

[aprendtech.com](#) >> [blog](#) >> [this post](#)

If you have trouble viewing this, try [the pdf of this post](#). You can [download the code](#) used to produce the figures in this post.

## A-space covariance from x-ray detector noise

My [last post](#) showed that detection performance is determined by the signal to noise ratio (SNR). I derived a formula for the SNR with multispectral measurements, which depends on the covariance of the A-space data. This post shows how to compute the A-space covariance from the x-ray data noise and the effective attenuation coefficient matrix. In general, this depends on the type of estimator used so instead I will use the Cramèr-Rao lower bound (CRLB), which is the minimum covariance for any unbiased estimator. This gives a general result independent of the specific estimator implementation. I will show that the constant covariance CRLB is sufficiently accurate for our purposes. These results will allow us to compute signal to noise ratios for limited energy resolution measurements that are directly comparable to the Tapiovaara-Wagner[1] optimal SNR with complete energy information.

### Maximum likelihood estimator with scalar normal data

In [a previous post](#) I showed that we could use a linearized model for the transmitted x-ray data

$$\delta \mathbf{L}_{with\_noise} = \mathbf{M} \delta \mathbf{A} + \mathbf{w} \quad (1)$$

where  $\mathbf{L} = -\log(I/I_0)$  is the log of the transmitted flux  $I$ ,  $I_0$  is the transmitted flux with no object in the beam,  $\mathbf{M} = \frac{\partial \mathbf{L}}{\partial \mathbf{A}}$  is the gradient of the measurements in A-space, and  $\mathbf{w}$  is a zero-mean multinormal random variable. Eq. 1 is linearized about an operating point  $(\mathbf{L}_o, \mathbf{A}_o)$  so  $\delta \mathbf{L} = \mathbf{L} - \mathbf{L}_o$  and  $\delta \mathbf{A} = \mathbf{A} - \mathbf{A}_o$ . To understand the estimator, I will start by assuming scalar data. The linearized model is then ( I will drop the  $\delta$  to simplify notation so  $\delta \mathbf{L} \rightarrow \mathbf{L}$  and  $\delta \mathbf{A} \rightarrow \mathbf{A}$  and the model is assumed to be used around the operating point)

$$L = MA + w \quad (2)$$

where  $w \sim \mathcal{N}(0, \sigma_L^2)$ . Since there is only one measurement, a reasonable way to proceed is to solve the deterministic part of (2) with the measurement so  $\hat{A} = L/M$ . The variance of the estimate will then be  $var(\hat{A}) = var(w/M) = \sigma_L^2/M^2$ .

This estimator is actually the maximum likelihood estimator (MLE). To see this, introduce the log-likelihood function  $\mathcal{L}$ , which is the logarithm of the probability distribution function evaluated with the actual measurement. The log-likelihood is a function of the data and the unknown parameter  $\mathcal{L} = \mathcal{L}(x; A)$ . The maximum likelihood estimator chooses the value of  $A$  that maximizes the log-likelihood  $\hat{A} = \max[\mathcal{L}(x; A)]_A$ . With scalar normal data, the log-likelihood is

$$\mathcal{L} = -\frac{(L - MA)^2}{2\sigma_L^2} + constant \quad (3)$$

Notice this reaches its maximum value when the first term is zero. Solving  $L - MA = 0$ , the estimator is  $\hat{A}_{MLE} = L/M$ , which is the solution of the non-random part of the model with the measurement.

The error in the estimate will be proportional to the radius of curvature  $\rho$  of  $\mathcal{L}$  at the maximum value. This is the negative of the second derivative evaluated at  $A = \hat{A}$

$$\text{radius of curvature} = \rho = - \left[ \frac{\partial^2 \mathcal{L}}{\partial A^2} \right]_{\hat{A}}$$

Differentiating (3), the radius of curvature is

$$\rho = \frac{M^2}{\sigma_L^2}$$

Notice that we can express the variance of the estimate as the inverse of the second derivative

$$\begin{aligned} \text{var}(\hat{A}) &= \frac{\sigma_L^2}{M^2} \\ &= \frac{1}{\rho} \\ &= - \frac{1}{\left[ \frac{\partial^2 \mathcal{L}}{\partial A^2} \right]_{\hat{A}}} \end{aligned}$$

In general, the second derivative may depend on the data, so we would use the expected value of the second derivative

$$\text{var}(\hat{A}) = - \frac{1}{\left\langle \left[ \frac{\partial^2 \mathcal{L}}{\partial A^2} \right]_{\hat{A}} \right\rangle}$$

Appendix 3A of Kay[2] proves that this is the minimum variance for any unbiased estimator, called the Cramèr-Rao lower bound (CRLB).

The MLE is the value of the parameter where the derivative is equal to zero. The first derivative of  $\mathcal{L}$  is

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial A} &= \frac{M}{\sigma_L^2} (L - MA) \\ &= \frac{M^2}{\sigma_L^2} \left( \frac{L}{M} - A \right) \\ &= I (g(L) - A) \end{aligned} \tag{4}$$

With this factoring, the maximum likelihood estimate is  $\hat{A}_{MLE} = g(L) = L/M$ . Notice that the variance of the estimate, from my discussion in the first paragraph of this section is  $\text{var}(\hat{A}) = \frac{\sigma_L^2}{M^2} = 1/I$ . The proof in Appendix 3A of Kay[2] shows that this is a general result. If we can factor the derivative as in (4), then the inverse of the leading factor,  $I$ , is the CRLB.

## Maximum likelihood estimator with vector multinormal data, constant covariance

We can parallel this derivation with vector data by following the two-step approach used in my [previous post](#) to derive the optimal detector: first solve the problem assuming “white” equal variance data and then use the whitening transform to convert the general case to the simplified one. If the covariance of the measured data is  $\mathbf{C}_L = \sigma_L^2 \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix with ones on its diagonal, then the log-likelihood function is

$$\mathcal{L}(\mathbf{L}; \mathbf{A}) = \frac{1}{2\sigma_L^2} (\mathbf{L} - \mathbf{MA})^T (\mathbf{L} - \mathbf{MA}) + \text{constant}$$

Expanding the product and gathering common terms utilizing the fact that  $\mathcal{L}$  is a scalar so we can take the transpose of any term without affecting the result

$$\mathcal{L}(\mathbf{L}; \mathbf{A}) = \frac{1}{2\sigma_L^2} \left[ \mathbf{L}^T \mathbf{L} - 2\mathbf{L}^T \mathbf{M} \mathbf{A} - \mathbf{A}^T (\mathbf{M}^T \mathbf{M}) \mathbf{A} \right] + \text{constant}$$

I will use the general formulas for matrix derivatives from [my post](#),

$$\begin{aligned} \frac{\partial \mathbf{b}^T \mathbf{x}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{x}^T \mathbf{b}}{\partial \mathbf{x}} = \mathbf{b} \\ \frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} &= (\mathbf{B} + \mathbf{B}^T) \mathbf{x} \\ &= 2\mathbf{B} \mathbf{x} \quad \text{if } \mathbf{B} \text{ symmetric} \end{aligned} \quad (5)$$

where  $\mathbf{b}$  is a constant vector and  $\mathbf{B}$  is a constant matrix and  $\mathbf{x}$  is the vector of variables. The derivative of  $\mathcal{L}$  is (since  $(\mathbf{M}^T \mathbf{M})$  is symmetric)

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{A}} &= \frac{1}{\sigma_L^2} \left[ \mathbf{M}^T \mathbf{L} - (\mathbf{M}^T \mathbf{M}) \mathbf{A} \right] \\ &= \frac{(\mathbf{M}^T \mathbf{M})}{\sigma_L^2} \left[ (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{L} - \mathbf{A} \right] \end{aligned} \quad (6)$$

The MLE is derived by setting the derivative (6) equal to zero so

$$\hat{\mathbf{A}}_{MLE} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{L}$$

and by analogy to the scalar case, the covariance of the estimate is the inverse of the leading factor of (6)

$$\text{cov}(\hat{\mathbf{A}}_{MLE}) = \sigma_L^2 (\mathbf{M}^T \mathbf{M})^{-1}$$

We can use these results for the case with a general covariance matrix  $\mathbf{C}_L$  by using the whitening transform  $\mathbf{V}$  defined so that  $\mathbf{C}_L^{-1} = \mathbf{V}^T \mathbf{V}$  and  $\mathbf{C}_L = \mathbf{V}^{-1} \mathbf{V}^{-T}$ . Transforming the measurement data so  $\mathbf{L}' = \mathbf{V} \mathbf{L}$ , the transformed linear model is

$$\begin{aligned} \mathbf{L}' &= \mathbf{V} (\mathbf{M} \mathbf{A} + \mathbf{w}) \\ &= \mathbf{M}' \mathbf{A} + \mathbf{w}' \end{aligned}$$

where  $\mathbf{M}' = \mathbf{V} \mathbf{M}$  and  $\mathbf{w}' = \mathbf{V} \mathbf{w}$ . The covariance of  $\mathbf{w}'$  is the identity matrix

$$\begin{aligned} \text{cov}(\mathbf{w}') &= \langle (\mathbf{V} \mathbf{w}) (\mathbf{V} \mathbf{w})^T \rangle \\ &= \mathbf{V} \mathbf{C}_L \mathbf{V}^T \\ &= \mathbf{V} \mathbf{V}^{-1} \mathbf{V}^{-T} \mathbf{V}^T \\ &= \mathbf{I} \end{aligned}$$

so we can apply the white case results. Using the MLE from the white case and noting that  $\mathbf{M}'^T \mathbf{M}' = \mathbf{M}^T \mathbf{V}^T \mathbf{V} \mathbf{M} = \mathbf{M}^T \mathbf{C}_L^{-1} \mathbf{M}$

$$\begin{aligned} \hat{\mathbf{A}}_{MLE} &= (\mathbf{M}'^T \mathbf{C}_L^{-1} \mathbf{M}')^{-1} \mathbf{M}'^T \mathbf{V}^T \mathbf{V} \mathbf{L} \\ &= (\mathbf{M}^T \mathbf{C}_L^{-1} \mathbf{M})^{-1} \mathbf{M}^T \mathbf{C}_L^{-1} \mathbf{L} \end{aligned} \quad (7)$$

Similarly, the covariance of the estimate is

$$\begin{aligned} \text{cov}(\hat{\mathbf{A}}_{MLE}) &= (\mathbf{M}'^T \mathbf{M}')^{-1} \\ &= (\mathbf{M}^T \mathbf{C}_L^{-1} \mathbf{M})^{-1} \end{aligned} \quad (8)$$

## Simple vector MLE example

We can see how these formulas work an example. Suppose we have two-spectrum data such as from two x-ray tube voltages or two bin PHA. In this case the  $\mathbf{M}$  matrix is

$$\mathbf{M} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

and the covariance is

$$\mathbf{C}_L = \begin{bmatrix} 1/N_1 & 0 \\ 0 & 1/N_2 \end{bmatrix}$$

where  $N_1$  and  $N_2$  are the mean values of the photon numbers of the measurements. Substituting these into (7), the MLE is

$$\hat{\mathbf{A}}_{MLE} = \frac{1}{(M_{11}M_{22} - M_{12}M_{21})} \begin{bmatrix} M_{22}L_1 - M_{12}L_2 \\ -M_{21}L_1 + M_{11}L_2 \end{bmatrix} \quad (9)$$

Notice the MLE is just the solution of the linear system  $\mathbf{MA} = \mathbf{L}$ . For example, we can write the solution for  $\hat{A}_1$  using determinants and cofactors as in elementary algebra

$$\hat{A}_1 = \frac{\begin{vmatrix} L_1 & M_{12} \\ L_2 & M_{22} \end{vmatrix}}{|\mathbf{M}|} = \frac{M_{22}L_1 - M_{12}L_2}{|\mathbf{M}|}$$

where  $||$  is the determinant.

The covariance of the estimate is

$$\text{cov}(\hat{\mathbf{A}}_{MLE}) = \frac{\begin{vmatrix} \frac{M_{22}^2}{N_1} + \frac{M_{12}^2}{N_2} & -\left(\frac{M_{21}M_{22}}{N_1} + \frac{M_{11}M_{12}}{N_2}\right) \\ -\left(\frac{M_{21}M_{22}}{N_1} + \frac{M_{11}M_{12}}{N_2}\right) & \frac{M_{21}^2}{N_1} + \frac{M_{11}^2}{N_2} \end{vmatrix}}{|\mathbf{M}|^2} \quad (10)$$

These results parallel those I derived in my dissertation and my 1976 paper[3]. The covariance (10) is Eqs. 14 and 15 of my paper and the MLE (9) with two spectra is just the solution of the deterministic equations with the measured data as I showed in my paper.

The matrix multiplications can get tedious so in my next post I will describe how to use the free symbolic mathematics program [Maxima](#) to carry them out. I am including Maxima scripts for some of the derivations in [the code](#) for this post.

## CRLB with general multinormal data

We know that with x-ray data the covariance is not constant. Indeed, the variance increases strongly with object thickness. So an immediate question is: What is the difference between the CRLB computed using constant covariance and the general result that takes the change into account? To answer this, I did a simulation that computes the error by using the constant covariance formula as a function of object thickness.

Appendix 3.C of Kay[2] derives the inverse of the CRLB  $\mathbf{I}(\mathbf{A})$  for the general multinormal case where the covariance can change. The results are (translated to my notation)

$$[\mathbf{I}(\mathbf{A})]_{ij} = \left[ \frac{\partial \mathbf{L}(\mathbf{A})}{\partial A_i} \right]^T \mathbf{C}^{-1}(\mathbf{A}) \left[ \frac{\partial \mathbf{L}(\mathbf{A})}{\partial A_j} \right] + \frac{1}{2} \text{tr} \left[ \mathbf{C}^{-1}(\mathbf{A}) \frac{\partial \mathbf{C}}{\partial A_i} \mathbf{C}^{-1}(\mathbf{A}) \frac{\partial \mathbf{C}}{\partial A_j} \right] \quad (11)$$

Recall from my [derivation of the linearized model](#) that  $\frac{\partial \mathbf{L}(\mathbf{A})}{\partial \mathbf{A}} = \mathbf{M}$ , so this is

$$[\mathbf{I}(\mathbf{A})]_{ij} = \mathbf{M}(:, i)^T \mathbf{C}^{-1} \mathbf{M}(:, j) + \frac{1}{2} \text{tr} \left[ \mathbf{C}^{-1}(\mathbf{A}) \frac{\partial \mathbf{C}}{\partial A_i} \mathbf{C}^{-1}(\mathbf{A}) \frac{\partial \mathbf{C}}{\partial A_j} \right]$$

where the notation  $\mathbf{M}(:, k)$  is the  $k$ 'th column of  $\mathbf{M}$ . Comparing this with Eq. 8, the first term is the constant covariance CRLB and the second term measures the effect of changes in  $\mathbf{C}$  as a function of  $\mathbf{A}$ .

We can see that the constant covariance term in (11) will be much larger than the second term as follows. With Poisson data, the variance of the logarithm is the inverse of the expected value and, since the PHA counts are independent, the covariance and its inverse are

$$\mathbf{C} = \begin{bmatrix} 1/N_1 & & \\ & \ddots & \\ & & 1/N_{nbins} \end{bmatrix}, \quad \mathbf{C}^{-1} = \begin{bmatrix} N_1 & & \\ & \ddots & \\ & & N_{nbins} \end{bmatrix},$$

where  $N_k$  is the number of counts in bin  $k$  and  $nbins$  is the number of bins. The first term in (11) is therefore proportional to the number of photons while the second term is essentially the product of  $\mathbf{C}$  and its inverse and is approximately constant. Since in medical imaging the number of counts is usually large, we would expect that the constant covariance CRLB is much larger than the derivative of the covariance term.

This is shown by the [code](#) for this post. The code has Matlab functions to compute the general and constant covariance CRLB for points along a diagonal line through A-space. In order to compute the general CRLB, we need to have formulas for the derivatives of the covariance in (11). Since  $\mathbf{C}$  is diagonal, we only need the derivative of each diagonal element

$$\frac{\partial}{\partial A_i} \left( \frac{1}{N_k} \right) = -\frac{1}{N_k^2} \frac{\partial N_k}{\partial A_i} = -\frac{M_{ki}}{N_k}, \quad (12)$$

which can be computed at each point from  $\mathbf{M}$  and  $N_k$ .

The path through A-space is shown in the left panel of Fig. 1. The right panel shows the norm of the difference of the constant covariance and general CRLB matrices divided by the norm of the general CRLB as a function of distance along the line.

## Conclusions

Fig. 1 shows that the errors in the CRLB using the constant covariance formula Eq. 8 are less than a few percent although, as expected, they become larger for thicker objects where the number of transmitted photons is smaller. We can therefore use the simpler formula Eq. 8 to compute the CRLB of the A-space data without incurring a significant error. Substituting in the formula from my [last post](#), the overall  $SNR^2$  is

$$SNR^2 = \delta \mathbf{A}^T \mathbf{C}_A^{-1} \delta \mathbf{A} = \delta \mathbf{A}^T \mathbf{M}^T \mathbf{C}_L^{-1} \mathbf{M} \delta \mathbf{A}$$

In the next posts, I will discuss computing the covariance of the measurements for different types of detectors. These require the derivation of the effective attenuation coefficient matrix  $\mathbf{M}$  and the covariance of the log data  $\mathbf{C}_L$ . Both of these depend on the properties of the detector and the energy resolution of the measurements. The derivations sometimes involve complex matrix computations, so I double-check them using Maxima and also Monte Carlo simulations.

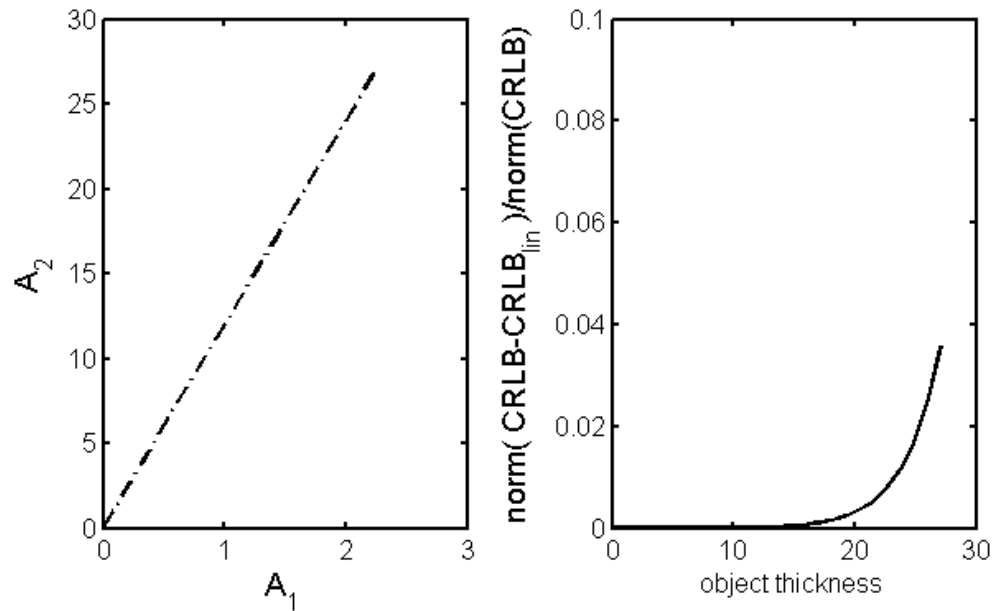


Figure 1: Errors with constant covariance CRLB vs object thickness. The object A-vectors lie along a diagonal line through A-space as shown in the left panel. The difference between the general and the constant covariance CRLB divided by the general formula is shown in the right panel.

## Bob Alvarez

Last edited September 24, 2013

©2012 by Aprend Technology and Robert E. Alvarez

Linking is allowed but reposting or mirroring is expressly forbidden.

## References

- [1] M. J. Tapiovaara and R. Wagner, "SNR and DQE analysis of broad spectrum X-ray imaging," *Phys. Med. Biol.* **30**, 519–529, (1985).
- [2] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Upper Saddle River, NJ: Prentice Hall PTR, 1993.
- [3] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerized tomography," *Phys. Med. Biol.* **21**, 733–44, (1976).