PENALIZED LIKELIHOOD TRANSMISSION IMAGE RECONSTRUCTION : UNCONSTRAINED MONOTONIC ALGORITHMS

Somesh Srivastava, Jeffrey A. Fessler *

EECS Dept., The University of Michigan
someshs@umich.edu, fessler@umich.edu

ABSTRACT

Statistical reconstruction algorithms in transmission tomography yield improved images relative to the conventional FBP method. The most popular iterative algorithms for this problem are the conjugate gradient (CG) method and ordered subsets (OS) methods. Neither method is ideal. OS methods "converge" quickly, but are suboptimal for problems with factored system matrices. Nonnegativity constraints are not imposed easily by the CG method. To speed convergence, we propose to abandon the nonnegativity constraints (letting the regularization discourage the negative values), and to use application-specific quadratic surrogates to choose the step size rather than using an expensive general-purpose line search. To ensure monotonicity, we develop a modification of the transmission log-likelihood. The resulting algorithm is suitable for large-scale problems with factored system matrices such as X-ray CT image reconstruction with afterglow models. Preliminary results show that the regularization ensures minimal negative values, and that the algorithm is indeed monotone.

1. INTRODUCTION

The dominant technique for X-ray computed tomography(CT) image reconstruction has been the filtered backprojection (FBP) algorithm. It is fast (FFT is used in its implementation), deterministic and its properties are well understood. In transmission tomography, for scans with low counts or those contaminated with significant background counts, the FBP method leads to attenuation maps with systematic biases due to the nonlinearity of the logarithm [1]. Even statistical techniques that intrinsically assume Gaussian noise lead to systematic biases. Thus, Poisson-like measruement statistics cannot be ignored. Statistical methods using the Poisson likelihood with suitable regularization have shown good performance [1–3].

The importance of compute time of the algorithms in commercial scanners cannot be over-emphasized. Incorporation of statistical methods became possible with the use of ordered subsets expectation maximization (OSEM) algorithms in PET scanners starting from 1997. Since typical clinical CT images are of sizes 512x512 or larger, statistical algorithms require very long compute times. So, there is a need to make algorithms faster for large image sizes.

Various methods have been proposed to accelarate iterative image reconstruction algorithms. One of the most important ones is the ordered subsets method (also known as the incremental gradient or the block iterative method) [4–6]. In these methods only a subset of projection views are used each "sub-iteration". Using suitable ordered subsets in the initial phases accelarates convergence by a factor equal to the number of the subsets used. Without relaxation these algorithms usually do not converge [7] and they are not monotonic.

However, OS methods are poorly matched to problems where system matrices are in a factored form, for example in X-ray CT scanners where detector afterglow is significant [8]. So, alternative accelaration methods are needed for such problems. Pre-conditioned conjugate gradient (PCG) algorithms have many desirable properties over OS algorithms like monotonicity (with a suitable line search) and capability of handling factored system matrices. They are however slower than OS in the initial iterations but not as slow as the older methods. Compared to coordinate descent algorithms, imposition of the non-negativity constraint on the attenuation constant in PCG algorithms is more difficult. Naive imposition of non-negativity constraint would cause PCG algorithms to lose their monotonicity property.

In this paper, we describe a method that preserves the monotonicity of the PCG algorithm while encouraging non-negativity constraint on the attenuation coefficients to some extent by suitably modifying the Transmission Penalized Likelihood (TPL) cost function. For simplicity, we handle the mono-enegetic case here; in the future this idea can be extended to the poly-energetic case.

2. THEORY

In transmission tomography, the means of the data are related exponentially to the projections (or line integrals) of

*Supported in part by NIH Grant R01 CA60711 and P01 CA59827.

the attenuation map through Beer's Law [9]. In addition, the measurements are contaminated by extra background counts, due mostly to random coincidences and scatter in PET and emission crosstalk in SPECT. Thus, we assume the following model (using the notation of [2]):

$$y_i \sim \mathsf{Poisson}\Big\{b_i\,\mathrm{e}^{-[\boldsymbol{A\mu}]_i} + r_i\Big\},\ i = 1,\ldots,N, \quad (1)$$

where N is the number of measurements, μ_j is the average linear attenuation coefficient in voxel j for $j = 1,\ldots,p$, and p denotes the number of voxels. The notation $[\boldsymbol{A\mu}]_i \triangleq \sum_{j=1}^{p} a_{ij}\mu_j$ represents ith line integral of the attenuation map $\boldsymbol{\mu}$ and $\boldsymbol{A} = \{a_{ij}\}$ is the $N \times p$ system matrix. We assume b_i, r_i and a_{ij} are known non-negative constants, where r_i is the mean number of background events, b_i is the blank scan factor and y_i represents the number of transmission events counted by the ith detector (or detector pair in PET).

Ideally, we seek to find a statistical estimate of the attenuation map $\boldsymbol{\mu}$ that agrees with the data and is physically reasonable, *i.e.* its elements are all non-negative. As in [2], we form a penalized-likelihood cost function $\Phi(\boldsymbol{\mu})$ and denote our estimate of the linear attenuation coefficients as

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &\triangleq \arg\min_{\boldsymbol{\mu}\geq\mathbf{0}} \Phi(\boldsymbol{\mu}), && (2)\\
\text{where } \Phi(\boldsymbol{\mu}) &= -L(\boldsymbol{\mu}) + \beta R(\boldsymbol{\mu})\\
-L(\boldsymbol{\mu}) &= \sum_{i=1}^{N} h_i([\boldsymbol{A\mu}]_i) && (3)\\
h_i(\ell) &= (b_i e^{-\ell} + r_i) - y_i \log\big(b_i e^{-\ell} + r_i\big) && (4)\\
R(\boldsymbol{\mu}) &= \sum_{j=1}^{p} \frac{1}{2} \sum_{k=1}^{p} w_{jk}\psi(\mu_j - \mu_k), && (5)
\end{aligned}$$

where $L(\boldsymbol{\mu})$ is the log of transmission Poisson likelihood, $R(\boldsymbol{\mu})$ is the regularizing roughness penalty function as described in [10].

To help minimize this cost function, we form paraboloidal surrogates $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^{(n)})$ at each iteration as described in [2]. The coordinate descent method of [2] is monotonic but unsuitable for large-scale problems like X-ray CT. Furthermore, the paraboloidal surrogates described in [2] rely on the non-negativity of the projections but such a condition can not be guaranteed to be true for gradient descent methods without a significant compute time overhead.

Our aim in this paper is to modify the cost function $\Phi(\boldsymbol{\mu})$ to $\tilde{\Phi}(\boldsymbol{\mu})$ and find a new minimum

$$\tilde{\boldsymbol{\mu}} \triangleq \arg\min_{\boldsymbol{\mu}} \tilde{\Phi}(\boldsymbol{\mu}) \qquad \text{such that} \qquad \Phi(\tilde{\boldsymbol{\mu}}) \approx \Phi(\hat{\boldsymbol{\mu}}). \quad (6)$$

If the condition in (6) is satisfied then we can perform the unconstrained minimization shown in (6) rather than the constrained minimization shown in (2), enabling the use of gradient based minimization methods.

2.1. Proposed Cost Function Modification

In (3) the negative of the log likelihood is a sum of functions, h_i, that depend on the values of b_i, r_i and the measurements y_i. The arguments of these functions are $[\boldsymbol{A\mu}]_i$. If $[\boldsymbol{A\mu}]_i$ is negative then at least one of the μ_j's is negative since the elements of the matrix A are non-negative. To state more concisely,

$$[\boldsymbol{A\mu}]_i < 0 \;\;\Rightarrow\;\; \exists\, j \text{ such that } \mu_j < 0. \qquad (7)$$

The condition $\mu_j < 0$ indicates that $\boldsymbol{\mu}$ under consideration is not physically possible. Thus, we argue that the value of h_i for $\ell < 0$ is somewhat arbitrary and it is not essential for it to match the usual log-likelihood function since negative values of ℓ are not physical.

We thus propose to replace the cost functions $h_i(\ell)$ for $\ell < 0$ with functions that are suited to our goal of preservation of monotonicity of the PCG algorithm. Now, consider the representative plots of h_i in Fig. 1 for the three possible cases of detector value y_i :

Case 1 $\quad y_i \leq r_i$
Case 2 $\quad r_i < y_i \leq r_i + b_i$
Case 3 $\quad r_i + b_i < y_i$

For $\ell < 0$ these functions rise exponentially. Therefore, it is impossible to find a true paraboloidal surrogate over $\mathbb{R}$. This is evident from the properties of $\dot{h}_i$ (first derivative of h_i) explained in Appendix A of [2] and a representative plot of $\dot{h}_i$ on [2, p. 807]. In [2], the optimal paraboloidal surrogate functions are found such that the surrogate majorizes h_i for $\ell \geq 0$ only.

We propose the following modification to h_i. In cases 1 and 2, for $\ell < 0$ we replace h_i with a straight line such that the continuity of the function is maintained and the slope of the line is equal to $\dot{h}_i(0)$. A straight line is chosen because it permits the surrogate to have a low curvature. (Low curvatures are advantageous as they increase the convergence speed of the algorithm [2].) In case 3, it is not possible to replace h_i for $\ell < 0$ with a straight line having negative slope. Having a line with negative slope would make $\tilde{\Phi}(\boldsymbol{\mu})$ non-differentiable; we restrict our cost functions to be differentiable so that conditions of (11) can be applied. For sake of simplicity we choose a parabola to replace h_i for $\ell < 0$. The parabola is chosen such that the continuity of h_i and $\dot{h}_i$ are maintained. For reasons of computational simplicity, the curvature of the parabola is computed using [2, eq. 29]. In cases 1 and 2, replacing h_i for $\ell < 0$ with a parabola would lead to higher curvatures in the surrogates when compared to replacing it with a straight line.
For cases 1 and 2, we define :

$$\tilde{h}_i(\ell) \triangleq \begin{cases} h_i(\ell) & \text{if } \ell \geq 0 \\ h_i(0) + \dot{h}_i(0)\ell & \text{if } \ell < 0. \end{cases} \qquad (8)$$

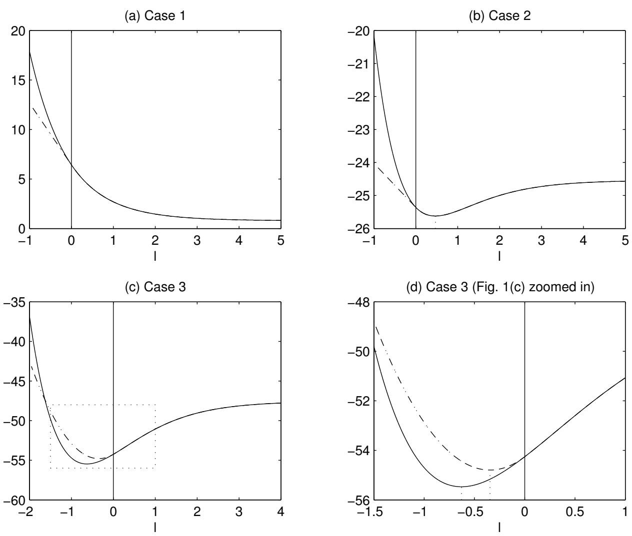

Fig. 1. Plots of $h_i(\ell)$ (solid line) and $\tilde{h}_i(\ell)$ (dashed line) for the three cases. $r_i = 10$, $b_i = 8$, and $y_i = 4, 15$ and 25 in (a), (b) and (c) above respectively. Note that $\tilde{h}_i = h_i$ for $\ell \geq 0$ and differs only for non-physical values of the attenuation coefficients.

For case 3, we define :

$$\tilde{h}_i(\ell) \triangleq \begin{cases} h_i(\ell) & \text{if } \ell \geq 0 \\ h_i(0) + \dot{h}_i(0)\ell + \frac{1}{2}\frac{(y_i - r_i)^2}{y_i}\ell^2 & \text{if } \ell < 0. \end{cases} \tag{9}$$

The new cost function can be thus written as follows :

$$\tilde{\Phi}(\boldsymbol{\mu}) \triangleq \sum_{i=1}^{N} \tilde{h}_i([\boldsymbol{A\mu}]_i) + \beta R(\boldsymbol{\mu}). \tag{10}$$

2.2. Construction of Paraboloidal Surrogates and their Minimization

The following conditions are sufficient for a function $\phi(\boldsymbol{\mu}; \boldsymbol{\mu}^{(n)})$ to be a surrogate of $\tilde{\Phi}(\boldsymbol{\mu})$ [2, eq. 7] :

$$\begin{aligned} \phi(\boldsymbol{\mu}^{(n)}; \boldsymbol{\mu}^{(n)}) &= \tilde{\Phi}(\boldsymbol{\mu}^{(n)}) \\ \phi(\boldsymbol{\mu}; \boldsymbol{\mu}^{(n)}) &\geq \tilde{\Phi}(\boldsymbol{\mu}),\ \forall \boldsymbol{\mu} \in \mathbb{R}^p. \end{aligned} \tag{11}$$

It can be proved that the following surrogate function satisfies all the above conditions :

$$\phi(\boldsymbol{\mu}; \boldsymbol{\mu}^{(n)}) = \sum_{i=1}^{N} q_i(l_i; l_i^{(n)}) + \beta\phi_R(\boldsymbol{\mu}; \boldsymbol{\mu}^{(n)}), \tag{12}$$

$$\text{where } l_i = [\boldsymbol{A\mu}]_i,\ l_i^{(n)} = [\boldsymbol{A\mu}^{(n)}]_i$$

$$q_i(l_i; l_i^{(n)}) \triangleq \tilde{h}_i(l_i^{(n)}) + \dot{\tilde{h}}_i(l_i^{(n)})(l_i - l_i^{(n)}) + \tfrac{1}{2}\breve{c}_i(l_i - l_i^{(n)})^2 \tag{13}$$

$$\breve{c}_i = \begin{cases} \frac{\dot{h}_i(l_i^{(n)}) - \dot{h}_i(0)}{l_i^{(n)}} & \text{if } l_i^{(n)} > 0, \text{cases 1 and 2} \\ \ddot{h}_i(0) & \text{if } l_i^{(n)} \leq 0, \text{cases 1 and 2} \\ \frac{(y_i - r_i)^2}{y_i} & \text{case 3} \end{cases}$$

where $\ddot{h}_i(\ell) = d^2 h_i(\ell)/d\ell^2$.

We use the Huber function as the potential function ψ in (5). $\phi_R(\boldsymbol{\mu}; \boldsymbol{\mu}^{(n)})$ is obtained using the surrogate of Huber function described in [11, p. 184].

3. SIMULATIONS

Size of the attenuation map of the phantom used in the simulations was 128x128, the number of angles was 80 and number of bins per angle was 136. Poisson noise was added to the sinogram. Initial estimate of the attenuation map, $\boldsymbol{\mu}_0$, for all the algorithms was obtained by first doing the FBP reconstruction and then setting the negative pixels to zero. The simulations were done in MATLAB.

We use the pre-conditioned steepest descent method to minimize the paraboloidal (quadratic) surrogate function (12) of the modified cost function $\tilde{\Phi}(\boldsymbol{\mu})$, which is named as QS-PSD-MOD. This is compared against unconstrained minimization of the pre-computed curvature [2] based paraboloidal (quadratic) surrogate function of the original cost function $\Phi(\boldsymbol{\mu})$ using the pre-conditioned steepest descent method called QS-PSD-PC here. We set the negative pixels to zero only after the last iteration and hope the solution is close to the one achieved when negative pixels are set to zero every iteration; this is done because the implementation in the latter case takes more CPU time as setting negative pixels to zero requires $\boldsymbol{A\mu}$ to be recomputed (an extra forward projection operation). We use the diagonal pre-conditioner mentioned in [12] in both the above algorithms. These algorithms are compared against the ordered subsets (with 5 subsets) separable paraboloidal surrogate with pre-computed curvatures (called, OS-SPS-PC [2]) minimization of the original cost function $\Phi(\boldsymbol{\mu})$. Without much loss of compute time, the non-negativity constraint can be applied easily each iteration of the OS-SPS-PC algorithm.

4. RESULTS AND DISCUSSION

Fig. 2 shows that the OS-SPS-PC algorithm stagnates early whereas the gradient based algorithms are decreasing. Though the OS-SPS-PC algorithm converges faster initially with respect to the number of iterations, it is actually slower as each iteration takes more CPU time than the gradient based algorithms. The gradient based algorithms achieve a lower cost than OS-SPS-PC. The condition of (6) has been found to be satisfied, thus validating the modification of the cost function $\Phi(\boldsymbol{\mu})$ to $\tilde{\Phi}(\boldsymbol{\mu})$. This modification helps us achieve a lower cost than OS-SPS-PC without losing on the image

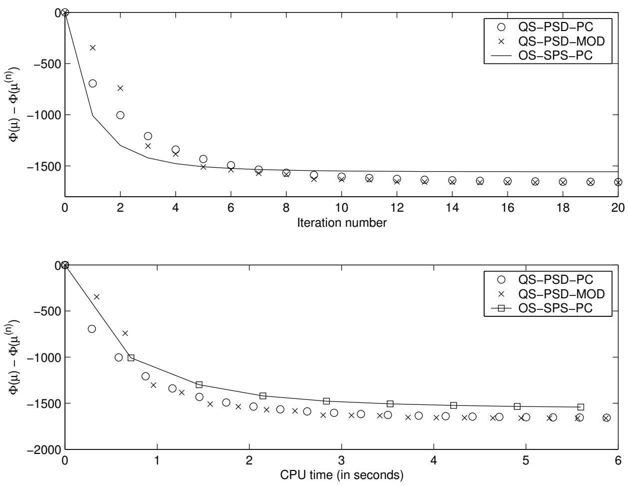

Fig. 2. Plots showing the variation of the original cost function $\Phi\ (\boldsymbol{\mu})$ with number of iterations and CPU time.

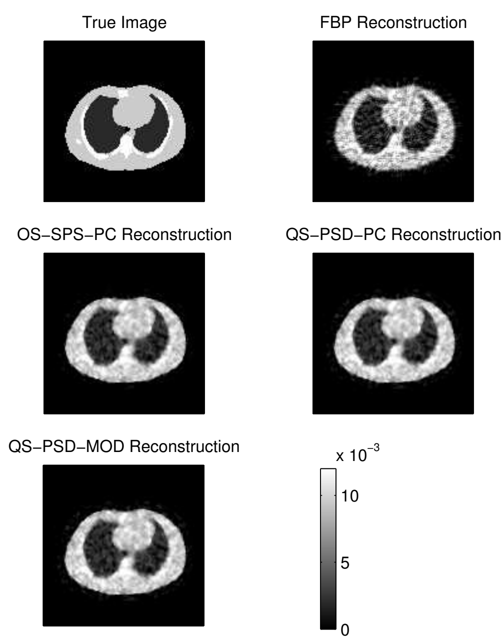

Fig. 3. Images showing the true phantom and its reconstructions by various algorithms.

quality. This is evident in Fig. 3. The ratio of magnitude of average of negative pixels in QS-PSD-PC and QS-PSD-MOD to the maximum value of the pixels in the true image of the phantom is approximately 5%. The sum of squared differences of corresponding pixels in the reconstructed images (with negative pixels set to zero) and the true image of the phantom divided by the sum total of the squared value of pixels in the true image of the phantom *i.e.* the quantity $\|\boldsymbol{\mu}_{reconstructed} - \boldsymbol{\mu}_{true}\|/\|\boldsymbol{\mu}_{true}\|$ is 0.2750, 0.1931, 0.1996 and 0.1976 in FBP, OS-SPS-PC, QS-PSD-PC and QS-PSD-MOD respectively.

5. FUTURE WORK

It is straight-forward to extend QS-PSD-PC and QS-PSD-MOD to their PCG versions *i.e.* QS-PCG-PC and QS-PCG-MOD respectively. Surrogates for QS-PSD-MOD with lower curvatures than those used in this paper can be derived. Performance of these algorithms with factored system matrices can be investigated.

6. REFERENCES

[1] J. A. Fessler, "Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans," *IEEE Tr. Im. Proc.*, vol. 4, no. 10, pp. 1439–50, Oct. 1995.

[2] H Erdoğan and J. A. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Tr. Med. Im.*, vol. 18, no. 9, pp. 801–14, Sept. 1999.

[3] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Tr. Med. Im.*, vol. 16, no. 2, pp. 166–75, Apr. 1997.

[4] C. Kamphius and F. J. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Tr. Med. Im.*, vol. 17, no. 6, pp. 1001–5, Dec. 1998.

[5] H Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, pp. 2835–51, Nov. 1999.

[6] S. Ahn and J. A. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Tr. Med. Im.*, vol. 22, no. 5, pp. 613–26, May 2003.

[7] A R De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography," *IEEE Tr. Med. Im.*, vol. 20, no. 4, pp. 280–8, Apr. 2001.

[8] J. Hsieh, O. E. Gurmen, and K. F. King, "Investigation of a solid-state detector for advanced computed tomography," *IEEE Tr. Med. Im.*, vol. 19, no. 9, pp. 930–40, Sept. 2000.

[9] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comp. Assisted Tomo.*, vol. 8, no. 2, pp. 306–16, Apr. 1984.

[10] J. A. Fessler, "Grouped coordinate descent algorithms for robust edge-preserving image restoration," in *Proc. SPIE 3170, Im. Recon. and Restor. II*, 1997, pp. 184–94.

[11] P. J. Huber, *Robust statistics*, Wiley, New York, 1981.

[12] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Tr. Im. Proc.*, vol. 8, no. 5, pp. 688–99, May 1999.