Comparison of Quadratic- and Median-Based Roughness Penalties for Penalized-Likelihood Sinogram Restoration in Computed Tomography

Patrick J. La Rivièere, Junguo Bian, and Phillip A. Vargas

Department of Radiology, The University of Chicago, Chicago, IL 60637, USA

Received 13 December 2005; Revised 3 March 2006; Accepted 5 March 2006

We have compared the performance of two different penalty choices for a penalized-likelihood sinogram-restoration strategy we have been developing. One is a quadratic penalty we have employed previously and the other is a new median-based penalty. We compared the approaches to a noniterative adaptive filter that loosely but not explicitly models data statistics. We found that the two approaches produced similar resolution-variance tradeoffs to each other and that they outperformed the adaptive filter in the low-dose regime, which suggests that the particular choice of penalty in our approach may be less important than the fact that we are explicitly modeling data statistics at all. Since the quadratic penalty allows for derivation of an algorithm that is guaranteed to monotonically increase the penalized-likelihood objective function, we find it to be preferable to the median-based penalty.

Copyright © 2006 Patrick J. La Rivièere et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

We have recently developed penalized-likelihood approaches to the problems of sinogram smoothing and sinogram restoration in computed tomography [1–4], with a particular eye to the low-dose regime being considered for screening studies for lung and colon cancer [5–9]. In both cases, we assume that the statistics of each detector measurement are given by the sum of a compound Poisson term, representing the photon counting statistics for polychromatic photons, and a Gaussian term, representing electronic noise. In the case of sinogram restoration, we generalize the measurement model to include blurring coefficients representing sinogram degradations such as off-focal radiation, detector afterglow, and detector crosstalk [2, 4]. From the noisy, degraded measurements, we then seek to estimate a set of “ideal,” undegraded line integrals by iteratively maximizing an objective function comprising a sum of a simple Poisson likelihood (an approximation to the measurement statistics assumed above) and a roughness penalty. The estimated line integrals can then be fed into an existing analytic reconstruction algorithm, such as those typically implemented in hardware on commercial CT scanners. The hope is that this iterative sinogram-domain approach would provide some of the statistical advantages of fully iterative approaches to image reconstruction at a lower computational cost.

In our previous studies of the smoothing and restoration approaches, we have made use of a quadratic roughness penalty applied in the line integral (log) domain to the difference between a given sample in sinogram space and its four adjoining neighbors, that is, between a given detector channel and its two neighboring channels (we assumed a single-row detector), as well as to its own reading at the preceding and following view angles. While the approaches performed better in resolution-variance studies at low doses than did Hsieh’s noniterative adaptive trimmed mean (ATM) filter [10], the ATM filter was surprisingly effective at reducing the influence of a small number of very noisy measurements without unduly compromising resolution. The ATM filter is only applied to measurements whose signal strength falls below a certain threshold, and it entails replacing the value in question with the trimmed mean of the values in a neighborhood of the sinogram around the measurement in question. The trimmed mean filter is a median-like filter based on order statistics, and it varies adaptively between applying a true median filter and a simple boxcar filter, depending on the signal level. The relatively strong performance of this filter suggested that it would be worthwhile to explore the use of a median-based roughness penalty in the context of the sinogram smoothing and restoration methods we have developed.
The use of median-based roughness penalties in fully iterative penalized-likelihood image reconstruction was pioneered by Alenius et al., who referred to them as median root priors (MRPs) [11, 12]. Unfortunately, it does not appear to be possible to derive an iterative algorithm that is guaranteed to monotonically increase an objective function based on the MRP of Alenius et al. However, they derive a heuristic update equation, based on Green’s one-step-late (OSL) strategy [13], that does not necessarily correspond to the maximization of a predefined objective function, but that does yield good results in practice.

In this work, we explore the use of a standard MRP penalty like those of Alenius et al. in the context of our sinogram-restoration approach and we make use of the heuristic, OSL strategy to derive the iterative update. We compare the images qualitatively and quantitatively to those obtained by use of our method with quadratic roughness penalties, as well as to those obtained by use of Hsieh’s ATM filter.

2. METHODS

2.1. Measurement model

We assume that the CT scan produces a set of measurements that are represented as a one-dimensional (1D) vector $y_{\text{meas}}$, with elements $y_{i,\text{meas}}$, where $N_y$ is the total number of measurements in the scan, and the index $i$ denotes a particular attenuation line through the patient (i.e., a specific combination of detector channel, detector row, and projection angle).

To review the model, we have been employing [2–4], we assume, when we simulate data, that each $y_{i,\text{meas}}$ is a realization of a random variable $Y_{i,\text{meas}}$ whose statistics are described by

$$y_{i,\text{meas}} \sim \text{Poisson} \left( \sum_{j=1}^{N_y} b_{ij} I_j \lambda_m^{(j)} \right) + \text{Normal} \left( d_i, \sigma_i^2 \right),$$

(1)

with the various terms in this equation defined in Table 1. The compound Poisson distribution in the first term has been derived and validated by Whiting [14] and rederived by Elbakri and Fessler [15, 16]. We assume that $I_j, G_i, d_i, \sigma_i^2$, the average energy $E_j = \sum_{m=1}^{M} E_m \lambda_m^{(j)}$ of the incident beam, and an energy-averaged and normalized estimated scatter term $s_i = (1/E_j) \sum_{m=1}^{M} E_m s_m^{(i)}$ are all known.

Our goal is to estimate a set of ideal, “monochromatic” attenuation line integrals:

$$I_i^{(\text{mono})} = \int_{L_i} \mu(x, E_r) d l,$$

(2)

$$i = 1, \ldots, N_y,$$

at some reference energy $E_r$ (usually $E_j$), from the set of measurements $y_{i,\text{meas}}$, $i = 1, \ldots, N_y$. These estimated line integrals can then be input to a standard analytic reconstruction algorithm as mentioned above.

Our strategy for estimating the line integrals entails maximizing a penalized-likelihood objective function. Because the model of (1) does not yield a tractable likelihood, we approximate it by defining a vector $y$ of new adjusted measurements with elements

$$y_i = \left\lfloor \left( \frac{y_{i,\text{meas}} - d_i}{E_j G_i} \right) + \frac{\sigma_i^2}{G_i^2 E_j^2} \right\rfloor,$$

(3)

where $\lfloor x \rfloor$ is $x$ for positive $x$ and zero otherwise, that are realizations of random variables $Y_i$ which we assume are approximately Poisson-distributed:

$$Y_i \sim \text{Poisson} \left\{ \sum_{j=1}^{N_y} I_j b_{ij} e^{-I_j^{(\text{poly})}} + s_i + \frac{\sigma_i^2}{G_i^2 E_j^2} \right\},$$

(4)

with $f_j(l)$ being an empirically determined function, typically polynomial, that adequately captures the effect of beam hardening in slices that do not contain substantial amounts of bone [17]. Our strategy is then to estimate the vector $I^{(\text{poly})}$, with elements $I_j^{(\text{poly})}$, from the vector of adjusted measurements $y$, since the needed $I_j^{(\text{mono})}$ can then be obtained by inverting (5). For simplicity, we drop the (poly) superscripts from $I^{(\text{poly})}$ in what follows.

2.2. Quadratic penalty approach

Our general strategy thus far [2–4] has been to maximize a penalized-likelihood objective function

$$\Phi(I; y) = \mathcal{L}(I; y) - \beta R(I),$$

(6)

where

$$\mathcal{L}(I; y) = \sum_{i=1}^{N_y} y_i \log \left[ \sum_{j=1}^{N_y} I_j b_{ij} e^{-I_j} + r_i \right] - \left[ \sum_{j=1}^{N_y} I_j b_{ij} e^{-I_j} + r_i \right],$$

(7)

is the Poisson log-likelihood for the random variables of (4) and where we have defined $r_i = s_i + \sigma_i^2/(G_i^2 E_j^2)$.

The roughness penalty $R(I)$ can be expressed in a general form as

$$R(I) = \sum_{k=1}^{K} \psi_k \left( \sum_{j=1}^{N_y} t_{kj} I_j \right),$$

(8)

given by Fessler [18], where $\psi_k$ is a potential function that assigns a cost to the $K$ combinations of attenuation line integral values represented by the linear combinations $\sum_{j=1}^{N_y} t_{kj} I_j$.

Our quadratic penalty approach entails choosing $\psi_k(t) = \omega_k t^2/2$ and constructing the $t_{kj}$ to create differences of a sinogram sample with its horizontal and vertical neighbors, with
\[ \omega_k = 1/2 \] for those neighbors. This is equivalent to

\[ R(I) = \frac{1}{4} \sum_{j} \sum_{k \in N_j} (l_j - l_k)^2, \tag{9} \]

where \( N_j \) denotes the neighborhood comprising the 4 nearest horizontal and vertical neighbors of measurement \( j \).

We have derived an algorithm that generates a sequence of estimates of \( I \) that are guaranteed to increase the objective function of (6) by making use of the optimization transfer principle [18], in which at each iteration one defines a surrogate to the likelihood function, such that the vector of line integrals maximizing this surrogate is guaranteed to have a higher penalized likelihood than the previous vector estimate. The resulting update is

\[ l_j^{(n+1)} = \left[ l_j^{(n)} - n_j + \beta \sum_{k=1}^{\hat{N}_j} \sum_{l=1}^{\hat{N}_k} t_j t_k \omega_k l_j^{(n)} \right] c_j^{(n)} + \beta v_j, \tag{10} \]

where the notation \( l_j^{(n)} \) denotes the estimate of \( l_j \) after the \( n \)th iteration, Here,

\[ n_j = \sum_{i=1}^{\hat{N}_j} b_{ij} \hat{g}_i \left( \sum_{i=1}^{\hat{N}_j} b_{ij} e^{-y_i^{(n)}} + r_j \right) e^{-y_i^{(n)}}, \tag{11} \]

and the \( c_j^{(n)} \) are the curvatures of paraboloidal surrogates constructed to give rise to an overall, easy-to-maximize quadratic surrogate to the objective function. One choice for the curvatures that guarantee monotonicity is

\[ c_j^{(n)} = l_j \sum_{i=1}^{\hat{N}_j} b_{ij}, \tag{13} \]

although in practice we make use of a different set of curvatures that do not guarantee monotonicity but that in practice lead to faster convergence [19].

### 2.3. Median penalty

For the median penalty approach, we employ a more heuristic approach based on producing a sequence of estimates that, in the absence of a penalty, would yield a maximum-likelihood estimate, but that incorporates a penalty term in each iteration that discourages deviations from the local median of the last iteration. This is consistent with the approach of Alenius et al. in fully iterative reconstruction [11, 12]. Specifically, the update is defined as

\[ l_j^{(n+1)} = \text{arg max}_{l_j \geq 0} \left[ S(I, I^{(n)}) - \beta R^{(\text{med})}(I, \hat{m}^{(n)}) \right], \tag{14} \]

where \( S(I, I^{(n)}) \) is a surrogate to the log likelihood to be described below and

\[ R^{(\text{med})}(I, \hat{m}^{(n)}) = \frac{1}{2} \sum_{k=1}^{\hat{N}_k} \frac{(l_k - \hat{m}_k^{(n)})^2}{\hat{m}_k^{(n)}} \tag{15} \]

is the median penalty, with

\[ \hat{m}_k^{(n)} = \text{Median} \left( I^{(n)}, \hat{m}_k^{(n)} \right) \tag{16} \]

denoting the median of the values of \( I^{(n)} \) in some neighborhood \( \hat{N}_k \) around value \( k \).

The surrogate \( S(I, I^{(n)}) \) is defined as

\[ S(I, I^{(n)}) = \sum_{k=1}^{\hat{N}_k} S_k(l_k, I^{(n)}), \tag{17} \]

where

\[ S_k(l_k, I^{(n)}) = \sum_{i=1}^{\hat{N}_j} \frac{l_k b_{ik} e^{-y_i^{(n)}}}{\bar{y}_i^{(n)}} \bar{g}_i \left( \frac{e^{-y_i^{(n)}}}{\bar{y}_i^{(n)}} \right) \bar{y}_i^{(n)} g_i \left( \bar{y}_i^{(n)} \right), \tag{18} \]

with

\[ \bar{y}_i^{(n)} = \sum_{k=1}^{\hat{N}_j} l_k b_{ik} e^{-y_i^{(n)}} + r_i, \tag{19} \]
and \( g(x) \equiv y_i \log x - x \). This surrogate satisfies both \( S(1^{(n)}) \leq L(1) \), \( \forall l_i \geq 0 \) and \( S(1^{(n)}, 1^{(n)}) = L(1^{(n)}) \), and thus in the absence of the penalty term, finding the \( l_i \) maximizing this surrogate necessarily increases the likelihood [4].

Substituting (17) and (15) into (14), our update is given by

\[
l_{j}^{(n+1)} = \arg \max_{l_j \neq 0} \left\{ \frac{N_y}{k=1} S_k \left( k, I^{(n)} \right) - \beta \frac{1}{2} \sum_{k=1}^{N_y} \left( \frac{l_i - \hat{m}_i^{(n)}}{\hat{m}_i^{(n)}} \right)^2 \right\}.
\]

(20)

We can solve for the maximum by setting the derivative with respect to \( l_j \) equal to zero. Doing so yields

\[
-\sum_{i=1}^{N_y} l_j b_{ij} \left( e^{-l_j y_i^{(n)}} - e^{-l_j y_i^{(n)}} \right) = 0.
\]

(21)

Solving for \( l_j \) yields the update

\[
l_{j}^{(n+1)} = \left[ -\ln \left( \frac{\sum_{i=1}^{N_y} l_j b_{ij} \left( y_i - y_i^{(n)} \right) e^{-l_j y_i^{(n)}}}{\sum_{i=1}^{N_y} l_j b_{ij}} \right) \right].
\]

(22)

3. RESULTS

3.1. Qualitative results

To compare the two penalties to each other as well as to the ATM filter, we simulated projections of a numerical ellipse phantom shown in Figure 1 that we have used previously and that is modeled on the physical phantom employed by Hsieh [10]. We computed a sinogram of 1024 angles \( \times 1024 \) bins of extent 0.5 mm at the isocenter with a source-to-isocenter distance of 540.0 mm. We simulated the data according to the forward model of (1) with discretization of a realistic CT spectrum into 1 keV bins. We assumed that the phantom was water-equivalent except for the three circular structures, which we assumed to be bone.

We simulated two exposure levels: a clinically typical \( I_i = 2.5 \times 10^6 \) and a low-dose level \( I_i = 2.5 \times 10^5 \). We chose \( G_i = 3.57 \times 10^{-3} \) pA/keV, such that \( G_i E_i = 0.25 \) pA/quanta for \( E_i = 70 \) keV, and \( \sigma_i^2 = 10.0 \) pA\(^2\) for all \( i \). We included the effects of off-focal radiation in the simulation by convolving with kernels having 13 nonzero values arrayed diagonally with slope \(-2\) in discrete sinogram space. The central value (corresponding to the zero point of the kernel) had relative value 1.0 and the six values on either side had relative value 0.02 (these thirteen values were normalized so that their sum was 1.0). We found through simulations with water-equivalent phantoms that the beam-hardening effect of the simulated tube spectrum was well represented as a second-order polynomial \( f(l) = l - 0.007l^2 \).

We reconstructed images by means of four different approaches.


2. ATM. Presmoothing of data by means of the ATM filter, followed by simple discrete deconvolution of the effect of off-focal radiation, beam-hardening correction, and reconstruction by FFBP with an unapodized ramp filter. The ATM filter was implemented with cut-off parameter \( \lambda = 75.0 \) pA and baseline parameter \( \delta = 0.05 \) pA, as in [10] by Hsieh, and with the filter length varying from 3 to 19.

3. Quadratic. Penalized-likelihood sinogram restoration using the quadratic neighborhood penalty followed by beam-hardening correction and reconstruction by FFBP with an unapodized ramp filter. The smoothing parameter \( \beta \) was varied from 0.01 to 50.

4. Median. Penalized-likelihood sinogram restoration using the quadratic neighborhood penalty followed by beam-hardening correction and reconstruction by FFBP with an unapodized ramp filter. The smoothing parameter \( \beta \) was varied from 0.01 to 50. The size of the neighborhood used in median calculation was \( 3 \times 3 \).

Figure 2 shows typical results of these reconstructions, where we have selected values of the smoothing parameters for the ATM and penalized-likelihood images that give approximately matched resolution at the center insert and approximately matched noise levels at the right insert, based on the resolution-variance results to be described below. It can be
seen that the noise level in the low-dose data leads to severe streaking artifacts in the Hanning filter reconstruction and that these are suppressed by the approaches under consideration, more so for the penalized-likelihood approaches than for the ATM approach. The results all appear to perform similarly at the higher-dose level. These qualitative impressions were explored quantitatively by use of resolution-variance studies.

3.2. Resolution-variance tradeoffs

To characterize resolution, we determined the local edge-spread function at the central and right high-attenuation inserts. The vertical profiles through these structures have profiles that are well fit by error functions parametrized by width $\sigma_b$ which implies that the effective blurring kernel is Gaussian with standard deviation $\sigma_b$. We employ the FWHM of the Gaussian, 2.35 $\sigma_b$, as our measure of resolution. To obtain an accurate fit, we performed targeted reconstructions of the central and right high-attenuation inserts with 0.25 mm pixel size from 10 different noise realizations. We then averaged the reconstructions together to obtain relatively low-noise profiles on which to perform the fitting. We characterized noise by calculating the average standard deviation of the pixel values in circular regions of interest (ROIs) of diameter 16.0 mm placed adjacent to, but not overlapping, the central and right high-attenuation inserts. We then plotted the resulting noise measure versus the resolution measure for the same location. Images reconstructed after smoothing with different values of the smoothing parameters $\alpha$ or filter length $\beta$ provide different combinations of such values and allowed us to sweep out a resolution-noise curve.

The results are given in Figure 3, where it can be seen that at the low-dose level, the two penalized-likelihood-based approaches both outperform the ATM filter in terms of resolution-noise performance. They perform very similarly to each other, with perhaps a slight advantage to the quadratic penalty. At high-dose levels, the approaches all perform relatively similarly.

4. CONCLUSIONS

We have compared two different penalty choices for a penalized-likelihood sinogram-restoration strategy we have been developing, one is a quadratic penalty we have employed previously and the other is a median-based penalty. We compared the approaches to a noniterative adaptive filter that loosely but not explicitly models data statistics. We found that the two approaches produced very similar resolution-variance tradeoffs to each other and that they outperformed the ATM filter in the low-dose regime, which suggests that the particular choice of penalty in our approach may be less important than the fact that we are explicitly modeling data statistics at all.

It is not possible to conclude, of course, that the penalized-likelihood approaches would outperform any noniterative adaptive filter. In generating the resolution-variance tradeoffs for the ATM filter, we made use of the parameters $\lambda$ and $\delta$ given by Hsieh in describing the filter in [10] and varied the filter length parameter $\beta$. The filter length $\beta$ is the most natural parameter to vary in sweeping out resolution-variance curves, but it is possible that further adjusting the parameters $\lambda$ and $\delta$ could further improve the achievable tradeoffs. In particular, it is possible that the version of the ATM filter implemented on GE scanners has been optimized beyond what was presented in [10].

Since the quadratic penalty allows for derivation of a monotonic algorithm guaranteed to increase the likelihood function while the median filter approach offers no such guarantee, we find it to be preferable to the median-based penalty. However, it might be worthwhile to explore the new class of median-like prior proposed by Hsiao et al. that does indeed involve maximization or minimization of a joint objective function involving the image of interest and an...
auxiliary field derived from the local medians of the image [20]. Another possibility for future work would be to explore a penalty that uses Hsieh’s ATM filter as an OSL prior much as the median prior was used here.

ACKNOWLEDGMENT

This work was supported by the Schweppe Foundation.

REFERENCES


Patrick J. La Rivière et al.

Science Symposium and Medical Imaging Conference (NSS-MIC ’04), vol. 5, pp. 3303–3307, Rome, Italy, October 2004, CD-ROM.


Patrick J. La Rivière received the A.B. degree in physics from Harvard University in 1994 and the Ph.D. degree from the Graduate Programs in Medical Imaging in the Department of Radiology at The University of Chicago in 2000. He is currently an Assistant Professor in the Department of Radiology at The University of Chicago, where his research interests include tomographic reconstruction in computed tomography, X-ray fluorescence computed tomography, and optoacoustic tomography.

Junguo Bian received the B.E. degree from Department of Engineering Physics at Tsinghua University, China, in 1998, and the M.S.N.E. degree from the Department of Nuclear Engineering at Purdue University in 2004. He is currently a Ph.D. student in the Graduate Programs in Medical Physics in the Department of Radiology at The University of Chicago.

Phillip A. Vargas received the B.S. degree in applied physics from Purdue University in 2002. He is currently a Research Scientist in the Department of Radiology at the University of Chicago, where his research interests include tomographic reconstruction in computed tomography, X-ray fluorescence computed tomography, and optoacoustic tomography.