paths in a 2D backprojection step. These include the pi-algorithms [8], [9], ASSR [10], or AMPR [11]. However, these approaches are not sufficient when the cone angle gets larger, and it becomes necessary to consider the exact acquisition geometry in a 3D backprojection step, such as in the Feldkamp (FDK) algorithm [12], which has been modified and generalized for general multi-slice CT [13]–[15]. All these approaches are approximate by nature and reduce rather than eliminate cone-beam artifacts. By contrast, the algorithms of Katsevich [16] provide an analytic solution to the helical cone beam scan inversion, with the hope of completely eliminating cone-beam artifacts. However, they were originally derived under the assumption of continuously sampled detector surfaces, not the discrete form necessary with hardware realizable in the foreseeable future. Although modifications have been proposed for a detector with discrete sampling [17], these still do not offer the possibility to easily incorporate scanner-specific characteristics in the system model. In addition, most single backprojection-based techniques depend on projection data interpolation schemes, which limit the resolution characteristics of the final output.

This work is considered proprietary and confidential until time of publication.
As an attempt to provide more flexibility in the reconstruction choices, iterative reconstruction (IR) algorithms have been recently introduced for multi-slice helical CT images [18]–[21]. Enabled by recent advances in computer processing hardware [22] or additional algorithm developments [23] necessary to handle the additional computational cost, IR methods are now considered an emerging reconstruction technique for clinical CT patient data, with the objective of improving image quality in situations challenging for state-of-the-art convolution backprojection. Modern medical CT imaging demands low-dose scans, high resolution reconstructions, and artifact reduction even when data is limited or incomplete. Fortunately, IR techniques are particularly well equipped to address these challenges, although at the cost of longer reconstruction time.

Successful methods adapted for iterative reconstruction from CT data are based on the statistics of random fluctuations in sinogram measurements [24]–[26]. Rather than manipulating data to force it to conform to traditional analytical reconstruction models, statistical methods attempt, to the degree possible, to explicitly include non-idealities in the problem description. This view of image reconstruction requires only the knowledge of a description of the way in which each measurement is influenced by unknown image values. Such deficiencies in data as Poisson counting noise and incomplete scan coverage are expected and their description is built into the reconstruction process. Rather than treating all measurements with equal weighting, a statistical model allows different degrees of credibility among data. This modeling provides statistical methods a robustness not easily incorporated into single backprojection techniques. Problem formulation for image reconstruction becomes the expression of a statistical measure relating sinogram data to the estimated image volume, in the form of an objective function, which may be solved iteratively. The addition of a stabilizing function in the form of a regularization prior may further reduce artifacts and image noise. A priori information about the distribution of the image space, such as an image smoothness penalty, provides another tool for controlling image quality.

This view of the reconstruction problem represents a significant departure from conventional analytical techniques. The non-quadratic nature of the objective function often resulting from modeled image statistics affects the usual trade-off between image noise and resolution. While the choice of appropriate filter kernels represents the principal method of adapting image characteristics to clinician’s expectations for single backprojection-based techniques, several elements of statistical IR methods may be tuned simultaneously to maximize image quality. However, the texture of IR images, a consequence of statistical modeling, requires careful investigation of the practical facets of image quality (contrast, noise, resolution) to warrant application to clinical diagnostics.

Since the introduction of IR methods to CT, much of the effort has been devoted to demonstrating the feasibility of the proposed techniques and illustrating some of its benefits in the general case. The results in this paper focus on demonstrating the value of iterative reconstruction in reducing artifacts, improving resolution, and lowering noise in reconstructed images. The great majority of artifacts we attack, as well as limitations in reconstruction resolution, result from the combination of detector aperture width and limited sampling rates. Section II introduces the general framework of the statistical reconstruction problem, and shows the formulation of the objective function. Section III presents more details on the forward modeling calculations, leading to accurate reconstructions. A novel approach to regularization is introduced in section IV, with the design of a general family of convex potential functions flexible enough to provide sufficient control over desired final image quality. An efficient technique for the iterative solution of the optimization problem is then reviewed in section V. Finally, using the scans and techniques introduced in section VI, the results presented in section VII demonstrate both superior spatial resolution performance and helical artifact reduction over analytical methods, thus positioning IR methods for future clinical use.

II. Statistical Model for Image Reconstruction

Let \( \mathbf{x} = \{ x_j; j \in \{1 \ldots M \} \} \) be the discrete vector of three-dimensional image space. Its elements represent attenuation coefficients, or unknown densities of the elements of space forming the 3D volume and are the object of the reconstruction. Let \( \mathbf{y} = \{ y_i; i \in \{1 \ldots N \} \} \) be the discrete vector of projection measurements, representing the line integrals through the imaged object for a variety of positions and projection angles. Our IR algorithm uses the same calibrated and pre-processed data as conventional FBP.

An underlying assumption fundamental to the statistical formulation of the reconstruction problem is that \( \mathbf{x} \) and \( \mathbf{y} \) are random vectors, according to a common probability density function determined by patient anatomy and x-ray physics. The CT transmission scan does not provide the \( y_i \) directly, but rather is formed of a collection of recorded detector measurements \( \{ \lambda_i; i \in \{1 \ldots N \} \} \) that are related to the line integral projections by Beer’s law of attenuation [27]. They represent the detected x-ray intensity after attenuation by the scanned object, and follow a Poisson distribution:

\[
\lambda_i \sim \text{Poisson} \{ I_i e^{-\mu_i} \},
\]

where \( I_i \) is the impinging x-ray photon intensity, and \( \bar{y} \) is the ideal noiseless projection integral computed from the true 3D attenuation values \( \mathbf{x} \). The reconstruction problem may be formulated in the Bayesian framework as the Maximum A Posteriori (MAP) estimate:

\[
\hat{x} = \arg \max_x P(\mathbf{x}|\mathbf{y}),
\]
where $P(.)$ denotes the probability, which is equivalent to
\[ \hat{x} = \arg \max_x \{ \log P(y|x) + \log P(x) \}. \] (1)

Frequently, a model of the form $y = Ax + n$ is used, linearizing the relation between $x$ and $y$ with the matrix $A$, an operator transforming the image space in a manner similar to the CT scanning system. The noise values in $n$ represent random fluctuations of the measurement about its mean as a result of photon and electronic noise.

The first term in the right hand side of equation (1) is the log-likelihood term. A good approximation to the log-likelihood for the x-ray transmission problem is based on a second order Taylor series expansion, in terms of the unknown image, and of the log of the Poisson probability mass function for the measurement counts $\lambda_i$ [28]. This yields the quadratic expression:
\[ \log P(y|x) \approx -\frac{1}{2} \sum_i d_i (y_i - [Ax]_i)^2 + f(y) \]
\[ = -\frac{1}{2} (y - Ax)^T D(y - Ax) + f(y) \] (2)
where $D$ is a diagonal matrix, and $f(y)$ is some function of the data. For transmission tomography, the coefficients $d_i$ of $D$ are proportional to detector counts $\lambda_i$, which are maximum likelihood estimates of the inverse of the variance of the projection measurements [28], [29]:
\[ d_i \propto \lambda_i = I_i e^{-y_i} \approx \frac{1}{\sigma_{y_i}^2}. \] (3)

The elements $d_i$ in the quadratic form of (2) reflect inherent variations in credibility of data. For example, if a particular measurement $y_i$ is photon-starved by some highly attenuating object, a problem which may cause artifacts in conventional images, the model reduces the weighting associated with that measurement by reducing the corresponding $d_i$. The dependence of the weighting matrix on the data differentiates this model from Gaussian approximations. The quality of the quadratic approximation in equation (2) improves as the signal-to-noise ratio grows, and is quite accurate for Poisson counts in the range of clinical CT [28]. We have recently proposed a more accurate noise model based on a compound Poisson-Gaussian model of the measurement counts for inclusion of electronic noise in the statistical model in cases where photon starvation occurs [30], but the simple model of equation (3) was used to generate all results in this paper.

Combining equation (2) with the MAP estimate of equation (1) yields:
\[ \hat{x} = \arg \min_x \left\{ \frac{1}{2} (y - Ax)^T D(y - Ax) + U(x) \right\}, \] (4)
where $U(x)$ is a scalar regularization term that is equal to $\log P(x)$ within an additive constant. The function $U(x)$ typically penalizes local differences between voxel elements, and in section IV we introduce a novel choice of $U(x)$ which is appropriate for our problem.

Note that FBP typically applies some kind of low-pass filtering of the noisy projection data to reduce noise in the low-signal regions. Whereas this is a form of statistical modeling, it is inaccurate, at best, as it does not consider the true distribution of the noise in the measurements. IR offers the opportunity to better model the physics of data acquisition. In practice, the measurement counts $\lambda_i$ are subject to a number of calibration pre-processing steps, including physical distortions such as scatter and beam hardening corrections, and other specific scanner corrections such as detector response and normalization. Although they could be included directly in the forward model and the noise model [31], in this paper we use the fully pre-corrected $y_i$ directly in the quadratic form of equation (4), and rely on the robustness of the quadratic approximation for the typical dose levels of clinical CT scans [28].

III. COMPUTATION OF THE FORWARD MODEL

The crucial advantage of statistical reconstruction methods is that they allow any choice of the matrix $A$. Any scanning geometry can be accurately modeled by proper computation of the entries in $A$, regardless of the three-dimensional sampling pattern. The model can be designed to realistically represent the scanner, although this may come at the cost of great computational expense. Because it is necessary to include the non-planar character of the measurements of the helical scan into the forward model, a fundamental component of our approach is to compute the coefficients in the three spatial dimensions. The details of the calculation of the elements of the forward model lie at the core of any efficient implementation of the iterative algorithm, and often drive computation time and reconstruction accuracy.

The majority of projection algorithms in the literature are optimized for the projection of the complete image volume into the sinogram space. This is primarily because iterative reconstruction methods such as Conjugate Gradient (CG) [32] or Ordered Subsets (OS) [33] require a full forward and/or backprojection for each iteration. Consequently, the computation and memory requirements of some forward and backprojection operators have been optimized for this situation. Siddon’s method [34] is one of the fastest algorithms recognized to date to directly compute the ray path through voxel space, using a parametric description of the ray. It has been optimized and improved upon with accelerations such as in [35]. Other fast techniques propose simple
incremental computations among image voxels for a single projection view in order to maximize performance. For instance, the distance-driven (DD) method [36] leads to fast implementation without degrading the frequency response for rectangular basis functions.

On the other hand, voxel-based iterative algorithms, such as Gauss-Seidel (GS) [37] or the algorithm of section V, may be preferred to projection-based techniques due to their convergence speed. Previously, we have calculated the forward projection using a solid voxel model [18], [19], which is based on the computation of the intersection between the ray path and 3D voxels in the helical cone-beam geometry using the Liang-Barsky line-clipping algorithm [38]. While this model can be very accurate, it is computationally expensive because it requires the projections of many rays per voxel / detector pair to account for the finite size of the detector and voxel elements. As an alternative, the DD kernel of [36] has been shown to produce images free of artifacts related to the forward model, so we propose here an implementation tailored to coordinate descent optimization algorithms. We operate directly in the native geometry of the scanner in order to avoid any loss of accuracy which might affect resolution performance. The computation for a single voxel consists of three steps repeated for each projection view:

1) For each view angle, project the voxel’s center onto the detector array.
2) Estimate the 2D footprint of the voxel onto the detector array by appropriately magnifying a “flattened” version of the voxel and placing it at the position computed in 1) above. Notice that this footprint may overlap several detector elements.
3) Apply the DD projection kernel per equations (11) and (12) to compute the coefficients of the forward model for each element of the detector array within this neighborhood.

Figure 1 illustrates the spatially-varying nature of the model in the native cone-beam geometry with a curved detector for two different positions of the source. The DD kernel can be considered as the convolution between the voxel response and the detector response. Let \( \theta \) be the ray angle in the \((x, y)\) plane parallel to the detector channel axis, and \( \phi \) the angle in the \((y, z)\) plane parallel to the detector row axis (see Figure 2). Intuitively, our objective is to “flatten” the voxel along the dimension most closely parallel to the detector face. By flattening the voxel, we simplify the computation of its projection. Define the angle \( \hat{\theta} \) which results from selecting the 45 degree rotation of \( \theta \) such that \(|\hat{\theta}| < \pi/4\):

\[
\hat{\theta} = (\theta + \pi/4) \mod \pi/4 - \pi/4
\]  

(5)

Figure 2 shows how \( \hat{\theta} \) measures the angle between a ray passing through the voxel center, and a normal to the flattened voxel surface. We use a separable expansion of the 2D projection in the 3D case. The projection coefficient of \( A \) for voxel \( i \), view \( j \), channel \( k \), row \( l \), is therefore:

\[
C_{i,j,k,l} = A_{i,j,k} \times B_{i,j,k,l},
\]

(6)

where, for rectangular basis functions in the cone-beam geometry, as illustrated in Figure 2,

\[
A_{i,j,k} = \frac{\Delta_{xy}}{\cos \hat{\theta}} V_c(\delta_c) \ast S_c(\delta_c)
\]

(7)

\[
B_{i,j,k,l} = \frac{1}{\cos \phi} V_r(\delta_r) \ast S_r(\delta_r),
\]

(8)

where \( \Delta \) is the voxel size, \( \delta \) is the distance between the center of the projected voxel and the center of the detector, \( V(\cdot) \) is the voxel window, \( S(\cdot) \) is the detector sensitivity function, and “\( \ast \)” denotes convolution. We use the subscripts \( c \) and \( r \) to denote the channel and the row dimensions of the multi-slice detector, respectively. With \( L \) as the length of the voxel projection onto the detector by magnification, and \( D \) as the size of the detector element, the DD kernel is:

\[
V(\delta) = \text{rect}(\delta/L)
\]

(9)

\[
S(\delta) = \frac{1}{D} \text{rect}(\delta/L).
\]

(10)

Note that the inclusion of both the detector and the voxel response is important for reconstructions with high spatial sampling where \( L < D \) is possible. The coefficients of (6) are therefore:

\[
A_{i,j,k} = \frac{\Delta_{xy}}{\cos \hat{\theta}} \text{clip} \left[ 0, \frac{D_c + L_c}{2} - |\delta_c|, \min(L_c, D_c) \right]
\]

(11)

\[
B_{i,j,k,l} = \frac{1}{\cos \phi} \text{clip} \left[ 0, \frac{D_r + L_r}{2} - |\delta_r|, \min(L_r, D_r) \right].
\]

(12)

The function \( \text{clip}[a, b, c] = \min(\max(a, b), c) \). This model also offers the possibility of easily including scanner-specific characteristics such as focal spot size and detector response non-uniformity, although, for the time-being, these effects are not taken into account.

Note that the inclusion of both the detector and the voxel response is important for reconstructions with high spatial sampling where \( L < D \) is possible. The coefficients of (6) are therefore:

\[
A_{i,j,k} = \frac{\Delta_{xy}}{\cos \hat{\theta}} \text{clip} \left[ 0, \frac{D_c + L_c}{2} - |\delta_c|, \min(L_c, D_c) \right]
\]

(11)

\[
B_{i,j,k,l} = \frac{1}{\cos \phi} \text{clip} \left[ 0, \frac{D_r + L_r}{2} - |\delta_r|, \min(L_r, D_r) \right].
\]

(12)

The function \( \text{clip}[a, b, c] = \min(\max(a, b), c) \). This model also offers the possibility of easily including scanner-specific characteristics such as focal spot size and detector response non-uniformity, although, for the time-being, these effects are not taken into account.
IV. STABILIZING THE OBJECTIVE

Without the regularization term $U(x)$ in equation (4), it is well known that the image estimates are excessively noisy and unstable [39], [40]. Regularization enforces smoothness in the reconstructed images by encouraging neighboring pixels to have similar values, and the parameters of the regularizing term are used to control the trade-off between noise and resolution in the final reconstruction.

In accordance with equation (1), we choose $U(x)$ as the negative log probability of a prior distribution. Typical choices for the prior distribution are Markov Random Fields (MRF’s) because they result in a function $U(x)$ with only local interaction terms. A commonly used general class of MRF’s takes the form of

$$U(x) = \frac{1}{f(\sigma)} \sum_{\{j,k\} \in C} \Psi_\sigma (x_j - x_k),$$

(13)

where $\Psi(\cdot)$ is the potential function, penalizing local differences, and $f(\cdot)$ is some monotone increasing function. The parameter $\sigma$ is a scalar that is determined empirically and controls the prior strength relative to the noise model over the local neighborhood defined by the set $C$ of all 26 pairs of neighboring voxels in 3D. It is advantageous to regularize over a 3D neighborhood for reconstruction from cone-beam measurements.

The Generalized Gaussian Markov Random Field (GGMRF) [41] is a class of MRF’s with desirable properties which result in a regularization term of the form

$$U(x) = \frac{1}{p\rho_p} \sum_{\{j,k\} \in C} b_{j,k}\rho (x_j - x_k)$$

(14)

$$\rho(\Delta) = |\Delta|^p \quad p \geq 1.$$  

(15)

The $b_{j,k}$ are directional weighting coefficients, which we choose as the inverse of the distance between the center voxel and the elements in $C$, normalized such that $\sum_{\{j,k\} \in C} b_{j,k} = 1$. The exponent parameter $p$ of the GGMRF in (15) allows one to control the degree of edge-preservation in the reconstruction. As long as $p > 1$, the resulting regularization term is strictly convex. Combined with equation (4), the resulting objective function is strictly convex as well, which guarantees a unique global minimum of the cost function and allows simpler optimization algorithms for the derivation of the solution. Convexity is desirable since it ensures that the reconstruction does not change discontinuously with the sinogram data [41]. When $p = 2$, the regularization term is quadratic, and the reconstructed image tends to have softer edges. As $p$ is reduced, the regularization term becomes non-quadratic and edge sharpness tends to be enhanced. Other choices of potential functions offer alternatives for varying edge sensitivity [39], [42]–[48]. In general, regularization affects noise and resolution, and may preserve high and low contrast detail differently. Such flexibility is important to clinical imaging.

We next introduce a novel prior formulation, designed to provide flexibility over the GGMRF. It is a family of convex functions, which encompasses GGMRF and Huber-like functions [45]. This new potential function takes the form:

$$\rho(\Delta) = \frac{|\Delta|^p}{1 + |\Delta/c|^{p-q}}$$

(16)

The corresponding derivative is known as the influence function:

$$\rho'(\Delta) = \frac{|\Delta|^{p-1}}{1 + |\Delta/c|^{p-q}} \left( p - \left( \frac{p-q}{c^{p-q}} \right) \frac{|\Delta|^{p-q}}{1 + |\Delta/c|^{p-q}} \right) \text{sign}(\Delta)$$

(17)

We assume that $p \geq q \geq 1$. The constants $p$ and $q$ determine the powers near, and distant from the origin, respectively. The constant $c$ determines the approximate threshold of transition between low and high contrast regions. We restrict ourselves to convex functions, that is $1 \leq q \leq p \leq 2$. The details of the convexity analysis can be found in the Appendix. We refer to this family of convex priors as the $q$-Generalized Gaussian MRF (q-GGMRF). It contains interesting special cases for some values of $p$ and $q$:

- $p = q = 2$: Gaussian prior
- $p = 2; q = 1$: Approximate Huber prior
- $1 < q = p \leq 2$: Generalized Gaussian MRF
- $1 \leq q < p \leq 2$: q-Generalized Gaussian MRF

Figure 3 compares the influence function of the Gaussian and edge-preserving GGMRF priors to several $q$-GGMRF priors.

In the Gaussian case, the influence function is linear around the origin, the region which controls textures in uniform regions. With $p$ fixed, smaller values of $q$ retain better edge-preserving characteristics, as the influence becomes constant for larger values of $\Delta$. The value of $c$ controls the inflexion point: higher $c$ pushes the edge-preserving behavior towards larger $\Delta$. We will examine suitable values of the parameters $p$, $q$, and $c$ for CT imaging in the results section (section VII).
V. Computation of the Solution

With the choice of a strictly convex prior potential function, the cost function defined in equation (4) is strictly convex as well, and therefore has a unique global minimum. For this reason, any number of different optimization algorithms will converge to the same reconstructed image corresponding to the global minimum of (4) [49]. Therefore, the choice of optimization algorithm should be based on the computational efficiency by which the algorithm reaches the unique solution.

Statistical methods have a great advantage in the multi-slice helical case, in having little dependence in their implementation on the geometry of data collection. We attack the estimation/optimization of (4) in the same manner regardless of the scan pattern represented by $A$ or the selected prior $U(x)$. We select an algorithm in the class of voxel-based techniques for their ability to quickly converge high-frequency components, provided a good initial condition, as can be quickly obtained with FBP. We propose optimization over the full 3D volume through a sequence of one-dimensional updates where the image estimate $\hat{x}$ is

$$\hat{x} = \arg\min_{x \in \Omega} \left\{ \frac{1}{2} (y - Ax)^T D (y - Ax) + \frac{1}{f(\sigma)} \sum_{(j,k) \in C} V_\sigma (x_j - x_k) \right\}, \quad (18)$$

and $\Omega$ is the convex set of non-negative reconstructions. The optimization of a strictly convex functional over a convex set yields a unique solution, if it exists, so the addition of this positivity constraint is valid. For image regularization, we use in practice the q-GGMRF of section IV, but to emphasize that this method is independent of the particular choice of the prior for implementation, we employ the general form of (13) in equation (18). For implementation, our approach to the solution is a sequence of one-dimensional optimization steps, where all other image elements are fixed during a single element’s update. Each one-dimensional optimization computes element $x_j$ at iteration $(n+1)$ from $x$ at iteration $(n)$ based on:

$$\hat{x}_j^{(n+1)} = \arg\min_{x_j \geq 0} \left\{ \sum_{i=1}^{N} \frac{d_i}{2} (y_i - A_{i,j} x^{(n)} + A_{i,j} (x_j^{(n)} - x_j))^2 + \frac{1}{f(\sigma)} \sum_{k \in C_j} V_\sigma (x_j - x_k^{(n)}) \right\}. \quad (19)$$

We perform single voxel updates in random order to minimize the correlation between adjacent updates and maximize convergence speed [50]. At each step, the one-dimensional cost function in equation (19) must be minimized by computing the root of its derivative:

$$\theta_1 + \theta_2 \left( x - x_j^{(n)} \right) - \frac{1}{f(\sigma)} \sum_{k \in C_j} I_\sigma \left( x - x_k^{(n)} \right) \bigg|_{x=x_j^{(n+1)}} = 0, \quad (20)$$

where the first and second derivatives for the left-hand side of (19) yield:

$$\theta_1 = \sum_{i=1}^{N} d_i A_{i,j} (y_i - A_{i,j} x^{(n)}) \quad (21)$$

$$\theta_2 = \sum_{i=1}^{N} d_i A_{i,j}^2 \quad (22)$$

$I_\sigma(\cdot)$ is the influence function, i.e. the first derivative of the prior potential, and its expression for the q-GGMRF is given in equation (17). Because in general the non-quadratic shape of the regularizer does not lead to a closed-form solution, a simple half-interval search is performed, to some tolerance around the root [29]. The general framework of this iterative solution can be summarized as:

1) Initialize $x^{(0)}$ from FBP images
2) Perform initial forward projection to compute $Ax^{(0)}$
3) Perform single image-space iteration
   a) Select voxel $x_j^{(n)}$ according to random schedule
   b) Calculate the elements of the forward model $A_{i,j}$
   c) Compute $\theta_1$ and $\theta_2$
   d) Perform half-interval search to find the root of (20)
   e) Update $x_j^{(n+1)}$
   f) Update forward projection estimate $Ax^{(n+1)} = Ax^{(n)} + A_{i,j} (x_j^{(n+1)} - x_j^{(n)})$
   g) Repeat steps 3a-3f until all voxels have been visited
4) Repeat full iteration until convergence is achieved: $\forall j \in \{1 \ldots M\} \ |x_j^{(n+1)} - x_j^{(n)}| < 1$ HU

This approach, called Iterative Coordinate Descent (ICD) [29] guarantees global and monotonic convergence for convex a posteriori log probability density functions [51], and has shown rapid convergence properties provided a good choice of initial conditions. We use the standard FBP images as initial conditions, offering a good estimate of the low frequency components.
While the cost of each iteration remains high relatively as compared to FBP, a full 3D ICD reconstruction typically converges in fewer than 10 iterations, to the point where successive image differences are smaller than the visible range of 1 HU.

We refer to the ICD approach to solving the MAP estimation problem posed in equation (4) as “MAP-ICD” in the remainder of this paper, to emphasize that the global solution to the reconstruction problem is in fact the combination of the choice of a cost function and the choice of an optimization algorithm, both of which may be made independently, within the limits of the ability of the optimization technique to handle the constraints of the statistical model.

VI. MATERIALS AND METHODS

We acquire data on the GE Lightspeed 16-slice third generation CT scanner to assess the performance of the algorithm. The Lightspeed geometry corresponds to 541.0 mm source-to-isocenter distance, and 949.075 mm source-to-detector distance. The elements of the multi-slice detector are formed on an arc concentric to the focal spot of the x-ray source. All the scans used here contain 984 views per rotation for a 1.0 sec gantry period. In the following, we refer to scan sets by describing the detector configuration and scanning trajectories. The $R \times W$ notation represents a multi-slice scan taken with $R$ detector rows, and where each row thickness is $W$ mm at the isocenter of the CT gantry. Total detector aperture corresponds to the product $R \times W$ in mm. The helical pitch of acquisition is described as $P/R : 1$, that is the number of detector rows $P$ traveled along the axis of the gantry during one full rotation. The quantity $P/R$ is often referred to as the normalized pitch of helical acquisition. In all images, the notation “WW” defines the window width in Hounsfield units (HU) selected to display the images.

First, we assess the performance of the q-GGMRF prior model and general image quality of the MAP-ICD algorithm. For this purpose, we use the GE Performance Phantom (GEPP) [27]. It is formed of a Plexiglas insert with resolution bars and tungsten wires in water, and supports quantitative measurements of noise and resolution. The phantom was scanned in 16x0.625mm helical mode at pitch 15/16:1, and 100 mA. Our objective is to select a set of parameters $p$, $q$, and $c$ based on visual inspections of reconstructed images with parameters varied systematically across of the range of convex prior behavior. We limit ourselves to convex priors in this discussion so as to preserve the global convergence properties of the algorithm. Our purpose is not to provide an extensive comparison to prior art, but rather to show that this prior is designed to perform well in the context of regularized iterative reconstruction of clinical CT images. To obtain comparable results for various shapes of the potential function, we match noise between the resulting images within a fixed ROI to less than 1 HU. Matching noise ensures that the strength of the prior relative to the noise model is similar in all images. While keeping two of the parameters $p$, $q$, and $c$ fixed, we vary the other to study the impact on image quality.

With a proper choice of the prior model, our major concern is then the comparison of the MAP-ICD algorithm against conventional analytical reconstruction relative to these image quality characteristics: in-plane resolution / noise trade-off, cross-plane resolution, and helical artifacts. For comparison, we will consider “FBP”, the 2D filtered backprojection with adaptive view-weighting [6], [27] used as baseline; “FDK”, a Feldkamp-based algorithm providing explicit modeling of the 3D cone-beam geometry during backprojection [2], “conjugate FDK”; a Feldkamp-based algorithm making use of conjugate samples during 3D backprojection to achieve better slice-sensitivity profile [52], and “Katsevich”, Katsevich’s exact analytic inversion algorithm for 3D reconstruction [16].

In the following experiments aimed at demonstrating resolution performance, and later the presence or absence of helical artifacts, we will use of an anthropomorphic head phantom. While all algorithms treat the cone-beam geometry with different degrees of accuracy, the limited sampling of the fine details in the bone structure due to discrete scanner characteristics also generates artifacts, generally described as “windmill” artifacts [3]. The cracks in the skull that vary rapidly from plane to plane in our head phantom are a particularly strong source of artifacts, and also provide some visual insights into in-plane and cross-plane resolution, with realistic imaging for CT diagnosis.

An important goal of this study of iterative reconstruction is understanding its effect on cross-plane resolution in CT systems. In conventional linear, spatially-invariant analysis, samples are placed in a uniform pattern, yielding a band of recoverable frequencies having an easily discernible shape and size. In contrast, complete analysis of potential helical CT resolution is necessarily a three-dimensional problem without such simple sampling patterns. To our knowledge, no such analysis has been published. However, it is known that displacement of samples from a rectangular pattern in other applications can improve the recovered bandwidth; one common example is interlaced television scanning [53].

In helical scan CT, each view of data corresponds to a complete set of measurements for each channel and row location in the detector array, with incremental displacements of the patient between successive views even with high pitch. Consequently, the minimum sample spacing along the z direction is typically much smaller than the size of the detector row elements. This high sampling rate in z means that, if the data includes only extremely low-frequency in-plane information, the Nyquist criteria may be satisfied even for very high spatial cross-plane resolution. In practice, frequency content will be more uniform in the three variables, and therefore less dramatic improvement will be possible than is indicated by minimum z-spacing. While the maximum sampling rate in z is high, the cross-plane resolution is also limited by detector aperture and focal spot size [27]. However, over-sampling has been used to recover resolution from spatially filtered data in other applications [54], so it is reasonable to believe that it can be used in our problem to improve over classical reconstruction resolution as well.
For simplicity, we will assume that the x-ray source is a point, and that the object is imaged about the isocenter. We denote the effective detector aperture at the isocenter by $W$. This is the physical spacing of the detector rows multiplied by the ratio (source-to-iso)/(source-to-detector). The $z$ measurements for the object are then approximately convolved with the function $h(z) = \frac{1}{W} \text{rect}(z/W)$, where $h(z)$ is a square pulse centered at $z = 0$ with width 1 and area 1. The frequency response in $z$ is then given by the continuous time Fourier transform of $h(z)$, which is sinc($f$/fc), where $f_c = 1/W$. The first null of the sinc function is then at $f = f_c$. For frequencies below this first null given by $f < f_c$, it is possible, in principle, to recover the desired resolution because the sinc function magnitude is greater than zero. For example, for a detector aperture size at iso of $W = 1.25$mm, the first null falls at $f_c = 8$ lp/cm, so we might expect to be able to recover resolutions up to 8 lp/cm. Alternatively, for $W = 0.625$mm, the first null falls at $f_c = 16$ lp/cm, so we might expect to be able to recover resolutions up to 16 lp/cm. However, direct reconstruction methods such as FBP typically fall far short of this resolution, as we will illustrate in section VII. Iterative reconstruction, which explicitly models the extents of the detector cells and image voxels, has the potential to improve on this.

In order to illustrate this potential benefit, we will compute the frequency response of the reconstruction algorithm using an empirical method. Using the head phantom scan above, we will add to the sinogram the forward projection of five synthetic points within a 10cm diameter from the isocenter in the center plane of the volume: one point at the isocenter, and two others on the vertical and the horizontal axis, respectively. The forward projection of these synthetic points is computed using the DD kernel of section III. Adding the synthetic data to the original sinogram will allow consideration of the real scan statistics in the experiment. We will then reconstruct those impulses within the head data with both MAP-ICD and FBP. For this experiment only, we will use a Gaussian prior ($p = q = 2$). We use these parameters so that the MAP reconstruction is linearly related to the data, and our results are therefore independent of the contrast level of the synthetic points. Varying the scale parameter, $\sigma$, of the a priori image model will produce varying resolution, with smaller $\sigma$ yielding smoother response and less spatial resolution. After taking the difference between the reconstructions with and without the added impulses for both MAP-ICD and FBP, and averaging the reconstructed point spread functions across the five locations, we then sample the response in the frequency domain to form comparison plots.

For a more complete analysis, we will also provide further visual and quantitative evidence of superior cross-plane resolution performance with MAP-ICD. We will compare the reconstructions of the head phantom scanned in 16x1.25mm at helical pitch 9/16:1 to 16x0.625mm at both helical pitch 15/16:1 and 9/16:1, where higher sampling resulting from $W = 0.625$mm will yield results closer to ground truth, with pitch 15/16:1 approaching the original scan trajectory of 16x1.25mm at 9/16:1, and pitch 9/16:1 providing even finer sampling. For another quantitative measure, we will use a wire-in-air phantom, containing 6 wires inside a 20cm circular phantom, with each wire sloped with a ratio of 4:1 relative to the z axis. Using partial volume effect in the reconstructed axial images, the profile through the wires will be computed and averaged over all wire locations to yield the measured slice sensitivity profile (SSP), comparing MAP-ICD to FBP for all three scan trajectories considered above. Finally, for a more systematic visual evaluation of cross-plane resolution, a suitable grid pattern aligned perpendicular to the z axis will be used. We will provide reconstructions of the high-resolution insert of the AAPM CT performance phantom\(^1\), which is generally accepted as a challenging case in visual resolution studies. It features low-contrast holes ranging from 0.4mm to 1.1mm at intervals of 0.1mm. We will repeat scans for all protocols considered above after placing the AAPM phantom upright on the CT table, and perform reconstructions followed by multi-planar reformats.

While improved cross-plane resolution performance with MAP-ICD is one of the major results we present in this paper, it is also important to confirm that helical “windmill” artifacts are controlled as well. To further emphasize the effect of limited sampling, we scan the head phantom with a wide detector pitch, using 16x1.25mm at helical pitch 15/16:1. We will compare the results of MAP-ICD with each of FBP, FDK, Conjugate FDK, and Katsevich to study robustness against helical artifacts. We will also consider a helical rib phantom scanned in 8x1.25mm and helical pitch 13.4/8:1. The Teflon ribs oriented to change very rapidly from plane to plane, as well as the tapered hole in the center of the phantom, typically increase the level of helical artifacts, especially at high helical pitch.

Before getting to the results, we define the filter kernels used in the FBP images as “standard” or “bone” with the following characteristics demonstrated on the 0.05 mm tungsten wire submerged in water of the GEPP and scanned axially at 120 kV, 200 mA and reconstructed at 5 mm thickness. Resolution is measured in line pairs per centimeter (lp/cm). The bone kernel is a high-frequency emphasis filter, designed as explained in [27], with a 50% MTF of 8.6 lp/cm and 10% MTF of 11.9 lp/cm for a corresponding standard deviation of noise of 11.7 HU at 200 mA, while the standard kernel offers a different compromise between image noise and resolution with 50% MTF of 4.3 lp/cm and a 10% MTF of 6.9 lp/cm for a standard deviation of noise of 3.2 HU also at 200 mA. These measurements are for in-plane resolution, and are taken with the GEPP scanned at 1.0 sec/rotation and imaged with 2D FBP at 0.625 mm slice thickness.

\(^1\)Described in a report by the AAPM Task Force on CT Scanner Phantoms, approved by the American Association of Physicists in Medicine [55].
VII. RESULTS

A. Performance of q-GGMRF for In-Plane Resolution/Noise Trade-Offs

The form of the q-GGMRF prior introduced in Section IV depends on three parameters: $p$ and $q$ control the degree of curvature in the potential function in two regions and $c$ determines the boundary between the two. Here we demonstrate the trade-offs in selecting those parameters and arrive at what appears to be a useful compromise.

Iteratively reconstructed CT images will appear different from the FBP images to which radiologists have become accustomed, depending on the parameter choices of the a priori image model. These images will gain clinical acceptance only if they avoid characteristics which are disturbing to those reading them. In order to arrive at a sensible choice for the q-GGMRF parameters, we qualitatively compare imagery in Figures 4-6. The first of these shows several values for $p$, under the assumption that $q$ assumes an edge-preserving, small value. The choice of $p = 2$ shows good edge preservation but also preservation of low contrast information. As the value of $p$ descends from 2, high contrast detail is increasing sharply rendered, but low contrast areas begin to show plateauing and sharp spikes to an objectionable degree. This is due to the strong character of the influence function of the GGMRF with small $p$ near the origin, as shown in Figure 3. In practice, we prefer the quadratic prior behavior in regions of soft tissue where the presence of lesions may be detected, and the higher resolution of the non-quadratic prior in other regions requiring more detail.

In Figure 5, we vary the parameter $q$. It appears that any value for $q$ above 1.4 in this case causes unacceptable levels of smoothing for our attempts at higher resolution reconstructions. At the other extreme, with $p = 2.0$ and $q = 1.0$, which corresponds to the Huber-Markov model, the degree of salt-and-pepper noise allowed by the strong edge-preserving property becomes objectionable. Again we see that the strongest forms of edge preservation have problematic side effects in other aspects of image quality.

Having evidence that differing exponents in the q-GGMRF for the low and high-contrast regions are desirable, we consider in Figure 6 the dependence on the threshold between the regions, $c$, measured in Hounsfield units. Clearly values of 100 and above place large edges into the quadratic penalty range, causing the sort of smoothing witnessed in the Gaussian case earlier. As $c$ approaches zero, the model approaches the GGMRF, with the attendant plateauing found earlier.

Figure 7 and Table I more quantitatively compare three forms of the q-GGMRF. Noise powers are again approximately matched to allow comparison of resolution as measured by the 50% level of the modulation transfer function (MTF). Notice that the GGMRF has the highest spatial resolution, but that the q-GGMRF achieves comparable resolution performance, much greater than with the Gaussian prior. Based on these results, it appears that the values $p = 2$, $q = 1.2$, and $c = 10$ represent a good compromise between resolution, low contrast sensitivity, and high contrast edge preservation at a fixed noise level. Figure 3 also illustrates how deviations from our preferred selection push the behavior of the prior model towards one of the extreme behaviors of either the Gaussian case or the edge-preserving GGMRF. This choice of parameters achieves visually pleasing image quality. Therefore, we will focus on these prior parameters in our future experiments.

For a comparison at equal resolution between FBP and MAP-ICD reconstruction, we apply the high-frequency “bone” kernel as well as the “standard” kernel in FBP to the wire section of the GEPP. It provides a means to accurately measure the in-plane MTF, while the standard deviation of noise can be measured in the homogeneous regions of the phantom (water and Plexiglas). The voxel size is decreased in the reconstructions to properly compute the MTF curves. Results are shown in Figure 8 and Table II. The measured MTF for the MAP-ICD image is comparable to that of the FBP image reconstructed with the bone kernel, while noise attenuation is close to 50% better in the MAP-ICD image than in the FBP image with the standard kernel.

B. Improvements in Cross-Plane Resolution

Figure 9 shows the frequency response plots resulting from adding synthetic point sources to the head phantom data set with $W = 1.25$mm and a helical pitch of 9/16 : 1, using the methodology described in section VI. Reconstructions were done with a voxel size of $\Delta z = 0.625$mm so that the maximum discrete-time frequency of $\pi$ corresponds to 8 lp/cm. Figure 10 shows the reconstructions associated with each of the three curves. Notice that the frequency response corresponding to MAP-ICD with $\sigma = 32$ has much wider MTF than that of FBP, but with comparable levels of noise. In fact, the MAP-ICD can recover frequencies much closer to $1/W$ or 8 lp/cm for this case. On the other hand, the reconstruction with $\sigma = 8$ has comparable resolution to FBP, but much lower noise.

The results above are generated with a Gaussian prior to linearly relate the reconstructed images to the data, so that the results are independent of the contrast level. It is possible to achieve even better results with the q-GGMRF and the parameters selected in section VII-A thanks to its edge-preserving behavior. In the next results, we apply the q-GGMRF to the three head scans introduced in section VI, comparing results from wide and small detector apertures. Figure 11 compares FDK reconstructions using the “bone” kernel to the MAP-ICD images. In the MAP-ICD reconstructions, the voxel size is decreased to take advantage of the high sampling rate of the helical scan along the z-axis, while conventional analytic approaches cannot easily reduce the slice sensitivity profile without kernel adjustments and enduring a noise penalty. The results demonstrate that the IR images made from the scan with $W = 1.25$mm achieve close to the cross-plane resolution of the FDK images with $W = 0.625$mm, while reducing artifacts and image noise at the same time. Details of the fissure in the bone creating an air gap inside the phantom are clearly visible in the MAP-ICD image while even the presence of the gap is not obvious.
in the FDK image obtained from the same scan. On the other hand, the details of the crack match very well those of FDK reconstructions from the scans with higher sampling, thus validating the MAP-ICD results. These results can also be visualized through multi-planar reformat (MPR) and maximum intensity projection (MIP) renderings of the reconstruction volume focused on the gap in the bone viewed in the sagittal direction, as presented in Figure 12. Again, the definition of the crack along the z axis with MAP-ICD roughly matches the results from FDK reconstruction of data obtained at twice the sampling rate, while concurrently reducing image noise and artifacts.

For quantitative results corroborating the visual study above, SSP measurements obtained from the wire-in-air phantom presented in section VI are shown in Figure 14 for the same scan protocols as used above. Figure 13 shows a zoom over the profile of one of the wires in the axial plane for each of the considered cases. The quantitative measurements taken at 50\% and 10\% of the maximum SSP intensity (FWHM and FWTM, respectively) shown in Table III confirm the visual results: an improvement of 40\% in SSP is achieved with MAP-ICD relative to FDK. Thanks to the edge-preserving behavior of the q-GGMRF model, the SSP curve also falls off to zero more rapidly than in the case of FDK applied to the same scan, and avoids overshoots or undershoots, thus reconstructing clean, well-defined edges.

For a more systematic visual evaluation of z-axis resolution and general image quality, reformats of the AAPM grid phantom referenced in section VI for each of the protocols above appear in Figure 15. The z direction corresponds to the vertical axis in the reformats. In the MAP-ICD images with \( W = 1.25\text{mm} \) (at the bottom), the smaller 0.4mm and 0.5mm holes on the left-hand side of the phantom are clearly visible, while only the 0.9mm holes appear clearly separated from FDK-based reconstructions using conjugate backprojection of the same scan (at the top). Interestingly, the 0.6mm bar pattern (third from the left) is not well resolved in the MAP-ICD reconstruction. This matches very well the prediction based on our model for the frequency response, since 0.6mm is very close to 0.625mm, which corresponds to 8 lp/cm, that is the first null in the frequency response for \( w = 1.25\text{mm} \). Therefore, while the frequencies at the nulls of the frequency response cannot be recovered, the results in Figure 15 demonstrate resolution recovery beyond the classical resolution of 8 lp/cm with MAP-ICD. The IR technique achieves only slightly lower resolution than that of FDK-based reconstructions with half the detector aperture at \( W = 0.625\text{mm} \) (middle images from Figure 15) and higher sampling rate than the case above, consistent with the observations in Figure 14 and Table III.

C. Reduction of Helical Artifacts

With the model parameters selected in VII-A and the z-axis resolution performance demonstrated in VII-B, it is important to show that helical artifacts remain under control. In fact, because these artifacts are caused in part by limited sampling dominates image quality. In this case, MAP-ICD can once again take advantage of smaller image voxels and avoids overshots or undershoots, thus reconstructing clean, well-defined edges.

In Figure 16, knowledge of the cross-section of the head phantom without “windmill” artifacts is provided in the top left with an axial scan. Significant artifacts remain in the FDK reconstruction, as well as in the Katsevich image. The latter indicates that even though the exact inversion formula treats the exact 3D geometry of data acquisition with a high degree of accuracy, limited sampling dominates image quality. In this case, MAP-ICD can once again take advantage of smaller image voxels and improved resolution to significantly reduce the artifact. This result is further confirmed with reconstructions of the rib phantom in Figure 17. In the FBP image, blurring of the ribs as well as adjacent shading are apparent due to the orientation of the ribs in three dimensions and the high helical pitch of the scan, which reduces sampling coverage. “Windmill” artifacts also surround the tapered hole. MAP-ICD removes nearly all these artifacts.

D. Clinical Results

Finally, the results of this study would not hold without successful application to real clinical data. For this, we reconstruct a human head scan, in order to observe both brain soft tissue and bone. Figure 18 confirms that on clinical data as well, MAP-ICD demonstrates significantly improved image quality, most dramatically achieved by reducing the level of noise. Small vessels and other structures present in the fat and soft tissue around the skull or near the orbits appear clearly in the IR image while they remain mostly hidden by noise in the FBP image. The improvement in resolution is particularly visible around the sinus area, where the thin walls between the sinus cavities are more clearly visible in the IR image, and in the detail of the air cells in the inner ear region. The reduction of noise in the posterior fossa and the temporal lobes allows better examination of the brain tissue, improving the low contrast differentiation between the cerebellum and the fourth ventricle, visible as a darker region in the center of the posterior fossa. Meanwhile, the helical artifacts that distort the brain tissue near the inner ear in the FBP image are quite satisfactorily removed from the MAP-ICD image. We note that beam hardening artifacts are still present in the brain stem region as well as at the base of the skull, which is not surprising since our model does not explicitly account for beam hardening beyond a simple pre-correction at this time. Nonetheless, one could argue that the novel texture of the image, although significantly different from conventional FBP, may provide better diagnostic value overall.

VIII. DISCUSSION

The general results presented above illustrate several major aspects of our iterative algorithm. First, good balance between carefully designed statistical noise and image regularization models needs to be achieved to provide an acceptable solution to
the CT reconstruction problem. The \( q \)-GGMRF analytical prior we introduced in this paper provides necessary flexibility in its parameters to control the behavior both around the origin and at the tails of the distribution, and appears better suited to clinical CT imaging than the conventional Gaussian or GGMRF priors. Its parameterization through \( p, q, \) and \( e \) is understood well enough to produce promising preliminary results. The last remaining parameter is \( \sigma \) which controls the prior strength over the noise model, and currently remains empirically adjusted for best image quality, when the prior strength is sufficient to control noise without leading to over-smoothing. A more systematic way of setting \( \sigma \) is needed. A consequence of the constant nature of this parameter on the reconstruction may be non-uniform resolution as the trade-off between varying confidence weighting in ray projections and non-linear image smoothing is not actively managed. With constant \( \sigma \), the result of a spatially-variant noise model is also spatially-variant resolution [56], which can lead to directional blurring of edges. Fessler recently proposed a framework to help with this issue in the quadratic case [57], relying on the global interaction among image model, noise model, and system forward model.

Another major differentiating aspect of IR relative to conventional analytical algorithms is precisely the quality of the description of the interaction between image and detector elements. Detector and voxel responses are explicitly included in the forward model. Nearly all analytical reconstruction techniques assume the presence of continuous sampling. In the discrete implementation, therefore, either projection domain or image domain interpolations have to be performed. Because the interpolation techniques often fail to preserve high-frequency contents, a loss of spatial resolution in the reconstructed images often results. In a FDK-type reconstruction algorithm with linear interpolation, for example, it has been shown that nearly 30% reduction in \( z \)-resolution may occur [58]. Higher order interpolations incur a significant penalty in image noise [59]. In other cases, to reduce aliasing artifacts, \( z \)-smoothing algorithms are often employed and further reduce the spatial resolution in \( z \). By contrast, the geometric model used in statistical methods is intrinsically spatially-variant and makes better use of the sinogram information. While analytical methods must use advanced techniques such as focal spot wobble [60] or conjugate-ray backprojection [52] to achieve greater resolution, IR methods may already recover frequencies closer to the maximum system resolution without this extra information. In the view of iterative reconstruction, conjugate samples do not provide additional information, contrary to the case of analytical backprojection, where they can help reduce interpolation errors due to discrete detector sampling. In addition, forward modeling may also include such second order effects as focal spot size, physical detector response and non-idealities in scanning trajectory introduced by the CT tube or table, all of which may result in better resolution.

The results presented above emphasize cross-plane resolution in the multi-slice helical geometry. Because of the continuous motion of the CT table and high DAS trigger frequencies, helical scanning implies a very high sampling rate in the \( z \) dimension. A quick calculation of the total number of available samples relative to the number of unknown image elements under a simplified view of the linear problem as \( y \approx Ax \) may show easily that reconstruction does not suffer from under-determination. In fact, to recover frequencies less than \( 1/W \) that corresponds to the first null of the sinc function we discussed in section VI, Nyquist dictates that the sampling frequency should be \( 2/W \), which is sample period \( W/2 \). Therefore, IR techniques should theoretically reconstruct images at half the detector width to recover the highest frequency. This is consistent with all the reconstruction experiments in section VII where MAP-ICD typically uses \( \Delta_x = 0.625 \text{mm} \) for \( W = 1.25 \text{mm} \). With these parameters, it appears that we may even be able to recover frequencies that lie beyond the first null in the frequency response, as illustrated in Figure 15.

Finally, we cannot compare iterative reconstruction to analytical methods without commenting on computation speed. High model accuracy leads to complexity in the calculations, and multiple passes over the data are needed to reach convergence. For a 16-slice high resolution case such as those considered in section VII, current serial implementation still requires in excess of 12 hours for reconstruction of the portion of the volume covered by the scanning trajectory over a handful of gantry rotations. We have proposed some acceleration methods for the algorithm, for instance improving the order of the ICD updates [61], or replacing the half-interval search with a one-step approach [62]. But overall reconstruction time remains a significant challenge, short of possible hardware optimization.

IX. Conclusion

We have presented a Bayesian framework for iterative CT image reconstruction that produces significant improvements over direct analytical methods in terms of noise, resolution, and helical artifacts. As the reconstruction technique remains independent from the exact form of the forward model, this method is applicable to any geometry and is particularly well suited to the multi-slice helical problem. We introduced a novel model for image statistics providing further control over image quality for clinical application. We also presented an analysis of cross-plane resolution performance to support the superior results of our iterative algorithm relative to conventional analytical reconstruction. Computational speed remains a particular challenge for the practical application of IR methods with high accuracy in clinical CT, because spatially-varying geometric models do not easily lend themselves to the kind of hardware optimizations which have allowed reconstruction performance of multiple frames per second for analytical algorithms on current commercial scanners. In addition, the non-linear problem formulation generates novel image appearance which may be disturbing at first for clinicians who are accustomed to the well-understood properties of linear reconstruction techniques. Ultimately, the success of IR methods in clinical applications will depend on the demonstration of potential improvement in diagnostic readability, or modified scanning protocols to the benefit of the patient.
The function $g$ is convex if and only if $g''(x, p, r) \geq 0$. To demonstrate this, we consider the second derivative:

$$g''(x, p, r) = g'(x, p - 1, r) [p - rg(x, r, r)] \text{sign}(x)$$

And

$$g''(x, p, r) = |x|^{p-r} \rho (x, r, r) \text{ and } g(x, p, r) = |x|g(x, r - 1, r)$$

So

$$g''(x, p, r) = |x|^{p-r-2} g(x, r, r) [p - 1 - rg(x, r, r)] [p - rg(x, r, r)]$$

$$- r^2 |x|^{p-r-1} g(x, r, r) \frac{1}{x} g(x, r, r) [1 - g(x, r, r)]$$

$$= r^2 |x|^{p-r-2} g(x, r, r) \times$$

$$\left[ \left( \frac{p-1}{r} - g(x, r, r) \right) \left( \frac{p}{r} - g(x, r, r) \right) - g(x, r, r) (1 - g(x, r, r)) \right]$$

The function $g(x, p, r)$ is convex if and only if $g''(x, p, r) \geq 0$, which therefore translates to:

$$\left( \frac{p-1}{r} - g(x, r, r) \right) \left( \frac{p}{r} - g(x, r, r) \right) - g(x, r, r) (1 - g(x, r, r)) \geq 0.$$  

This is true if and only if for all $x$, we have:

$$\frac{\left( \frac{p-1}{r} - g(x, r, r) \right) \left( \frac{p}{r} - g(x, r, r) \right)}{g(x, r, r) (1 - g(x, r, r))} \geq 1.$$  

(23)

Some properties of $g(x, r, r)$ are:

1) $g(0, r, r) = 0$
2) $\lim_{x \to \infty} g(x, r, r) = 1$
3) $\forall x \geq 0$, $g(x, r, r)$ is a monotone increasing function of $x$
4) $\{g(x, r, r) : x \in \mathbb{R} \} = [0, 1)$

Using property 4 and criterion (23), we have the necessary and sufficient condition that for all $\rho \in [0, 1)$:

$$\frac{\left( \frac{p-1}{r} - \rho \right) \left( \frac{p}{r} - \rho \right)}{\rho (1 - \rho)} \geq 1.$$  

(24)
Remember that $1 \leq p \leq 2$ and $0 \leq r \leq 1 - p$. Using this constraint, we have that
\[
\frac{p - 1}{r} \geq \frac{p - 1}{p - 1} = 1
\]
Similarly, we have that
\[
\frac{p}{r} \geq \frac{p}{p - 1} \geq 2
\]
So therefore
\[
\frac{(p - 1 - \rho)(p - \rho)}{\rho(1 - \rho)} \geq \frac{(2 - \rho)(1 - \rho)}{\rho(1 - \rho)} = \frac{2 - \rho}{\rho}
\]
\[
\geq 1
\]
(25)
This concludes the proof that the q-GGMRF is convex for $1 \leq q \leq p \leq 2$.

REFERENCES


### TABLE I

Noise and in-plane resolution performance of the Q-GGMRF prior for the image results in Figure 7. With $p = 2.0$, $q = 1.2$, and $c = 10$, the Q-GGMRF presents a good compromise between edge preservation for high contrast and low contrast imaging free of plateaupping, compared to either the Gaussian or the GGMRF priors.

<table>
<thead>
<tr>
<th>Prior</th>
<th>$p$</th>
<th>$q$</th>
<th>$c$</th>
<th>$\sigma$</th>
<th>Std. Dev. (HU)</th>
<th>50% MTF (lp/cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>2.0</td>
<td>2.0</td>
<td>n/a</td>
<td>10</td>
<td>11.15</td>
<td>4.24</td>
</tr>
<tr>
<td>q-GGMRF</td>
<td>2.0</td>
<td>1.2</td>
<td>10</td>
<td>8</td>
<td>10.88</td>
<td>7.23</td>
</tr>
<tr>
<td>GGMRF</td>
<td>1.1</td>
<td>1.1</td>
<td>n/a</td>
<td>0.1</td>
<td>10.78</td>
<td>8.16</td>
</tr>
</tbody>
</table>
Fig. 1. Forward projection of a voxel $V$ in 3D space using the native cone-beam geometry of the detector. A separable kernel between the channel and row directions is used to compute the coefficients. The resulting forward model is spatially-varying, as illustrated by the size of the kernel which changes between source positions $S_1$ and $S_2$ as a function of the distance between voxel and detector.

Fig. 2. Forward model computation by kernel-based magnification in multi-slice native cone-beam geometry. The model computes the projection of rectangular basis functions onto the detector along the detector channel dimension (left) and detector row dimension (right). A convolution model between voxel response and detector response is used to compute the coefficients.

<table>
<thead>
<tr>
<th></th>
<th>FBP Standard</th>
<th>FBP Bone</th>
<th>MAP-ICD</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% MTF (lp/cm)</td>
<td>4.39</td>
<td>8.53</td>
<td>8.66</td>
</tr>
<tr>
<td>10% MTF (lp/cm)</td>
<td>7.04</td>
<td>11.90</td>
<td>13.20</td>
</tr>
<tr>
<td>Water Std. Dev. (HU)</td>
<td>20.76</td>
<td>85.09</td>
<td>12.76</td>
</tr>
<tr>
<td>Plexiglass Std. Dev. (HU)</td>
<td>24.99</td>
<td>90.94</td>
<td>13.01</td>
</tr>
</tbody>
</table>

**TABLE II**

Comparison of FBP and IR for measurement of in-plane MTF and noise, for the images in Figure 8.
Fig. 3. Influence function for the q-GGMRF prior, plotted for different parameters $p$, $q$, and $c$, and across a range of local voxel differences $\Delta$ relevant in a range of Hounsfield units relevant to clinical imaging: solid line: q-GGMRF $p = 2$, $q = 1.2$, $c = 10$; * line: GGMRF $p = 1.3$, $q = 1.3$; ◦ line: Gaussian $p = 2$, $q = 2$; △ line: q-GGMRF $p = 2$, $q = 1.2$, $c = 5$; + line: q-GGMRF $p = 2$, $q = 1.5$, $c = 10$.

Fig. 4. Influence of the exponent parameter $p$ of the q-GGMRF on image quality at matched noise with MAP-ICD on a GE Performance Phantom: 16×0.625mm, helical pitch 15/16:1, 100mA, 1 sec/rotation, WW=350; $\Delta x = \Delta y = 0.488mm$, $\Delta z = 0.625mm$; Top left: q-GGMRF $p = 2.0$, $q = 1.2$, $c = 10$; Top right: q-GGMRF $p = 1.6$, $q = 1.2$, $c = 10$; Bottom left: GGMRF $p = 1.2$, $q = 1.2$, $c = 10$; Bottom right: GGMRF $p = 1.1$, $q = 1.1$, $c = 10$. For $q$ and $c$ fixed, smaller $p$ favors plateauing and salt-and-pepper noise in homogeneous regions.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>FWHM (mm)</th>
<th>FWTM (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16×1.25mm P9/16:1</td>
<td>FDK Conj</td>
<td>1.45</td>
</tr>
<tr>
<td>16×1.25mm P9/16:1</td>
<td>MAP-ICD</td>
<td>0.89</td>
</tr>
<tr>
<td>16×0.625mm P9/16:1</td>
<td>FDK Conj</td>
<td>0.63</td>
</tr>
<tr>
<td>16×0.625mm P15/16:1</td>
<td>FDK Conv</td>
<td>0.72</td>
</tr>
</tbody>
</table>

TABLE III
SSP MEASUREMENTS AT 50% (FWHM) AND 10% (FWTM) OF THE NORMALIZED SSP MAGNITUDE, OBTAINED FROM THE PLOTS IN FIGURE 14.
Fig. 5. Influence of the exponent parameter $q$ of the q-GGMRF on image quality at matched noise with MAP-ICD on a GE Performance Phantom: $16 \times 0.625 \text{mm}$, helical pitch 15/16:1, 100mA, 1 sec/rotation, WW=350; $\Delta x = \Delta y = 0.488 \text{mm}$, $\Delta z = 0.625 \text{mm}$; Top left: Gaussian $p = 2.0$, $q = 2.0$, $c = 10$; Top right: q-GGMRF $p = 2.0$, $q = 1.4$, $c = 10$; Bottom left: q-GGMRF $p = 2.0$, $q = 1.2$, $c = 10$; Bottom right: q-GGMRF $p = 2.0$, $q = 1.0$, $c = 10$; For $p = 2.0$ and $c = 10$ fixed, smaller $q$ leads to higher resolution.

Fig. 6. Influence of the threshold parameter $c$ of the q-GGMRF on image quality at matched noise with MAP-ICD on a GE Performance Phantom: $16 \times 0.625 \text{mm}$, helical pitch 15/16:1, 100mA, 1 sec/rotation, WW=350; $\Delta x = \Delta y = 0.488 \text{mm}$, $\Delta z = 0.625 \text{mm}$; Top left: q-GGMRF $p = 2.0$, $q = 1.4$, $c = 1000$; Top right: q-GGMRF $p = 2.0$, $q = 1.2$, $c = 100$; Bottom left: q-GGMRF $p = 2.0$, $q = 1.2$, $c = 10$; Bottom right: q-GGMRF $p = 2.0$, $q = 1.2$, $c = 1$; For $p$ and $q$ fixed, large values of $c$ result in excessive smoothing.
Fig. 7. Assessment of q-GGMRF image quality with MAP-ICD on a GE Performance Phantom: $16 \times 0.625\text{mm}$, helical pitch 15/16:1, 100mA, 1 sec/rotation, WW=350; $\Delta x = \Delta y = 0.488\text{mm}$, $\Delta z = 0.625\text{mm}$; Left: Gaussian prior $p = 2.0$, $q = 2.0$, $\sigma = 10$; Center: q-GGMRF prior $p = 2.0$, $q = 1.2$, $c = 10$, $\sigma = 8$; Right: GGMRF prior $p = 1.1$, $q = 1.1$, $\sigma = 0.1$.

Fig. 8. Comparison of FBP vs MAP-ICD performance on the GE Performance Phantom: $16 \times 0.625\text{mm}$, helical pitch 15/16:1, 100mA, 1 sec/rotation, WW=400; Left: FBP “standard” kernel; Center: FBP “bone” kernel; Right: MAP-ICD; MAP-ICD parameters: $\Delta x = \Delta y = 0.24\text{mm}$, $\Delta z = 0.625\text{mm}$, $p = 2.0$, $q = 1.2$, $c = 10$, $\sigma = 16$.

Fig. 9. Cross-plane frequency response comparing FBP to MAP-ICD, corresponding to the images in Figure 10. For each case, the frequency response is computed by projecting into the sinogram five synthetic points placed in a 10cm diameter around the isocenter area, and sampling the average of the reconstructed point spread functions in the frequency domain. For MAP-ICD, the calculation is done with a Gaussian prior to match the analytic model and study the contrast-independent response. The result shows that MAP-ICD can dramatically improve cross-plane resolution by approaching the theoretical limit more than FBP.
Fig. 10. Head phantom illustrating cross-plane resolution performance and comparing FBP to MAP-ICD, corresponding to the frequency response plots in Figure 9. The scan is 16×1.25mm, helical pitch 9/16:1, 100mA, WW=400. Left: 2D FBP “bone” kernel; Center: MAP-ICD with $\sigma = 8$; Right: MAP-ICD with $\sigma = 32$; MAP-ICD parameters: $\Delta x = \Delta y = 0.244\text{mm}$, $\Delta z = 0.625\text{mm}$, $p = 2.0$, $q = 2.0$, $c = 10$. The result shows that MAP-ICD can further improve cross-plane resolution relative to 2D FBP, even with a Gaussian a priori model.

Fig. 11. Qualitative example of improved cross-plane resolution with MAP-ICD. Head Phantom, 320mA, 1 sec/rotation, bone kernel, WW=400, $\Delta x = \Delta y = 0.488\text{mm}$. Top left: 16×1.25mm pitch 9/16:1 conjugate FDK $\Delta z = 1.25\text{mm}$; Top right: 16×1.25mm pitch 9/16:1 MAP-ICD $\Delta z = 0.625\text{mm}$, $p = 2.0$, $q = 1.2$, $c = 10$; Bottom left: 16×0.625mm pitch 9/16:1 conjugate FDK $\Delta z = 0.625\text{mm}$; Bottom right: 16×0.625mm pitch 15/16:1 FDK $\Delta z = 0.625\text{mm}$. 
Fig. 12. Qualitative example of improved cross-plane resolution with MAP-ICD (sagittal view). Head Phantom. Top: Multi-Planar Reformat (MPR) images; Bottom: Maximum Intensity Projections (MIP) images. With, in each image, from top to bottom: 16×1.25mm pitch 9/16:1 conjugate FDK; 16×0.625mm pitch 9/16:1 conjugate FDK; 16×1.25mm pitch 9/16:1 MAP-ICD. Image spacing is 0.625mm in the z direction for all images in the reconstructed volume.

Fig. 13. Wire-in-air phantom used for quantitative measurements of slice sensitivity profile (SSP) based on a 4:1 slope relative to the z axis, with the “bone” kernel, WW=400, $\Delta x = \Delta y = 0.488mm$. Top left: 16×1.25mm pitch 9/16:1 conjugate FDK $\Delta z = 1.25mm$, Top right: 16×1.25mm pitch 9/16:1 MAP-ICD $\Delta z = 0.625mm$, $p = 2.0, q = 1.2, c = 10$, Bottom left: 16×0.625mm pitch 9/16:1 conjugate FDK $\Delta z = 0.625mm$, Bottom right: 16×0.625mm pitch 15/16:1 FDK $\Delta z = 0.625mm$. 
Fig. 14. Normalized slice sensitivity profiles (SSP) plots based on the wire phantom reconstructions in Figure 13 providing another measure of cross-plane resolution. Horizontal axis represents millimeters; $\Delta x = \Delta y = 0.488\text{mm}$. * line: $16 \times 1.25\text{mm}$ pitch $9/16:1$ conjugate FDK $\Delta z = 1.25\text{mm}$; + line: $16 \times 0.625\text{mm}$ pitch $9/16:1$ conjugate FDK $\Delta z = 0.625\text{mm}$; ◦ line: $16 \times 0.625\text{mm}$ pitch $15/16:1$ FDK $\Delta z = 0.625\text{mm}$; Solid line: $16 \times 1.25\text{mm}$ pitch $9/16:1$ MAP-ICD $\Delta z = 0.625\text{mm}$, $p = 2.0, q = 1.2, c = 10$.

Fig. 15. Visual comparison of cross-plane resolution using reformatted images of the AAPM grid phantom placed upright onto the CT table. FDK uses the “bone” kernel, WW=400. $\Delta x = \Delta y = \Delta z = 0.2\text{mm}$. From top to bottom: $16 \times 1.25\text{mm}$ pitch $9/16:1$ conjugate FDK; $16 \times 0.625\text{mm}$ pitch $9/16:1$ conjugate FDK; $16 \times 1.25\text{mm}$ pitch $9/16:1$ MAP-ICD, $p = 2.0, q = 1.2, c = 10$. 
Fig. 16. Comparison of various analytical reconstruction algorithms and MAP-ICD relative to helical artifacts due to limited sampling, on a Head Phantom, 16×1.25mm, helical pitch 15/16.1, 320mA, 1 sec/rotation, WW=400; Top left: reference axial FBP; Bottom left: Feldkamp-based; Top right: Katsevich-based; Bottom right: MAP-ICD; MAP-ICD parameters: $\Delta x = \Delta y = 0.488$mm, $\Delta z = 0.625$mm, $p = 2.0$, $q = 1.2$, $c = 10$. MAP-ICD can reduce helical artifacts as well or better than approximate or even exact analytical inversion approaches.

Fig. 17. Illustration of the reduction in helical artifacts for MAP-ICD (right) vs 2D FBP (left) on a Rib Phantom scanned in 8×1.25mm helical mode, pitch 13.4/8:1, 320mA, 0.5 sec/rotation, WW=400; FBP parameters: $\Delta x = \Delta y = 0.488$mm, $\Delta z = 1.25$mm, “standard” kernel; MAP-ICD parameters: $\Delta x = \Delta y = 0.488$mm, $\Delta z = 0.625$mm, $p = 2.0$, $q = 1.2$, $c = 10$. 
Fig. 18. Qualitative clinical results on a human head: standard FBP (top) vs MAP-ICD (bottom) 32×0.625mm helical scan, pitch 17/32:1, 140kV, 280mA, WW=300; Reconstruction parameters: $\Delta x = \Delta y = 0.586\text{mm}$, $\Delta z = 1.2\text{mm}$, $p = 2.0$, $q = 1.2$, $c = 10$. The MAP-ICD image is not fully corrected for beam hardening artifacts.