SYSTEM PARAMETER ESTIMATION IN TOMOGRAPHIC INVERSE PROBLEMS

A.ALESSIO AND K.SAUER

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46616
USA[†]

AND

A. MOHAMMAD-DJAFARI

Laboratoire des Signaux et Systèmes (L2S)
Plateau de Moulon
91192 Gif sur Yvette
France

Abstract. Inverse problems are typically solved under the assumption of known geometric system parameters describing the forward problem. Should such information be unavailable or inexact, estimation of these parameters from only observed sensor data may be necessary prior to reconstruction of the desired signal. We demonstrate the feasibility of such estimation via maximum-likelihood methods for the system parameters with expectation-maximization as an optimization mechanism within a Bayesian estimation framework for the final reconstruction problem.

Key words: Expectation-Maximization,Calibration, Parameter Estimation, Computed Tomography

1. Introduction

Bayesian approaches to inverse problems require knowledge or estimation of a number of parameters which determine the statistical description for the forward model and the *a priori* density of the function to be estimated. Among these are the "hyperparameters" related to variances of both elements and the form of the prior, which have been widely studied [1–12]. It is usually taken for granted that the description of the system underlying the forward problem is well known.

[†] Email: aalessio@nd.edu sauer@nd.edu

In some problems, particularly those involving measurements in the field, the geometric description of the data collection may include significant uncertainty. Some non-destructive testing applications, for example, have elements which are invariably imprecise [13,14]. Calibration may be possible on-site, but may also be too costly if it forces duplication of measurement. Archived data may lack such information either through omission at the time of capture or subsequent loss.

In the absence of test measurements from known objects for calibration, this task of estimating parameters from only corrupted data may be termed a "missing data" problem. We utilize techniques based on the expectation-maximization algorithm [15–17] to find the maximum-likelihood (ML) values for the system parameters, similar to approaches for statistical hyperparameter estimation. The expectation operation is computed or estimated over realizations of the random vector X representing the object of Bayesian reconstruction.

In this work we apply EM-type algorithms for geometric parameter estimation to X-ray tomography in which a single parameter, independent of other geometric elements, is unknown. Examples include both parallel and fan-beam data in two dimensions. For the parallel beam case, more direct methods of estimating certain parameters, based on the moments of projections [18], are available. The present demonstration is intended only as a feasibility study for a method applicable to more general problems.

The complexity of the expectation and maximization steps of EM varies widely among problems, with closed forms often unattainable. Either or both operations are commonly approximated, through Monte-Carlo methods for the former and iterative ascent methods for the latter[19,20,12]. The examples in Section 3 include approximations of both operations, resulting in relatively simple backprojection at the expectation step and Newton-style optimization for the maximization. In both cases, parameter estimates appear to converge rapidly and yield relatively accurate reconstructions, though low signal-to-noise ratio cases may require more robust optimization than simple ascent methods used here.

2. Estimation Description

2.1. EM ALGORITHM

An ideal system calibration setting includes a known object with discretized representation in the vector X and measurements from the instrument in question in the vector Y. In the Bayesian context, an *a priori* distribution $p(x)$ may be applied to describe the statistics of the random object X. Given $\{X = x, Y = y\}$ the challenge of finding the system parameters (θ) can be framed as a straightforward maximum likelihood problem with

$$\hat{\theta} = \arg\max_{\theta} \ln p(x, y|\theta).$$

The distribution of Y conditioned on $\{X = x\}$ is often of a common, tractable form dictated by physics of the forward problem while an *a priori* distribution incorporates what may be viewed as regularization to estimates of X.

In the present problem, we wish to maximize $p(y|\theta)$ in the absence of knowing X. The likelihood function can be obtained from:

$$\hat{\theta} = \arg\max_{\theta} \ln \int p(x, y|\theta) dx.$$

Unfortunately, this integral poses a challenge and a closed form solution may not exist. Such problems, in which observations may be termed "incomplete," are frequently solved by the EM algorithm. Using the terminology of [17], the "complete data" for this problem consists of the vector (x, y) and the EM algorithm may be applied with the following formulation:

$$\theta^{n+1} = \arg\max_{\theta} E[\ln p(X, y|\theta)|Y = y, \theta^n] = \arg\max_{\theta} \tilde{Q}(\theta; \theta^n). \tag{1}$$

$\tilde{Q}(\theta; \theta^n)$ represents the conditional expectation. This expectation may be approximated through Markov chain Monte Carlo (MCMC) methods in the general case of intractable integrals in (1). The maximization step may also be non-trivial.

Because the geometric system parameters in θ do not influence $p(x)$, the *a priori* term does not affect the maximization step and our application of EM can be re-written with

$$Q(\theta, \theta^n) = E[\ln P(y|X, \theta)|Y = y, \theta^n] \tag{2}$$

for Poisson data in which the density is a probability mass function. The influence of the *a priori* density $p(x)$ remains implicit in the form of the conditional distribution over which the expectation is computed.

In what follows we we first propose an appropriate approximation $Q_a(\theta, \theta^n)$ for $Q(\theta, \theta^n)$. Then to obtain an estimate $\hat{\theta}$, we propose two approaches:

$$a) \ \hat{\theta} \ = \ \lim_{n \to \infty} \theta^n \text{ with } \theta^{n+1} = \arg\max_{\theta} Q_a(\theta, \theta^n),$$
$$b) \ \hat{\theta} \ = \ \arg\max_{\theta} Q_a(\theta, \theta)$$

which we call respectively "EM" and "Direct" algorithms. Finally we compare their relative performances in both parallel and fan beam tomography.

2.2. EXPECTATION APPROXIMATIONS FOR TRANSMISSION TOMOGRAPHY

In this paper, we apply the process above to the geometric parameters of X-ray tomographic systems, with the goal of subsequent Bayesian reconstruction of X. In most applications, the log likelihood is well approximated by [21]

$$\ln P(Y = y|x, \theta) \approx -\frac{1}{2} \sum_i (y_i - A_{i*}(\theta)x)^2 d_i e^{-y_i} + c(y), \tag{3}$$

with $A_{i*}x$ representing the i-th discretized forward projection via a row of the projection matrix A, d_i signifying the i-th X-ray input dosage, and $c(y)$ a function

of y only which does not affect subsequent steps. A is directly dependent on the geometric parameters in θ. Thus the conditional expectation of (2) is approximately that of the total weighted mean squared difference between data in y and forward projections of x, dependent on θ. While quadratic, this expression maintains the statistically appropriate weighting of the data's reliability by $\{y_i\}$.

The conditional expectation of (3) remains formidable in general. En route to simplifying the approximation without resorting to Monte-Carlo methods, we separate the total squared sinogram error into a bias and variance term. (Here we use the term "bias" as the expected difference between y_i and $A_{i*}X$.) This yields

$$E[\ln P(y|X,\theta)|Y=y,\theta^n] \approx$$
$$-\tfrac{1}{2}\sum_i d_i e^{-y_i}\left\{(y_i - A_{i*}(\theta)E[X|Y=y,\theta^n])^2 + Var(A_{i*}(\theta)X|Y=y,\theta^n)\right\} \quad (4)$$

Given a stationary model for $p(x)$ and the fact that perturbations of θ will not typically strongly affect the length of integral projection paths, the conditional variance of the forward projections should not be highly sensitive to θ. It is likely to remain most dependent on the $\{d_i e^{-y_i}\}$ due to the weighting in (3). We therefore further posit that this expression's dependence on θ (through A) will be dominated by the bias term.

Provided we can find the conditional mean of X, the maximization step of EM now appears manageable. However, this "estimate" of X represents a minimum mean-squared error inverse problem, itself known to be relatively difficult and frequently found through stochastic estimation methods [22,23]. Under a Gaussian model for $p(x)$, however, X becomes *a posteriori* Gaussian with its mean a linear function of y. Applying this modeling and using filtered backprojection (FBP) as the linear mapping, we use an FBP image under parameter θ^n as $Q_a(\theta|\theta^n) = E[X|Y=y,\theta^n]$. The obvious gap in this argument is the sub-optimality of shift-invariant, one-dimensional filtering in FBP in computing the MMSE estimate of X [24]. Subsequent experiments will verify the utility of these simplifications, which reduce EM to the same computation as a deterministic non-linear least squares problem.

2.3. MAXIMIZATION STEP

Many different approaches have been studied to maximize $Q(\theta;\theta^n)$. Each particular application may warrant a different maximization method. For instance, some applications would call for a simulated annealing approach [23], while others, which have a concave Q, could be maximized via ascent methods. If an ascent method is viable, one must compute at least the first derivative of the expectation. Using the previously discussed approximations for the tomographic problem and $\mu_{x|y,\theta^n}$ as the conditional expectation,

$$\frac{\partial}{\partial\theta}Q_a(\theta|\theta^n) = \sum_i d_i e^{-y_i}\left\{\left(y_i - A_{i*}(\theta)\mu_{x|y,\theta^n}\right)\sum_j \mu_{x_j|y,\theta^n}\frac{\partial}{\partial\theta}A_{ij}(\theta)\right\}_{\theta=\theta_0}. \quad (5)$$

In most system geometries, the given derivative (and the Hessian for low-dimensional θ) will be tractable if not trivial.

3. Implementation

3.1. OVERVIEW

This algorithm was tested with both parallel beam and fan beam projection data. The synthetic parallel data had varying signal to noise ratio while the fan beam data, from a conventional commercial scanner, had a high SNR. For these initial trials, we choose to use ascent methods for optimization. In most of the tests, we employed a numerical approximation of the first and second derivatives through sampling the objective functions about the current θ. Relatively wide sampling intervals help make these numerical approximations more robust to noise on the Q function. In the parallel case only, we computed the true derivatives of $Q(\theta; \theta^n)$ for EM. The derivative values, whether approximated through sampling or computed exactly, were used in Newton-Raphson's method until convergence.

Due to the approximations introduced above, the function $Q_a(\theta; \theta)$, directly in the parameter θ, is accessible as well.

$$Q_a(\theta; \theta) = -\frac{1}{2} \sum_i d_i e^{-y_i} (y_i - A_{i*}(\theta)\mu_{x|y,\theta})^2 \qquad (6)$$

This allows consolidation of the two EM steps of expectation and maximization. Though this simplification seems appealing, it is not clear whether optimizing $Q_a(\theta, \theta)$ is in general advantageous. While the EM algorithm seeks the maximum likelihood solution ($\arg\max_\theta \ln P(y|\theta)$), our direct optimization maximizes the approximation to $E[\ln P(y|X, \theta)|Y = y, \theta]$ instead. It is not apparent that these two functions of θ will have maxima at the same point, which may cause discrepancies between our EM and direct estimates. In short, for both the parallel and fan beam situations, we implemented two algorithms. In one case, denoted as the "EM algorithm," we maximize our approximation to (2) in θ for each θ^n. In the other method, labeled as "direct," we maximize the same function along the line $\theta^n = \theta$.

3.2. PARALLEL BEAM

A parallel beam scanner with a rotation of 180° was simulated to create a synthetic phantom, resulting in a sinogram of dimension 128x128. The phantom represented realistic X-ray attenuation values, with input dosage for this example set to 2,000 or 20,000 photons per ray. Input dosage (d_i) was in each case constant among rays. Higher dosages produce higher quality image reconstructions, with noise variance approximately inversely proportional to input dosage. The estimated parameter was the offset of the center of rotation of the detectors from the origin. This offset, t_0, was measured from the first-indexed detector to the center of rotation. If the detectors were perfectly centered, t_0 would equal -99.219mm. The phantom has a diameter of roughly 200mm and attenuation densities which range from 0.01mm^{-1} to 0.04mm^{-1}.

Three dimensional representations of the parallel beam Q_a functions under the approximation in (3) appear in figures 1 (a) & (b). Points A, B, and C in these figures summarize the estimated values for the EM and "direct" methods. Note that
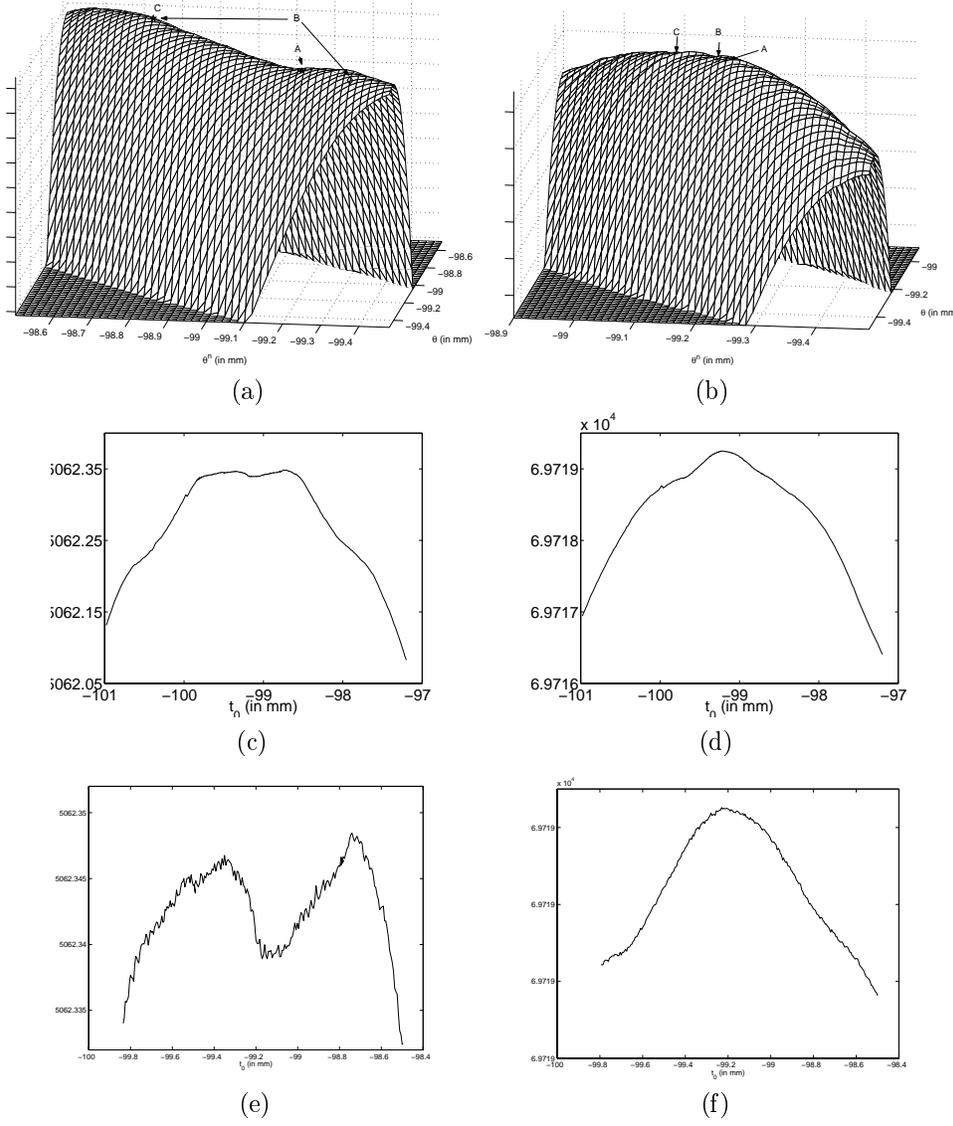
Figure 1.   Parallel beam $Q_a(\theta; \theta^n)$; point A is the true value for theta, t_0, point B is the estimated value from direct maximization of $Q_a(\theta; \theta)$, point C is the estimated value from EM algorithm. (a) Low SNR ($d = 2,000$), (b) High SNR ($d = 20,000$). Plots (c) and (d) are the same cases for $Q_a(\theta, \theta)$, along the diagonal of (a) and (b), with greater range in t_0, while (e) and (f) show a magnified portions of (c) and (d) near their maximum.

while both the high and low SNR $Q_a(\theta; \theta^n)$ have well defined ridges near $\theta = \theta_n$, the low SNR case does not have a unique well-defined local maximum. In the low SNR case ($d_i = 2000$), the EM algorithm converged to a value between -98.7mm
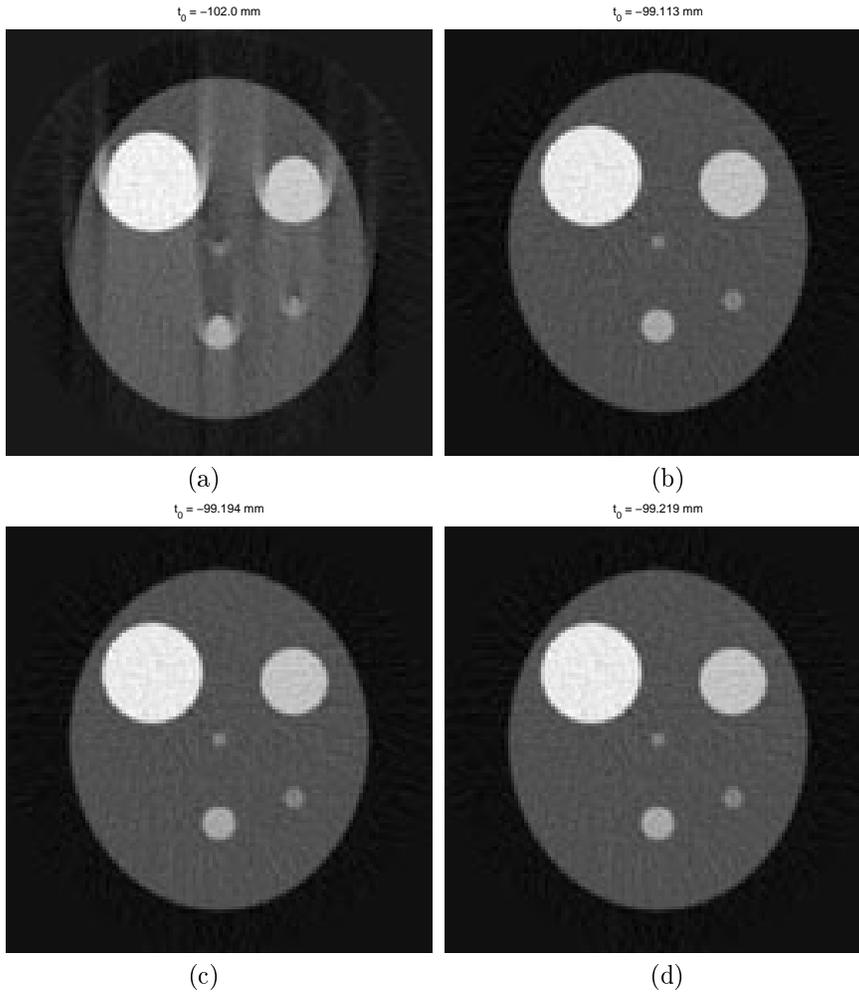
Figure 2. Parallel beam reconstructions for varying t_0, (a) reconstructed image with t_0 far from true value, (b) Image with value estimated by EM, (c) Image with value estimated by direct maximization of Q_a, (d) Reconstructed with true value.

to -98.9mm, depending on the initial estimate. The direct method converged to -99.35mm or -98.72mm due to the presence of two local maxima in the function (see fig. 1(a)). All of these low SNR estimates are relatively far from the desired value and result in a reconstructed image with undesirable variations from the ideal reconstruction. The plot in Figure 1(e) shows that the direct method clearly benefitted from the robust, sampled derivative approximation used here in place of exact derivatives. Low SNR cases may well frustrate common ascent methods for this problem. Should the log-likelihood and Q_a functions be non-concave and/or have excessive noise, solutions depend heavily on the numerical method. Stochas-

tic optimization methods such as simulated annealing would clearly be of interest for such cases.

Figure 2 displays reconstructed images for different values of t_0 for the high SNR case ($d = 20,000$), which has a significantly better-behaved Q_a shown in Figure 1(b), (d). Newton-Raphson's method, using the true derivative values, caused the EM algorithm to converge to t_0 =-99.113mm, yielding the FBP reconstruction in fig. 2(b) in under 19 iterations. Using approximations of the derivative via sampling to maximize $Q_a(\theta; \theta)$ directly provides an estimate of -99.194mm. The reconstruction with the true offset value (t_0=-99.219mm), Figure 2(d), has no discernable differences from the high SNR, estimated value reconstructions (figures 2(b,c)). In sum, the two different methods, EM and "direct", yield different, but acceptable estimates in the high SNR case which vary slightly from the true value (see Fig. 1 A,B,&C). These discrepancies may result from slight noise and the approximations accepted for Q.

3.3. FAN BEAM

This estimation method was also applied to an equiangular, fan beam geometry. A single source transmits through an object and counts are registered on a bank of detectors forming equiangular beams with the source. The backprojection reconstruction routine utilized a simple low-pass filter and linear interpolation of filtered data as described in [25]. The projection information for parameter estimation in this case was obtained from 128 detectors collecting photon counts at 128 different rotations with a high SNR, extracted from a set of 1024×1024. For our tests, we estimated the arc span formed between the two farthest detectors. The simulations discussed here have the system parameters of D, the distance from the source to the origin, set to 900mm and w, the width and height of the image, set to 600mm. D will serve only as a scaling factor.

With the 128x128 projection data, the value of the arc span converges to its final value of 0.74 radians in approximately 5 iterations of the EM algorithm. The plot of $Q_a(\theta, \theta)$ in Figure 3(a) is unimodal and yet more smooth than the high SNR case for the synthetic phantom. Direct optimization of $Q_a(\theta; \theta)$ (Fig. 3(a)) produced 0.73 radians as the final estimate, slightly closer to a value chosen manually to yield the best FBP reconstruction in purely visual evaluation without precise knowledge of X itself or arc span. Figure 4 shows the convergence of this algorithm when the initial estimate for the arc is 0.6 radians. The final arc values were used to reconstruct the images in figures 3(c) and (d). These final FBP reconstructions were made with 1024x1024 projection data for better visual evaluation.

Figure 3 shows the difference in the reconstructed image for a small variation in arc span. Note the artifacts around the circular objects in fig 3(b) when the arc is in error. The arc values obtained from the two variants of the estimation algorithm resulted in reconstructed images that were indistinguishable in quality from the reconstructed image generated by the scanner's custom software. Only small differences in background variations apparently due to filter choice were perceptible. However, the difference between the parameter values suggests that
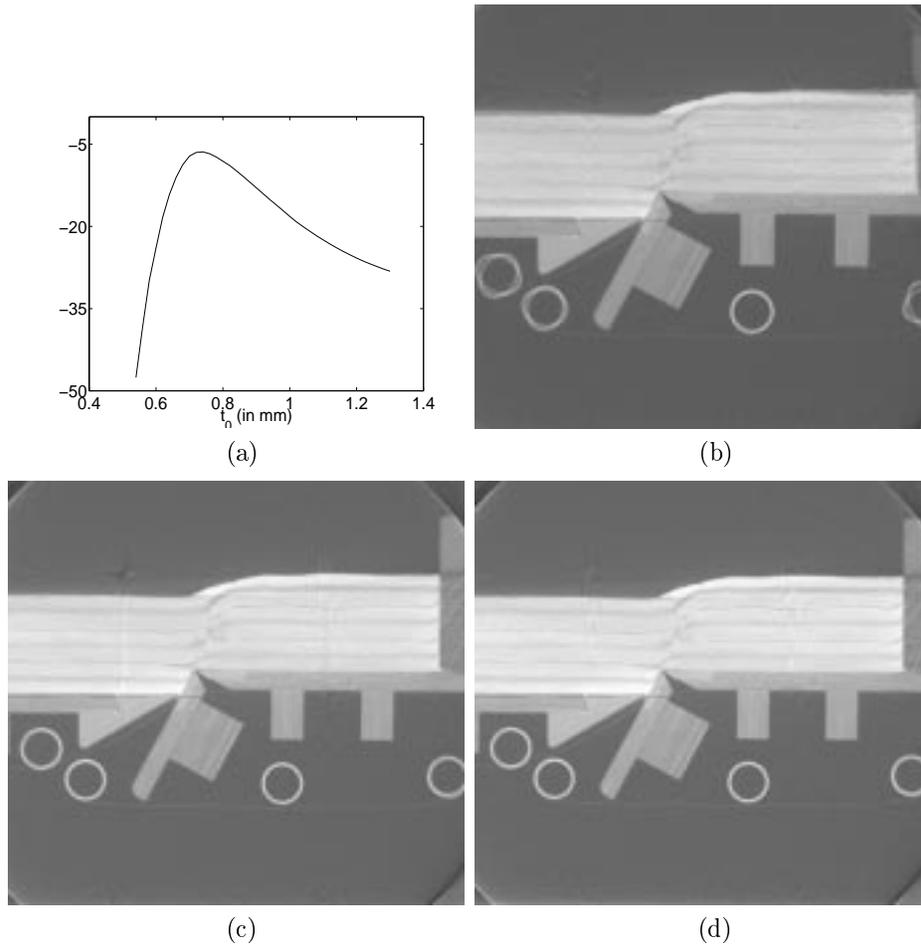
Figure 3. Plot in (a) shows $Q_a(\theta, \theta)$ for the high SNR fan-beam problem. Fan beam reconstructions (b)$D = 900$mm w=600mm arc = 0.810 radians, (c) $D = 900$mm w=600mm arc = 0.744 radians- EM Value, (d)$D = 900$mm w=600mm arc = 0.733 radians-"Direct" Value. Data courtesy of Institut Francais de Petrol.

even in this high SNR, apparently concave case, there exists a small difference under our approximations between solution of EM and direct optimization of Q_a.

4. Conclusion

The EM algorithm was applied to estimate certain system parameters in two scanner applications. In both high SNR cases, the algorithm converged to near its appropriate value in few iterations and yielded final reconstructions free of visible artifacts from geometric distortion. In the process of designing a simple algorithm, we accepted several approximations. The initial success of the algorithm reinforces
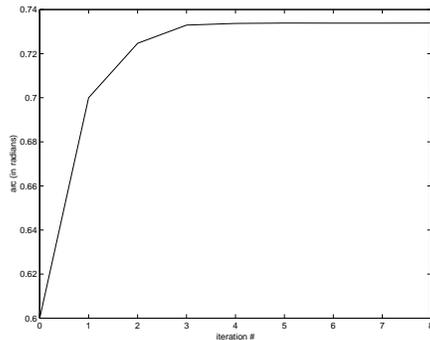
Figure 4.   Fan beam arc length values for successive iterations of EM algorithm.

the utility of these approximations, but their applicability in low SNR remains questionable. Further study will be aimed at resolving optimization difficulties and potential bias in estimates.

In principle, EM may be applied to a great variety inference tasks. The results presented here show that this approach has promise for simple geometric parameterization of tomographic systems. We intend to apply this technique to other systems which have larger numbers of unknown parameters, such as more general uncertainty in source and detector locations, and in gantry rotations.

References

1. J. Rice, "Choice of smoothing parameter in deconvolution problems," in *Contemporary Mathematics*, J. S. Maron, ed., vol. 59, pp. 137–151, American Math. Soc., Providence, RI, 1986.
2. L. Younes, "Parametric inference for imperfectly observed Gibbsian fields," in *Probability Theory and Related Fields*, vol. 82, pp. 625–645, Springer Verlag, 1989.
3. S. Reeves and R. Mersereau, "Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration," *Optical Engineering*, **29**, pp. 446–454, May 1990.
4. A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, "A study of methods of choosing the smoothing parameter in image restoration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **13**, (4), pp. 326–339, 1991.
5. N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. on Image Processing*, **1**, pp. 322–336, July 1992.
6. A. Mohammad-Djafari, "On the estimation of hyperparameters in Bayesian approach of solving inverse problems," in *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, (Minneapolis, Minnesota), pp. 495–498, April 27-30 1993.
7. R. Schultz, R. Stevenson, and A. Lumsdaine, "Maximum likelihood parameter estimation for non-Gaussian prior signal models," in *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 2, (Austin, TX), pp. 700–704, November 1994.
8. J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. on Image Processing*, **3**, pp. 404–420, July 1994.
9. Z. Zhou and R. Leahy, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," in *Proc. of IEEE Int'l Conf. on Image Proc.*, (Washington, D.C.), pp. 284–287, October 23-26 1995.

10. B. D. Jeffs and W. H. Pun, "Simple shape parameter estimation from blurred observations for a generalized Gaussian MRF image prior used in MAP image restoration," in *Proc. of IEEE Int'l Conf. on Image Proc.*, (Lausanne, Switzerland), pp. 465–468, September 16-19 1996.

11. V. Solo, "A SURE-fired way to choose smoothing parameters in ill-conditioned inverse problems," in *Proc. of IEEE Int'l Conf. on Image Proc.*, (Lausanne, Switzerland), pp. 89–92, September 16-19 1996.

12. S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields, with applications to Bayesian tomography," *IEEE Trans. on Image Processing*, **7**, pp. 1029–1044, July 1998.

13. C. Klifa and B. Lavayssière, "3D reconstruction using a limited number of projections," in *Proc. of the SPIE Conference on Visual Communications and Image Processing*, vol. SPIE-1360, (Lausanne, Switzerland), pp. 443–454, 1990.

14. J. J. Sachs and K. Sauer, "Reconstruction from sparse radiographic data," in *Discrete Tomography: Foundations, Algorithms, and Applications*, G. Herman and A. Kuba, eds., pp. 363–383, Birkhäuser, Boston, MA, 1999.

15. L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statistics*, **37**, pp. 1554–1563, 1966.

16. L. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statistics*, **41**, (1), pp. 164–171, 1970.

17. A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, **39**, (1), pp. 1–38, 1977.

18. S. Azevedo, D. J. Schneberk, P. Fitch, and H. E. Martz, "Calculation of the rotational centers in computed tomography sinograms," *IEEE Trans. on Nuclear Science*, **NS-37**, pp. 1525–1540, August 1990.

19. T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, **8**, pp. 194–202, June 1989.

20. P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. on Medical Imaging*, **9**, pp. 84–93, March 1990.

21. C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. on Image Processing*, **5**, pp. 480–492, March 1996.

22. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, **21**, pp. 1087–1091, 1953.

23. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **PAMI-6**, pp. 721–741, Nov. 1984.

24. J. Prince and A. Willsky, "A projection space MAP method for limited angle reconstruction," in *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, (New York), pp. 1268–1271, April 11-14 1988.

25. A. Rosenfeld and A. Kak, *Digital Picture Processing*, vol. 1, Academic Press, San Diego, 1982.