

# A study of four minimization approaches for iterative reconstruction in X-ray CT

Bruno De Man, *Member, IEEE*, Samit Basu, *Member, IEEE*, Jean-Baptiste Thibault, Jiang Hsieh, *Member, IEEE*, Jeffrey A Fessler, *Member, IEEE*, Charles Bouman, *Fellow, IEEE*, and, Ken Sauer, *Member, IEEE*

## I. INTRODUCTION

**M**OST iterative reconstruction algorithms are based on a cost function combined with an optimization approach to iteratively find the optimum (we'll assume *minimum* for convenience) of that cost function. The cost function includes a data fit term, requiring the calculated sinogram to be similar to the measured sinogram, and often also a prior term or a regularization term, requiring the image to satisfy some desired properties based on prior knowledge. Ideally, the cost function completely defines the reconstructed image, meaning that the actual minimization approach finds the global minimum of the cost function. Even though this is theoretically obvious in the absence of local minima and for convergent algorithms, it has (to our knowledge) never been demonstrated for non-trivial problems, that multiple completely different optimization approaches indeed result in the same reconstructed images, and in a reasonable number of iterations. In this work, we compare an iterative coordinate descent approach (ICD) [1], a conjugate gradient approach (CG) [2-3], a separable parabolic surrogate approach with ordered subsets (OS) [4], and a convergent ordered subsets approach (COS) [5]. In addition to showing that - if applied with care - all approaches result in the same final image, we also give an indication of the number of iterations and time to convergence for the studied approaches.

## II. ALGORITHMS

Statistical tomographic reconstruction is based on the maximization of the log-posterior probability (also called maximum a posteriori or MAP):

$$\text{MAP} = \log P(\text{measurement} \mid \text{image}) + \log P(\text{image}). \quad (1)$$

The first term is called the likelihood and can be formulated by combining a model for the acquisition  $P(\text{calculated} \mid \text{image})$  and a noise model  $P(\text{measurement} \mid \text{calculated})$ . The second term is called the prior (or regularizer). We use the following acquisition model:

$$\hat{p}_i = \sum_{j=1}^J l_{ij} \mu_j \quad (2)$$

Bruno De Man and Samit Basu are with GE Global Research, Niskayuna, NY. Jean-Baptiste Thibault and Jiang Hsieh are with GE Healthcare, Milwaukee, WI. Jeffrey A. Fessler is with University of Michigan, Ann Arbor, MI. Charles Bouman is with Purdue University, West Lafayette, IN. Ken Sauer is with University of Notre-Dame, South Bend, IN.

where  $\hat{p}_i$  is the calculated sinogram value for projection line  $i$ ,  $\mu_j$  is the reconstructed value in image pixel  $j$ , and  $l_{ij}$  is the projection coefficient for pixel  $j$  and projection line  $i$ . We used a distance-driven projector and backprojector [6], which are matched, so  $l_{ij}$  is also the backprojection coefficient. We used a Gaussian noise model :

$$P(p_i \mid \hat{p}_i) = \frac{1}{2\pi\sigma_i} \exp\left(-\frac{(p_i - \hat{p}_i)^2}{2\sigma_i^2}\right) \quad (3)$$

where  $p_i$  is the real measurement for projection line  $i$  and  $\sigma_i$  is the corresponding standard deviation. We used a quadratic prior with strength  $\beta$  and neighborhood weights  $n_{jk}$ . These assumptions reduce the MAP problem into a penalized weighted least squares minimization problem with the following cost :

$$C = \frac{1}{2} \sum_i w_i (p_i - \hat{p}_i)^2 + \beta \sum_j \sum_{k \neq j} n_{jk} \frac{(\mu_j - \mu_k)^2}{2} \quad (4)$$

where each weight  $w_i$  is the inverse of the standard deviation on measurement  $i$ .

The iterative coordinate descent approach (ICD) [1] can be described as follows. In every iteration  $n$  we sequentially address every image pixel  $j$  in a random (but fixed) order and we compute its attenuation  $\mu_j^n$  as

$$\arg \min_{\mu_j^n} C(\mu_1^n, \dots, \mu_{j-1}^n, \mu_j^n, \mu_{j+1}^n, \dots, \mu_J^{n-1}). \quad (5)$$

The solution is easily obtained by requiring

$$\frac{\partial C}{\partial \mu_j^n} = 0. \quad (6)$$

resulting in

$$\mu_j^n = \mu_j^{n-1} + \frac{\sum_i l_{ij} w_i (p_i - \hat{p}_i) + \beta \sum_{k \neq j} n_{jk} (\mu_k - \mu_j)}{\sum_i l_{ij}^2 w_i + \beta \sum_{k \neq j} n_{jk}} \quad (7)$$

We used no under- or over-relaxation.

The conjugate gradients approach (CG) [2-3] can be summarized as follows. The direction of each new update step is computed as

$$d^n = \nabla C + \gamma d^{n-1} \quad (8)$$

where

$$\gamma = \frac{\|\nabla C^n\|^2}{\|\nabla C^{n-1}\|^2}. \quad (9)$$

The actual update step is given by

$$x^n = x^{n-1} + \hat{\alpha} d^n \quad (10)$$

with

$$\hat{\alpha} = \arg \min_{\alpha} C(x^{n-1} + \alpha d^n). \quad (11)$$

We did not use a pre-conditioner.

The ordered subsets (OS) approach can be derived by defining separable parabolic surrogate functions for the cost function [4] and is described as follows.

$$\mu_j^n = \mu_j^{n-1} + \alpha \frac{\sum_{i \in S} l_{ij} w_i (p_i - \hat{p}_i) + \beta \sum_{k \neq j} n_{jk} (\mu_k - \mu_j)}{\sum_{i \in S} l_{ij} w_i \hat{p}_i + \beta \sum_{k \neq j} n_{jk}} \quad (12)$$

We studied various numbers of subsets, including the limiting case of *one* subset, and we studied different relaxation factors.

The convergent ordered subset approach (COS) [5], also referred to as transmission reconstruction by incremental optimization transfer (TRIOT), can be summarized as follows. First the cost function is expressed as the sum of a number of terms, one for each subset. For every subset iteration, the gradient for that particular subset is re-computed. Finally, the image is updated by finding an optimal combination of the most recent gradients.

### III. EXPERIMENTS

For evaluation, we used an axial measurement of a thorax phantom on a GE Lightspeed CT scanner and we reconstructed only the central slice to limit the problem to 2D. We used FBP as initial estimate for the reconstruction. We performed 200 iterations and we analyzed the cost and the reconstructed image after convergence. We also estimated the number of iterations and the computation time required to reach certain convergence criteria.

Most experiments were performed with a non-negativity constraint. Because our CG implementation did not allow for a non-negativity constraint, we did a separate experiment without this constraint, resulting in a similar but different cost (as expected). For easy comparison, all the costs *without* non-negativity are shown shifted by a small DC offset to match the results *with* non-negativity.

The first series of experiments includes 200 iterations of ICD, CG, OS with 41 subsets and COS. The results in Figure 1 show the relative convergence properties. OS gets stuck in a limit cycle behavior, resulting in a higher cost and a discrepancy of about 7HU. All other approaches converge to the same minimum cost. We also analyzed the images and the difference was smaller than 1HU. The final images are shown in Figure 2 for ICD and OS, and their difference is shown in Figure 3. The images look visually identical but the difference image shows the small (less than 7HU) discrepancy for OS.

In a second series of experiments, we showed that OS can be made to converge by switching to one single subset. This is illustrated in Figure 4 (solid line). We showed that switching to a lower relaxation factor of 0.1 results in a lower cost much closer to the global minimum (not shown). We also performed 25 iterations of OS followed by 175 iterations of COS and showed that this combination results in favorable convergence properties (dashed line). We performed additional

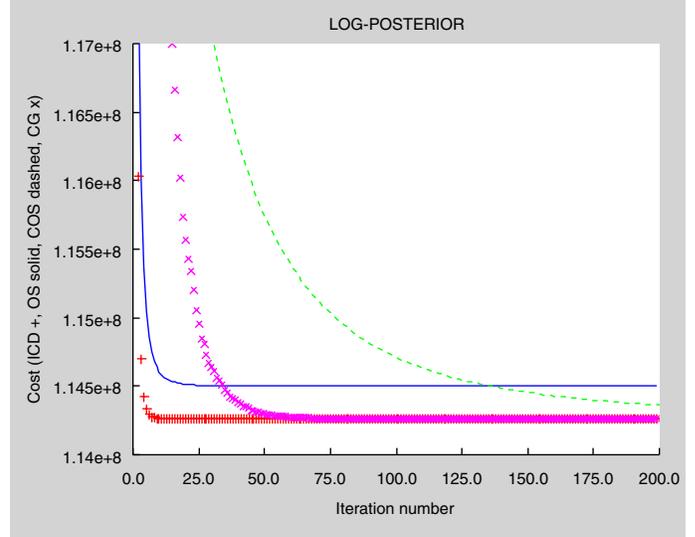


Fig. 1. Total cost as a function of iteration number for 200 iterations of ICD (red +), CG (purple x), OS with 41 subsets (blue solid), and COS (green dashed).

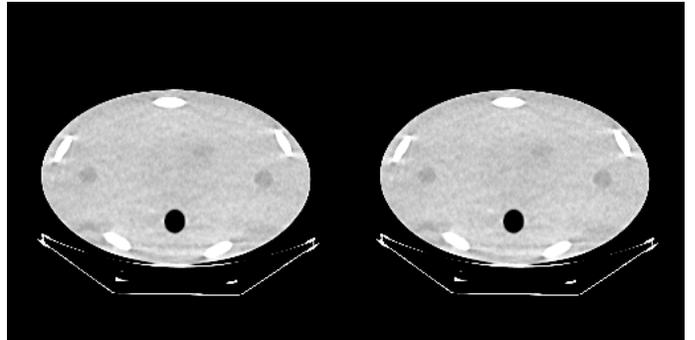


Fig. 2. Image after 200 iterations of ICD (left) and OS (right). The level is 930HU and the window 116HU. No attempt was made to optimize cost function or image quality. The OS image has a discrepancy of about 7HU (not visible).

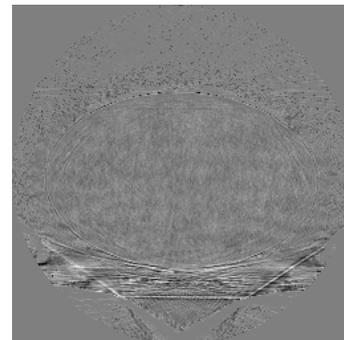


Fig. 3. Difference between OS and ICD after 200 iterations each. The level is 0HU and the window 20HU. There is a clear difference mainly in the neighborhood of the patient table edges.

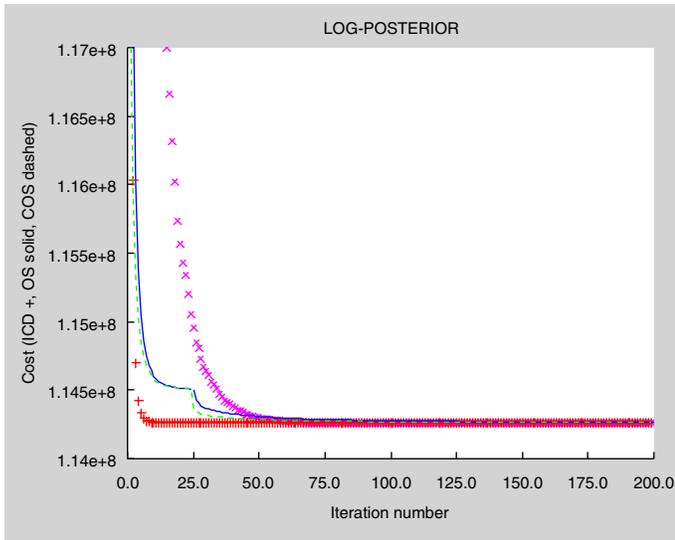


Fig. 4. Total cost as a function of iteration number for 200 iterations of ICD (red +), CG (purple x), OS switching from 41 subsets to 1 subset (blue solid), and COS initialized with 25 iterations of OS (green dashed).

OS reconstructions with an over-relaxation factor of 1.95 and with 82 subsets. In both cases, the required number of iterations for convergence decreases but the final cost and discrepancy increase, as expected. These results suggest that a gradually decreasing subset and relaxation schedule would be desirable.

In an additional experiment, we used a blank image as initial estimate instead of FBP. Only ICD significantly suffered from this. This is consistent with the understanding that ICD results in fast convergence of the high frequencies, while the low frequencies converge more slowly.

ICD required about 10 to 20 iterations (depending on the convergence criterion). OS with 41 subsets required 3 times more iterations, and CG 6 times more. Even though the number of required iterations was very different for all approaches, this was compensated by the respective times per iteration, and the actual computation times were of the same order of magnitude.

However, none of the implementations were optimized for performance, and the timing results are only included as a rough indication. For all approaches, the reconstruction time for a 512x512 image based on a 888x984 sinogram was about 10 minutes on a 3GHz Pentium Xeon, using distance-driven projection-backprojection.

#### IV. CONCLUSIONS

We studied four different minimization approaches for iterative reconstruction in CT and showed that all four result in the same reconstructed image after convergence. No attempt was made to optimize the cost function or the corresponding image quality. The regular ordered subsets approach was indeed non-convergent (limit cycle behavior), unless we reduced the number of subsets or the relaxation factor, or unless we switched to the convergent ordered subsets method. While the number of required iterations varies between the different methods, the time per iteration usually goes in the other direction, and consequently the computation times were all of the same order of magnitude.

#### REFERENCES

- [1] K. Sauer, C. Bouman "A Local Update Strategy for Iterative Reconstruction from Projections," *IEEE Trans. Signal Proc.*, 41(2):534-548, Feb. 1993.
- [2] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, Z. Zhou "Fast Gradient-Based Methods for Bayesian Reconstruction of Transmission and Emission PET Images," *IEEE Trans. Med. Imag.*, 13(4):687-701, Dec 1994.
- [3] J. A. Fessler, S. D. Booth "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Im. Proc.*, 8(5):688-699, May 1999.
- [4] H. Erdogan, J. A. Fessler "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, 44(11):2835-2851, Nov. 1999.
- [5] S. Ahn, J. A. Fessler, D. Blatt, A. O. Hero "Convergent Incremental Optimization Transfer Algorithms: Application to Tomography," *IEEE Trans. Med. Imag.*, (to appear).
- [6] B. De Man, S. Basu "Distance-driven Projection and Backprojection in Three Dimensions," *Phys. Med. Biol.*, 49: 2463-2475, May 2004".