Tomographic Reconstruction of Images from Noisy Projections - A Preliminary Study

A.P. Dalgleish[1], D.L. Dowe[1] and I.D. Svalbe[2]

[1] Clayton School of IT, Monash University, Australia
[2] School of Physics, Monash University, Australia

Abstract. Although Computed Tomography (CT) is a mature discipline, the development of techniques that will further reduce radiation dose are still essential. This paper makes steps towards projection and reconstruction methods which aim to assist in the reduction of this dosage, by studying the way noise propagates from projection space to image space. Inference methods Maximum Likelihood Estimation (MLE), Akaike's Information Criterion (AIC) and Minimum Message Length (MML) are used to obtain accurate models obtained from minimal data.

1 Introduction

The optimal reconstruction of images from traditional Computer Tomographic (CT) projections, in the presence of noise, is a real-world problem. High-resolution CT scans deliver a large amount of radiation to the patient. CT scans are used frequently by physicians to diagnose different forms of cancer and many other conditions. The number of CT scans performed has increased, from 3 million in 1980, to 62 million in 2006 [1]. CT scans now account for half the radiation dose that an American receives per year.

We are interested in understanding how noise propagates through projections, to the reconstructed images, through a known reconstruction method, the Finite Radon Transform. It is hoped that through this understanding, we can take steps towards obtaining lower dose CT scans with the same resolution.

Minimum Message Length (MML), Maximum Likelihood Estimation (MLE) and Akaike's Information Criterion (AIC) are used in this context to model the noise in the CT reconstruction data, as well as modelling the underlying data. It is hoped that the use of these modelling techniques will allow a means to reconstruct images from minimal data.

Current CT methods of reconstructing images from projection data include Filtered Back Projection and Fourier Inversion. These methods use approximations in the methods they use for the reconstruction of projection data, the Fourier Inversion for example interpolates between projection values to obtain the reconstructed image. This masks how the noise propagates through the projections to the reconstructed image.

The reconstruction method used here is an exact reconstruction method, called the Finite Radon Transform (FRT). This method does not use any approximations or any other a priori methods of data manipulation. There are

however two constraints that are imposed on this method of reconstruction; they are that the data is assumed to be in form of a discrete matrix, and that this matrix is mapped to a torus.

The Finite Radon Transform is used in this preliminary study because it has been shown [2] that projections from real X-ray CT data can be mapped into the FRT projection form via the use of the Mojette transform [3], as an intermediary stage. This approach bypasses the need for approximation or interpolation in the reconstruction method, as the FRT is exactly invertible. All of the required data processing can be done on one-dimensional projection data. Also, as the CT data that is acted upon is digital data, it can be assumed to be discretized.

Understanding the propagation of noise from FRT through Mojette to real CT projection data and the reconstructed image is an essential prelude to the practical application of the FRT for tomographic reconstruction.

Noise in real CT projections arises from many different sources, these include X-ray scatter, Quantum noise of the X-rays (X-ray photon statistics), the data acquisition system - the electronic noise produced by this system, can combine with the photon noise.

By studying the effect that noise has on the inversion of projection data, it is hoped that we can use this information to assist in the correction of noisy FRT projection data. This information could then be used to assist in the reconstruction of projection data from other reconstruction methods.

The FRT has not been used in this context before, and that makes this research new and novel. Whilst the FRT is not comparable with any of the current methods of CT projection and reconstruction, it is possible to convert the projection data from conventional CT projections to another projection-reconstruction method, the Mojette method. This method is much closer to conventional CT methods than the FRT method. Once the data is in the form of the Mojette projection-reconstruction method, this data can be converted to the Finite Radon Transform.

The beauty of FRT is the fact that it has a fixed pre-defined set of projections, so there is no need for us to choose projection angles. For a noise free set of projection data of a PxP matrix, when P is a prime number, then the inverse also has a well defined algorithm, which makes the inverse simple for noise free data [4]. The exactness of FRT reconstruction method is the property that we have exploited for this research, in conjunction with the data estimation techniques Maximum Likelihood Estimation (MLE), Akaike's Information Criterion (AIC) and Minimum Message Length (MML).

2 Discrete Projection Techniques

The Finite Radon Transform was used for this research because it provides a means for exact reconstruction without the need for any approximations. The Mojette Projection and Reconstruction methods provide the link between the Finite Radon Transform, and conventional CT methods, since there is a known mapping between the Finite Radon Transform and the Mojette method, and

there is also a known mapping between the Mojette method and conventional CT data.

2.1 Finite Radon Transform (FRT) Projections

The data used for the Finite Radon Transform (FRT) is in the form of a discrete, square matrix of size $P \times P$, called an 'Image Matrix', where P is a prime number. As stated before, the image matrix is also assumed to be mapped to a torus, allowing projections of the image data to wrap around the sides of the image matrix [5]. The next 'pixel' of the image array that is part of the same projection to be sampled is determined by moving m pixels across and 1 pixel down. Since the image matrix is a torus, the movements can wrap around the sides.

The definition of the Finite Radon Transform given here is quite concise. For an in-depth discussion of the Finite Radon Transform and the projection and reconstruction methods, see [4, 5].
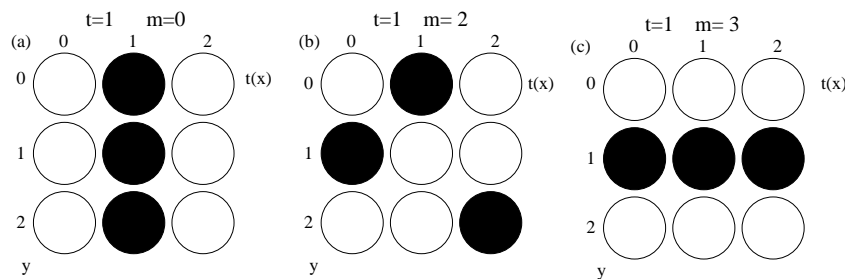

Fig. 1. FRT projections for a $P \times P = 3 \times 3$ image matrix (a) A projection where $t = 1$ and $m = 0$. (b) A projection where $t = 1$ and $m = 2$. (c) Projection $m = P$ and $t = 1$ sums the pixel values along row 1.

2.2 Mojette Projections

The Mojette technique is an asymmetric projection technique, where the projection angles are specified by discrete rational angles $\tan^{-1}(q/p)$ and the beam spacing by $1/\sqrt{p^2 + q^2}$, where p and q are integers and $q > 0$. Thus the projection value becomes the sum of the pixels, whose centres lie on the line $t = qx - py$ [4]. This means that image matrix no longer has to be mapped to a torus. The Mojette projections are close in form to that of the real projection data acquired in a traditional CT scan.

For further information on the Mojette Transform see [4, 6].

3 Modelling Techniques

Recall from the above sections that the FRT and Mojette projection and reconstruction methods are efficient and effective, projection and reconstruction methods. However they both suffer problems in the presence of noise. There are several corner based Mojette inversion methods [6], in which noise in the projection matrix is propagated in from the corners, and as the corners of the projection matrix tend to have larger amounts of noise, the propagation of noise can be substantial. In the Finite Radon Transform noise from projection data is carried from projection space to image space. The FRT is an exact reconstruction method that spreads a uniform amount of noise from projections into the image.

The poor performance of these methods in the presence of noise provides the need for a method of noise estimation, outlined below are three proposed methods for noise estimation, namely Maximum Likelihood Estimation (MLE), Akaike's Information Criterion (AIC) and Minimum Message Length (MML).

The research consisted of two main noise models that were added to the image matrix, the image matrix which was a fixed binary image. In the first of these models, Gaussian noise was added to the image matrix, with known $\mu = 0$, we varied the sigma values from $0 < \sigma < 1$. The value of μ was assumed to be known when using the modelling techniques, this model is known as the 'No Cut' model. The data modelling techniques were used to infer the value of σ.

For the other noise model, Gaussian noise was added to the image in two distinct regions, that is two different values for σ, namely σ_l and σ_r, once again we set a known $\mu_l = \mu_r = \mu = 0$. These regions were separated via a straight line, whose centre lay in the centre pixel, the line had associated with it an angle θ, allowing us to specify different regions of noise. This is referred to as the 'Cut' method, since the line 'cuts' the image into two noise regions. In this case, data modelling techniques were used to infer the value of θ as well as the value of the two σ values, from the two distinct noise regions. Figure 2 demonstrates an example 'cut' of a 7×7 image, it also shows the assumption that was used in our testing, this was that any pixel through which the line passes through the centre is counted as part of the σ_r side.

The data modelling methods that are outlined below all use the following method for encoding θ, $A = \sum_{i=1}^{N}(\phi(i))$, where A represents the total number of angles that have been examined for a particular matrix size P, $N = \frac{P-1}{2}$, and ϕ is the Euler's Totient. If we assume that the angles are ordered by size, then to refer to a certain angle, we can simply refer to the angle via an index i, where $0 < i < A$. Therefore we encode θ as $\log{(A)} + \log{(8)}$, where $\log{(8)}$ encodes the octant, which the line passes through. The value of A is fixed for a particular value of P, due to the fact that we only examine rational angles in the range $0 \leq \theta < 45$. The FRT and Mojette both use rational angles for constructing the projection space.
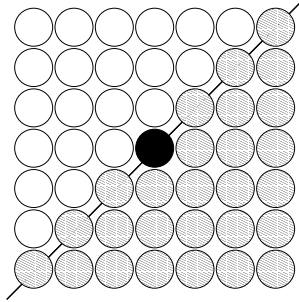
Fig. 2. An example cut, $\theta = 45°$, σ_l applies to unshaded pixels, σ_r applies to shaded pixels, μ is set to 0 and is known.

3.1 Maximum Likelihood Estimation (MLE)

Maximum Likelihood is based on the concept of conditional probability. The 'Likelihood' of parameter d given an event E is $P(d|E)$, in this case we use the observed data to estimate the parameters.

A drawback of Maximum Likelihood is that it will always prefer a more complex model, thus in our tests it always chooses the model with the 'cut', as per (2), over the model with no 'cut', as per (1),

$$L(x|\mu,\sigma) = \tfrac{1}{2} \times P^2 \times \log(2\pi) + P^2 \times \log(\sigma) + \tfrac{1}{2\sigma^2} \times \sum_{i=0}^{P^2-1}(x_i - \mu)^2 \quad (1)$$

where P is the size of the image matrix, and x_i is a pixel in the matrix and

$$L(x|\mu,\sigma_l,\sigma_r) = n_l \times \log(\sqrt{2\pi}\sigma_l) + \tfrac{1}{2\sigma_l^2}\sum_{i=0}^{n_l-1}(x_i-\mu)^2 + n_r \times \log(\sqrt{2\pi}\sigma_r) +$$
$$\tfrac{1}{2\sigma_r^2}\sum_{i=n_l}^{P^2-1}(x_i-\mu)^2 + \log(A) + \log(8)$$

$$(2)$$

where n_l and n_r refer to the number of pixels to the left and right of the line which separates the two noise regions, $n_l + n_r = P^2$ and $n_l \leq n_r$.

In (1) there is assumed to be one Normal noise distribution to which all pixels belong, and as the image matrix is square, and of size $P \times P$, there are P^2 pixels. For (2), the noise is assumed to come from two separate noise distributions that are partitioned by the 'cut'.

3.2 Akaike's Information Criterion (AIC)

Maximum Likelihood Estimation can be improved by using the 'Akaike Information Criterion' (AIC), this was proposed by Akaike in 1970, and is a method of performing *penalised maximum likelihood*. The addition of this penalty factor means that, although there may be a model with a higher [negative] log likelihood, this may not be chosen if its model is sufficiently complex [7].

Akaike's Information Criterion attempts to minimize the quantity

$$AIC = 2 \times L + 2K \qquad (3)$$

where L refers to the [negative] log-likelihood, such as either (1) or (2) above, and K refers to the number of free parameters.

3.3   Minimum Message Length (MML) Models

MML is a technique for modelling data, originally proposed by Wallace and Boulton [8], and based on Occam's Razor, 'the simplest explanation is the best'. This means that with MML if there are two messages that model the data equally well, the one with the shortest length will be chosen. If we consider our data as a string S, then a message (I) is defined as being two parts, $I = H \; : \; A$, where ':' is concatenation, H is a hypothesis about the source of the data string S, and A is an encoding of the data using a code the would be optimally efficient if the hypothesis H, were true [9].

For further discussion of the MML coding technique see [8–10], for a detailed comparison between AIC and MML, see [11]. The approximation to the message length that was explored in this research, is the Quadratic Approximation to Message Length method, for a full discussion of this method see [12, 10].

The general form of this approximation is

$$MessageLength \approx -\log\left(\boldsymbol{\theta}\right) + L + \frac{1}{2}\log F + \frac{D}{2}(1 - \log\left(12\right)) \qquad (4)$$

where $\boldsymbol{\theta}$ is the prior, L is the [negative] log-likelihood, F is the determinant of the expected Fisher Information Matrix and D is the dimensionality.

Minimum Description Length (MDL) is a method that is very similar to MML, for more information on MDL see [13]. For a discussion of the relationship between MML and MDL see [14, sec. 11.4] and [10, sec. 10.2].

There were two possible models for MML to use.
These message length for the 'no cut' model (with $D = 1$) is given by

$$MessageLength_{noCut} = \log\left(2R\sigma\right) + \frac{1}{2} \times P^2 \times \log\left(2\pi\right) + P^2 \times \log\left(\sigma\right) + \frac{1}{2\sigma^2} \times \sum_{i=0}^{P^2-1}(x_i - \mu)^2 \times \frac{1}{2}\log\left(\frac{3}{\sigma^2} \times \frac{P^2}{\sigma^2}\right) + \frac{D}{2}(1 - \log\left(12\right)) \qquad (5)$$

where P refers to the size of the image matrix, and $R = 6$.
In this message length the prior is $\frac{1}{2R\sigma}$ (for $e^{-R} \leq \sigma \leq e^R$) , L is given by (1) and the expected Fisher Information is $\left(\frac{3}{\sigma^2} \times \frac{P^2}{\sigma^2}\right)$.

The message length for the 'cut' model (with $D = 2$) is given by

$$MessageLength_{cut} = \log\left(2R\sigma_l\right) + \log\left(2R\sigma_r\right) + \log\left(A\right) + \log\left(8\right) + n_l \times \log\left(\sqrt{2\pi}\sigma_l\right) + \frac{1}{2\sigma_l^2}\sum_{i=0}^{n_l-1}(x_i - \mu)^2 + n_r \times \log(\sqrt{2\pi}\sigma_r) + \frac{1}{2\sigma_r^2}\sum_{n_l}^{P^2-1}(x_i - \mu)^2 + \frac{1}{2} \times \log\left(\frac{3}{\sigma_l^2} \times \frac{3}{\sigma_r^2} \times \left(\frac{n_l}{\sigma_l^2} + \frac{n_r}{\sigma_r^2}\right)\right) + \frac{D}{2}(1 - \log\left(12\right))$$

$$(6)$$

where n_l and n_r refer to the number of pixels to the left and right of the line which separates the two noise regions, $n_l + n_r = P^2$ and $n_l \leq n_r$, P refers to the size of the matrix, and $R = 6$.

In this Message Length the prior is given by $\frac{1}{2R\sigma_l} \times \frac{1}{2R\sigma_r}$ (for $e^{-R} \leq \sigma_l \leq e^R$ and $e^{-R} \leq \sigma_r \leq e^R$), L is given by (2) and the expected Fisher Information is given by $\left(\frac{3}{\sigma_l^2} \times \frac{3}{\sigma_r^2} \times \left(\frac{n_l}{\sigma_l^2} + \frac{n_r}{\sigma_r^2} \right) \right)$.

4 Results

Testing involved varying values of σ for the 'No Cut' model, where $0 < \sigma < 1$, and varying σ_l, σ_r and θ, for the 'Cut' model, where $0 < \sigma_l < 1$, $0 < \sigma_r < 1$, and $0 \leq \theta < 45$. Tables 1 and 2 show sample results from our testing. All tests were run on an AMD Athlon X2 4400, with 2Gb RAM, with run-times in the order of 35 seconds for $P = 101$.

The Kullback-Leibler (KL) Distance [10, Ch. 4] [11, sec. 6.1.4] for two different Normal distributions, from $N(\mu_1, \sigma_1^2)$ to $N(\mu_2, \sigma_2^2)$ is given by (7).

$$KL = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} \tag{7}$$

Table 1. Results of an average over 100 runs per entry, with varied matrix (P) sizes, with one region of noise having value of $\sigma = 0.9$ and known $\mu = 0$, the true model in this case has no 'cut'. In Tables 1 and 2 bold entries denote minimum at each P.

Method	P	Mean KL $\times 10^3$	σ_{KL} $\times 10^4$
MLE	7	3.3775	9.0906
AIC	7	3.1871	8.0018
MML	7	**2.8444**	2.8233
MLE	13	3.0014	4.5181
AIC	13	3.4823	4.7807
MML	13	**2.7876**	1.3227
MLE	23	2.9298	3.0811
AIC	23	2.9376	3.5868
MML	23	**2.7816**	0.8471
MLE	29	2.9090	2.6427
AIC	29	2.8756	2.4625
MML	29	**2.7771**	0.5874
MLE	31	2.8610	2.5891
AIC	31	2.8588	2.5151
MML	31	**2.7746**	0.6432
MLE	101	2.8050	0.9532
AIC	101	2.7964	1.3412
MML	101	**2.7767**	0.1920

Table 2. Results of an average over 100 runs per entry, with varied matrix (P) sizes, with two distinct regions of noise having values of $\sigma_l = 0.7$, $\sigma_r = 0.4$, $\theta = atan(0.5)$ and known $\mu = 0$. σ_{KL} refers to the standard deviation of the distribution of KL-Distances, averaged over 100 runs per entry, the true model in this case has a 'cut'.

Method	P	Mean KL $\times 10^3$	σ_{KL} $\times 10^4$
MLE	7	3.6339	4.9118
AIC	7	**3.6287**	4.2386
MML	7	22.9901	4.4659
MLE	13	3.4470	2.2414
AIC	13	3.5218	1.9877
MML	13	**3.4436**	2.1782
MLE	23	**3.3648**	1.2160
AIC	23	3.3754	1.0974
MML	23	3.3769	0.8456
MLE	29	**3.3706**	0.9451
AIC	29	3.3777	0.9969
MML	29	3.3769	0.8456
MLE	31	3.3658	0.9798
AIC	31	3.3770	0.8570
MML	31	**3.3584**	0.8657
MLE	101	3.3549	0.2566
AIC	101	3.3572	0.2738
MML	101	**3.3525**	0.6010

The results in Table 1 refer to experiments in which Gaussian noise was added to a predefined image, in this case the noise was assumed to be Gaussian noise spread across one region. Unlike the previous experiments, there is no 'cut' and no θ and there is only one σ value. These results highlight the predisposition of MLE and AIC to prefer the more complicated model. Given the chance of a model that has two distinct regions of noise associated with it (therefore two separate σ values and a θ value) and a model with only one region of noise (so only a single σ), the results show that AIC and MLE will choose the more complicated model more often than MML. The benefits of this choice are highlighted in the KL Distance values for the various data modelling techniques.

It should be noted that it is not inherently clear for AIC whether the number of free parameters refers to both discrete and continuous free parameters, or just continuous free parameters, testing was done on both, and the results of Tables 1 and 2 present the results when only continuous free parameters were considered, however the difference in KL values for the two separate cases were minimal.

The results in Table 2 refer to experiments in which Gaussian noise was added to a predefined image, however in this case there was assumed to be two regions of noise that may or may not have the same σ value. The noise was assumed to be from two separate normal distributions, with a known μ value common to the two normal distributions, and a separate unknown σ value for each of the distributions. The regions were partitioned by a line, that went through the

center pixel of the image, and has associated with it an angle θ, hence the noise model has a 'cut'.

The noise in both the 'cut' and 'no cut' cases was added directly to the image itself and not to the projections. This was done as a prelude to the addition of noise to projection data. This circumvents the need to 'untangle' the projections before performing the noise estimation methods, but work will be done on this at a later stage.

Unlike MLE, both AIC and MML can be used as a means of model selection, that is, these noise estimation methods can be used to choose between two different noise models. In our tests, these methods were used to estimate the values of the σ or the two σ values and the θ value. MLE will always choose the model with two σ values and the θ value, as this will maximise the likelihood. AIC and MML will attempt to choose between this model and the model with only one σ value.

5 Conclusion

This research is a preliminary study into how noise propagates through projections from projection space to image space. The use of data modelling techniques should allow for the use of our understanding of noise propagation, to help us model noisy projection data. Minimum Message Length (MML) is well suited to this, since it works well when there are small amounts of data from which we are attempting to infer parameters.

At this point however the study has been limited to the study of noise estimation in image space, with Maximum Likelihood Estimation (MLE), Akaike's Information Criterion (AIC) and Minimum Message Length (MML) as methods for data modelling.

The Finite Radon Transform is used in this context since it provides a means of projecting and reconstructing image data exactly, as it does not use any approximations or interpolation, so it is clear how the noise is propagating. MLE, AIC and MML are being used in this context as it for the data modelling methods. For these initial tests they all have a similar performance, however these tests are concerned with data that is much simpler than would be found in real world CT data.

It is our hope that through continued study of this we will be able to construct a noise tolerant method of projection and reconstruction of Tomographic data, that will allow for CT scans to have the same anatomical detail with a reduction in the amount of radiation that the patient receives.

The MML work would be enhanced by more properly modelling the spatial correlation within images [15, 16].

References

1. Rabin, R.C.: With Rise in Radiation Exposure, Experts Urge Caution on Tests. The New York Times (19 June 2007)

2. Svalbe, I., van der Spek, D.: Reconstruction of Tomographic Images using Analog Projections and the Digital Radon Transform. Linear Algebra and its Applications **339**(15) (2001) 125–145

3. Normand, N., Guédon, J., Philippé, O., Barba, D.: Controlled Redundancy for Image Coding and High-speed Transmission. Proc. SPIE Visual Communications and Image Processing **2727** (1996) 1070–1081

4. Kingston, A., Svalbe, I.D.: Projective Transforms on Periodic Discrete Image Arrays. Advances In Imaging and Electron Physics **139** (2006) 75–177

5. Matus, F., Flusser, J.: Image Representations via a Finite Radon Transformation. IEEE Transactions: Pattern Analysis and Machine Intelligence **15**(10) (October 1993) 996–1006

6. Normand, N., Kingston, A., Éuenou, P.: A Geometry Driven Reconstruction Algorithm for the Mojette Transform. Discrete Geometry for Computer Imagery, 13th International Conference (2006) 122–133

7. Akaike, H.: Statistical prediction information. Annals of the Institute of Statistical Mathematics **22** (1970) 203–217

8. Wallace, C.S., Boulton, D.M.: An Information Measure for Classification. Computer Journal **11**(2) (1968) 185–194

9. Wallace, C.S., Dowe, D.L.: Minimum Message Length and Kolmogorov Complexity. Computer Journal **42**(4) (1999) 270–283

10. Wallace, C.S.: Statistical and Inductive Inference by Minimum Message Length. Springer, New York (2005)

11. Dowe, D.L., Gardner, S., Oppy, G.R.: Bayes not Bust! Why Simplicity is no problem for Bayesians. British Journal for the Philosophy of Science (December 2007) Forthcoming.

12. Wallace, C.S., Freeman, P.R.: Estimation and inference by compact coding. Journal of the Royal Statistical Society. Series B (Methodological) **3**(49) (1987) 240–265

13. Rissanen, J.: Modelling by the shortest data description. Automatica **14** (1978) 465–471

14. Comley, J.W., Dowe, D.L.: Minimum Message Length and Generalized Bayesian Nets with Asymmetric Languages. Advances in Minimum Description Length: Theory and Applications (April 2005) 265–294 Chapter 11.

15. Wallace, C.S.: Intrinsic Classification of Spatially Correlated Data. Computer Journal **41**(8) (1998) 602–611

16. Visser, G., Dowe, D.L.: Minimum Message Length Clustering of Spatially-Correlated Data with Varying Inter-Class Penalties. 6th IEEE International Conference on Computer and Information Science (ICIS 2007) (July 2007) 17–22