Energy Dependent Information in X-Ray Imaging: Part 2. Information Extraction and Noise

Robert E. Alvarez and Leonard A. Lehmann

April 11, 2013

The information in the energy spectrum of X-ray photons transmitted through the body can be extracted by using vector space techniques. This information can be used to reduce errors due to the two important sources of extraneous detail in X-ray images: overlying anatomical structures and random noise. The effect of intervening anatomy can be reduced by techniques that use the energy dependent information to selectively remove the effect of materials of a given chemical composition from the image, The effect of random noise is reduced by energy selective systems because they extract more information for the same dose than conventional systems. Energy selective systems can produce conventional images with the same noise variance as conventional systems and they can detect small objects with a better signal to noise ratio than conventional systems.

Contents

1 Introduction

The first paper of this series (Alvarez 1983[2]) presented a vector space method for representing the energy dependence of X-ray attenuation and for extracting this information from simple measurements. In this paper, we present techniques that use this information to derive medically useful images and study the noise in these images.

One of the most important sources of diagnostic errors in radiological imaging is extraneous detail. There are two principal sources of this extraneous detail in medical images. The most commonly noted is X-ray quantum noise and other random noise in the measurement apparatus. However, another equally important source of noise is "clutter" due to overlying anatomical features, which can mask the diagnostically significant indicia. Energy selective methods can be used to significantly reduce the effects of both types of noise.

In the first part of this paper, we describe techniques that can remove the effect of a material of a given composition from the image. This type of processing is unique to energy selective radiography and is effective in removing anatomical clutter from images. It can be used, for example, to produce chest radiographs with either bone or soft tissue removed. It is also useful in computed tomography for producing data that are independent of the effects of a particular material.

In the second half of the paper, we discuss the effect of noise in energy selective systems. It might seem that the extra information produced by these systems would result in increased patient dose. The results presented indicate that this is not true. The noise in the conventional images produced from the energy selective information can be made equal to that in a conventional system using the same total dose. Furthermore, the signal to noise ratio for a medically important class of imaging tasks will be shown to be better than (or at worst equal to) the signal to noise ratio for a conventional system performing the same task.

The information produced by energy selective systems can lead to improved medical diagnoses. The results presented in these papers form a sound theoretical basis for extracting the energy selective information as well as a framework for developing new techniques for usefully presenting it to the user.

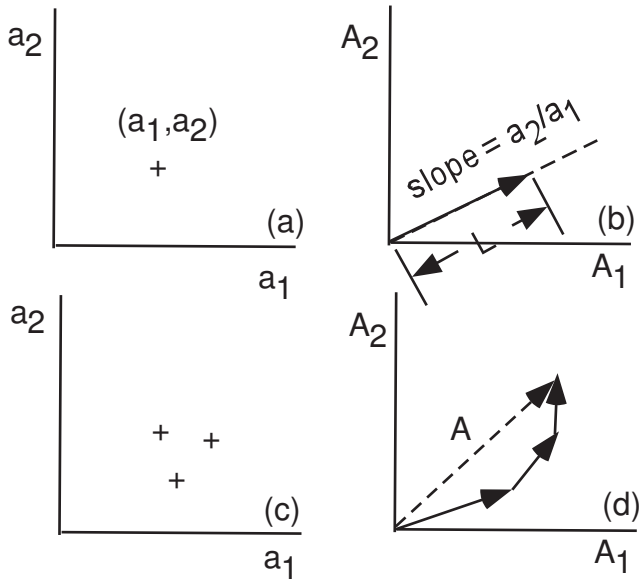

Figure 1: Vector space representation of simple materials. Part (a) shows the basis set coefficients for a pure material composed on one substance. Part (b) shows the line integrals. These lie along a line with slope a_2/a_1 and length proportional to the object thickness, L. Part (c) shows an object composed of three materials, resulting in three points in (a_1, a_2) space. The total vector A in Part (d) is the sum of three vectors from each of the three materials.

2 Techniques for Information Extraction

The data produced directly by energy selective X-ray systems are images of the basis set coefficients (in a computed tomography system) or their line integrals (in a single projection system). These images can be used directly but we have developed methods to extract information from the data and make it more useful for diagnosis. In this section we describe the physical bases for these techniques and show how the vector space description can be used to understand them and form the basis for developing new ones.

2.1 Vector Space Representation of Simple Materials

In order to understand the signal processing techniques it is instructive to consider the vector space representation for simple types of materials.

The simplest case is an object composed of a single compound or a mixture of fixed proportions with particles smaller than the resolution of the system. In this case the basis set coefficients are fixed and the material can be represented as a single point (a_1, a_2) in a two dimensional plot as in Figure 1a. The line integrals of the basis set coefficients will be

$$A_i = a_i L \quad (i = 1, 2) \tag{1}$$

where L is the thickness of the material along the path of the X-ray beam. Thus the two dimensional representation of the line integrals in this case will be a straight line through the origin with slope s,

$$s = \frac{a_2}{a_1} \tag{2}$$

and length proportional to the thickness as shown in Figure 1b.

Any real object will be composed of more than one material. The basis set coefficients for this case will be a set of points on the two dimensional representation, one for each type of material as shown in Figure 1c. The line integrals for this case are

$$A_i = \int a_i(x, y, z)dl \quad (i = 1, 2) \tag{3}$$

The total vector in the two dimensional representation is the vector sum of the contributions due to the individual materials, Figure 1d.

2.2 Synthesized Monoenergetic Images

Using the models developed in the previous section, signal processing techniques can be devised to extract medically useful information. Perhaps the simplest form of processing is to form images representing the same physical quantities imaged in conventional systems but at an adjustable display energy, E_d. In CT, we can form images of the linear attenuation coefficient, $\mu(E_d)$, while in single projection systems we can form images of the line integral

$$L(E_d) = \int \mu(x, y, z; E_d)dl. \tag{4}$$

These images have physical characteristics similar to those of conventional systems but the display energy is not determined by the physical characteristics of the system but is a parameter under our control. These images represent a single energy so they are called synthesized monoenergetic images. Since the basis functions are known a-priori, any desired display energy can be used as E_d so the conventional image is a subset of the energy selective data and an energy selective system extracts more information than a conventional system.

In CT, the calculation of synthesized monoenergetic images is based on the fundamental vector space representation

$$\mu(E) = a_1 f_1(E) + a_2 f_2(E). \tag{5}$$

In an energy selective CT system, where (a_1, a_2) are determined at points in the object cross section, a display image can be calculated by carrying out the mathematical operation indicated in equation (5) at every point in the image using the basis functions evaluated at the display energy $f(E_d), (f_2(E_d)$. Note that the energy selective CT image is not subject to beam hardening artifacts (Alvarez and Seppi 1979[3]).

Analogous results can be obtained for the line integral of the attenuation coefficient $L(E)$ in equation (4), which can be expressed in terms of the line integrals of the basis set coefficients A_i in equation (3) as

$$L(E_d) = A_1 f_1(E_d) + A_2 f_2(E_d). \tag{6}$$

The synthesized monoenergetic image has a useful vector interpretation. The fundamental decomposition, equation (5), can be considered a dot product between two vectors. One, with components (a_1, a_2), depends on the material characteristics while the other, with components $f_1(E_d), f_2(E_d)$ depends only on the chosen energy. The operation of calculating a conventional image from the energy selective information is then equivalent to projecting the vector representing the basis set coefficients onto a vector representing the values of the basis functions at the display energy and then adjusting the scale.

2.3 Selective Material Images

Selective material imaging relies on the observation, discussed in the first part of this section, that the vector representation for a given material always lies along a single line. If a generalized projection is formed perpendicular to this line at every point in the image, then variations in that material will not contribute to the resultant image. It will, effectively, have been canceled.

Suppose that in a single projection system the object consists of a feature of interest plus a constant background. Then, as shown in figures 4a and 4b, the total vector will consist of a constant vector $\mathbf{A}_b$ plus a variable vector $\mathbf{A}_f$. Making a generalized projection at an angle ϕ_b perpendicular to $\mathbf{A}_b$, will cancel the background material while a generalized projection at an angle ϕ_f perpendicular to $\mathbf{A}_f$, will cancel the feature.

If the object consists of a feature of interest completely surrounded by a uniform background the results are similar. This case can be transformed to the previous case by defining an effective material with basis set coefficients equal to the difference between the feature coefficients and the background coefficients (Lehmann 1982[4]):

$$\mathbf{a}_d = \mathbf{a}_f - \mathbf{a}_b. \tag{7}$$

By forming a generalized projection perpendicular to the line integrals of the effective material $\mathbf{a}_d$, the embedded material will be canceled. This situation is shown in figures 4c and 4d. Note that in this case the cancellation angle depends on the basis set coefficients of both materials. Thus it depends not only on their ratio but on their magnitude.

2.4 Generalized Projection Signal Processing

Thus far we have discussed two uses of energy selective data: synthesized monoenergetic imaging and selective material cancellation, Although these two uses produce very different appearing images, they are both generalized projections of basis set data and therefore are closely related.

Lehmann (1982[4]) shows that with the proper choice of basis functions there are two disjoint regions of projection angle. The first quadrant corresponds to synthesized monoenergetic images and the second quadrant corresponds to selective material removal images. Projection vectors within the third and fourth quadrants generate images which are contrast reversed from those in the first and second quadrants, respectively, but contain no new information. Thus synthesized monoenergetic images and selective material

cancellation images are the only types of images that can be calculated by a generalized projection.

3 Noise Variance and Covariance

The measurements used by energy selective systems are random quantities and the estimation of the energy dependent information should be based on statistical techniques. Previous results (Alvarez and Macovski 1976[1]) derived a maximum likelihood estimator for the basis set coefficient line integrals that leads to a simple and intuitively appealing procedure: The estimator solves the deterministic equations relating the line integrals and the transmitted flux using the actual measurements as an estimate of the flux. The previous results also derived expressions for the variances of the line integral estimates based on the Cramer-Rao lower bound for maximum likelihood estimators. In this section, we derive these expressions using matrix methods. This derivation is useful in its own right and also introduces notation that will be used in following sections. We also derive matrix expressions for the variance of a linear combination of line integrals for later use. Finally, we discuss the relationship of the noise to the x-ray imaging system properties.

As discussed in the first paper in this series, a dual energy system computes the line integrals of the basis set coefficients from the flux measurements with two different effective spectra. These are related by two integral equations

$$I_1(A_1, A_2) = \int S_1(E) exp[-A_1 f_1(E) - A_2 f_2(E)] dE \tag{8}$$

$$I_2(A_1, A_2) = \int S_2(E) exp[-A_1 f_1(E) - A_2 f_2(E)] dE. \tag{9}$$

Introducing logarithms in these equations approximately linearizes them, which is convenient for the analysis of noise where the deviations are relatively small. Equations (8 and 9) are then expressible in matrix form as

$$L = \begin{bmatrix} \log(I_1(A_1, A_2)) \\ \log(I_2(A_1, A_2)) \end{bmatrix} = \begin{bmatrix} L_1(A_1, A_2) \\ L_2(A_1, A_2) \end{bmatrix}. \tag{10}$$

Expressing small deviations of L from a mean value using a Taylor's series

$$L(A_0 + \delta A) \approx L(A_0) + \frac{\partial L}{\partial A} \delta A. \tag{11}$$

Defining

$$M = \frac{\partial L}{\partial A} = \begin{bmatrix} \frac{\partial L_1}{\partial A_1} & \frac{\partial L_1}{\partial A_2} \\ \frac{\partial L_2}{\partial A_1} & \frac{\partial L_2}{\partial A_2} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \tag{12}$$

we can express the linear terms in the Taylor's series expansion in equation (11) as

$$\delta L = L(A_0 + \delta A) - L(A_0) \approx M\delta A. \tag{13}$$

The covariance of L is, by definition

$$R_L = cov(L) = Ex\left[(L - \overline{L})^2\right] \tag{14}$$

where $Ex(\cdot)$ is the mean value. For small deviations $\overline{L} \approx L(A_0)$ and $\overline{A} \approx A_0$ so

$$R_L \approx Ex\left[(\delta L)(\delta L)^T\right] \quad and \quad R_A \approx Ex\left[(\delta A)(\delta A)^T\right] \tag{15}$$

where T designates the matrix transpose operation. Solving equation (13) for δA,

$$\delta A = M^{-1}\delta L \tag{16}$$

so the covariance of A is

$$R_A = Ex\left[\left(M^{-1}\delta L\right)\left(\delta L^T M^{-T}\right)\right]. \tag{17}$$

where $M^{-T} = \left(M^{-1}\right)^T$. Since the M matrix is deterministic, we apply the expectation operator only to the δL factors and

$$R_A = M^{-1}R_L M^{-T}. \tag{18}$$

We assume that the measurements are statistically independent so

$$R_L = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \tag{19}$$

where $\sigma_i^2 = var(\log(I_i)) \quad i = 1, 2$. The inverse matrix M^{-1} is

$$M^{-1} = \frac{1}{J}\begin{bmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{bmatrix} \tag{20}$$

where

$$J = M_{11}M_{22} - M_{12}M_{21}. \tag{21}$$

Substituting into equation (18)

$$R_A = \frac{1}{J^2}\begin{bmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{bmatrix}\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\begin{bmatrix} M_{22} & -M_{21} \\ -M_{12} & M_{11} \end{bmatrix}. \tag{22}$$

Multiplying this out results in the expressions:

$$Var(A_1) = \frac{M_{22}^2\sigma_1^2 + M_{12}^2\sigma_2^2}{J^2} \tag{23}$$

$$Var(A_2) = \frac{M_{21}^2\sigma_1^2 + M_{11}^2\sigma_2^2}{J^2} \tag{24}$$

$$Cov(A_1, A_2) = -\frac{M_{22}M_{21}\sigma_1^2 + M_{11}M_{12}\sigma_2^2}{J^2} \tag{25}$$

where σ_i^2 is the variance of the transmitted flux with spectrum I_i and

$$M_{ij} = \frac{\partial L_i}{\partial A_j} = \frac{\partial log(I_i)}{\partial A_j} \quad i,j = 1,2. \tag{26}$$

We can also express the covariance matrix as

$$R_A = \frac{1}{J^2}\begin{bmatrix} M_{22}^2\sigma_1^2 + M_{12}^2\sigma_2^2 & -(M_{22}M_{21}\sigma_1^2 + M_{11}M_{12}\sigma_2^2) \\ -(M_{22}M_{21}\sigma_1^2 + M_{11}M_{12}\sigma_2^2) & M_{21}^2\sigma_1^2 + M_{11}^2\sigma_2^2 \end{bmatrix}. \tag{27}$$

3.1 Variance of a Linear Combination

A matrix expression for the variance of a linear combination will be useful in later discussions. Suppose we form a linear combination of line integrals

$$P = p_1A_1 + p_2A_2 = p^TA \ , \ p = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}. \tag{28}$$

Let $\delta P = P - \tilde{P}$ and $\delta A = A - \tilde{A}$ so $\delta P = p^T\delta A$. Then,

$$Var(P) = Ex\left[(\delta P)(\delta P)^T\right] = Ex\left[(p^T\delta A)(\delta A^T p)\right] = p^TR_Ap. \tag{29}$$

3.2 Relationship of Noise to Physical System Properties

In each of the expressions for the noise variance and covariance the denominator is the square of a determinant that gives the conditioning of the two simultaneous equations that must be solved for the line integrals. The terms M_{ij} can be interpreted to be average values of the basis set functions over the two spectra used to measure the data. This may be seen by using the expressions for the flux I in equations (8 and 9),

$$I_i = \int S_i(E)e^{-A_1f_1(E)-A_2f_2(E)}dE \quad i = 1,2. \tag{30}$$

Thus M_{ij} is

$$M_{ij} = \frac{\partial log(I_i)}{\partial A_j} = \frac{1}{I_i}\frac{\partial I_i}{\partial A_j} \tag{31}$$

$$M_{ij} = -\frac{\int f_j(E) S(E) e^{-A_1 f_1(E) - A_2 f_2(E)} dE}{\int S(E) e^{-A_1 f_1(E) - A_2 f_2(E)} dE}.$$

Note that by equation (26), the M_{ij} are negative. To simplify the subsequent discussions, we will introduce new positive coefficients $m_{ij} = -M_{ij}$. Notice that we can use the positive coefficients in the noise equations (23, 24, and 25) without changing them. Finally, the term in the denominator normalizes the transmitted spectrum $S(E)e^{-A_1 f_1(E) - A_2 f_2(E)}$ so its integral is one. Thus m_{ij} is the average value of $f_j(E)$ over the spectrum transmitted through the body in measurement i.

The numerators of the expressions for the noise in equations (23 and 24) have terms which become smaller as the variance of the measurements become smaller. Thus, there are two factors which determine the overall noise. One is the conditioning of the equations expressed through the Jacobian determinant in equation (21), which is determined by the effective energy of the two spectra used in the measurement. The other factor is the noise in the individual measurements.

4 Noise Optimal Generalized Projections

In previous sections of this paper, generalized projections for a synthesized monoenergetic or for the cancellation of a particular material were described. While these projections may produce the image with desirable properties for a particular imaging task, they do not take noise into account. Depending on the a-priori knowledge of the object composition and the type of imaging task, particular projections may give better performance from the point of view of noise. In this section we discuss these noise optimal projections. Two types will be described. The first type produces a synthesized monoenergetic image at a display energy which gives a minimum variance. The second type is an optimal generalized projection image which maximizes the signal to noise ratio for the task of distinguishing a material from a background.

In this section we will be calculating optima of matrix expressions using matrix calculus. Some useful formulas are shown in Table 1

4.1 Minimum Variance Synthesized Monoenergetic Image

First, we describe how to compute an optimal display energy E_d that minimizes the variance of a synthesized monoenergetic image.

Suppose the image is a linear combination $P = p_1 A_1 + p_2 A_2 = p^T A$ where

$$p = \begin{bmatrix} f_1(E_d) \\ f_2(E_d) \end{bmatrix} = p(E_d). \tag{32}$$

As discussed previously in equation (29), the variance is

$$Var(P) = p^T R_A p. \tag{33}$$

Table 1: Matrix Calculus Formulas[5]

vector-scalar derivative $\frac{\partial Y}{\partial t} = \begin{bmatrix} \frac{\partial y_1}{\partial t} \\ \vdots \\ \frac{\partial y_n}{\partial t} \end{bmatrix}$	vector-vector derivative $\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}$
matrix-scalar derivative $\frac{\partial F}{\partial t} = \begin{bmatrix} \frac{\partial F_{11}}{\partial t} & \cdots & \frac{\partial F_{1m}}{\partial t} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_{n1}}{\partial t} & \cdots & \frac{\partial F_{nm}}{\partial t} \end{bmatrix}$	chain rule $\frac{\partial Z(Y(X))}{\partial X} = \frac{\partial Z}{\partial Y}\frac{\partial Y}{\partial X}$
Product rule $\frac{\partial\left(Y^T Z\right)}{\partial X} = \left(Z^T\right)\frac{\partial Y}{\partial X} + \left(Y^T\right)\frac{\partial Z}{\partial X}$	derivative of quadratic form $\frac{\partial\left(x^T Ax\right)}{\partial x} = x^T\left(A + A^T\right)$

So, applying the chain rule

$$\frac{\partial Var(P)}{\partial E_d} = \frac{\partial Var(P)}{\partial p}\frac{\partial p}{\partial E_d}. \tag{34}$$

Applying the rules for the derivative of a quadratic form and a vector from Table 1 and noting that R_A is symmetric so $R_A^T = R_A$, $\frac{\partial Var(P)}{\partial p} = 2p^T R_A$ and

$$\frac{\partial Var(P)}{\partial E_d} = 2p^T R_A p' \tag{35}$$

where

$$p' = \left[\begin{array}{c} \frac{\partial f_1}{\partial E} \\ \frac{\partial f_2}{\partial E} \end{array} \right]_{E=E_d}. \tag{36}$$

As shown in Appendix A the optimal display energy is given implicitly as the energy such that the basis functions take on the following values:

$$p_1 = f_1(E_{d_{opt}}) = \frac{m_{21}\sigma_1^2 + m_{11}\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{37}$$

$$p_2 = f_2(E_{d_{opt}}) = \frac{m_{22}\sigma_1^2 + m_{12}\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{38}$$

Physically, these expressions may be interpreted as defining the optimal display energy as a suitably defined average energy over the two spectra used in the measurement process. This may be seen more clearly for the case of a counting detector. Here the noise variance is equal to the average flux. Substituting in the equations for the optimal display energy yields (after some rearrangement):

$$f_1(E_{d_{opt}}) = \frac{m_{21}I_2 + m_{11}I_1}{I_1 + I_2} \tag{39}$$

$$f_2(E_{d_{opt}}) = \frac{m_{22}I_2 + m_{12}I_1}{I_1 + I_2} \tag{40}$$

where $I_1$and I_2 are the average transmitted fluxes with the two measurement spectra. Recall that the m_{ij} can be interpreted as the average value of f_i over spectrum i.

The expressions for the optimal display energy must be regarded as an approximation since there is no guarantee that a single value of E_d will satisfy both expressions in equation (37). However, computer simulation shows that these expressions are very close to the actual optimal value (Alvarez and Seppi 1979[3]). Furthermore, we show in Appendix A that the second derivative is positive at the optimal energy so the variance at the optimal display energy is a minimum.

The variance of the synthesized monoenergetic image at the optimal display energy may be calculated by substituting the optimal coefficients, equation (37), into equation (29).

$$Var(P) = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{41}$$

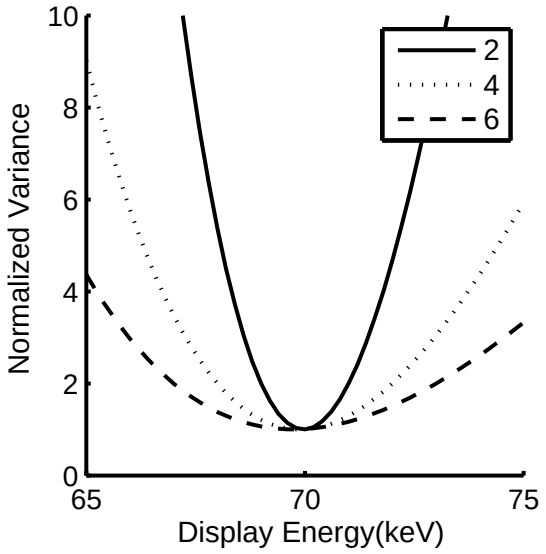

Figure 2: Variance *vs.* display energy for measurement spectra separated by 2, 4, and 6 Kev. The variance is normalized by dividing by the variance of a conventional image using the same total dose. Note that the minimum value is 1 for all the spectra. That is the dual energy and the conventional system have the same variance. Note also that as the energy difference of the measurement spectra gets smaller, the sensitivity of the results increases.

The implication of this result is clearer if the variance for a counting detector system $\sigma_i^2 = \frac{1}{I_i}$ is substituted. Then

$$Var(P) = \frac{1}{I_1 + I_2} \tag{42}$$

This is the variance of the line integral for a single spectrum with average flux equal to the sum of the fluxes in the two measurement spectra. The variance in equation (42) does not depend on the separation of the average energies, and hence the conditioning of the measurement process. This result can be used in computed tomography to produce a beam hardening corrected synthesized monoenergetic image from two spectra that are too ill-conditioned to produce low noise basis set coefficient images (Rutt and Fenster 1980[6]). The noise in the optimal image will be as small as for a well-conditioned measurement set. However, as shown in Appendix A, the second derivative, and hence the rate of the changes in the variance, at the optimal energy gets larger as the conditioning gets worse. Since due to beam hardening the transmitted spectrum will vary across the object, if the measurement set is too ill-conditioned then it may not be possible to define a single optimal energy for the whole object. This is illustrated in Figure 2, where the variance *vs.* display energy is plotted for several cases with varying separation in average energy between the two spectra.

4.2 Maximum Signal to Noise Ratio Projection

Consider the simple but widely applicable case of a feature of interest over a uniform background. If the composition of the feature is known *a-priori* then this information can be used to form a generalized projection image which maximizes the signal to noise ratio (SNR). At first glance it may seem that the angle which cancels the background material would be the optimum. However, this is not true. The optimal angle depends on the noise properties and on the composition of the feature of interest and of the background. In the next section, we show the perhaps surprising result that the SNR of an energy selective system with an optimal projection is greater than that of a conventional system except in degenerate situations such as the feature having the same composition as the background. Even in these degenerate situations, the SNR of the energy selective system is equal to that of the conventional system.

The derivation of the projection with optimal signal to noise ratio begins with a precise definition of signal to noise ratio. The signal to noise ratio for the imaging task described above is the difference in the generalized projection values of the background and the background plus feature of interest divided by the standard deviation of the noise in the generalized projection.

$$SNR = \frac{\Delta P}{\sigma_P} \tag{43}$$

where the signal

$$\Delta P = |P_{b+f} - P_b| \tag{44}$$

and the generalized projection P is

$$P = p_1 A_1 + p_2 A_2 \tag{45}$$

$$P = K^T A. \tag{46}$$

with $K^T = [p_1, p_2]$ and $A = [A_1, A_2]$. In these equations b denotes a measurement through the background material while $b + f$ denotes a measurement through the background plus feature of interest, as shown in Figure 3 . Therefore, the signal $\Delta P = K^T (A_{b+f} - A_b) = K^T D$ with $D = A_{b+f} - A_b = [d_1, d_2]^T$. Using the formula for the variance of a linear combination, equation (29), the signal to noise ratio (squared) is

$$SNR^2 = \frac{(\Delta P)^2}{Var(P)} = \frac{(K^T D)^2}{K^T R_A K} \tag{47}$$

where the noise is assumed to be the same over the background and background plus feature regions. That is the feature of interest is assumed to have low overall attenuation.

The optimal projection K_{opt} satisfies

$$\frac{\partial (SNR)^2}{\partial K}|_{K_{opt}} = 0. \tag{48}$$

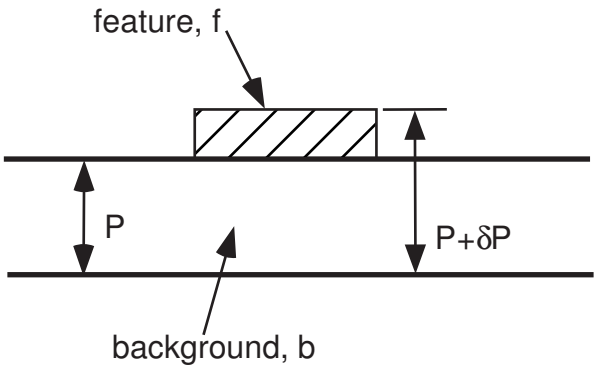

Figure 3: Imaging task used in definition of signal to noise ratio. The object consists of a feature superimposed on a background.

We show in Appendix B that setting the derivative equal to 0 yields two solutions. One is

$$K = \begin{bmatrix} 1 \\ -d_1/d_2 \end{bmatrix}, \tag{49}$$

which generates the generalized projection with zero SNR, clearly not the solution desired. The second solution

$$K_{opt} = R_A^{-1} D. \tag{50}$$

gives the maximum signal to random noise ratio. Note that the result depends on the difference vector of the two materials and on the noise properties.

The signal to noise ratio at the optimal projection K_{opt} can be found by substituting in equation (47). The result is

$$SNR_{opt} = \frac{D^T R_A^{-1} D}{\left(D^T R_A^{-1} D\right)^{1/2}} = \left(D^T R_A^{-1} D\right)^{1/2} \tag{51}$$

$$= \left[\frac{d_1^2 var(A_2) + d_2^2 var(A_1) - 2d_1 d_2 cov(A_1, A_2)}{var(A_1)var(A_2) - cov^2(A_1, A_2)}\right]^{1/2} \tag{52}$$

This optimal signal to noise ratio can also be expressed in terms of the original measurement variances, $\sigma_i^2 \quad i = 1, 2$. By equation (18), $R_A = M^{-1} R_L M^{-T}$. Using the general result from matrix theory that $(ABC)^{-1} = C^{-1} B^{-1} A^{-1}$, $R_A^{-1} = M^T R_L^{-1} M$ so

$$SNR_{opt} = \left(D^T M^T R_L^{-1} M D\right)^{1/2}. \tag{53}$$

Since $R_L = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ then $R_L^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix}$ so subsituting in this equation

$$SNR_{opt} = \left[\frac{1}{\sigma_1^2}(d_1 m_{11} + d_2 m_{12})^2 + \frac{1}{\sigma_2^2}(d_1 m_{21} + d_2 m_{22})^2 \right]^{1/2}. \tag{54}$$

We may interpret this as two distinct signal to noise ratios adding orthogonally as independent quantities. Each is the partial SNR attributable to a measurement. This result will be used in the next section where we compare the signal to noise ratio of conventional and energy selective systems.

5 Comparison of Noise in Conventional and Energy Selective Systems

Since energy selective systems extract more information than conventional systems, it might seem that they require higher dose. In fact, the opposite is true. As shown in this section, energy selective systems extract more information for the same dose than conventional systems. Care must be exercised in comparing the two types of systems since they produce physically different types of information. Two measures of noise will be discussed, noise variance and signal to noise ratio. The noise variance will be compared in synthesized monoenergetic images while the signal to noise ratios will be compared for systems performing the same imaging task, the detection of a small feature overlying a uniform background.

In order to assure the same dose, we will assume that the energy selective system uses the same spectrum as the conventional system. This is shown in Figure 4. The conventional system uses the complete spectrum shown in Part (a). The dual energy system creates the low and high energy spectra (Parts (b) and (c)) using a threshold so all photons with energy below the threshold are in the low energy spectrum and those above the threshold are in the other. With this technique, we can always create a dual energy system with exactly the same dose as a conventional system.

5.1 Comparison of Noise Variance

The comparison from the point of view of noise variance is quite simple. As discussed in the section on optimal projections, the optimal synthesized monoenergetic single projection image in a system with Poisson distributed noise has a variance equal to

$$var(P) = \frac{1}{I_1 + I_2} \tag{55}$$

This is the variance of a conventional system using the same total flux as the sum of the flux in both measurements used in an energy selective system. Since, by our assumption, the sum of the energy selective spectra is the conventional spectrum, there is

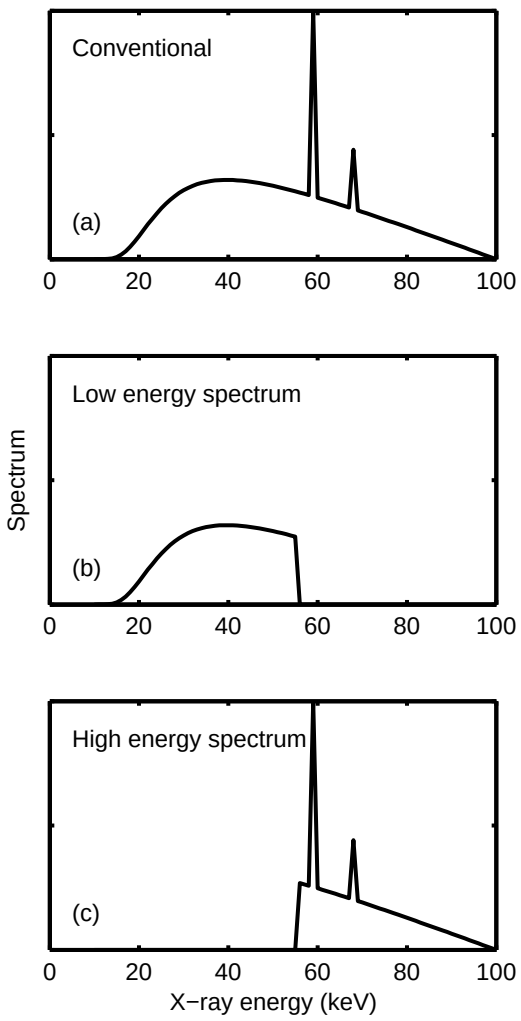

Figure 4: X-ray spectra used for the comparison of conventional and dual energy systems. As discussed in the test, by using these spectra, we guarantee that both conventional and dual energy systems have the same dose.

no increase in noise caused by carrying out the decomposition process and then forming the linear combination of a synthesized monoenergetic image. But, by carrying out the decomposition process, the basis set coefficient line integrals are available for signal processing such as material cancellation. Thus, from the point of view of variance, the energy selective system extracts more information for the same dose as a conventional system.

5.2 Comparison of Signal to Noise Ratio

The comparison from the point of view of signal to noise ratio is more complex. In order to carry it out, a suitable imaging task must be found (so that signal can be defined) and expressions for the signal to noise ratio of energy selective and conventional systems must be derived.

The imaging task is the same as discussed in the previous section on optimal signal to noise ratio: distinguish between a region containing only background material and another region containing background plus a small feature of interest. Figure 3 illustrates the geometry. For the comparison to be valid, both systems should have the same X-ray technique factors. This will be assured by assuming that the conventional and dual energy spectra are as shown in Figure 4. The only difference between the conventional and the dual energy systems will be in the processing of the data. The conventional system simply sums the measurements and then forms the logarithm. The energy selective system uses the two measurements to calculate the line integrals of the basis set coefficients and then forms the optimal projection discussed in the previous section,

There are then two sets of measurements: $(I_1, I_2)_b$ and $(I_1, I_2)_{b+f}$ where I_i is the transmitted flux with spectrum i and b denotes a measurement only through the background material while $b+f$ denotes a measurement through the background plus feature. The signal to noise ratio for the energy selective system was previously derived as equation (54). What remains is to derive the signal to noise ratio for a conventional system performing the same task.

We will assume that the conventional image consists of the logarithm of the sum of the transmitted fluxes with the two spectra. That is,

$$L = \log(I_1 + I_2). \tag{56}$$

The signal for the conventional image is the difference of this quantity between regions containing background plus feature and regions containing only background:

$$\Delta L = \log(I_1 + I_2)_{b+f} - \log(I_1 + I_2)_b \tag{57}$$

In terms of the measurement spectra, the first term in the equation for the conventional signal is:

$$\log(I_1 + I_2)_{b+f} = \int [S_1(E) + S_2(E)]\, exp\,[-A_{1,b+f} f_1(E) - A_{2,b+f} f_2(E)]\, dE \tag{58}$$

with a similar expression for $(I_1 + I_2)_b$. Using the definitions of the previous section,

$$A_{i,b+f} = A_{i,b} + d_i \quad i = 1, 2 \tag{59}$$

where, by assumption, d_1 and d_2 are small quantities. As shown in Appendix C, the signal to noise ratio of a conventional system is

$$SNR_{conventional} = \frac{|(d_1 m_{11} + d_2 m_{12})I_1 + (d_1 m_{21} + d_2 m_{22})I_2|}{(I_1 + I_2)^{1/2}} \tag{60}$$

Comparing this result with that in equation (54) for the optimal signal to noise ratio of an energy selective system shows that

$$SNR_{\text{conventional}} \leq SNR_{\text{energy selective}} \tag{61}$$

with equality if and only if

$$(d_1 m_{11} + d_2 m_{12}) = (d_1 m_{21} + d_2 m_{22}) \tag{62}$$

The m_{ij} are for the spectra transmitted through the background region and they are assumed to be the same for the spectra transmitted through the background plus feature region since the feature has low attenuation.

The condition for equal signal to noise ratio has an interesting physical interpretation. Using the average basis function interpretation of the m_{ij}, the condition states that the difference of the line integral vectors (between the background plus feature and background only regions) must be equal for the two measurement spectra. Thus the energy selective and conventional systems will have the same signal to noise ratio only if the effective energies of the two measurement spectra are the same (and the system gathers essentially no energy dependent information) or the feature has zero attenuation (and there is no signal for either system). Neither of these cases is important, so for practically useful situations, the energy selective system always has a better signal to noise ratio than the conventional system.

6 Conclusions

The energy spectrum of X-rays transmitted through the body contains a great deal of information. The information can be used to reduce the effects of two important sources of noise in diagnostic imaging: overlying anatomical detail and quantum random noise, The effects of anatomical detail are reduced by using the energy selective information to produce images with the effects of specific materials selectively canceled. These images have less clutter and therefore enhance the conspicuity of medically important features. The energy selective data can also be used to produce computed tomography images in which the extraneous details of beam hardening artifacts have been removed.

The effects of X-ray quantum noise are reduced by energy dependent techniques. The energy selective systems extract more information for the same dose than conventional systems. They can form images with the same noise as conventional systems while at the same time extracting the energy dependent information. Furthermore, they can detect small features with a signal to noise ratio better than conventional systems. From either the point of view of anatomical noise or quantum noise, energy selective systems extract

information more efficiently than conventional systems. This is due to the use of the a-priori knowledge of the physics of X-ray interactions with matter expressed in the vector space description of the energy dependence of these interactions. This description is sufficiently accurate for even the most rigorous quantitative diagnostic applications yet it reflects the fundamental simplicity of the physics and allows the information to be extracted with practically useful apparatus.

Appendix A:The Optimal Synthesized Monoenergetic Image

In this appendix the display energy for a synthesized monoenergetic image, defined implicitly by equation (37), is shown to yield the minimum variance. An expression is also derived for the second derivative at the optimal energy. The second derivative becomes increasingly large as the average energies of the two measurement spectra become equal. Thus, the optimum becomes increasingly critical and it becomes harder to achieve the optimal condition at all points in an image.

The display energy defined by equation (37) is shown to be optimal by substituting these expressions in the general formula for the derivative of variance with respect to display energy and showing that the derivative is equal to zero. Suppose that

$$c_1 = f_1(E_d) \tag{63}$$

and

$$c_2 = f_2(E_d) \tag{64}$$

are substituted in the general expression for noise variance of a linear combination of the basis set line integrals, equation (29). If this is then differentiated with respect to the display energy E_d using equation (35) the result is

$$\frac{dVar}{dE_d} = \frac{2}{J^2}\left[\sigma_1^2(f_1 m_{22} - f_2 m_{21})(f_1' m_{12} - f_2' m_{21}) + \sigma_2^2(f_1 m_{12} - f_2 m_{11})(f_1' m_{12} - f_2' m_{21})\right] \tag{65}$$

Substituting the expressions for the optimal values of $f_1(E_d)$ and $f_2(E_d)$ yields

$$\frac{dVar}{dE_d} = \frac{-2\sigma_1^2\sigma_2^2}{J^2(\sigma_1^2+\sigma_2^2)}\left[f_1'(m_{22} - m_{12}) - f_2'(m_{21} - m_{11})\right] \tag{66}$$

The expression in the bracket in equation (66) can be further simplified by using the definition of m_{ij} from equation (26)

$$m_{ij} = \frac{\partial log(I_i)}{\partial A_j} = \frac{1}{I_i}\frac{\partial I_i}{\partial A_j} = -\frac{\int f_j(E) S_i(E) e^{-f_1(E)A_1 - f_2(E)A_2} dE}{\int S_i(E) e^{-f_1(E)A_1 - f_2(E)A_2} dE} = -\langle f_j \rangle_i \tag{67}$$

The bracket notation $\langle f \rangle_i$ should be interpreted as the average value of function f in transmitted spectrum i. We can use the bracket notation to see that, for example,

$$m_{22} - m_{12} = \langle f_2 \rangle_2 - \langle f_2 \rangle_1 \approx f_2(\overline{E_2}) - f_2(\overline{E_1}) \approx (\overline{E_2} - \overline{E_1}) f_2' \tag{68}$$

Similarly $m_{21} - m_{11} \approx (\overline{E_2} - \overline{E_1})f_1'$. Substituting these expressions in equation (66) results in

$$\frac{dVar}{dE_d} = \frac{-2\sigma_1^2\sigma_2^2}{J^2(\sigma_1^2+\sigma_2^2)}\left[f_1'f_2' - f_2'f_1'\right](\overline{E_2} - \overline{E_1}). \tag{69}$$

The term $[f_1'f_2' - f_2'f_1']$ is identically equal to 0 so the derivative is also equal to 0.

Differentiating equation (66) again and evaluating at the optimal display energy yields

$$\frac{d^2Var}{dE_d^2} = \frac{2\sigma_1^2\sigma_2^2}{J^2(\sigma_1^2+\sigma_2^2)}\left[f_1''f_2' - f_2''f_1'\right](\overline{E_2} - \overline{E_1}) + \frac{2(\sigma_1^2+\sigma_2^2)}{(\overline{E_2} - \overline{E_1})^2}.$$

This value is positive, so the extremum is a minimum. As the measurement spectra become more ill-conditioned, then $(\overline{E_2} - \overline{E_1})^2 \to 0$ and the second term becomes large. Thus the radius of curvature at the minimum becomes smaller and the variance changes rapidly with the display energy so a small error will result in greatly increased noise.

Appendix B: Optimal Signal to Noise Ratio Generalized Projection

In this Appendix, we derive the formula for the generalized projection coefficients that yield the maximum signal to noise ratio. As formulated in Section 4.2, we want to solve

$$\frac{\partial(SNR)^2}{\partial K}|_{K_{opt}} = 0 \tag{70}$$

where

$$SNR^2 = \frac{(K^TD)^2}{K^TR_AK}. \tag{71}$$

.

Using the following results from Table 1

$$\frac{\partial(K^TD)^2}{\partial K} = 2(K^TD)\frac{\partial(K^TD)}{\partial K} = 2(K^TD)D^T \tag{72}$$

$$\frac{\partial(K^TR_AK)}{\partial K} = K^T(R_A^T + R_A) = 2K^TR_A \tag{73}$$

where the last step in equation (73) follows from the fact that R_A is symmetric so $R_A = R_A^T$. Finally, applying the quotient rule,

$$\frac{\partial}{\partial x}\left(\frac{f}{g}\right) = \frac{g\frac{\partial f}{\partial x} - f\frac{\partial g}{\partial x}}{g^2}, \tag{74}$$

we can write

$$\frac{\partial(SNR)^2}{\partial K} = \frac{(K^TR_AK)2(K^TD)D^T - (K^TD)^2 2K^TR_A}{(K^TR_AK)^2}. \tag{75}$$

Factoring the numerator

$$\frac{\partial(SNR)^2}{\partial K} = 2(K^T D)\frac{(K^T R_A K)D^T - (K^T D)(K^T R_A)}{(K^T R_A K)^2} = 0, \tag{76}$$

it is clear that one solution is $(K^T D) = 0$, which is equivalent to

$$K = \begin{bmatrix} 1 \\ -d_1/d_2 \end{bmatrix}. \tag{77}$$

Recall that, by definition, $D = [d_1, d_2]^T$.

We can show that the other solution is $K = R_A^{-1} D$ by substituting in the numerator of equation (76). First note that, by the symmetry of R_A,

$$K^T = D^T R_A^{-T} = D^T R_A^{-1}. \tag{78}$$

Substituting the expression for K in the numerator of equation (76) shows that the numerator and hence the derivative is equal to 0,

$$(K^T R_A K)D^T - (K^T D)(K^T R_A) = (D^T R_A^{-1} R_A R_A^{-1} D)D^T - (D^T R_A^{-1} D)(D^T R_A^{-1} R_A) = 0. \tag{79}$$

Substituting the optimal K in the expression for the signal to noise ratio squared

$$SNR_{opt}^2 = \frac{(K^T D)^2}{K^T R_A K} = \frac{\left(D^T R_A^{-1} D\right)^2}{D^T R_A^{-1} R_A R_A^{-1} D} = D^T R_A^{-1} D. \tag{80}$$

Appendix C: Signal to Noise Ratio in a Conventional Imaging System

In this Appendix, an expression is derived for the signal to noise ratio of a conventional imaging system in terms of the m_{ij} coefficients. The imaging task is as assumed in Section 4.2.

First, we need an expression for the signal, as defined in that section. Substituting equation (59) into equation (58) yields

$$(I_1 + I_2)_{(b+f)} = \int S^*(E) e^{-d_1 f_1(E) - d_2 f_2(E)} dE \tag{81}$$

where

$$S^*(E) = [S_1(E) + S_2(E)]\, e^{-A_{1,b} f_1(E) - A_{2,b} f_2(E)}. \tag{82}$$

Note that

$$(I_1 + I_2)_b = \int S^*(E) dE. \tag{83}$$

Approximating the exponential in equation (81) for small values of its argument as

$$e^{-d_1 f_1(E) - d_2 f_2(E)} \approx 1 - d_1 f_1(E) - d_2 f_2(E), \tag{84}$$

the equation becomes

$$(I_1 + I_2)_{(b+f)} \approx (I_1 + I_2)_b - d_1 \int S^*(E) f_1(E) dE - d_2 \int S^*(E) f_2(E) dE. \tag{85}$$

We need the logarithm of equation (85). Since the d_i are assumed small, the logarithm can be approximated as

$$\log (I_1 + I_2)_{(b+f)} \approx \log (I_1 + I_2)_b - d_1 \langle f_1(E) \rangle_{1+2} - d_2 \langle f_2(E) \rangle_{1+2} \tag{86}$$

where $\langle . \rangle_{1+2}$ denotes an average using spectrum $S^*(E)$ as a weighting function. The signal δL in a conventional system is thus closely approximated by

$$\delta L = \log (I_1 + I_2)_{(b+f)} - \log (I_1 + I_2)_b \approx -d_1 \langle f_1(E) \rangle_{1+2} - d_2 \langle f_2(E) \rangle_{1+2} \tag{87}$$

Next, we need an expression for the noise. Assuming a Poisson distributed noise source for simplicity, the variance is

$$Var(L) = \frac{1}{I_1 + I_2}. \tag{88}$$

Substituting this in equation (87) the signal to noise ratio is

$$SNR = \left| d_1 \langle f_1(E) \rangle_{1+2} + d_2 \langle f_2(E) \rangle_{1+2} \right| (I_1 + I_2)^{1/2} \tag{89}$$

This result must now be expressed in terms of the M_{ij} coefficients. This can be done by using the result stated in equation (26),

$$M_{ij} = \langle f_j(E) \rangle_i . \tag{90}$$

Using the linearity of the averaging operation,

$$\langle f_1(E) \rangle_{1+2} = \frac{M_{11} I_1 + M_{21} I_2}{I_1 + I_2} \tag{91}$$

and

$$\langle f_2(E) \rangle_{1+2} = \frac{M_{12} I_1 + M_{22} I_2}{I_1 + I_2}. \tag{92}$$

After some rearrangement of terms, the signal to noise ratio of a conventional system is

$$SNR = \frac{|(d_1 M_{11} + d_2 M_{12}) I_1 + (d_1 M_{21} + d_2 M_{22}) I_2|}{(I_1 + I_2)^{1/2}}. \tag{93}$$

References

[1] R. E. Alvarez and A. Macovski. Energy-selective reconstructions in x-ray computerized tomography. *Physics in Medicine and Biology*, 21(5):733–44, 1976.

[2] Robert E. Alvarez. Energy dependent information in x-ray imaging: Part 1. the vector space description. *TBP*, 1983.

[3] Robert E. Alvarez and Edward J. Seppi. A comparison of noise and dose in conventional and energy selective computed tomography. *IEEE Transactions on Nuclear Science*, NS-26:2853–2856, 1979.

[4] Leonard Alan Lehmann. *Utilization of multi-spectral measurements in. radiology.* PhD thesis, Department of Electrical Engineering, Stanford University, 1982.

[5] Michael Syskind Petersen, Kaare Brandt Pedersen. *The Matrix Cookbook.* http://matrixcookbook.com, 2008.

[6] B. Rutt and A. Fenster. Split-filter computed tomography: a simple technique for dual energy scanning. *J Comput Assist Tomogr*, 4(4):501–509, Aug 1980.