Comparing implementations of penalized weighted least squares sinogram restoration

Peter Forthmann, Thomas Koehler, Michel Defrise, Patrick La Riviere

Abstract

We have explored and compared two implementation strategies for PWLS sinogram restoration: (1) A direct matrix-inversion strategy based on the closed-form solution to the PWLS optimization problem and (2) an iterative approach based on the conjugate-gradient algorithm. Obtaining optimal performance from each strategy required modifying the naive off-the-shelf implementations of the algorithms to exploit the particular symmetry and sparseness of the sinogram-restoration problem. For the closed-form approach, we subdivided the large matrix inversion into smaller coupled problems and exploited sparseness to minimize matrix operations. For the conjugate gradient approach, we exploited sparseness and preconditioned the problem to speed convergence. Despite the acceleration strategies, the direct matrix-inversion approach was found to be uncompetitive with iterative approaches, with a computational burden an order of magnitude or more higher. The iterative conjugate-gradient approach, however, does appear promising, with computation times half that of our previous penalized-likelihood implementation.

I. Introduction

A. Motivation

A CT scanner measures the energy that is deposited in each channel of a detector array by X-rays that have been partially absorbed on their way through the object. The measurement process is complex, and quantitative measurements are always and inevitably associated with errors [1].

Sources for errors in absorption measurements are manifold. First, the CT measurement system (comprising power supply, gantry mechanics, detector scintillators, and readout electronics) rarely yields data that conform to the idealized models of the imaging process typically used for image reconstruction [1]. The second kind of error lies in the physics of detecting X-rays. No matter

how exact the measurement instruments are, the particle nature of the X-rays will unavoidably cause statistical fluctuations, which is known as noise.

The data that are obtained from a CT scanner, sometimes called sinogram data or raw data, are therefore not readily usable for reconstruction and need to be “restored” by removing or diminishing the traces of corruption effects during the measurement. Currently on commercial scanners, many of these effects are corrected through the use of a variety of preprocessing steps that are independent of one another and that do not, in general, explicitly model the statistics of the data noise [2], [3].

In recent years, we have formulated CT sinogram preprocessing as a statistical restoration problem in which the goal is to obtain the best estimate of the line integrals needed for reconstruction from the set of noisy, degraded measurements [4], [5], [6], [7]. We presented a general imaging model relating the degraded measurements to the sinogram of ideal line integrals and developed a method to estimate these line integrals by iteratively optimizing a statistically based objective function. We introduced this method as a middle ground between current practice and fully iterative algorithms, which iteratively refine the image estimate so that it agrees well in a predefined sense with the measured data when a model of the imaging process is applied to the image estimate. Fully iterative algorithms are extremely computationally intensive, and while some are beginning to approach clinically practical reconstruction times, these generally do not include complete models of the degradations discussed here [8], [9], [10].

While CT noise is most accurately modeled as compound Poisson, since the energy integrating detectors measure an energy-weighted sum of independent Poisson random variables [11], [12], [13], we have examined approximations based of the simple Poisson distribution, which gives rise to a penalized likelihood (PL) objective function, and based on the normal distribution, which gives rise to a penalized weighted least squares (PWLS) objective function [4]. The PWLS approach was found to give similar results to PL except for very low-SNR sinograms [4]. However, as implemented, there was little computational benefit to using the PWLS approach rather than the PL approach, even in the higher SNR regime. Computational burden is a critical factor in CT image reconstruction schemes. Clinicians expect reconstructed CT images to be available immediately after completion of the scan, which is important for verifying scan quality and for trauma cases.

In this work we develop and compare two different methods for implementing PWLS sinogram

restoration with the hope of improving computational performance. The first approach takes advantage of the fact that the PWLS estimator has a closed-form solution, albeit one that involves a large matrix inverse. We exploit the sparse and banded structure of the matrices to introduce a tiling strategy that allows us to break the problem into a large number of smaller, slightly overlapping problems. Secondly we consider an accelerated iterative approach to finding the PWLS solution, based on a preconditioned conjugate gradient optimization method.

It should be mentioned that Li *et al.* have explored the use of PWLS for sinogram *smoothing* with promising results. Their methods are applied to pre-processed, log-transformed measurements, assuming normally distributed measurement noise with an empirically determined, signal-dependent variance [14]. They do not explicitly consider the problem of sinogram restoration.

B. Effects of interest

In this paper, we focus on two specific sinogram degradations: non-uniform detector gain and crosstalk.

1) Non-uniform detector gain: In a real system, the responses of the single detector elements usually vary from channel to channel due to gain variations. Figure 1 shows the effect on the reconstructed image of uncorrected simulated random channel response variations of 3%. The left image was reconstructed from an undegraded data set, the right image was reconstructed from the corrupted sinogram data.

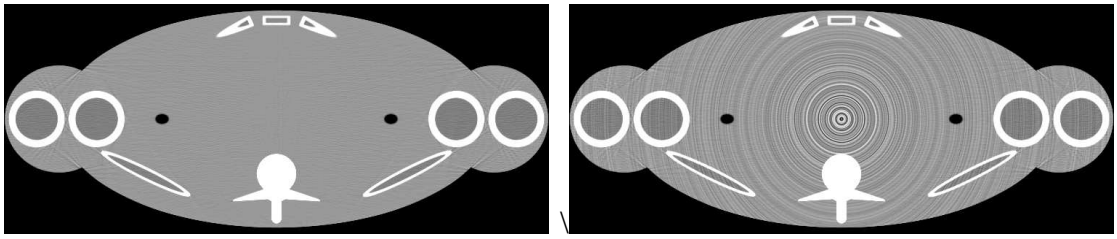

Figure 1. The effects of small (3%) random deviations in the air scan. *Left:* The forbild thorax phantom reconstructed from clean data. *Right:* The same phantom reconstructed from a corrupted sinogram. (L/W: -20/200 HU)

2) Crosstalk: When X-ray photons impinge on a CT detector, the scintillator in a detector element emits optical photons, which are detected by a photodiode. The scintillator light and the resulting charge generated by the photodiode are proportional to the deposited photon energy, and

give an estimate of the X-ray intensity seen by an individual detector pixel. Most of the deposited energy stays inside the detector channel, but there are leakage phenomena into other channels, which produce so-called *crosstalk*. CT detector crosstalk is ascribed to effects like optical and *K*-escape photon migration through the detector spacers, as well as electronic crosstalk in the readout circuitry [15].

Detector crosstalk can be modeled sufficiently accurately by assuming that each pixel obtains contributions from its immediate neighborhood only, corresponding to convolution with a 3×3 kernel. The kernel used in this paper was measured on a Philips Brilliance-16 CT system (Philips Medical Systems, Cleveland) [16].

Building a monolithic detector with, for example, 672×64 detector elements is not practicable, and one usually resorts to combining modular building blocks. Depending on the assembly of the modules, crosstalk will not typically be homogeneous across the entire detector. Although the kernel may be invariant within a module, the situation is different at the module boundaries, across which signal leak may be much smaller. An example of this kind of behavior is shown in the upper part of Figure 2, where one can see the results of missing crosstalk at the module boundaries as slightly brighter lines. Apart from a general loss of spatial resolution in the images due to crosstalk, its absence in distinct places on the detector leads to strong ringing in the images. This is shown in the bottom right image of Figure 2.

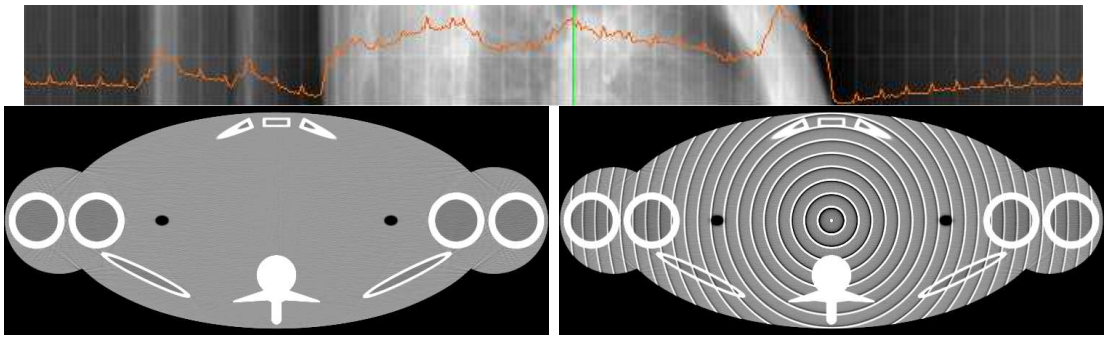

Figure 2. *Top:* A CT raw projection taken from a thorax scan. The peaks in the profile are due to regularly interrupted crosstalk at detector module boundaries. The increasing line integral values towards the edges of the detector are caused by the bow tie filter. *Bottom left:* The forbild thorax phantom reconstructed from clean data. *Bottom right:* Image reconstructed from crosstalk-corrupted sinogram. (L/W: -20/200 HU)

II. METHODS

A. Measurement Model

We start from a somewhat simplified version of a measurement model we have introduced previously [4]. We assume that the CT scan produces a set of measurements that are organized into a one-dimensional (1D) vector $\mathbf{y}$, with elements y_i^{meas}, $i = 1, \ldots, N_Y$, where N_Y is the total number of measurements in the scan, given by the product of the number of detector elements and the number of projection views acquired. So for a conebeam system $N_Y = N_u N_v N_\alpha$, where N_u is the number of detector columns, N_v is the number of detector rows, and N_α is the number of projection views.

We assume that each y_i^{meas} is a realization of a random variable Y_i^{meas} whose statistics are described by

$$Y_i^{meas} \approx G_i \overline{E}_i \text{Poisson} \left\{ \sum_{j=1}^{N_y} I_j b_{ij} e^{-f_j(l_j^{(mono)})} \right\}, \tag{1}$$

Here, G_i is the gain in the channel making the measurement, $\overline{E}_i$ the average energy of the photons incident on the patient along that line, I_j the number of photons incident along the line, b_{ij} is the degradation kernel, f_j is a function capturing the effect of beam-hardening, and the $l_j^{(mono)}$ are the set of ideal line integrals that we seek to estimate from the measurements. All of these quantities are assumed to be known, except, of course, the $l_j^{(mono)}$. For simplicity we have omitted terms for scatter, dark current, and electronic noise.

Of particular interest for sinogram restoration is the factor b_{ij}, which can model both literal signal leakage from channel to channel (crosstalk) or from a given channel to itself at a later time (afterglow), as well as "geometric leakage" resulting in photons traveling along response lines other than those assumed in the naive point source-point detector model. Effects such as off-focal radiation and anode angulation can be modeled this way, for example [4], [17].

B. PWLS objective function

We have previously discussed how to formulate penalized likelihood solutions to the estimation of the ideal line integrals. We derived a separable paraboloidal surrogates strategy yielding an iterative update guaranteed to monotonically increase the PL objective function. We have also explored PWLS approaches to estimating the ideal transmitted intensities (from which the line

integrals can be readily obtained) [4]. In this work, we again pursue a PWLS approach, with a slightly different formulation, with the intent of exploring alternative and hopefully faster implementations.

We frame the problem as one of estimating ideal *exponentiated line integrals*, which we define as

$$z_i \equiv e^{-f_i(l_i^{(mono)})}. \tag{2}$$

Naturally, the line integrals of interest can easily be computed from an estimate $\hat{z}_i$ of z_i by use of

$$\hat{l}_i^{(mono)} = f_i^{-1}\left[\ln\left(\hat{z}_i\right)\right]. \tag{3}$$

We assume we can write $G_i = \overline{G} g_i$,where G is an overall gain common to all detectors and g_i is the small channel-to-channel deviation, typically just a few percent at most. We define adjusted measurements

$$\tilde{Y}_i \equiv \left(\frac{Y_i^{meas}}{\overline{E}_i \overline{G}}\right). \tag{4}$$

These adjusted measurements then have mean

$$\left\langle \tilde{Y}_i \right\rangle \simeq \sum_{j=1}^{N_y} b_{ij} I_j g_j z_j = BD\boldsymbol{z}, \tag{5}$$

where $\boldsymbol{B}$ is a matrix with elements b_{ij} and $\boldsymbol{D}$ is a diagonal matrix, $\boldsymbol{D} = \mathrm{Diag}\left[I_i g_i\right]$. We have made a small approximation by bringing the factor g_i inside the sum as g_j. These adjusted measurements have approximate variance

$$\sigma^2_{\tilde{Y}_i} \simeq \boldsymbol{BDz}. \tag{6}$$

We seek to find z by minimizing a PWLS objective function

$$\Phi^{PWLS}(\mathbf{z}; \tilde{\mathbf{y}}) \equiv \mathcal{L}^{PWLS}\left(\mathbf{z}; \tilde{\mathbf{y}}\right) + \beta R\left(\mathbf{z}\right), \tag{7}$$

where

$$\mathcal{L}^{PWLS}\left(\tilde{\mathbf{z}}; \tilde{\mathbf{y}}\right) = \sum_{i=1}^{N_Y} w_i \left[\tilde{y}_i - \left[\boldsymbol{BD}z\right]_i\right]^2. \tag{8}$$

It is traditional in PWLS formulations to choose the weights to be the reciprocal of an unbiased estimate of the variances. However, in low-dose CT, where this kind of noise modeling is likely to be most significant, we have found that inverse-variance weighting leads to undersmoothing

of low-count measurements and persistence of noise-induced streak artifacts, since for Poisson-distributed measurements low-count measurements actually have lower variance (albeit higher relative uncertainty).[4] We wish to choose weights that will smooth the least reliable (lowest count) measurements more heavily and thus we set the weights equal to the unbiased estimate of the variances, rather than their reciprocal:

$$w_i = \tilde{y}_i. \tag{9}$$

An alternative, and potentially more principled, approach to achieve a similar effect would be to use a spatially varying penalty or smoothing parameter, but here we simply use the variance weighting. The specific choice of weights does not affect the conclusions of this paper since the computational burden depends principally on the dynamic range of the weights, which is the same for either choice of weights.

The roughness penalty $R(\mathbf{z})$ can be expressed very generally in the form

$$R(\mathbf{z}) = \sum_{k=1}^{K} \psi_k \left([Tz]_k \right), \tag{10}$$

given by Fessler [18], where T is a matrix and ψ_k a potential function that assigns a cost to the K combinations of exponentiated line integral values represented by the matrix product $T\mathbf{z}$. The smoothing or regularization parameter β in Eqn. 7 determines the relative influence of the likelihood and smoothness terms. In this work, we employ a quadratic penalty $\psi_k(t) = \omega_k t^2/2$ applied to the simple difference of a sinogram sample with its horizontal and vertical neighbors, with $\omega_k = 1/2$. Thus $R(\mathbf{z}) = ||\boldsymbol{Rz}||^2$ for an appropriate choice of matrix $\boldsymbol{R}$.

Overall, in matrix form, we seek

$$\hat{\mathbf{z}} = \arg\min \Phi^{PWLS}(\mathbf{z}; \tilde{\mathbf{y}}) = \arg\min(||\mathbf{W}^{\frac{1}{2}}\mathbf{y} - \mathbf{W}^{\frac{1}{2}}\mathbf{B}\mathbf{D}\mathbf{z}||^2 + \beta||\mathbf{R}\mathbf{z}||^2), \tag{11}$$

where

$$\mathbf{W}^{\frac{1}{2}}\mathbf{W}^{\frac{1}{2}} = \mathbf{W}, \tag{12}$$

and $\mathbf{W} = \mathrm{Diag}\,[\tilde{y}_i]$. Note that the objective function does not include an explicit positivity constraint. This allows for a closed-form solution. An explicit positivity constraint is unlikely to be active since the SNR is typically good even for the most attenuated lines in CT.

C. Closed-form solution

Equation (11) has a closed-form solution given by the solution to the following linear system:

$$((\mathbf{BD})^{\mathbf{T}}\mathbf{WBD} + \beta\mathbf{R}^{\mathbf{T}}\mathbf{R})\mathbf{z} = (\mathbf{BD})^{\mathbf{T}}\mathbf{Wy}. \tag{13}$$

Numerous well-established methods for solving this system are available. One of the most straightforward methods for this purpose is the LU decomposition with forward- and back-substitution. The LU method solves a system of equations with N unknowns using an operation count proportional to N^3, which includes forward- and back-substitutions [19]. This method requires that the system matrix be non-singular, which is the case for realistic system matrices arising in sinogram restoration. One can use straightforward Gaussian elimination as well, but the LU method is more flexible as it allows to solve for multiple right-hand-sides (RHS).

There are two issues with using a standard method like LU for the problem of sinogram restoration. The first is the matrix size. With realistic detector sizes of 672×64 elements, even if the problem were restricted to only crosstalk, the corresponding system matrix would still have $43008^2 = 1.9 \cdot 10^9$ elements corresponding to 16 GB of data if the matrix elements are stored as double precision floating point numbers.[1] Even with a lot of memory, these sizes are cumbersome to handle, especially because for an efficient implementation it is necessary to hold more than one matrix in memory, for example $\mathbf{BD}$ as well as $(\mathbf{BD})^{\mathbf{T}}\mathbf{WBD}$.

Also, algorithms like Gaussian elimination or LU decomposition with forward- and back-substitution require $\mathcal{O}(\mathrm{N}^3)$ operations. For the above example detector size, this is an unreasonably large number. To reduce the computational effort, we introduce two further implementation strategies: a software tiling strategy aimed at breaking the problem into smaller, albeit coupled units, and a strategy to exploit matrix sparseness.

1) Software tiling strategy: In order to reduce the problem of excessive operation counts, one can partition the detector in tiles of reasonable size (see Figure 3). More specifically, these should be called software tiles to distinguish them from the hardware modules that comprise a CT detector (see Section I-B2). Each tile is then processed independent of the others. Unfortunately, the software tiles cannot be treated completely independently because crosstalk couples the

[1]Making use of the extended precision that doubles offer compared with floats is highly advisable. Experience shows that this can improve the restoration results.

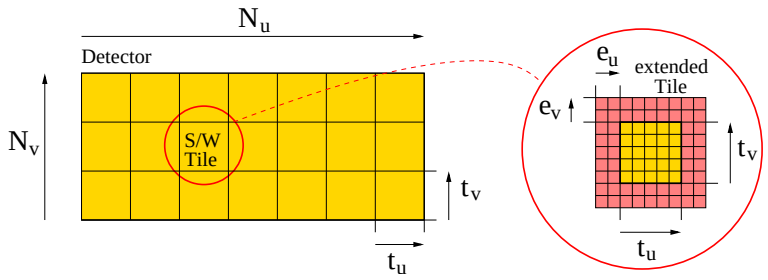

Figure 3. Tiling principle. The detector is virtually partitioned into rectangular tiles. Each tile is extended with a margin of surrounding pixels. The algorithm is applied to the extended tiles, although only the unextended region is used.

resulting equations. Even if the software tile boundaries are chosen so as to coincide with the zero-crosstalk boundaries between the detector modules (see Section I-B2), the roughness penalty leads to further coupling that cannot be ignored without inducing intolerable edge effects.

Since both the coupling due to crosstalk and the penalty are highly local, it is possible to effectively eliminate edge effects by using larger tiles that overlap. This means one extends a chosen tile by a suitable margin, does the processing on the extended tile, and then stores only the inner part of the restored tile.

2) Exploiting sparseness: It was mentioned previously that the system matrix is very sparse because of the highly local nature of the coupling between detector elements. The LU method is used to invert general matrices, so that when applied to a sparse matrix, it spends most of the computation time on multiplying zeros.

We can further reduce computational load and storage by exploiting the banded structure of the matrices, which can be seen in Fig. 4. Due to the 3×3 crosstalk kernel (see Sec. I-B2), $\mathbf{BD}$ contains entries on three bands with three diagonals each, as shown in Fig. 4. The center band is centered around the main diagonal of the matrix, the other two bands show a displacement of $\pm N_u$ from the center band. The matrix $\mathbf{A} = (\mathbf{BD})^{\mathbf{T}}\mathbf{W}(\mathbf{BD})$ must therefore contain five bands with five diagonals each, with band displacements of $\pm N_u$ and $\pm 2N_u$ from the central band. The penalty matrix R is also a 3×3 kernel, so that the entire expression $((\mathbf{BD})^{\mathbf{T}}\mathbf{W}(\mathbf{BD}) + \beta\mathbf{R}^{\mathbf{T}}\mathbf{R})\mathbf{z}$ has the same shape as $\mathbf{A}$.

This means that for an optimally compressed storage where only the non-zero bands are kept, the memory requirements are low. For a detector with 672×64 elements, for example, 25 bands with 672×64 elements each must be stored, which amounts to $25 \cdot 672 \cdot 64 = 1075200$ elements,

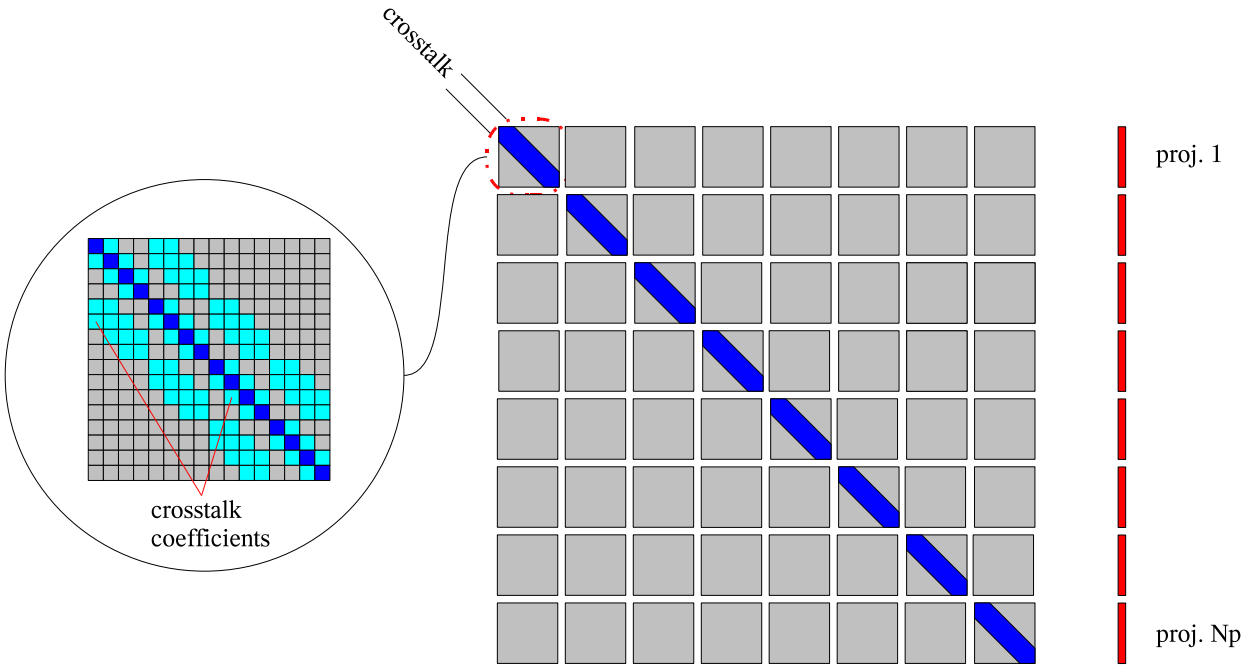

Figure 4. Graphic representation of the model containing only crosstalk. The large square structure represents the system matrix with band diagonal structure, the column on the right stands for the vector of uncontaminated intensities $\boldsymbol{z}$ and contains all $N_Y = N_u N_v N_\alpha$ measured samples.

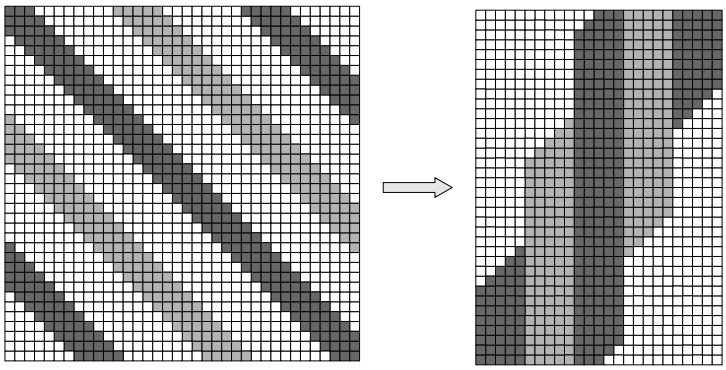

Figure 5. Matrix compression. All non-zero diagonals of the uncompressed matrix on the left are stored vertically in the compressed matrix on the right.

corresponding to 8 MB when stored as double precision floating point numbers. Figure 5 serves the purpose of illustrating the compact storage scheme and shows the compression of a 36×36 matrix with five bands.

For banded matrices the standard linear algebra libraries offer efficient storage schemes, in

which only the inner diagonal part of the matrix, i.e. a number of contiguous k_l subdiagonals and k_u superdiagonals are stored. The corresponding solvers operate efficiently by neglecting the lower left and upper right zero regions of the matrix during decomposition.

However, a closer look at the structure of the system matrix reveals that most of the k_l subdiagonals and k_u superdiagonals are also zero, which means that a simple compact band storage scheme still incorporates many zeros. Since the standard resources (Numerical Recipes routines, GNU Scientific Library (GSL), Intel Math Kernel Library (MKL)) do not offer solutions that are optimized to our special matrix structure, it was necessary to implement a dedicated library.

Efficient LU decomposition is done according to *Crout's method* [19]. Since the matrix shape of our PWLS problem is exactly known, it was possible to adapt Crout's method to this structure and obtain a speedup of two as opposed to the highly optimized LU routine that the MKL library offers. The nature of the adaptation is such that the adapted version of Crout's method touches only those matrix elements – both of the original and L and U matrices – which are known to be non-zero. In addition, one can make use of the fact that when the system matrix is decomposed, the values in the resulting L and U matrices are located on the same diagonals. The values on the additional diagonals that do appear are smaller by seven orders of magnitude, which makes them negligible.

D. Iterative solution

An alternative strategy for solving the system of linear equations that is given by the PWLS formulation of the sinogram restoration problem is the use of a fast iterative solution algorithm. In our previous work, we used a separable algorithm that was readily parallelizable but slow to converge, offering little computational gain over the penalized-likelihood approach. However, since the matrix $\mathbf{A} = (\mathbf{BD})^{\mathbf{T}}\mathbf{W}(\mathbf{BD}) + \beta\mathbf{R}^{\mathbf{T}}\mathbf{R}$ is symmetric and positive definite, it is possible to use the conjugate gradient (CG) algorithm [20] to solve the system of equations.

The advantage of this algorithm is that each iteration consists only of a matrix-vector multiplication with the original system matrix and a few additional vector operations. There are three points that make the CG approach attractive.

1) Since the system matrix is so sparse, only a small number of matrix elements need to be considered for building the product.

2) There is no need to store any intermediate matrices, which may potentially be less sparse than the one to invert.
3) The system matrix has so few bands, and also only in fixed positions, that it is possible to devise an optimal storage scheme so that detector tiling becomes unnecessary, as discussed in the previous section.

The CG algorithm was first tested to solve the unweighted equation

$$((\mathbf{BD})^{\mathbf{T}}\mathbf{BD} + \beta\mathbf{R}^{\mathbf{T}}\mathbf{R})\mathbf{z} = (\mathbf{BD})^{\mathbf{T}}\mathbf{y} \tag{14}$$

on the Forbild thorax data set. Without the weighting matrix W and assuming no intensity modulation, i.e. the values D_{jj} of matrix $\mathbf{D}$ being close to one, the matrix

$$((\mathbf{BD})^{\mathbf{T}}\mathbf{BD} + \beta\mathbf{R}^{\mathbf{T}}\mathbf{R}) \tag{15}$$

is diagonal dominant with constant values along the diagonal. In this case, the matrix is relatively close to the unit matrix, and the CG algorithm needs five iterations to converge to the solution. With the weighting matrix included, the situation is different. Due to the large dynamic range of the measurement values that compose $\mathbf{W}$, the range of the matrix elements that compose $((\mathbf{BD})^{\mathbf{T}}\mathbf{WBD} + \beta\mathbf{R}^{\mathbf{T}}\mathbf{R})$ also increases, so that for the Forbild thorax example about 400 iterations are needed to obtain a properly restored data set.

It is well known that the CG algorithm converges more slowly as the matrix to invert grows different from the unit matrix. For this reason one usually performs so-called *preconditioning*, which has been implemented in fully iterative tomographic reconstruction as well [21], [22]. In the general case, if the problem

$$\mathbf{Ax} = \mathbf{b} \tag{16}$$

is not well conditioned in the sense of a fast solution with the CG method, and an approximate inverse $\tilde{\mathbf{A}}^{-1}$ of $\mathbf{A}$ is known, one can precondition the problem by setting $\hat{\mathbf{A}} = \tilde{\mathbf{A}}^{-\mathbf{1}}\mathbf{A}$ and $\hat{\mathbf{b}} = \tilde{\mathbf{A}}^{-\mathbf{1}}\mathbf{b}$. Instead of (16) one then solves

$$\hat{\mathbf{A}}\mathbf{x} = \hat{\mathbf{b}}. \tag{17}$$

In order to remain compliant with the band storage format, admissible preconditioning matrices $\tilde{\mathbf{A}}^{-1}$ must be diagonal so that the multiplication does not generate any new non-zero elements. It is simple but efficient to choose

$$\tilde{a}_{ij}^{-1} = \delta_{ij}\, 1/a_{ij}, \tag{18}$$

with a_{ij} and $\tilde{a}_{ij}^{-1}$ being the elements of $\mathbf{A}$ and $\tilde{\mathbf{A}}^{-1}$, respectively.

We tested this preconditioning strategy by comparing the appearance of images obtained after varying numbers of iterations with and without preconditioning.

III. Results

A. Data simulation

We simulated CT data using the Radonis CT simulation environment (Philips Research). This software also allows specifying a subdivision of the source, detectors and trajectory into given numbers of sourcelets, detectorlets and trajectorylets to model the limited spatial resolution of the data. It then calculates the analytic line integrals through numerical phantoms defined as superpositions of geometric primitives such as ellipsoids and averages the exponentiated line integrals to obtain the simulated measurements. We simulated the parameters of a standard Philips 16-slice scanner, with a focal length of 570.0 mm and a 16-row detector of 672 channels of transverse extent 1.4 mm each and of longitudinal extent 1.14 mm (both measured at the detector). We generated Poisson realizations of the resulting exponentiated line integrals assuming a blank scan intensity of 1.4×10^5 photons per channel, with random (but assumed known) 3% channel-to-channel variation in the blank scan. Crosstalk was simulated by applying the following kernel to the data

$$\frac{1}{124} \begin{pmatrix} 1 & 5 & 1 \\ 5 & 100 & 5 \\ 1 & 5 & 1 \end{pmatrix},$$

with no crosstalk applied across detector module boundaries. The detector was assumed to comprise modules of dimensions 16×16 channels. We simulated a helical trajectory of pitch 4 mm, with 1160 angles per turn. A total of 5510 angles over 1710 degrees (4.75 turns) were simulated. For image reconstruction after sinogram restoration, we used the aperture-weighted wedge method [23].

B. Comparison of PWLS and PL

Disregarding the computational effort at this point, the tiled PWLS method yields very nearly identical results to the PL algorithm, which is shown in Figure 6. The noise-resolution plots are obtained by measuring the MTFs and standard deviations on the Forbild thorax phantom shown in Figure 7, where the MTFs are measured at the upper wire.

C. Effect of tiling on image quality

Figure 7 shows the influence of edge effects between software tiles on the reconstructed images when 0, 2, or 4 channels of overlap is used. The partitioning of the problem through tiling reduces the overall computational complexity, and we wish to use the minimum possible overlap to minimize computational time. The results indicate that a 4-pixel overlap eliminates artifacts, but that any less overlap leads to image artifacts.

D. Effect of preconditioning of the conjugate gradient method

Figure 8 shows the effects of preconditioning in terms of resulting images. Without preconditioning it takes 400 iterations to converge to a solution whereas preconditioning allows convergence in 20 iterations. To further examine convergence behavior, we show in Fig. 9 a plot of the total absolute change in image pixel values

$$\Delta_k = \sum_i \left| x_i^{(k+1)} - x_i^{(k)} \right| \tag{19}$$

as a function of iteration. Here $x_i^{(k)}$denotes the ith pixel value at the kth iteration. It is seen that the preconditioned algorithm achieves negligible levels of change (~0.01) to the image after about 20 iterations, where the unconditioned algorithm takes 400 iterations to reach a similar level.

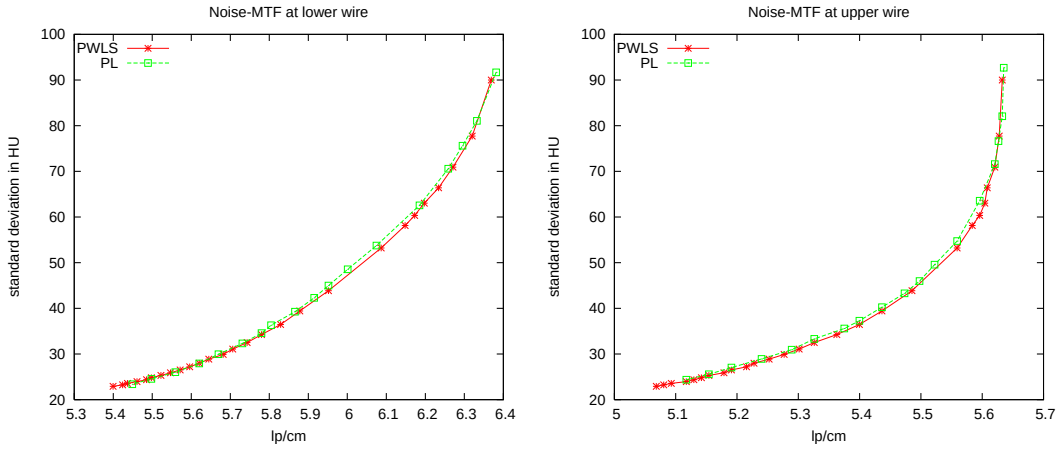

Figure 6. Noise-resolution curves comparison between the PL and PWLS methods.

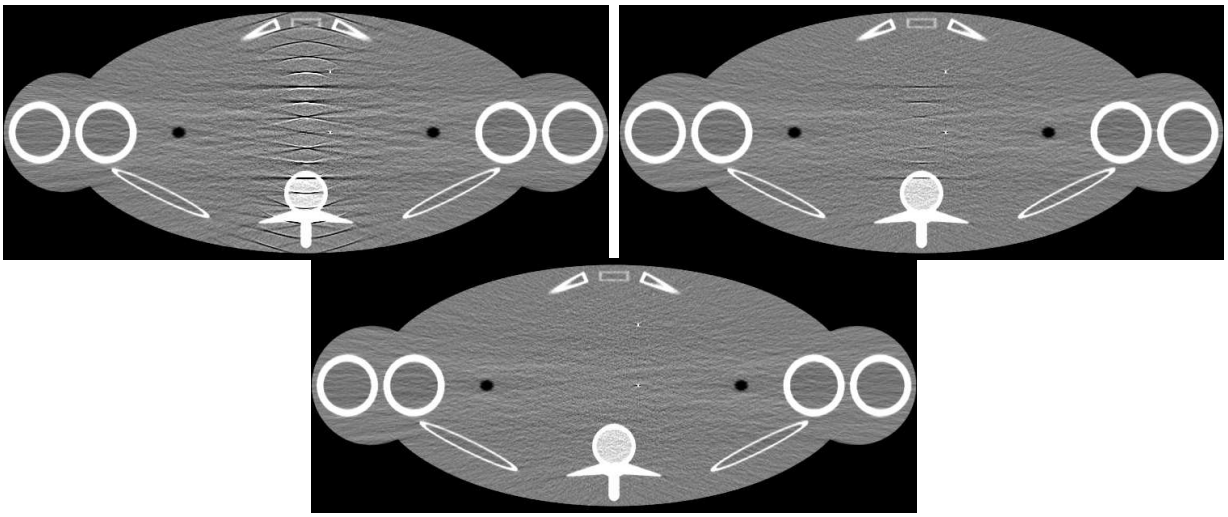

Figure 7. Reconstruction from data restored on separate tiles with overlap of 0, 2, and 4 pixels (top to bottom). (L/W: 0/500 HU)

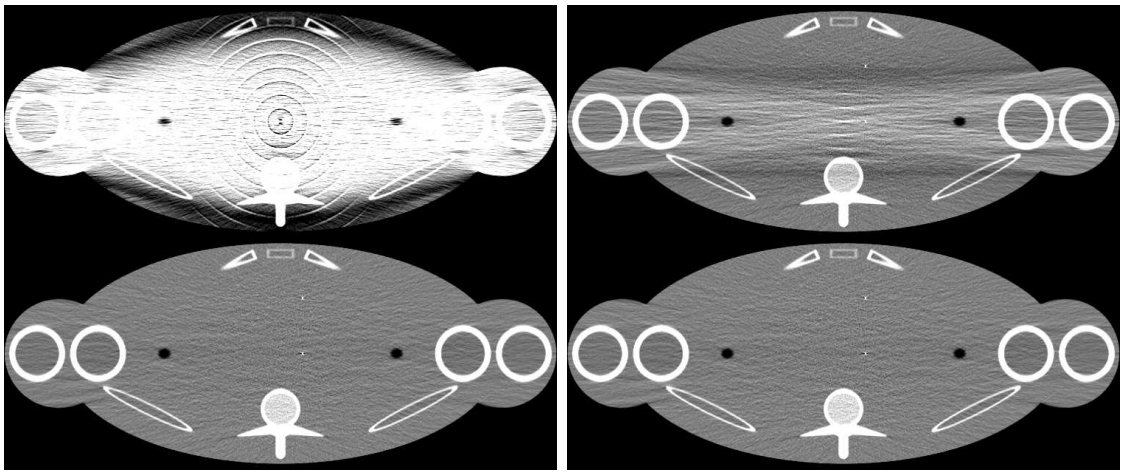

Figure 8. Effects of preconditioning. From upper left to lower right: Reconstructed data after 20, 150, and 400 CG iterations without preconditioning and 20 iterations after preconditioning.

E. Computational performance results

The computational performance results are given in Table 1. These results are for well-optimized C++ code running on a single core of a Intel Core Duo 6600 at 2.40 GHz with 2 GB of RAM.

We see that even with the use of tiling and sparseness, the closed-form strategy is still not competitive with well-optimized iterative approaches, which are an order of magnitude or more

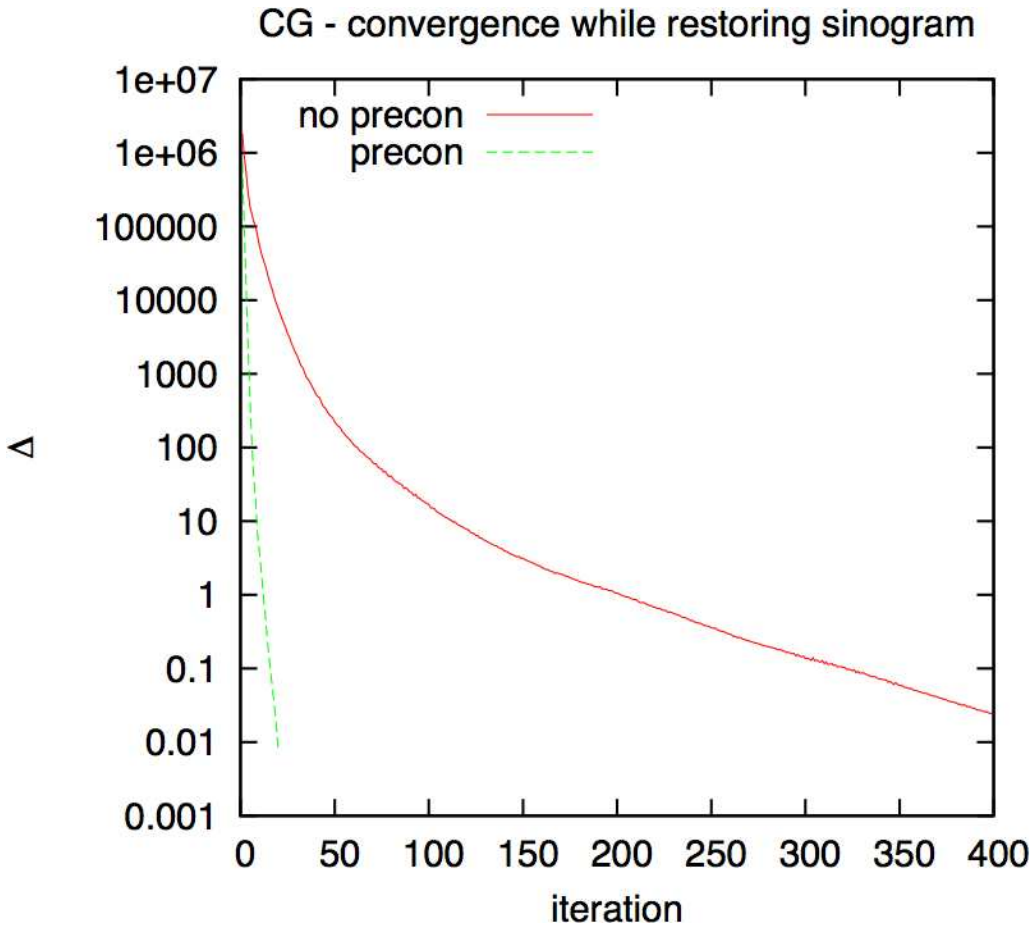

Figure 9. Plot of the absolute total change in image pixel values (Δ) as a function of iteration for preconditioned ("precon 1") and unconditioned ("no precon") CG algorithms.

Method	Computation time (mins:secs)
PL Sinogram restoration (20 iters)	5:35
Closed form with tiling, sparseness	45:00
Conjugate gradient, no preconditioning	81:20
Conjugate gradient, preconditioning	2:02

Table I

COMPUTATIONAL PERFORMANCE RESULTS

faster. The preconditioning CG algorithm does lead to a computation speed advantage of a factor of two as compared to the Poisson-based PL method.

IV. Conclusion

A PWLS sinogram-restoration method based on a Gaussian noise model was studied. Using the Gaussian model has the advantage that the problem can be reduced to solving a system of linear equations, for which there are many efficient methods available in the literature. Some of these have been used and discussed in this paper. The main difficulty with the formal approach of trying to invert the matrices that result lies in their size, which is due to the large amount of data that need to be considered. Straightforward matrix inversion schemes therefore do not seem to be an option.

We have explored and compared two implementation strategies for PWLS sinogram restoration: (1) A direct matrix-inversion strategy based on the closed-form solution to the PWLS optimization problem and (2) an iterative approach based on the conjugate-gradient algorithm. Obtaining optimal performance from each strategy required modifying the generic off-the-shelf implementations of the algorithms to exploit the particular symmetry and sparseness of the sinogram-restoration problem. For the closed-form problem, we subdivided the large matrix inversion into smaller coupled problems and exploited sparseness to minimize matrix operations. For the conjugate-gradient approach, we exploited sparseness and preconditioned the problem to speed convergence.

Despite the acceleration strategies, the direct matrix-inversion approach was found to be uncompetitive with iterative approaches, with a computational burden an order of magnitude or more higher. The iterative conjugate gradient approach, however, does appear promising, with computation times half that of our previous penalized likelihood implementation.

A point that has been mentioned but not stressed so far is the significance of the fact that the problem to be solved is a weighted least squares problem. In the absence of weighting, the system matrix would be data-independent and could be inverted once and for all. Unfortunately this is not the case because $\mathbf{A} = (\mathbf{BD})^{\mathbf{T}}\mathbf{W}(\mathbf{BD})$ must be inverted, and $\mathbf{W}$ contains measured data. Therefore, repeated matrix inversion is required for each patient and view.

The effects considered in this paper are arguably the two most significant needing restoration in sinograms. Off-focal radiation and afterglow issues have been reduced in new tube and detector

designs. However, these could be readily accommodated by the strategies described here. They would lead to somewhat less sparse system matrices as they would introduce new couplings among detector channels and projection views, but the structure would be regular and amenable to tiling and compression strategies described here.

ACKNOWLEDGMENT

The work of Patrick La Riviere was supported in part by NIH grant R01-134680CA.

REFERENCES

[1] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. Bellingham: SPIE Press, 2003.

[2] J. Hsieh, O. E. Gurman, and K. F. King, "Investigation of a solid-state detector for advanced computed tomography," *IEEE Trans. Med. Imag.*, vol. 19, pp. 930–940, 2000.

[3] R. P. Luhta, K. M. Brown, and S. J. Utrup, "Off-focal radiation corrrection in CT," *U.S. Patent*, no. 6,628,744, 2003.

[4] P. J. La Rivière, J. Bian, and P. Vargas, "Penalized-likelihood sinogram restoration for computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, pp. 1022–1036, 2006.

[5] P. Forthmann, T. Köhler, M. Defrise, and P. J. La Rivière, "Comparison of three sinogram restoration methods," vol. 6142, (San Diego, California, USA), pp. 61420Y–1 – 61420Y–12, 2006.

[6] P. Forthmann, T. Köhler, P. Begemann, and M. Defrise, "Penalized maximum-likelihood sinogram restoration for dual focal spot computed tomography," vol. 52, pp. 4513 – 4523, 2007.

[7] P. J. La Rivière and P. Vargas, "Correction for resolution non-uniformities caused by anode angulation in computed tomography," in *Nuclear Science Symposium Conference Record,*, vol. 5, pp. 2919–2923, 2006. CD-ROM.

[8] M. Kachelriess, M. Knaup, and O. Bockenbach, "Hyperfast parallel-beam and cone-beam backprojection using the cell general purpose hardware," *Med Phys*, vol. 34, pp. 1474–1486, Apr 2007. Evaluation Studies.

[9] J. Thibault, K. Sauer, C. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Med Phys*, vol. 34, pp. 4526–4544, Nov 2007.

[10] F. J. Beekman and C. Kamphuis, "Ordered subset reconstruction for x-ray CT," *Phys. Med. Biol.*, vol. 46, pp. 1835–1855, 2001.

[11] I. A. Elbakri and J. A. Fessler, "Efficient and accurate likelihood for iterative image reconstruction in x-ray computed tomography," in *Proc. SPIE*, vol. 5032, pp. 1839–1850, 2003.

[12] I. A. Elbakri, *X-ray CT image reconstruction using statistical methods*. PhD thesis, The University of Michigan, 2003.

[13] B. R. Whiting, "Signal statistics in X-ray computed tomography," in *Proc. SPIE*, vol. 4682, pp. 53–60, 2002.

[14] T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang, "Nonlinear sinogram smoothing for low-dose X-ray CT," *IEEE Trans. Nucl. Sci.*, vol. 51, pp. 2505–2513, 2004.

[15] S. Utrup, M. Chappo, B. Harwood, T. Krecic, R. Luhta, R. Mattson, D. Salk, and C. Vrettos, "Design and performance of a 32 slice CT detector system using back illuminated photodiodes," vol. 5368, pp. 40 – 51, 2004.

[16] R. Carmi, O. Shapiro, and D. Braunstein, "Resolution enhancement of x-ray CT by spatial and temporal MLEM deconvolution correction," 2004. Proceedings of the IEEE Medical Imaging Conference.

[17] P. J. La Rivière and P. Vargas, "Correction for resolution non-uniformities caused by anode angulation in computed tomography," *IEEE Trans. Med. Imag.*, vol. 27, pp. 1333–1341, 2008.

[18] J. A. Fessler, "Statistical image reconstruction methods for transmission tomography," in *Handbook of Medical Imaging, Vol. 2* (J. M. Fitzpatrick and M. Sonka, eds.), pp. 1–70, Bellingham, WA: SPIE Press, 2000.

[19] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C, Second edition*. Cambridge, UK: Cambridge University Press, 1992.

[20] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," vol. 49, pp. 409 – 436, 1952.

[21] N. H. Clinthorne, T. S. Pan, P. C. Chiao, L. Rogers, and J. A. Stamos, "Preconditioning methods for improved convergence rates in iterative reconstructions," vol. 12, no. 1, pp. 78 – 83, 1993.

[22] J. A. Fessler and S. Booth, "Conjugate-gradient preconditioning methods for shift-variant pet image reconstruction," *IEEE Trans. Image Processing*, vol. 8, pp. 688–699, 1999.

[23] K. M. Brown, D. J. Heuscher, and P. Kling, "Conebeam computed tomography imaging," *U.S. Patent*, no. 6,775,346, 2002.