Statistical X-ray computed tomography image reconstruction with beam hardening correction

Idris A. Elbakri, Jeffrey A. Fessler
Electrical Engineering and Computer Science Department
4240 EECS Bldg., University of Michigan, Ann Arbor, MI 48109-2122

ABSTRACT

This paper describes two statistical iterative reconstruction methods for X-ray CT. The first method assumes a mono-energetic model for X-ray attenuation. We approximate the transmission Poisson likelihood by a quadratic cost function and exploit its convexity to derive a separable quadratic surrogate function that is easily minimized using parallelizable algorithms. Ordered subsets are used to accelerate convergence. We apply this mono-energetic algorithm (with edge-preserving regularization) to simulated thorax X-ray CT scans. A few iterations produce reconstructed images with lower noise than conventional FBP images at equivalent resolutions. The second method generalizes the physical model and accounts for the poly-energetic X-ray source spectrum and the measurement nonlinearities caused by energy-dependent attenuation. We assume the object consists of a given number of non-overlapping tissue types. The attenuation coefficient of each tissue is the product of its unknown density and a known energy-dependent mass attenuation coefficient. We formulate a penalized-likelihood function for this poly-energetic model and develop an iterative algorithm for estimating the unknown densities in each voxel. Applying this method to simulated X-ray CT measurements of a phantom containing both bone and soft tissue yields images with significantly reduced beam hardening artifacts.

Keywords: X-ray CT, statistical reconstruction, ordered subsets, poly-energetic, beam hardening

1. INTRODUCTION

X-ray computed tomography (CT) provides structural information about tissue anatomy by imaging the tissue attenuation characteristics. CT scanners record 'projection' data that measure the amount of X-ray attenuation through tissue at different angles.

The linear attenuation coefficient $\mu(x, y, z, \mathcal{E})$ characterizes the overall attenuation property of tissue. It depends on the spatial coordinates and the beam energy, and has units of inverse distance. For a ray of infinitesimal width, the expected photon flux detected along a particular projection line L_i is given by:

$$E[Y_i] = \int I_i(\mathcal{E}) e^{-\int_{L_i} \mu(x,y,z,\mathcal{E}) dl} d\mathcal{E}, \quad i = 1, ..., N, \tag{1}$$

where Y_i represents the photon flux measured at the ith detector, and there are N detectors. The integral in the exponent is taken over the line L_i and $I_i(\mathcal{E})$ incorporates the energy dependence of the incident ray and detector sensitivity. The goal of any CT algorithm is to reconstruct the attenuation map μ from the measured data $[Y_1, ..., Y_N]$.

Filtered back projection (FBP) is the standard reconstruction technique for CT. Under situations with truncated or missing data, such as when metallic implants are present or when nonstandard scanning geometries (cone-beam, multi-slice or helical) are used, FBP-type algorithms require innovative rebinning and interpolation, and lead to biased estimators since they do not take the statistics of the data into account.[1]

Statistical techniques have several attractive features.[2–5] They can be well suited for arbitrary geometries and situations with truncated data. They can also model such phenomena as scatter and energy dependence leading to more accurate and artifact-free reconstruction. Statistical methods also easily incorporate the system geometry, detector response, object constraints and any prior knowledge. Their main drawback (when compared to FBP) is longer computation times. A lot of research effort is being invested in developing accelerated versions of statistical algorithms.[6,7,5]

The next section discusses the mono-energetic model and algorithm. Section 3 generalizes the model and the resulting algorithm to the more realistic poly-energetic case. In Section 4 we present preliminary results. Section 5 summarizes the results and outlines potential future work.

Email: ielbakri@umich.edu, WWW: http://www.eecs.umich.edu/∼fessler, Voice: 734-763-1434.

2. MONO-ENERGETIC X-RAY CT

In this section we describe the physical and statistical models for the problem of X-ray CT reconstruction. For simplicity, we assume a mono-energetic X-ray beam. The statistical reconstruction method involves maximizing a penalized-likelihood objective function. Maximizing the objective function by some appropriate iterative algorithm yields the reconstructed image.

2.1. Mono-energetic Model and Assumptions

Under the assumption of a mono-energetic beam, i.e. $I_i(\mathcal{E}) = I_i(\mathcal{E}_o)\delta(\mathcal{E} - \mathcal{E}_o)$, the expected photon flux simplifies as follows:

$$E[Y_i] = I_i(\mathcal{E}_o)e^{-\int_{L_i}\mu(x,y,z,\mathcal{E}_o)dl}. \tag{2}$$

We parameterize the image in object space (attenuation coefficient) using square pixels. The goal of the algorithm becomes to estimate the value of the (discretized) attenuation coefficient at those pixels. Let $\mu = [\mu_1, ..., \mu_p]'$ be the vector of unknown attenuation coefficients having units of inverse length, where $'$ stands for transpose.

The measurements in a photon-limited counting process such as X-ray CT are reasonably modeled as independently distributed Poisson random variables.[4] In transmission tomography, the mean number of detected photons is related exponentially to the projections (line integrals) of the attenuation map. The measurements are also contaminated by extra background counts, caused primarily by scatter. In addition, the detector introduces additive Gaussian read-out noise that can be accounted for in a several ways.[8] We assume the following model for our measurements:

$$Y_i \sim \text{Poisson}\{b_i e^{-[\mathbf{A}\mu]_i} + r_i\}, \quad i = 1, ..., N \tag{3}$$

where $b_i = I_i(\mathcal{E}_o)$ is the blank scan factor and N is the number of measurements (or, equivalently, the number of detector bins). The notation $[\mathbf{A}\mu]_i = \sum_{j=1}^{p} a_{ij}\mu_j$ represents the ith line integral. The $N \times p$ matrix $\mathbf{A} = \{a_{ij}\}$ is the system matrix which accounts for the system geometry as well as any other significant physical effects such as detector response. For ray i and pixel j, a_{ij} is the (normalized) area of overlap between the ray beam and the pixel. The term r_i accounts for the mean number of background events and can incorporate an estimate for read-out noise as well. The photon flux measured by the ith detector is represented by Y_i. We assume the Y_i's are independent and that b_i, r_i and $\{a_{ij}\}$ are known non-negative constants.[4]

To find a statistical estimate for the attenuation coefficient vector μ that is anatomically reasonable, we use a likelihood-based estimation approach. This is a natural choice since the likelihood is based on the statistical properties of the problem.[9,10] The Poisson log likelihood for independent measurements is given by:

$$L(\mu) = \sum_{i=1}^{N} \left\{ Y_i \log(b_i e^{-[\mathbf{A}\mu]_i} + r_i) - (b_i e^{-[\mathbf{A}\mu]_i} + r_i) \right\} \tag{4}$$

ignoring constant terms. When $\mathbf{A}$ has full column rank and the data is noise-free, maximizing the likelihood function will give a perfect result. In reality, the data is noisy and maximum likelihood (ML) will give a very noisy reconstruction due to the ill-posedness of the problem, hence the need for regularization.

We regularize by adding a penalty term to the likelihood function. A general form for the regularizing penalty is the following:

$$R(\mu) = \sum_{k=1}^{K} \psi([\mathbf{C}\mu]_k) \tag{5}$$

where the ψ's are potential functions acting on the soft constraints $\mathbf{C}\mu \approx 0$ and K is the number of such constraints. We use the convex edge-preserving Huber penalty

$$\psi(x;\delta) = \begin{cases} \frac{x^2}{2}, & x < \delta \\ \delta|x| - \frac{\delta^2}{2}, & x \geq \delta. \end{cases} \tag{6}$$

Combining the likelihood with a penalty gives a penalized-likelihood (PL) objective function:

$$\Phi(\mu) = L(\mu) - \beta R(\mu) \tag{7}$$

where β is a scalar that controls the tradeoff between the data fit and the penalty terms. The goal of the reconstruction technique becomes to maximize (7) subject to certain object constraints such as non-negativity:

$$\hat{\mu} = \underset{\mu \geq 0}{\text{argmax}} \ \Phi(\mu). \tag{8}$$

The next section will discuss an approach to finding a solution for (8) using an iterative technique.

2.2. Penalized Weighted Least Squares with Ordered Subsets

We formulate our algorithm by deriving a quadratic approximation to the Poisson likelihood, which leads to a simpler objective function. This is equivalent to assuming an additive noise Gaussian model for the data. This model is reasonable because of the large SNR in clinical X-ray CT.[11,12] The quadratic approximation leads to a penalized weighted-least-squares (PWLS) estimate. With high photon flux, PWLS leads to negligible bias[1] and the simpler objective function reduces computation time.

For convenience, we write the negative log likelihood for transmission data

$$-L(\mu) = \sum_{i=1}^{N} h_i([\mathbf{A}\mu]_i) \tag{9}$$

$$= \sum_{i=1}^{N} \left\{ -Y_i \log(b_i e^{-[\mathbf{A}\mu]_i} + r_i) + (b_i e^{-[\mathbf{A}\mu]_i} + r_i) \right\}. \tag{10}$$

Applying a second-order Taylor's expansion to $h_i(l)$ around some value $\hat{l}_i$ yields[13]:

$$h_i(l) \approx h_i(\hat{l}_i) + \dot{h}_i(\hat{l}_i)(l - \hat{l}_i) + \frac{\ddot{h}_i(\hat{l}_i)}{2}(l - \hat{l}_i)^2 \tag{11}$$

where $\dot{h}_i$ and $\ddot{h}_i$ are the first and second derivatives of h_i. Assuming $Y_i > r_i$, we can estimate the line integral with

$$\hat{l}_i = \log\left(\frac{b_i}{Y_i - r_i}\right). \tag{12}$$

Substituting this estimate into (11) gives the following approximation for h_i:

$$h_i(l) \approx (Y_i - Y_i \log Y_i) + \frac{w_i}{2}(l - \hat{l}_i)^2. \tag{13}$$

The first term in (13) is independent of l and can be dropped. The weight is $w_i = \frac{(Y_i - r_i)^2}{Y_i}$. The resulting cost function is

$$\Phi_q(\mu) = \sum_{i=1}^{N} \frac{w_i}{2}([\mathbf{A}\mu]_i - \hat{l}_i)^2 + \beta R(\mu). \tag{14}$$

The subscript $_q$ indicates that this objective function is based on a quadratic approximation to the log likelihood. We retain the penalty term, which may be non-quadratic.

According to the optimization transfer principle,[14,15] rather than minimizing the cost function in (14), one can resort to replacing Φ_q with a surrogate function that is easier to minimize at each iteration. The process is repeated iteratively, using a new surrogate function at each iteration.

The surrogate we seek for Φ_q is one that will make the cost function separable, so that all pixels can be updated simultaneously. Towards that end, we exploit the convexity of the data-fit term in (14). Rewrite the line integral[14,15]

$$[\mathbf{A}\mu]_i = \sum_{j=1}^{p} a_{ij}\mu_j = \sum_{j=1}^{p} \alpha_{ij} \left\{ \frac{a_{ij}}{\alpha_{ij}}(\mu_j - \mu_j^n) + [\mathbf{A}\mu^n]_i \right\} \tag{15}$$

where

$$\sum_{j=1}^{p} \alpha_{ij} = 1, \ \forall i, \alpha_{ij} \geq 0. \tag{16}$$

By convexity,

$$([\mathbf{A}\mu]_i - \hat{l}_i)^2 \leq \sum_{j=1}^{p} \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^n) + [\mathbf{A}\mu^n]_i - \hat{l}_i \right)^2. \tag{17}$$

The left hand side of (17) satisfies the conditions of the optimization transfer principle for a surrogate for the data-fit term.[4] We thus get the following separable surrogate:

$$Q(\mu; \mu^n) = \sum_{i=1}^{N} \sum_{j=1}^{p} \alpha_{ij} \frac{w_i}{2} \left(\frac{a_{ij}}{\alpha_{ij}} (\mu_j - \mu_j^n) + [\mathbf{A}\mu^n]_i - \hat{l}_i \right)^2. \tag{18}$$

A similar development can be pursued for the convex penalty term $R(\mu)$ yielding a separable penalty surrogate, denoted $S(\mu; \mu^n)$. We now seek to minimize the new separable global surrogate:

$$\phi(\mu; \mu^n) \stackrel{\triangle}{=} Q(\mu; \mu^n) + \beta S(\mu; \mu^n). \tag{19}$$

Since the surrogate is a separable paraboloid, it can be easily minimized by zeroing the first derivative. This leads to the following simultaneous update algorithm:

$$\mu_j^{n+1} = \left[\mu_j^n - \frac{\left. \frac{\partial \phi(\mu; \mu^n)}{\partial \mu_j} \right|_{\mu=\mu^n}}{\left. \frac{\partial^2 \phi(\mu; \mu^n)}{\partial \mu_j^2} \right|_{\mu=\mu^n}} \right]_+, \quad j = 1, ..., p. \tag{20}$$

The first and second derivatives of the surrogate are easily shown to be:

$$\left. \frac{\partial \phi(\mu; \mu^n)}{\partial \mu_j} \right|_{\mu=\mu^n} = \sum_{i=1}^{N} a_{ij} w_i ([\mathbf{A}\mu^n]_i - \hat{l}_i) + \beta \left. \frac{\partial S}{\partial \mu_j} \right|_{\mu=\mu^n} \tag{21}$$

$$\left. \frac{\partial^2 \phi(\mu; \mu^n)}{\partial \mu_j^2} \right|_{\mu=\mu^n} = \sum_{i=1}^{N} \frac{a_{ij}^2 w_i}{\alpha_{ij}} + \beta \left. \frac{\partial^2 S}{\partial \mu_j^2} \right|_{\mu=\mu^n} \tag{22}$$

To make the denominator in (20) small (and hence the step size large), we want $\{\alpha_{ij}\}$ to be large. We also want $\{\alpha_{ij}\}$ to facilitate convergence, and to be independent of the current iterate (so that it can be pre-computed).[3] One convenient choice is

$$\alpha_{ij} = \frac{a_{ij}}{\sum_{j=1}^{p} a_{ij}}. \tag{23}$$

It is possible that better choices exist.

Note that both the numerator and denominator in (20) involve summations over sinogram indices (i.e. back projection). In the spirit of similar work[7,6] we use ordered subsets to accelerate convergence. Ordered subsets subsample the sinogram in the angular domain. The back-projection process over the complete sinogram is replaced with successive back-projections over these subsets. One iteration is complete after going through all of the subsets. Ordered subsets accelerate algorithm convergence by a factor proportional to the number of subsets.[7,2] We call this method the ordered subsets penalized weighted least squares (OS-PWLS) algorithm for transmission tomography. To summarize, the OS-PWLS algorithm flows as follows:

- **for** each iteration $n = 1, ..., \text{niter}$
  - **for** each subset $S = 1, ..., M$
    - $\mu_{\text{old}} = \hat{\mu}$
    - $\hat{\mu}_j = \left[\hat{\mu}_j - \frac{M \sum_{i \in S} a_{ij} w_i ([\mathbf{A}\hat{\mu}]_i - \hat{l}_i) + \beta \left. \frac{\partial S}{\partial \mu_j} \right|_{\mu=\mu_{old}}}{M \sum_{i \in S} \frac{a_{ij}^2 w_i}{\alpha_{ij}} + \beta \left. \frac{\partial^2 S}{\partial \mu_j^2} \right|_{\mu=\mu_{old}}} \right]_+ \quad j = 1, \ldots, p.$

  - **end**

- **end**

The next section presents a poly-energetic model and generalizes OS-PWLS accordingly.

3. POLY-ENERGETIC X-RAY CT

In the previous section, we ignored the poly-energetic nature of the X-ray beam and the energy dependence of the attenuation coefficient. With a poly-energetic source, the detected photon flux along path L_i is given by (1). Nonlinear beam hardening artifacts result if one were to ignore the energy dependence of the measurements.[16–18] The nonlinear behavior generally leads to a reduction in the attenuation coefficient. Thick bones also generate dark streaks. In soft tissue, the values are depressed in a non-uniform manner, leading to what has been termed 'cupping'. In addition, bone areas can 'spill over' into soft tissue, leading to a perceived increase in the attenuation coefficient. In this section we develop a model for poly-energetic X-ray CT and develop an iterative algorithm that generalizes OS-PWLS.

3.1. Poly-energetic Statistical Model for X-ray CT

We assume that the object is comprised of K known non-overlapping tissue types. We also assume that the tissue class of each voxel is known. These classes can be determined by segmenting a FBP reconstruction that has been pre-processed for soft-tissue beam hardening effects.[18] For the kth tissue type, the attenuation coefficient is modeled as the product of the energy-dependent *mass* attenuation coefficient $m_k(\mathcal{E})$ (cm^2/g) and the energy-independent density $\rho^k(x,y)$ (g/cm^3) of the tissue.[19,18,20] Expressed mathematically,

$$\mu(x,y,\mathcal{E}) = \sum_{k=1}^{K} m_k(\mathcal{E}) \rho^k(x,y) r^k(x,y) \tag{24}$$

where $r^k(x,y) = 1$ if $(x,y) \in$ tissue k and $r^k(x,y) = 0$ otherwise.

We again denote the system matrix with $\mathbf{A} = \{a_{ij}\}$ and make the following definitions:

$$R_k \stackrel{\triangle}{=} \{\text{set of pixels classified as tissue type } k\}, \tag{25}$$

$$\underline{r}^k = \left(r_1^k, \ldots, r_p^k\right) \text{ where } \begin{cases} r_j^k = 1, & j \in R_k \\ r_j^k = 0, & \text{otherwise,} \end{cases} \tag{26}$$

$$s_i^k(\rho) \stackrel{\triangle}{=} \int_{L_i} \rho^k(x,y) r^k(x,y)\, dl, \tag{27}$$

$$\underline{v}_i(\rho) = (s_i^1, s_i^2, ..., s_i^K). \tag{28}$$

We also assume that the mass attenuation coefficients $\{m_k(\mathcal{E})\}_{k=1}^K$ of the K tissue types are known. Discretization aside, from (1) the mean of the measured data along path L_i is

$$\begin{aligned} E[Y_i] &= \int I_i(\mathcal{E}) e^{-\sum_{k=1}^{K} m_k(\mathcal{E}) \int_{L_i} \rho^k(x,y) r^k(x,y)\, dl} d\mathcal{E} \\ &= \int I_i(\mathcal{E}) e^{-\sum_{k=1}^{K} m_k(\mathcal{E}) s_i^k(\rho)} d\mathcal{E} \\ &= \int I_i(\mathcal{E}) e^{-\underline{m}'(\mathcal{E}) \underline{v}_i(\rho)} d\mathcal{E} \\ &\stackrel{\triangle}{=} \bar{Y}_i(\underline{v}_i(\rho)) \end{aligned} \tag{29}$$

where $\underline{m}'(\mathcal{E}) = [m_1(\mathcal{E}), ..., m_K(\mathcal{E})]$. We have expressed the measurements as a function of the vector $\underline{v}_i$ which has as its elements the line integrals of the K different tissue densities. From knowledge of the X-ray spectrum, we tabulate the values of $\bar{Y}_i(\underline{v}_i)$ and its gradient $\nabla\bar{Y}_i(\underline{v}_i)$. In the discrete domain,

$$s_i^k(\rho) = \sum_{j=1}^{p} a_{ij} r_j^k \rho_j. \tag{30}$$

The goal of the algorithm is to estimate the density coefficient vector $\rho = [\rho_1, ..., \rho_p]$. Note that rather than estimating K vector quantities of length p, each representing the density of one kind of tissue, the assumption of non-overlapping tissue types enables us to keep the number of unknowns equal to p, as is the case in the mono-energetic model. This is possible only if prior segmentation of the object is available. This can be obtained from a good FBP image, for example.

3.1.1. Poly-Energetic Model Cost Function

We now setup the Poisson log likelihood in terms of the density ρ and the vector function $\underline{v}_i$. To get a quadratic cost function, we again use the second-order Taylor's expansion.

Recall that the function $\bar{Y}_i(\underline{v}_i(\rho))$ represents the expected value of the measurement Y_i at the ith detector. Using $\bar{Y}_i$ in (4) gives the following negative log likelihood:

$$\begin{aligned} -L(\rho) &= \sum_{i=1}^{N} h_i(\underline{v}_i(\rho)) & (31) \\ h_i(\underline{v}_i(\rho)) &\triangleq -Y_i \log\left[\bar{Y}_i(\underline{v}_i(\rho)) + r_i\right] + (\bar{Y}_i(\underline{v}_i(\rho)) + r_i) & (32) \end{aligned}$$

The problem now is to find an estimate $\hat{\rho}$ such that:

$$\hat{\rho} = \underset{\rho \geq 0}{\text{argmin}} \ \Phi(\rho) \tag{33}$$

where

$$\Phi(\rho) = -L(\rho) + \beta R(\rho). \tag{34}$$

The regularization term can be treated exactly as before or it can be modified to avoid smoothing between different tissue types. For simplicity in this presentation we focus on the likelihood term.

Suppose we have an initial estimate of $\underline{v}_i(\rho)$, denoted $\hat{\underline{v}}_i = (\hat{s}_i^1, ..., \hat{s}_i^K)$, obtained from forward projecting an initial density image $\hat{\rho}$. We expand $h_i(\underline{v}_i(\rho))$ in a second-order Taylor series around $\hat{\underline{v}}_i$:

$$h_i(\underline{v}) \approx h_i(\hat{\underline{v}}_i) + \nabla h_i(\hat{\underline{v}}_i)(\underline{v} - \hat{\underline{v}}_i) + \frac{1}{2}(\underline{v} - \hat{\underline{v}}_i)'\nabla^2 h_i(\hat{\underline{v}}_i)(\underline{v} - \hat{\underline{v}}_i). \tag{35}$$

Taking the first and second derivatives of $h_i(\underline{v}) = -Y_i \log \bar{Y}_i(\underline{v}) + \bar{Y}_i(\underline{v})$ yields the following:

$$\begin{aligned} \nabla h_i(\underline{v}) &= \left(1 - \frac{Y_i}{\bar{Y}_i(\underline{v})}\right) \nabla\bar{Y}_i(\underline{v}) & (36) \\ \nabla^2 h_i(\underline{v}) &= \left(1 - \frac{Y_i}{\bar{Y}_i(\underline{v})}\right) \nabla^2\bar{Y}_i(\underline{v}) + \frac{Y_i}{\bar{Y}_i^2(\underline{v})} \nabla'\bar{Y}_i(\underline{v}) \nabla\bar{Y}_i(\underline{v}). & (37) \end{aligned}$$

The gradient ∇h_i is a row vector and the Hessian operator ∇^2 gives a $K \times K$ matrix of partial derivatives. To simplify the algorithm and maintain the desirable property of separability, we assume that Y_i is close enough to $\bar{Y}_i(\hat{\underline{v}}_i)$ for us to drop the first term on the right of (37). This also ensures that the resulting Hessian approximation is nonnegative definite.

In the Taylor expansion (35), the first term is constant and does not affect minimization, so we drop it. We now have the following quadratic approximation to the negative log likelihood:

$$\begin{aligned} -L(\rho) \approx \Phi_q(\rho) &\triangleq \sum_{i=1}^{N} \nabla h_i(\hat{\underline{v}}_i)(\underline{v}_i(\rho) - \hat{\underline{v}}_i) + \frac{1}{2Y_i}(\underline{v}_i(\rho) - \hat{\underline{v}}_i)'\nabla'\bar{Y}_i(\hat{\underline{v}}_i)\nabla\bar{Y}_i(\hat{\underline{v}}_i)(\underline{v}_i(\rho) - \hat{\underline{v}}_i) & (38) \\ &= \sum_{i=1}^{N} \nabla h_i(\hat{\underline{v}}_i)(\underline{v}_i(\rho) - \hat{\underline{v}}_i) + \frac{1}{2Y_i}\left(\nabla\bar{Y}_i(\hat{\underline{v}}_i)(\underline{v}_i(\rho) - \hat{\underline{v}}_i)\right)^2. & (39) \end{aligned}$$

Substituting (30) into (39) and expanding the vector inner product yields:

$$\Phi_q(\rho) = \sum_{i=1}^{N} \left\{ \sum_{k=1}^{K} \frac{\partial h_i}{\partial s_i^k}(\underline{\hat{v}}_i)(s_i^k(\rho) - \hat{s}_i^k) + \frac{1}{2Y_i} \left[\sum_{k=1}^{K} \frac{\partial \bar{Y}_i}{\partial s_i^k}(\underline{\hat{v}}_i)(s_i^k(\rho) - \hat{s}_i^k) \right]^2 \right\} \quad (40)$$

$$= \sum_{i=1}^{N} \left\{ \sum_{k=1}^{K} \nabla_k h_i(\underline{\hat{v}}_i) \left(s_i^k(\rho) - \hat{s}_i^k \right) + \frac{1}{2Y_i} \left[\sum_{k=1}^{K} \nabla_k \bar{Y}_i(\underline{\hat{v}}_i) \left(s_i^k(\rho) - \hat{s}_i^k \right) \right]^2 \right\}, \quad (41)$$

where ∇_k denotes the kth element of the gradient vector. To simplify the above equation, we make the following definitions:

$$\begin{aligned} Z_i &\triangleq \sum_{k=1}^{K} \nabla_k \bar{Y}_i(\underline{\hat{v}}_i) \hat{s}_i^k = \nabla \bar{Y}_i(\underline{\hat{v}}_i) \underline{\hat{v}}_i, \\ b_{ij} &\triangleq \sum_{k=1}^{K} \nabla_k \bar{Y}_i(\underline{\hat{v}}_i) a_{ij} r_j^k, \\ \mathbf{B} &\triangleq \sum_{k=1}^{K} \mathbf{D}\left(\nabla_k \bar{Y}_i(\underline{\hat{v}}_i) \right) \mathbf{A} \mathbf{D}\left(\underline{r}^k \right). \end{aligned}$$

Recall that $\mathbf{A} = \{a_{ij}\}$ is the geometrical system matrix. The matrix $\mathbf{B} = \{b_{ij}\}$ is a weighted system matrix, with the weights expressed as the non-zero elements of a diagonal matrix $\mathbf{D}(\cdot)$, to the left of $\mathbf{A}$. The term Z_i combines constants independent of ρ. With the above definitions, expressing the line integrals explicitly in terms of the image pixels yields the following form of the cost function:

$$\Phi_q(\rho) = \sum_{i=1}^{N} \left\{ \sum_{k=1}^{K} \nabla_k h_i(\underline{\hat{v}}_i) \left(\sum_{j=1}^{p} a_{ij} r_j^k (\rho_j - \hat{\rho}_j) \right) + \frac{1}{2Y_i} \left(\sum_{j=1}^{p} b_{ij} \rho_j - Z_i \right)^2 \right\} + \beta R(\rho). \quad (42)$$

This cost function is convex, so we can easily derive a separable surrogate and an iterative update, along the lines of Section 2. We state the results of the algorithm derivation below.

3.1.2. OS-PWLS Algorithm for Poly-Energetic CT

The separable paraboloidal surrogate for $\Phi_q(\rho)$ is given by:

$$\begin{aligned} Q(\rho; \rho^n) &= \sum_{j=1}^{p} \sum_{i=1}^{N} \sum_{k=1}^{K} \nabla_k h_i(\underline{\hat{v}}_i) a_{ij} r_j^k \rho_j - \sum_{i=1}^{N} \sum_{k=1}^{K} \nabla_k h_i(\underline{\hat{v}}_i) a_{ij} r_j^k \hat{\rho}_j \\ &\quad + \sum_{j=1}^{p} \sum_{i=1}^{N} \frac{1}{2Y_i} \alpha_{ij} \left(\frac{b_{ij}}{\alpha_{ij}} (\rho_j - \rho_j^n) + [\mathbf{B}\rho^n]_i - Z_i \right)^2 + \beta S(\rho; \rho^n) \end{aligned} \quad (43)$$

where we choose

$$\alpha_{ij} \triangleq \frac{b_{ij}}{\sum_{j=1}^{p} b_{ij}}.$$

Setting the point of linearization of the Taylor series at ρ^n, and evaluating the first and second derivative of Q at the same point gives:

$$\left. \frac{\partial Q(\rho; \rho^n)}{\partial \rho_j} \right|_{\rho=\rho^n} = \sum_{i=1}^{N} \sum_{k=1}^{K} a_{ij} r_j^k \nabla_k h_i(\underline{\hat{v}}_i) + \beta \left. \frac{\partial S}{\partial \mu_j} \right|_{\rho=\rho^n} \quad (44)$$

$$\left. \frac{\partial^2 Q(\rho; \rho^n)}{\partial \rho_j^2} \right|_{\rho=\rho^n} = \sum_{i=1}^{N} \frac{1}{Y_i} \frac{{b_{ij}}^2}{\alpha_{ij}} + \beta \left. \frac{\partial^2 S}{\partial \mu_j^2} \right|_{\rho=\rho^n}. \quad (45)$$

Here is the overall algorithm:

- initialize with $\hat{\rho}$.
- **for** each iteration $n = 1, ..., \text{niter}$
  - **for** each subset $S = 1, \ldots, M$
    - compute $\hat{s}_i^k = \sum_{j=1}^{p} a_{ij} r_j^k \hat{\rho}_j$ for $k = 1, \ldots, K$, $\underline{\hat{v}}_i = [\hat{s}_i^1, ..., \hat{s}_i^K]$
    - compute $\bar{Y}_i(\underline{\hat{v}}_i)$, its gradient vector $\nabla \bar{Y}_i(\underline{\hat{v}}_i)$ and $h_i(\underline{\hat{v}}_i)$
    - compute $b_{ij} = \sum_{k=1}^{K} \nabla_k \bar{Y}_i(\underline{\hat{v}}_i) a_{ij} r_j^k$
    - compute $d_j = \sum_{i \in S} \frac{1}{Y_i} \frac{b_{ij}^2}{\alpha_{ij}}$
    - compute $\hat{N}_j = \sum_{i \in S} \sum_{k=1}^{K} a_{ij} r_j^k \nabla_k h_i(\underline{\hat{v}}_i)$
    - compute

$$\hat{\rho}_j = \left[\hat{\rho}_j - \frac{M\hat{N}_j + \beta \left. \frac{\partial S}{\partial \rho_j} \right|_{\rho=\hat{\rho}}}{M d_j + \beta \left. \frac{\partial^2 S}{\partial \rho_j^2} \right|_{\rho=\hat{\rho}}} \right]_+ \quad j = 1, \ldots, p \tag{46}$$

  - **end**
- **end**

We expect this algorithm to globally converge to the minimizer of the cost function $\Phi_q(\rho)$ when one subset is used, provided the penalty is chosen so that Φ_q is strictly convex. When two or more subsets are used, we expect it to be monotone in the initial iterations.

4. PRELIMINARY RESULTS

We now present results for the reconstruction algorithms outlined earlier. A preliminary comparison between FBP and OS-PWLS shows the superiority of the latter in terms of noise reduction for mono-energetic X-ray CT. Poly-energetic reconstruction results show that our model is effective in eliminating most beam hardening effects. Some artifacts persist, perhaps due to deficiencies in the Taylor approximation (35).

We assume a fan-beam system geometry with a source-to-detector distance of 1000 mm. The source-to-isocenter distance is 500 mm and the object is centered at isocenter. The field of view is 500 mm and the fan angle is 90^o. We assume a flat detector with a uniform response for the detector bins. The background r_i is set to zero.

4.1. Comparison with FBP

We compare reconstructions from FBP and OS-PWLS (5 iterations, 10 subsets). We use data from the simulated thorax phantom shown on the left in figure 1. The regions A (heart) and B (spine) ideally have uniform values and we compare noise behavior therein.

The noisy sinogram has size 250×250 and the reconstructed image is 128×128 pixels at a resolution of 3.91 mm/pixel. The regularization term in OS-PWLS is a Huber penalty, with $\beta = 120$ and $\delta = 0.0005$. We deduced the values for the parameters β and δ by trial and error. They give comparable resolution to FBP. Figure 2 shows the FBP and OS-PWLS images. The FBP image is used to initialize OS-PWLS.

We compute the standard deviation in regions A and B. The number of pixels used to compute the standard deviation in each region is almost 10×10. Since the two reconstructions have comparable resolution, the noise comparison is justified. Overall, the results show superior noise reduction for the OS-PWLS case. After 2 iterations, the standard deviation is reduced

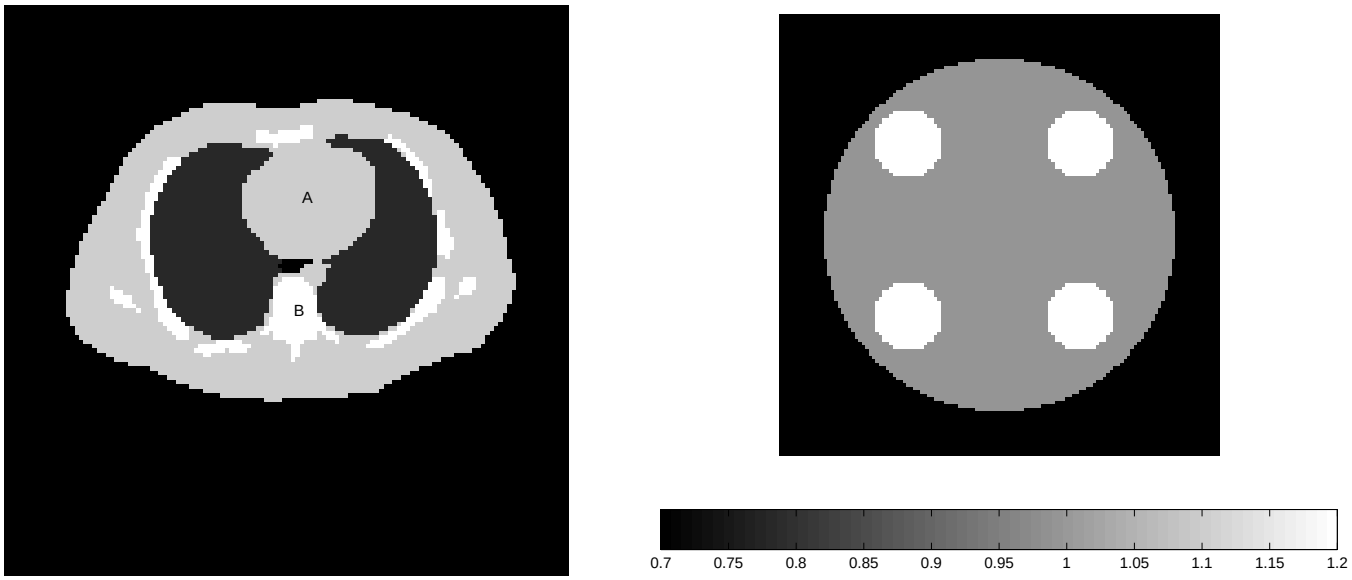

Figure 1. Thorax phantom (left) and bone/water phantom (right)

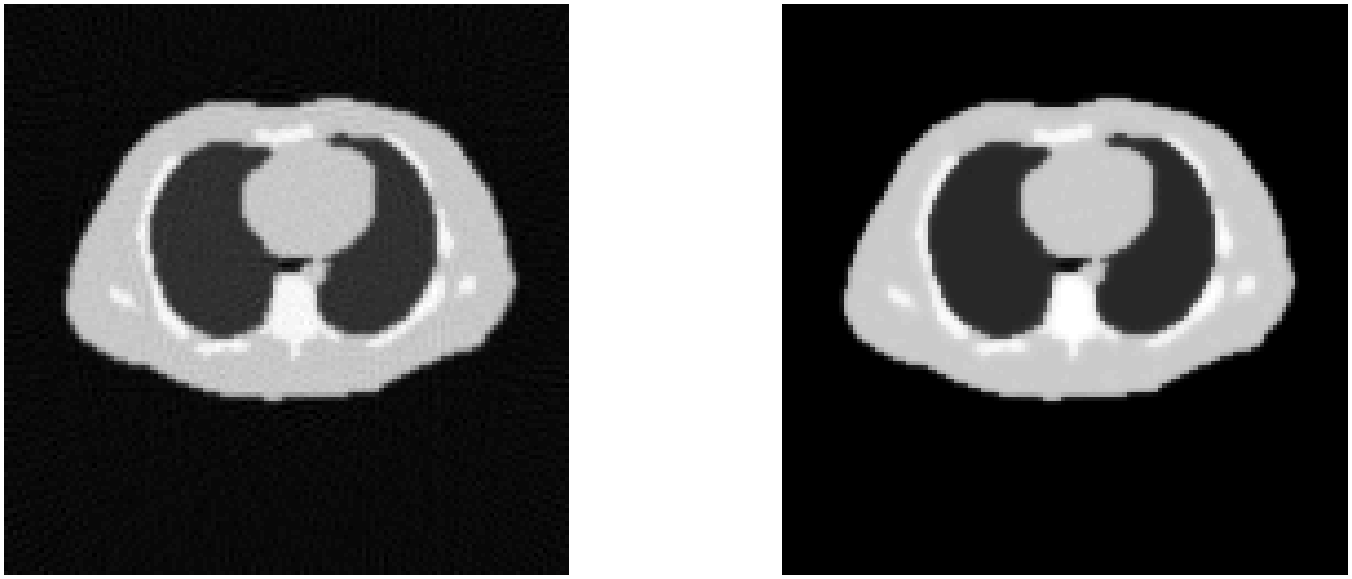

Figure 2. FBP reconstruction (left) and OS-PWLS reconstruction (right)

by 60% in the heart and 49% in the spine. After 5 iterations, the standard deviation is reduced by 78% and 53% in the heart and spine, respectively.

The plots in figure 3 illustrate algorithm convergence. The left plot illustrates the fact that the quadratic approximation to the likelihood is indeed a good one. On the right, we plot the objective function at every iteration with and without subsets. Subsets clearly offer significant acceleration.

4.2. Beam Hardening Correction

To test our poly-energetic model algorithm, we use the density phantom shown on the right in figure 1 and simulate transmission data with a poly-energetic spectrum with mean of 68 keV and standard deviation of 16 keV. Four high-density circles ('bone', $\rho = 2$ gm/cm^3) are immersed in water ($\rho = 1$ gm/cm^3), which is surrounded by air ($\rho = 0$ gm/cm^3). The phantom size is 128×128 and we use the geometry described earlier. We simulate a noise-free sinogram of size 150×150. The object is classified into bone or soft tissue by segmenting the original image using a density threshold of 1.2. The left image in figure

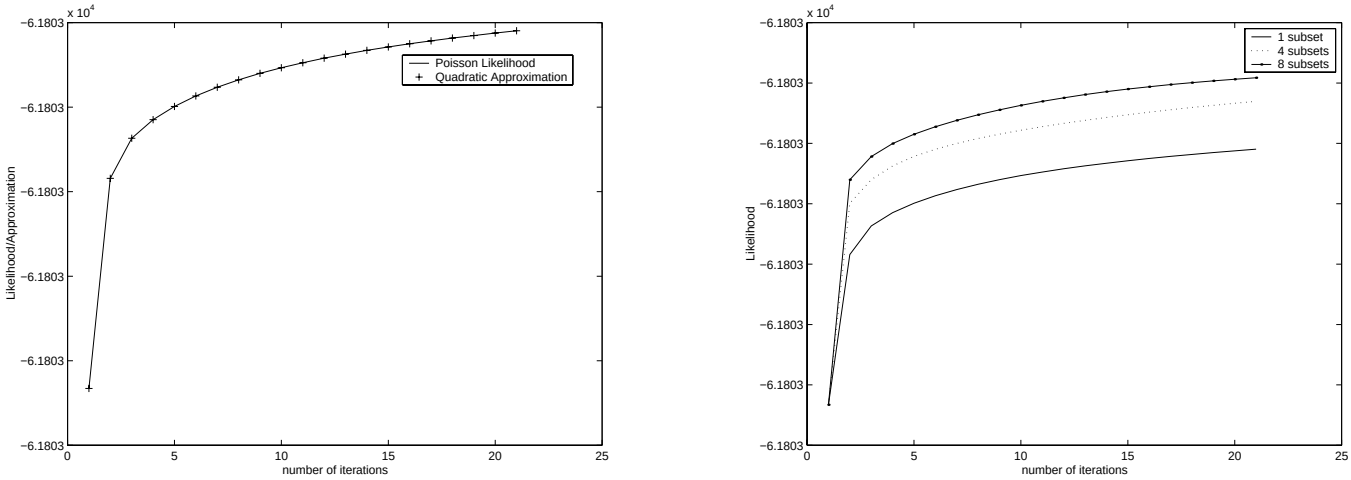

Figure 3. Quadratic approximation to the Poisson likelihood (left) and effect of subsets on convergence (right)

4 is reconstructed with FBP without any beam hardening correction. The right image in the same figure illustrates the result of using the mono-energetic OS-PWLS algorithm with poly-energetic data. Figure 5 shows the image reconstructed with the algoritm derived from the poly-energetic model. Both OS-PWLS results were reconstructed using 10 iterations and 4 subsets. OS-PWLS.

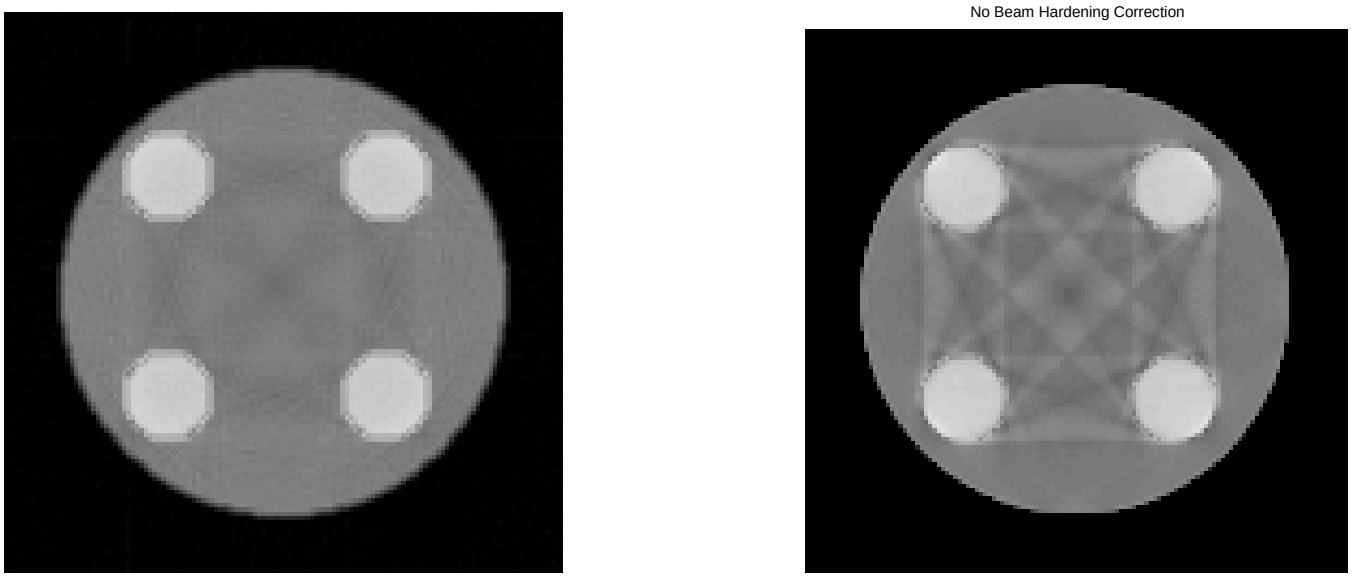

Figure 4. Uncorrected FBP (left) and OS-PWLS (right) images

The (FBP and OS-PWLS) images reconstructed without correction exhibit typical beam hardening artifacts. There is an overall reduction in the values of the reconstructed density for both high and low-density materials. Also, dark streaks caused by the high density circles dominate the central region of the image. The poly-energetic model OS-PWLS adequately addresses the reduction in density values. Some streaks persist, however, and we conjecture that they maybe due to the Taylor approximation.

With the knowledge of the mass attenuation coefficient, one can scale the resulting density images to obtain attenuation coefficient information at any energy using (24).

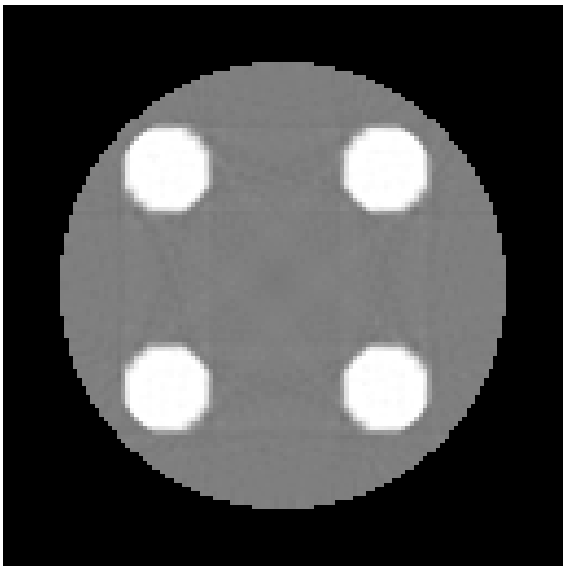

Figure 5. Corrected OS-PWLS reconstruction

5. CONCLUSIONS

We described a statistical iterative reconstruction method for X-ray CT and illustrated that it is superior to FBP in terms of noise reduction. The method is also generalized for energy dependent X-ray attenuation and produces images with significantly reduced beam hardening artifacts for a bone and soft tissue object.

The beam hardening artifacts were not completely eliminated. This perhaps is due to the limitations of the Taylor approximation in (35) and is the subject of further investigation.

Future work will include applying the poly-energetic approach to objects with three or more tissue types. Also, the current approach requires prior knowledge of the spatial distributions of the different tissues. Although this can be approximated from a good FBP image, it is desirable to jointly estimate the object density and tissue classes.[21] It is also necessary to carry out a comparison between our method and more conventional post-processing beam hardening correction algorithms.[18]

ACKNOWLEDGEMENTS

This work was supported in part by NIH grant R01CA/HL60711-06.

REFERENCES

1. J. A. Fessler, "Hybrid poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Transactions on Image Processing* **4**, pp. 1439–1449, October 1995.
2. J. Li, R. J. Jaszczak, K. L. Greer, and R. E. Coleman, "Implementation of an accelerated iterative algorithm for cone-beam SPECT," *Physics in Medicine and Biology* **39**, pp. 643–653, 1994.
3. J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized likelihood transmission image reconstruction," *IEEE Transactions on Medical Imaging* **16**, pp. 166–175, April 1997.
4. H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Transactions on Medical Imaging* **18**, pp. 801–814, September 1999.
5. C. Kamphius and F. J. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Transactions on Medical Imaging* **17**, pp. 1101–1105, December 1998.
6. H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Physics in Medicine and Biology* **44**, pp. 2835–2851, 1999.
7. H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transactions on Medical Imaging* **13**, pp. 601–609, December 1994.

8. D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White, "Compensation for read-out noise in charge-coupled-device images," *Journal of the Optical Society of America* **12**, pp. 273–83, February 1995.
9. K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *Journal of Computer Assisted Tomography* **8**, pp. 306–316, April 1984.
10. H. Erdogan and J. A. Fessler, "Accelerated monotonic algorithms for transmission tomography," *ICIP* **2**, pp. 680–684, 1998.
11. N. Srinivasa, K. R. Ramakrishnan, and K. Rajgopal, "Adaptive noise canceling in computed tomography," *Proc. of the Symposium of the Engineering of Computer-based Medical Systems* , pp. 65–68, June 1988.
12. H. P. Hiriyannaiah, W. E. Snyder, and G. L. Bilbro, "Noise in reconstructed images in tomography: parallel, fan and cone beam projections," *Proc. of $3rd$ Annual IEEE Symposium on Computer-based Medical Systems* , pp. 81–88, June 1990.
13. K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Transactions on Signal Processing* **41**, pp. 553–548, February 1993.
14. A. R. DePierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," *IEEE Transactions on Medical Imaging* **12**, pp. 38–333, June 1993.
15. A. R. DePierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Transactions on Medical Imaging* **14**, pp. 132–137, March 1995.
16. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, 1988.
17. R. A. Brooks and G. D. Chiro, "Beam hardening in X-ray reconstruction tomography," *Physics in Medicine and Biology* **21**(3), pp. 390–398, 1976.
18. P. M. Joseph and R. D. Spital, "A method for correcting bone induced artifacts in computed tomography scanners," *Journal of Computer Assisted Tomography* **2**, pp. 100–108, January 1978.
19. P. Sukovic and N. H. Clinthorne, "Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography," Submitted to the IEEE Transactions on Medical Imaging.
20. R. E. Alvarez and A. Macovski, "Energy-selective reconstruction in X-ray computerized tomography," *Physics in Medicine and Biology* **21**(5), pp. 733–744, 1976.
21. Y. Zhang, J. A. Fessler, N. H. Clinthorne, and W. L. Rogers, "Joint estimation for incorporating MRI anatomic images into SPECT reconstruction," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 1256–1260, 1994.