Optimal Energy Thresholds and Weights for Separating Materials Using Photon Counting X-Ray Detectors with Energy Discriminating Capabilities

Adam S. Wang[a,b] and Norbert J. Pelc[a,b,c]

[a]Departement of Electrical Engineering, Stanford University, Stanford, CA 94305;
[b]Department of Radiology, Stanford University, Stanford, CA 94305;
[c]Department of Bioengineering, Stanford University, Stanford, CA 94305

ABSTRACT

It is well known that decomposing an object into attenuation or material basis functions provides additional imaging benefits such as contrast enhancement or material subtraction. This can be accomplished with photon counting x-ray detectors (PCXDs) with energy discriminating capabilities, which enable us to count x-ray photons and classify them based on their energies. The richness of the information contained in these measurements can depend heavily on how these photons are binned together.

In this paper, our goal is to identify a method that yields the optimal energy thresholds and/or weights for binning data from energy discriminating PCXDs. Additional energy information from these PCXDs allows us to use maximum-likelihood to estimate the amount of the basis materials penetrated by the beam. However, due to inherent quantum noise, these estimates are themselves noisy. We show that for PCXDs that discriminate between low and high energy photons, it is beneficial to have a gap between the thresholds. Photons with energies that fall into this gap should either be discarded or counted separately to improve material separability. Furthermore, if the PCXD can discern the energy of each photon, we show that when estimating the amount of each of two material basis functions, two appropriately weighted sums of the photon counts provide as much information as knowing the number of counts at each energy.

Keywords: Dual Energy, Photon Counting, Detector Response Function, Maximum-Likelihood, Estimation, Material Basis Functions

1. INTRODUCTION

Recent advances in CT source and detector technologies have brought back much attention to dual- and multi-energy imaging.[1–6] While there are advantages and disadvantages to each of the many spectral imaging techniques, photon counting x-ray detectors (PCXDs) with energy discriminating capabilities ideally allow us to extract as much information as possible from the photons that are transmitted through the scanned object.[5] These advanced detectors have the capability of providing more information than the same scan acquired with traditional energy integrating detectors since the energy of each photon can be discerned. However, their full potential in reducing noise or separating materials depends on how information is measured, combined, and extracted. Therefore, our goal is to identify methods that yield the optimal energy thresholds and/or weights for binning photons.

Like many dual-energy tasks, we are interested in estimating the unknown amounts of two known materials. We use a known incident polychromatic x-ray spectrum whose transmission through the object is measured by a PCXD with energy discriminating capabilities. Even if the composition of the object is unknown, our analysis holds under the weaker assumption that attenuation as a function of x-ray energy is a property of two mechanisms, or basis functions: photoelectric and Compton.[7,8] While the traditional view of PCXDs is that they distribute photon counts into two or more energy bins based on simple thresholds, we will generalize the notion of ideal PCXDs to several forms. For each form, we first construct the theory behind evaluating the performance of an estimator that uses the measured data, and then optimize the detector parameters for a specific example.

Further author information: (Send correspondence to A.S.W.)
A.S.W.: E-mail: adamwang@stanford.edu, Telephone: 1 650 724 8030

Medical Imaging 2009: Physics of Medical Imaging, edited by Ehsan Samei, Jiang Hsieh,
Proc. of SPIE Vol. 7258, 725821 · © 2009 SPIE
CCC code: 1605-7422/09/$18 · doi: 10.1117/12.811454

2. MODEL

Consider an incident x-ray spectrum $I_0(E)$ with maximum energy M and attenuation basis functions $\mu_1(E)$ and $\mu_2(E)$ that are functions of x-ray energy E. Throughout this paper, we consider energies to the nearest keV. For an ideal detector with 1keV resolution, the expected number of photons of energy j that fall on the detector is $\lambda_j = I_{0,j} \exp(-(t_1\mu_{1,j} + t_2\mu_{2,j}))$ where $I_{0,j} = I_0(j)$, $\mu_{1,j} = \mu_1(j)$, $\mu_{2,j} = \mu_2(j)$, and t_1 and t_2 are the thicknesses of material 1 and 2, respectively. If we consider quantum statistics, then the number of photons of energy j that fall on the detector, r_j, is a Poisson distribution with mean λ_j; that is $r_j \sim \text{Poisson}(\lambda_j)$

Suppose that for each photon counted, our detector can increment the count in one of N bins (Fig. 1). For instance, if we have 2 bins, then our detector could increment the count in a "low energy" bin or a "high energy" bin depending on whether an incident photon is below or above some cutoff threshold energy. Therefore, d_i, the number of detected counts in bin i is given by:

$d_i = \sum_{j=\tau_{i-1}+1}^{\tau_i} r_j$, for $i = 1, \ldots, N$, and where $\{\tau_i\}$ are the cutoff threshold energies, with $\tau_0 = 0$ and $\tau_N = M$.

Because d_i is a sum of Poisson random variables, d_i is itself a Poisson random variable with mean $\gamma_i = \sum_{j=\tau_{i-1}+1}^{\tau_i} \lambda_j$.

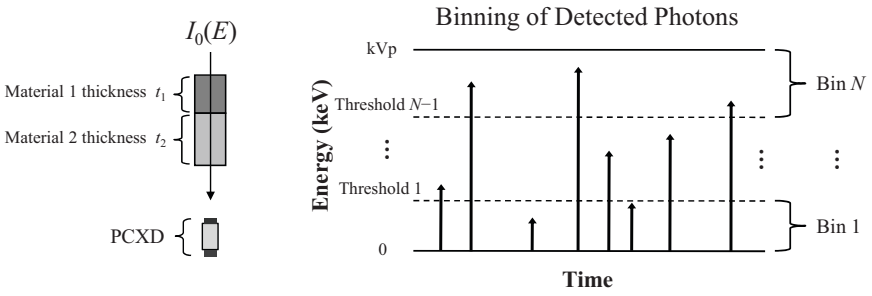


Figure 1. A PCXD sorts and counts photons by their energy. Each impulse on the right represents the energy of a detected photon and is counted in the appropriate bin by our idealized detector.

3. MAXIMUM-LIKELIHOOD ESTIMATOR FOR BINNED DATA

We utilize a maximum-likelihood estimator (MLE) as our method to estimate $t = [t_1\ t_2]^T$ directly from the raw data d.[6,9] This allows us to avoid performing a log-normalization, which is not effective when any $d_i = 0$, whereas MLE is well-suited for such measurements. Since each d_i is independent of any other and each is a Poisson distribution, the likelihood function is

$$\mathcal{L}_P(t) = f(d \mid t) = \prod_i \frac{\gamma_i^{d_i} e^{-\gamma_i}}{d_i!}. \tag{1}$$

We take our estimate of t to be $\hat{t} = \underset{t}{\text{argmax}}\, \mathcal{L}_P^*(t)$, where $\mathcal{L}_P^*(t)$ is the log-likelihood function

$$\mathcal{L}_P^*(t) = \log f(d \mid t) = \sum_i \left(-\gamma_i + d_i \log \gamma_i - \log d_i!\right). \tag{2}$$

3.1 Estimator Performance

Because the measurements $\{d_i\}$ are noisy, the estimate $\hat{t}$ will be noisy. For a given t, we will show that the variance (a measure of imprecision, noise, or uncertainty) of $\hat{t}$ is a function of $\{\tau_i\}$, the energy bin cutoff thresholds, and that the variance can be minimized by selecting the appropriate $\{\tau_i\}$.

Error propagation allows us to approximate the variance in the estimate by using a first-order Taylor series expansion of $\hat{t}$ about t.[3, 10, 11]

$$\mathrm{Cov}\left(\hat{t}_i, \hat{t}_j\right) \approx \sum_k \sum_l \left(\frac{\partial \hat{t}_i}{\partial d_k}\right) \left(\frac{\partial \hat{t}_j}{\partial d_l}\right) \mathrm{Cov}\left(d_k, d_l\right) \tag{3}$$

In our setup, the $\{d_i\}$ are independent of each other, so

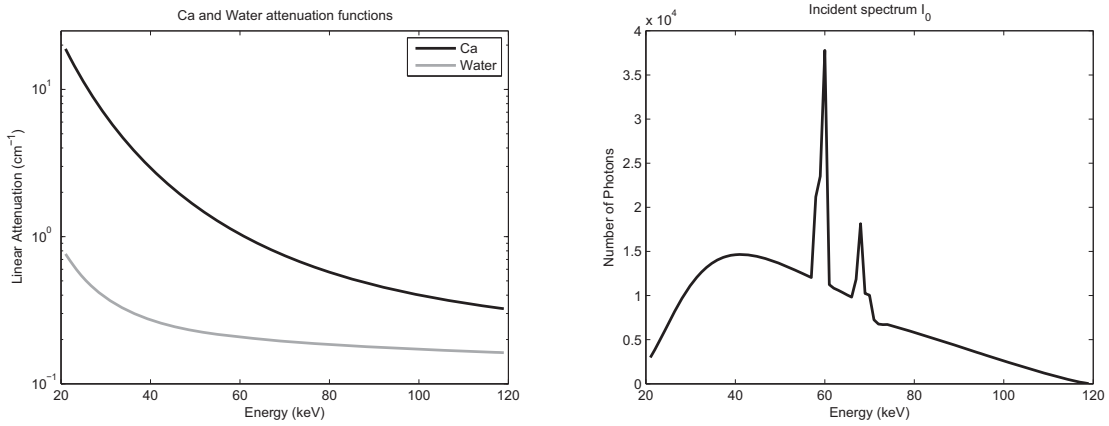
$$\mathrm{Cov}\left(\hat{t}_i, \hat{t}_j\right) \approx \sum_k \left(\frac{\partial \hat{t}_i}{\partial d_k}\right) \left(\frac{\partial \hat{t}_j}{\partial d_k}\right) \mathrm{Var}\left(d_k\right) \tag{4}$$

In Appendix A, we find the partial derivative terms in the above equation so that the covariance can be explicitly computed. Note that if the photon counts were completely disaggregated by energy, i.e., the number of counts at each energy can be measured independently, then we have the special case of $N = M$, and $\tau_0 = 0, \tau_1 = 1, \tau_2 = 2, \ldots, \tau_M = M$. We take this case of 1keV bin sizes to be the ideal measurement, resulting in the least noisy estimate.

For a given I_0, μ_1, μ_2, and t, the precision of estimate $\hat{t}$ is a function of the bin energy cutoffs used. Given an objective, such as to minimize $\mathrm{Var}(t_1)$, which allows us to most precisely estimate t_1, we can search through the space of all bin cutoffs since our formulation only considers energies to the nearest keV. Hence, for a fixed N, we can find the $\{\tau_i\}$ that minimize the target objective function, where $0 < \tau_1 < \tau_2 < \ldots < \tau_{N-1} < M$ and $\tau_i \in \mathbb{N}, \forall i = 1, \ldots, N$. For $N < 5$, searching through the space of possible $\{\tau_i\}$ is trivial, while it may take only a few minutes for $N = 7$.

3.2 Numerical Example

To put this theoretical formulation to practice, we need to first select an I_0, μ_1, μ_2. Take I_0 to be a 120 kVp beam with an intensity equivalent to that of the photons that would fall on a single detector in a single view out of a 400 mAs exposure (Fig. 2). We chose calcium and water as our two materials because decomposing attenuation into these two basis materials can have direct applications, such as bone densitometry or arterial plaque characterization. Furthermore, if indeed the underlying mechanisms for dual-energy material decomposition are the photoelectric and Compton attenuations functions, we can easily convert our results from using calcium and water into any other two materials via a linear transformation.[8]



(a) Attenuation functions (b) Incident spectrum

Figure 2. Attenuation functions and incident spectrum used for numerical calculations.

We ignore energies below 20 keV since they are heavily filtered at the source and have very little penetrating ability for materials of non-trivial thickness. Furthermore, because the attenuation of photons below 20 keV rises

rapidly, including this part of the curve in the analysis can result in poorly conditioned matrices that quickly drive our methods into problems with numerical errors. Finally, we take as our objective to predict as precisely as possible the amount of calcium, i.e. minimize $\mathrm{Var}(\hat{t}_{\mathrm{Ca}})$. Although more general objective functions could be selected, this objective is reasonable given that typically $t_{\mathrm{Ca}} \ll t_{\mathrm{H_2O}}$ and that our tasks require precise estimation of calcium levels.

Figure 3 shows the normalized standard deviation of $\hat{t}_{Ca}$ for the case of two energy bins, 5mm Ca, and water thicknesses of 20cm and 40cm. We can see that setting $\tau_1 = 54$keV provides for the least noisy estimate of the amount of calcium for 5mm Ca and 20cm water, while at 5mm Ca and 40cm water, the optimal τ_1 is 63keV. Clearly, when we only have two energy bins, the selection of τ_1 can significantly change the performance of our MLE, and the optimal threshold depends on the object being studied.

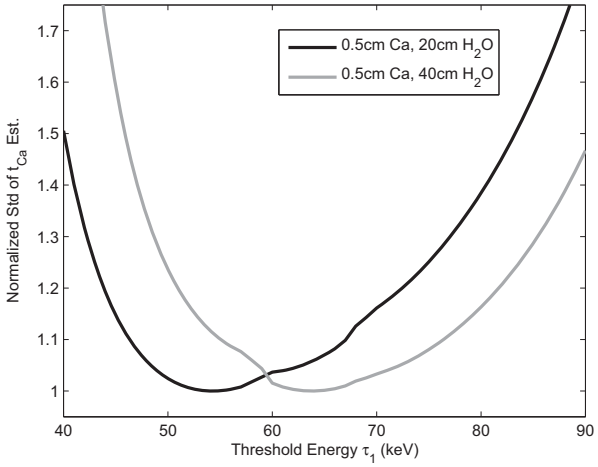


Figure 3. For two energy bins, the normalized standard deviation of $\hat{t}_{\mathrm{Ca}}$ is shown as a function of threshold energy τ_1.

4. RELAXATION OF ABUTMENT CONSTRAINT

Above, we assumed that the N bins are separated by cutoff thresholds, suggesting that each bin i collects photons from the interval of energies $(\tau_{i-1}, \tau_i]$. Adjacent bins have adjacent, non-overlapping intervals that abut each other. However, we can relax this constraint by allowing bin i to collect photons from energies $(l_i, u_i]$, where $0 \leq l_1 \leq u_1 \leq l_2 \leq u_2 \leq \ldots \leq l_N \leq u_N \leq M$. Then, adjacent bins have non-overlapping intervals, but these intervals do not necessarily abut each other.

For our numerical example, we can again perform an exhaustive search through the space of $\{l_i, u_i\}$ to find the optimal thresholds. The resulting thresholds are shown in Fig. 4 for various thicknesses of calcium and water, as well as for two and three bins. Each case compares the optimal thresholds when the abutment constraint is relaxed (thick bars) to when abutment is enforced (thin bars beneath). The standard deviation of $\hat{t}_{Ca}$, $\mathrm{Std}(\hat{t}_{\mathrm{Ca}})$, is listed for each case to the right of the chart. The most important result to note is that in the relaxed case, a significant portion of photons does not contribute to either binned measurement. When we consider two bins, the void between u_1 and l_2 represents energies where photons are ignored. Of course, adding a third bin can make use of some of the information contained here, but then there exists a void between u_2 and l_3. Without allowing such a void (by enforcing $u_1 = l_2$ and $u_2 = l_3$), the minimum $\mathrm{Std}(\hat{t}_{\mathrm{Ca}})$ increases 10.8% from 0.355mm to 0.398mm in the first case. Adding a third bin brings $\mathrm{Std}(\hat{t}_{\mathrm{Ca}})$ closer to the ideal-case minimum of 0.308mm. Lastly, we note that for thicker objects, the thresholds of the bins move to higher energies due to a beam hardening effect.

5. WEIGHTED MEASUREMENTS

We now generalize the concept of binning by allowing each detected photon to contribute a real-value amount to each "bin," depending on the photon's energy. Formally, for weights $w_{(i),j} \in \mathbb{R}$, the contribution of photons

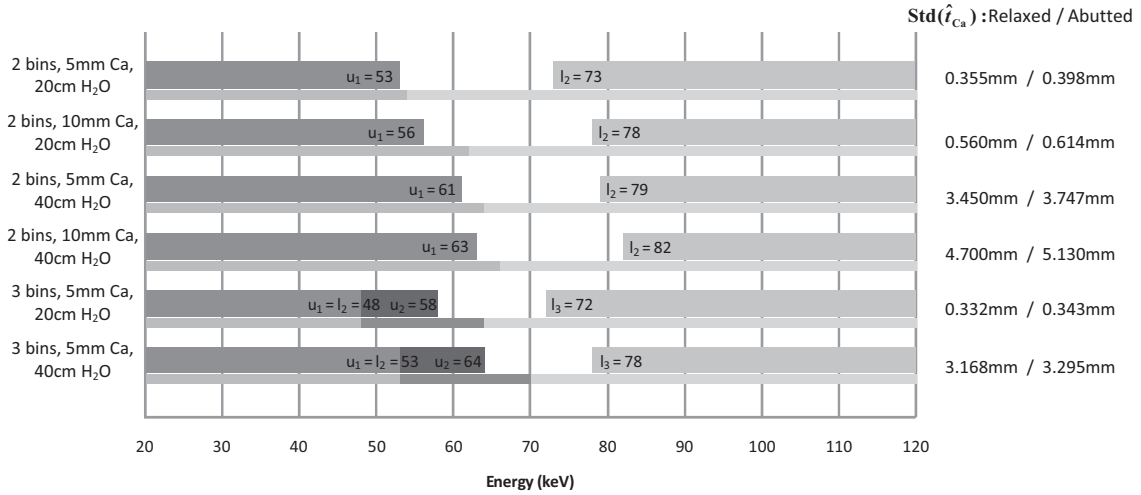


Figure 4. A comparison of thresholds when the abutment constraint is relaxed (thick bars) to when abutment is enforced (thin bars).

of energy j to bin i, the value of each bin measurement d_i is given by

$$d_i = \sum_j w_{(i),j} r_j = w_{(i)}^T r. \tag{5}$$

These weights can also be interpreted as the detector bins' response functions, as a function of energy. Binning counts as was done in the previous sections is a special case within the weighted measurements framework, with weights $w_{(i),j} = 1$ if $l_i < j \leq u_i$ and 0 elsewhere.

Because in general the measurements d are no longer Poisson distributed or uncorrelated, we model d with a multivariate Gaussian distribution: $d \sim \mathcal{N}(e, \Sigma)$, where $e = \mathrm{E}[d]$ and Σ is the covariance matrix of d. This model is sound when the number of transmitted photons is not too low because each d_i is the weighted sum of M independent Poisson random variables, where M is on the order of 100. Also, the first- and second-order moments of the model match those of d.

This leads us to a new likelihood function

$$\mathcal{L}_G(t) = f(d \mid t) = \frac{1}{(2\pi)^{(N/2)}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(d-e)^T\Sigma^{-1}(d-e)\right). \tag{6}$$

The log-likelihood function is

$$\mathcal{L}_G^*(t) = \log \mathcal{L}_G(t) = -\frac{1}{2}\left(\log|\Sigma| + (d-e)^T\Sigma^{-1}(d-e)\right), \tag{7}$$

where the constant term is dropped. Again, we can use error propagation to compute the variance of the estimate $\hat{t}$. The full expressions are shown in Appendix B.

5.1 Finding Optimal Weights

Finding the optimal weights for a given configuration $\{I_0, \mu_1, \mu_2, t\}$ requires a little more sophistication than before. We first note two aspects of the problem. First, we expect the optimal weights $w_{(1)}, \ldots, w_{(N)}$ to be smooth over the energies since μ_1 and μ_2 are themselves smooth as a function of energy, barring a K-edge. When the attenuation curves are smooth, there is little difference in the ability of photons of nearby energies to discriminate the two materials. Therefore, we would not expect the weights of nearby energies to vary much.

Secondly, for weights $W = [\ w_{(1)} \quad \cdots \quad w_{(N)}\ ]$, our MLE performance is no different than if detector readings d' were acquired with weights $W' = WY^T$, where Y is a full-rank $N \times N$ matrix. This is because the readings

that would have been acquired by using weights W can be recovered exactly through the relation $d = Y^{-1}d'$. Therefore, when searching for optimal weights, we must be careful to constrain them in such a way that we are searching for a unique solution.

Depending on the task or goal of the material decomposition, we can write a suitable objective function that we seek to minimize by varying the weights. We regularize the smoothness of the weights by adding a penalty proportional to the quadratic smoothness:

$$\phi(W) = \sum_{i=1}^{N} \sum_{j=1}^{M-1} \left(w_{(i),j+1} - w_{(i),j} \right)^2 . \tag{8}$$

For example, if our goal is to estimate t_1 as precisely as possible, then we would like to find weights W that minimize $\mathrm{Var}(\hat{t}_1) + \eta\phi(W)$, where η is the penalty proportionality constant. Furthermore, as mentioned before, we must constrain the weights to arrive at a unique solution. One simple approach is to arbitrarily fix N of the weights in $w_{(i)}$ for $i = 1, \ldots, N$. After finding the solution with these constraints, we can of course form alternate weights W' through any full-rank matrix Y.

Although in general the objective function may not be convex, we use an approximation of Newton's method for optimization.[12] While it is straightforward to evaluate the objective function for any set of weights, analytically finding expressions for its gradient vector and Hessian matrix can be tedious. Hence, we use numerical differentiation by evaluating the objective function at W and $W \pm \varepsilon\delta_{ij}$ for all i, j, where ε is a small constant and δ_{ij} is a matrix whose element (i, j) is 1 and is zero elsewhere. Furthermore, to reduce computational time, the Hessian matrix is approximated by its diagonal. Finally, we relax the smoothness contraint after each iteration by reducing η so that the smoothness constraint ultimately plays a small part compared to the intended objective.

5.2 Numerical Example

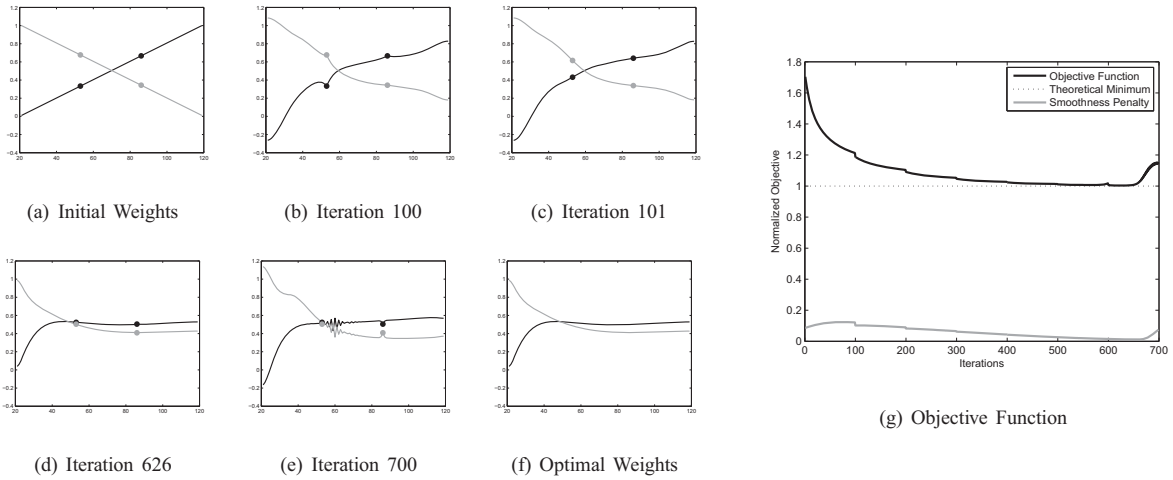
We consider two "bins" whose weights are initialized as two normalized ramp functions: $w_{(1)}$ goes linearly from 0 to 1, while $w_{(2)}$ goes linearly from 1 to 0. We fix the weights at energies 52 keV and 87 keV, approximately 1/3 of the total energy spread apart, so that the curves will be unique. The smoothness regularization parameter, η, is chosen so that $\eta\phi(W)$ is 10% of the initial objective function's value and is exponentially decreased with every iteration so that it halves with every 200 iterations. For comparison purposes, we are interested in finding optimal weights at 5mm Ca and 20cm water. Since 20cm of water significantly attenuates the lower energy photons, we add an additional term to the objective that prevents the contribution of the low energy weights from being dominated by the smoothness penalty. Our objective function is then:

$$\psi(W) = \mathrm{Var}(\hat{t}_{\mathrm{Ca}})|_{t=[.5\ 20]^T} + \mathrm{Var}(\hat{t}_{\mathrm{Ca}})|_{t=[0\ 0]^T} \frac{\mathrm{Var}_{Ideal}(\hat{t}_{\mathrm{Ca}})|_{t=[.5\ 20]^T}}{\mathrm{Var}_{Ideal}(\hat{t}_{\mathrm{Ca}})|_{t=[0\ 0]^T}}, \tag{9}$$

where $\mathrm{Var}(\hat{t}_{\mathrm{Ca}})$ is evaluated at the respective t with weights W, the fraction is a constant term that scales the second term to the same level as the first, and $\mathrm{Var}_{Ideal}(\hat{t}_{\mathrm{Ca}})$ is the variance of the ML estimate with ideal measurements.

Applying our approximate Newton's method to minimizing $\psi(W) + \eta\phi(W)$, we can quickly find smooth weights that begin minimizing the objective function. However, after 100 iterations, it is clear that the weight curves are deviating from the four fixed initial weights and now exhibit cusps (Fig. 5). Since these cusps seem to be artificially induced, we "shave" them off to obtain smooth weights by estimating the weights in the cusp regions with a local polynomial fit. Doing this every 100 iterations appears to be an effective technique since there is a rapid drop in both the calcium estimate variance $\psi(W)$ and the smoothness penalty $\phi(W)$ after each "shave" (Fig. 5g).

If we continue to run Newton's method for many iterations, high frequency oscillations in the weights eventually begin to appear, possibly due to numerical or round-off errors. These errors make the smoothness penalty and objective function rise rapidly, at which point our method is no longer updating the weights in a useful direction. Fortunately, before this occurs, the minimum $\psi(W)$ at iteration 626 is a sufficiently close solution to the ideal minimum. After smoothing out the cusps at the fixed points, we arrive at an optimal set of weights $\tilde{W} = [\tilde{w}_{(1)}\ \tilde{w}_{(2)}]$.



(a) Initial Weights (b) Iteration 100 (c) Iteration 101

(d) Iteration 626 (e) Iteration 700 (f) Optimal Weights

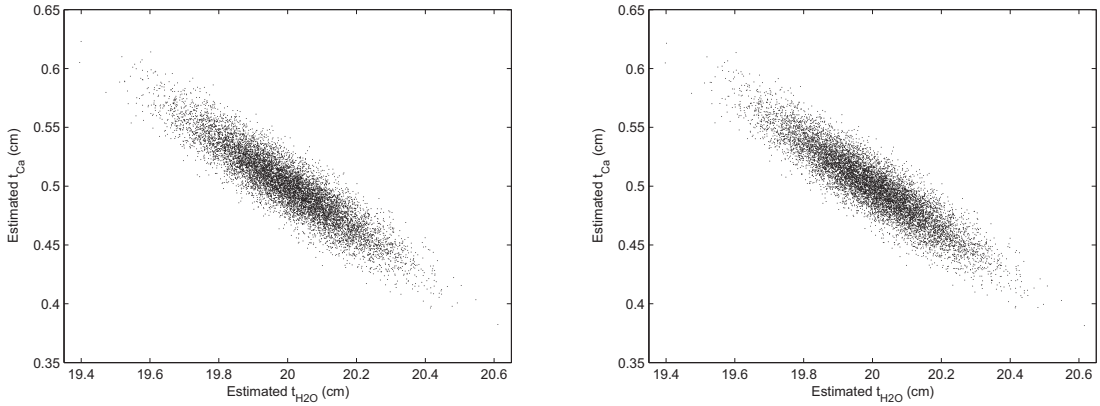
(g) Objective Function

Figure 5. Solving for optimal weights using an approximation of Newton's method.

5.3 Optimal Weights Performance

We tested this method of weighted binning of PCXD data by simulating noisy data and examining the performance of our MLE in estimating t. The system of equations in (12) in Appendix B describes the MLE solution for estimating t from the weighted binning data. Solving this non-linear system of equations analytically is not straightforward. Instead, we again use Newton's method to find the maximum of $\mathcal{L}_G^*$. The expression for the gradient of the MLE was already found for (12), and the Hessian matrix can be similarly found. Although the MLE function is not convex in t, we found that it is well-behaved, and the Newton method works well in finding the maximum.

For $t = [.5\ 20]^T$, we generated 10,000 realizations of ideally measured noisy data $\{r_1, r_2, \ldots, r_M\}$ and the corresponding weighted measurements $\{\tilde{d}_1, \tilde{d}_2\}$ using optimal weights $\tilde{W}$. The resulting MLE estimates of t for each sample are displayed in Figure 6. The sample variances and covariance of the estimates (Table 1) show that the empirical performance of the optimal weights match that of the ideal 1keV bins, all well within 1% of each other.



(a) Optimal weights (b) Ideal 1keV bins

Figure 6. Material decomposition estimation of 10,000 realizations for $t = [.5\ 20]^T$.

While these weights have near-optimal performance for estimating calcium thickness at 5mm Ca, 20cm water, it is not clear how they will perform at estimating water or for other thicknesses. We can predict this by finding

Table 1. Comparison of empirical performance to predicted theory for $t_{\mathrm{Ca}} = 0.5$cm, $t_{\mathrm{H_2O}} = 20$cm.

	$\mathrm{Var}(\hat{t}_{\mathrm{Ca}})$	$\mathrm{Var}(\hat{t}_{\mathrm{H_2O}})$	$\mathrm{Cov}(\hat{t}_{\mathrm{Ca}}, \hat{t}_{\mathrm{H_2O}})$
Empirical, Ideal	9.511×10^{-4}	2.241×10^{-2}	-4.196×10^{-3}
Empirical, Weights	9.519×10^{-4}	2.241×10^{-2}	-4.198×10^{-3}
Predicted, Ideal	9.464×10^{-4}	2.256×10^{-2}	-4.200×10^{-3}
Predicted, Weights	9.457×10^{-4}	2.255×10^{-2}	-4.197×10^{-3}

the theoretical variance of the estimates at various thicknesses using $\tilde{W}$ and comparing it to the theoretically predicted variance of the estimates resulting from our ideal 1keV binning that completely disaggregates the photon counts of different energies (Table 1). Furthermore, for the variances and the covariance of the estimates $\hat{t}_{\mathrm{Ca}}$ and $\hat{t}_{\mathrm{H_2O}}$, we normalize the performance of optimal weights $\tilde{W}$ by dividing the equivalent figure of merit for the ideal measurements and plot these over a wide range of diagnostically relevant t (Fig. 7). As can be seen, the values for weighted binning are essentially equal to unity over most of the range of thicknesses. Thus, even though the weights were computed for one object (5mm Ca, 20cm water), theory predicts that weighted binning with these weights performs as well as the original data over a wide range of objects. Note that for thick objects the analytical prediction of performance predicts variances lower than those of the ideal case, which cannot be correct. We believe that the Gaussian model used in (6) may be violated when very few photons penetrate the object, causing the predicted variance from the optimally weighted measurements to be less than that of the ideal case. In support of this hypothesis, we found that if we scale the incident intensity I_0 by a factor of 20, this problem largely disappears (Fig. 7d).

Nonetheless, our results suggest that using weights $\tilde{W}$ to measure only two parameters, $\{\tilde{d}_1, \tilde{d}_2\}$, for any thicknesses of calcium and water, we can estimate $\{t_{\mathrm{Ca}}, t_{\mathrm{H_2O}}\}$ as well as if we knew the detected photon counts at every energy. Even though weights $\tilde{W}$ are optimized for 5mm Ca, 20cm water and to minimize $\mathrm{Var}(\hat{t}_{\mathrm{Ca}})$, they are in fact optimal for any amount of any two materials, simply as long as the attenation of these materials can be expressed as a linear combination of the attenuation of calcium and water because estimates from data measured with weights $\tilde{W}$ have the same variances and covariance as the ideal measurements. However, note that finding optimal weights $\tilde{W}$ also depends on the shape of the incident spectrum. If I_0 is changed, then the weights we computed may no longer be optimal.

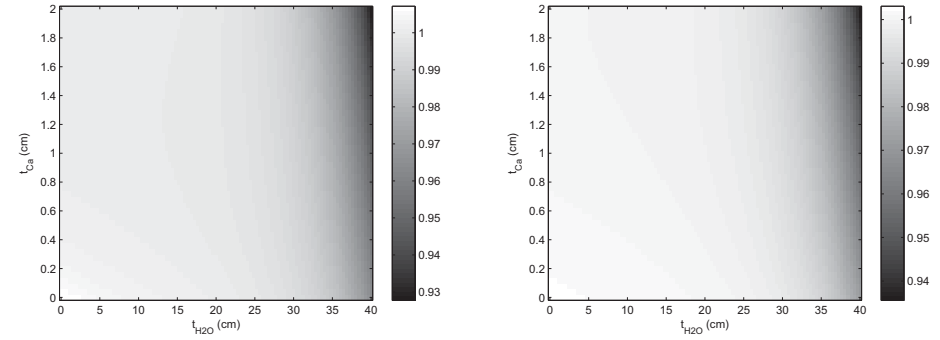
6. DISCUSSION

Our first main result is that relaxing the abutment constraint for binning thresholds can improve material decomposition performance by allowing some photons to be discarded. This is reminiscent of the use of a filter in between the two layers of a front-back detector to improve performance.[10] The benefit of this void region can be understood by examining the sensitivity of each material thickness estimate to an additional photon at various energies with ideal measurements (Fig. 8). When we only have two energy bins, additional photons in d_2 decrease the water estimate and increase the calcium estimate while doing the opposite for additional photons in d_1 because calcium has a relatively higher attenuation than water at lower energies than at higher energies. Our MLE for ideal measurements has very little sensitivity to photons in the void region (shaded region in the figure). Including photons in the shaded region into either of two bins adds noise while providing little sensitivity to material separation. Hence, the precision of the estimator using $\{d_1, d_2\}$ is improved if we do not count certain photons into either bin.

Our second main result is that optimal weights $\tilde{W}$ can be found that condense all of the photon counts and energy information into two measurements $\{\tilde{d}_1, \tilde{d}_2\}$ and still allow for t to be estimated as well as if we had ideal measurements of all photons. As mentioned before, since any full-rank matrix Y can transform weights $\tilde{W}$ into another set of optimal weights, it is not immediately clear what these weights represent.

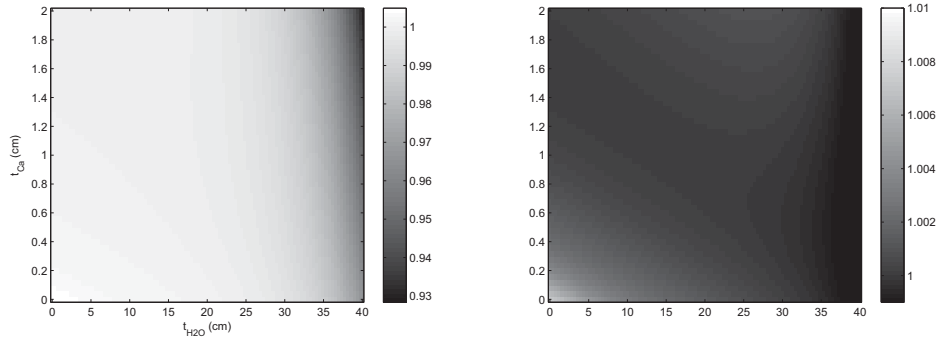
7. CONCLUSION

PCXDs offer a wealth of new information about the object being measured that traditional energy integrating detectors cannot assess. Our work offers new insight into increasing the precision of material or basis function



(a) Normalized Var($\hat{t}_{\mathrm{Ca}}$)

(b) Normalized Var($\hat{t}_{\mathrm{H_2O}}$)



(c) Normalized Cov($\hat{t}_{\mathrm{Ca}}, \hat{t}_{\mathrm{H_2O}}$)

(d) Normalized Var($\hat{t}_{\mathrm{Ca}}$), $I_0' = 20I_0$

Figure 7. Performance of optimal weights as a function of t, normalized to ideal performance.

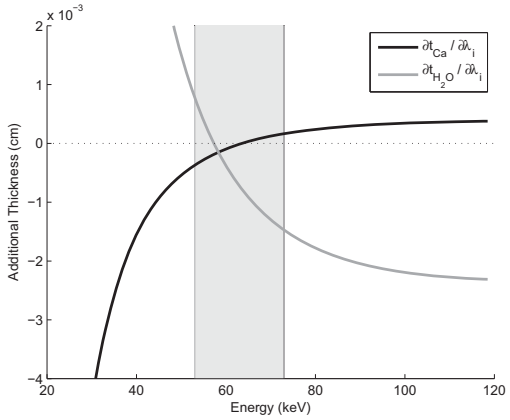


Figure 8. The change in estimated thickness using ideal measurements for an additional photon detected, at 5mm Ca and 20cm water.

decomposition from binned or weighted measurements by providing the theoretical groundwork to predict estimator performance and then the methods to optimize the performance. Interestingly, we found that having a gap between energy thresholds can improve the precision of our system. Photons with energies that fall into this gap should either be discarded or counted separately to improve material separability.

Moreover, when estimating the amount of each of two material basis functions, two appropriately weighted sums of the photon counts provide as much information as knowing the number of counts at each energy. Through theoretical predictions and empirical measurements, we showed that for a given x-ray spectrum, these weights are optimal for any amount of any two materials, as long as these materials can be expressed as a linear combination

of photoelectric and Compton attenuation.

ACKNOWLEDGEMENT

The authors would like to thank GE Healthcare and the Lucas Foundation for their financial support.

REFERENCES

[1] Flohr, T. G., McCollough, C. H., et al., "First performance evaluation of a dual-source CT (DSCT) system," Eur. Radiol. 16(2), 256-268 (2006).
[2] Tkaczyk, J. E., Rodrigues, R., et al., "Atomic number resolution for three spectral CT imaging systems," Proc. SPIE 6510, 651009 (2007).
[3] Roessl, E., Ziegler, A., and Proksa, R. "On the influence of noise correlations in measurement data on basis image noise in dual-energy like x-ray imaging," Med. Phys. 34(3), 959-966 (2007).
[4] Taguchi, K., Zhang, M., et al., "Image-domain material decomposition using photon-counting CT," Proc. SPIE 6510, 651008 (2007).
[5] Frey, E. C., Wang, X., et al., "Investigation of the use of photon counting x-ray detectors with energy discrimination capability for material decomposition in micro-computed tomography," Proc. SPIE 6510, 65100A (2007).
[6] Xu, J., Frey, E. C., Taguchi, K., and Tsui, B. M. "A Poisson likelihood iterative reconstruction algorithm for material decomposition in CT," Proc. SPIE 6510, 65101Z (2007).
[7] Alvarez, R. E. and Macovski, A., "Energy-selective reconstructions in x-ray computerized tomography," Phys. Med. Biol. 21(5), 733-744 (1976).
[8] Lehmann, L. A., Alvarez, R. E., Macovski, A., et al., "Generalized image combinations in dual KVP digital radiography," Med. Phys. 8(5), 659-667 (1981).
[9] Fessler, J. A., Elbakri, I., Sukovic, P., and Clinthorne, N. H., "Maximum-likelihood dual-energy tomographic image reconstruction," Proc. SPIE 4684, 38-49 (2002).
[10] Stevens, G. M. and Pelc, N. J., "Depth-segmented detector for x-ray absorptiometry," Med. Phys. 27(5), 1174-1184 (2000).
[11] Cowan, G., [Statistical Data Analysis] Oxford University Press, Oxford, p. 21, eq. (1.54) (2004).
[12] Boyd, S. and Vandenberghe, L., [Convex Optimization] Cambridge University Press, Cambridge, p. 487, §9.5.2, (2004).

APPENDIX A. POISSON MODEL

To find the maximum of the Poisson model log-likelihood function, we take the derivative of $\mathcal{L}_P^*$ with respect to t and set that equal to zero:

$$\begin{aligned} 0 &= \sum_i \gamma_i^{(1)} \left(d_i/\gamma_i - 1\right) \\ 0 &= \sum_i \gamma_i^{(2)} \left(d_i/\gamma_i - 1\right), \end{aligned} \tag{10}$$

where $\gamma_i^{(1)} = \sum_{j=\tau_{i-1}+1}^{\tau_i} \mu_{1,j}\lambda_j$ and $\gamma_i^{(2)} = \sum_{j=\tau_{i-1}+1}^{\tau_i} \mu_{2,j}\lambda_j$. The solution to (10), $\hat{t}$, is our ML estimate.

To find the necessary terms to estimate $\mathrm{Cov}(\hat{t}_i, \hat{t}_j)$ in (4), we take the partial of (10) with respect to d_i. For the first equation, we get:

$$0 = \sum_{j\neq i} \left[\left(\frac{\partial \gamma_j^{(1)}}{\partial d_i} \right) \left(d_j/\gamma_j - 1\right) - \gamma_j^{(1)} \left(d_j/\gamma_j^2\right) \left(\frac{\partial \gamma_j}{\partial d_i} \right) \right] + \left[\left(\frac{\partial \gamma_i^{(1)}}{\partial d_i} \right) \left(d_i/\gamma_i - 1\right) + \gamma_i^{(1)}/\gamma_i \left(1 - d_i/\gamma_i \left(\frac{\partial \gamma_i}{\partial d_i} \right) \right) \right]$$

Since we are expanding (3) about $\hat{t} = t$, we can substitute $d_i = \gamma_i$. Also, applying the chain rule, we find

$$\frac{\partial \gamma_j}{\partial d_i} = \left(\frac{\partial \hat{t}_1}{\partial d_i}\right)\gamma_j^{(1)} + \left(\frac{\partial \hat{t}_2}{\partial d_i}\right)\gamma_j^{(2)}.$$

Substituting these two things, we find:

$$\gamma_i^{(1)}/\gamma_i = \left(\sum_j \left(\gamma_j^{(1)}\right)^2/\gamma_j\right)\left(\frac{\partial \hat{t}_1}{\partial d_i}\right) + \left(\sum_j \left(\gamma_j^{(1)}\gamma_j^{(2)}\right)/\gamma_j\right)\left(\frac{\partial \hat{t}_2}{\partial d_i}\right)$$

We arrive at a similar expression from taking the partial of the second equation of (10) with respect to d_i. Solving these new expressions for the partial derivatives yields:

$$\begin{bmatrix} \frac{\partial \hat{t}_1}{\partial d_i} \\ \frac{\partial \hat{t}_2}{\partial d_i} \end{bmatrix} = \begin{bmatrix} \sum_j \left(\gamma_j^{(1)}\right)^2/\gamma_j & \sum_j \left(\gamma_j^{(1)}\gamma_j^{(2)}\right)/\gamma_j \\ \sum_j \left(\gamma_j^{(1)}\gamma_j^{(2)}\right)/\gamma_j & \sum_j \left(\gamma_j^{(2)}\right)^2/\gamma_j \end{bmatrix}^{-1} \begin{bmatrix} \gamma_i^{(1)}/\gamma_i \\ \gamma_i^{(2)}/\gamma_i \end{bmatrix} \tag{11}$$

Since d_k is a Poisson random variable, $\mathrm{Var}(d_k) = \gamma_k$, and we arrive at an expression for $\mathrm{Var}(\hat{t}_1)$, $\mathrm{Var}(\hat{t}_2)$, and $\mathrm{Cov}(\hat{t}_1, \hat{t}_2)$ by substituting (11) into (4).

APPENDIX B. GAUSSIAN MODEL

We defined $e = \mathrm{E}[d]$ and Σ to be the covariance matrix of d. Additionally, define matrix W and vector $w_{(k,l)}$ as

$$W = \begin{bmatrix} w_{(1)} & \cdots & w_{(N)} \end{bmatrix}, \quad w_{(k,l)} = \begin{bmatrix} w_{(k),1}w_{(l),1} & \cdots & w_{(k),M}w_{(l),M} \end{bmatrix}^T.$$

Then $e = W^T\lambda$ and $\Sigma_{ij} = w_{(i,j)}^T\lambda$. To find the maximum, we set the partial of the log-likelihood function with respect to t equal to 0. For the partial with respect to t_1,

$$\begin{aligned} 0 = \frac{\partial \mathcal{L}_G^*}{\partial t_1} &= -\frac{1}{2}|\Sigma|^{-1}\left(\frac{\partial |\Sigma|}{\partial t_1}\right) - \frac{1}{2}\left[-2\left(\frac{\partial e}{\partial t_1}\right)^T\Sigma^{-1}(d-e) + (d-e)^T\left(\frac{\partial \Sigma^{-1}}{\partial t_1}\right)(d-e)\right] \\ &= -\frac{1}{2}\left[\mathrm{tr}\left(\Sigma^{-1}\Sigma^{(1,0)}\right) - 2e^{(1,0)T}\Sigma^{-1}(d-e) - (d-e)^T\Sigma^{-1}\Sigma^{(1,0)}\Sigma^{-1}(d-e)\right], \end{aligned}$$

where we substituted the following:

$$\frac{\partial |\Sigma|}{\partial t_1} = |\Sigma|\mathrm{tr}\left(\Sigma^{-1}\left(\frac{\partial \Sigma}{\partial t_1}\right)\right), \quad \frac{\partial \Sigma^{-1}}{\partial t_1} = -\Sigma^{-1}\left(\frac{\partial \Sigma}{\partial t_1}\right)\Sigma^{-1}$$

and defined vector $\lambda^{(k,l)}$ to have elements $\lambda_i^{(k,l)} = (\mu_{1,i})^k(\mu_{2,i})^l\lambda_i$, matrix $\Sigma^{(k,l)}$ to have elements $\Sigma_{ij}^{(k,l)} = w_{(i,j)}^T\lambda^{(k,l)}$, and vector $e^{(k,l)} = W^T\lambda^{(k,l)}$. Then, for instance, matrix $\frac{\partial \Sigma}{\partial t_1} = \Sigma^{(1,0)}$ since element $\left(\frac{\partial \Sigma}{\partial t_1}\right)_{ij} = w_{(i,j)}^T\left(\frac{\partial \lambda}{\partial t_1}\right) = w_{(i,j)}^T\lambda^{(1,0)}$.

Similarly, we can find the partial of $\mathcal{L}_G^*$ with respect to t_2. Therefore, the ML estimate, $\hat{t}$, solves the two equations:

$$\begin{aligned} 0 &= -\tfrac{1}{2}\left[\mathrm{tr}\left(\Sigma^{-1}\Sigma^{(1,0)}\right) - 2e^{(1,0)T}\Sigma^{-1}(d-e) - (d-e)^T\Sigma^{-1}\Sigma^{(1,0)}\Sigma^{-1}(d-e)\right], \\ 0 &= -\tfrac{1}{2}\left[\mathrm{tr}\left(\Sigma^{-1}\Sigma^{(0,1)}\right) - 2e^{(0,1)T}\Sigma^{-1}(d-e) - (d-e)^T\Sigma^{-1}\Sigma^{(0,1)}\Sigma^{-1}(d-e)\right]. \end{aligned} \tag{12}$$

To find $\frac{\partial \hat{t}_1}{\partial d_i}$, we take the partial of the equations in (12) with respect to d_i. For the first equation,

$$\begin{aligned}
0 = &\operatorname{tr}\left(-\Sigma^{-1}\left(\frac{\partial \Sigma}{\partial d_i}\right)\Sigma^{-1}\Sigma^{(1,0)} + \Sigma^{-1}\left(\frac{\partial \Sigma^{(1,0)}}{\partial d_i}\right)\right) \\
&- 2\left(\left(\frac{\partial e^{(1,0)}}{\partial d_i}\right)^T \Sigma^{-1}(d-e) - e^{(1,0)T}\Sigma^{-1}\left(\frac{\partial \Sigma}{\partial d_i}\right)\Sigma^{-1}(d-e) + e^{(1,0)}\Sigma^{-1}\left(\delta_i - \left(\frac{\partial e}{\partial d_i}\right)\right)\right) \\
&- \left[2\left(\frac{\partial}{\partial d_i}\left(\Sigma^{-1}(d-e)\right)\right)^T \Sigma^{(1,0)}\left(\Sigma^{-1}(d-e)\right) + \left(\Sigma^{-1}(d-e)\right)^T\left(\frac{\partial \Sigma^{(1,0)}}{\partial d_i}\right)\left(\Sigma^{-1}(d-e)\right)\right]
\end{aligned}$$

where δ_i is the unit vector in the i-th dimension. Since we are interested in finding $\frac{\partial \hat{t}}{\partial d_i}$ about $d = e$, we can set $d - e = 0$. Also, $\frac{\partial \Sigma^{(k,l)}}{\partial d_i} = \Sigma^{(k+1,l)}\left(\frac{\partial \hat{t}_1}{\partial d_i}\right) + \Sigma^{(k,l+1)}\left(\frac{\partial \hat{t}_2}{\partial d_i}\right)$ and $\frac{\partial e^{(k,l)}}{\partial d_i} = e^{(k+1,l)}\left(\frac{\partial \hat{t}_1}{\partial d_i}\right) + e^{(k,l+1)}\left(\frac{\partial \hat{t}_2}{\partial d_i}\right)$.

Hence, we get

$$\begin{aligned}
0 = &\operatorname{tr}\left(\Sigma^{-1}\Sigma^{(2,0)} - \left(\Sigma^{-1}\Sigma^{(1,0)}\right)^2\right)\left(\frac{\partial \hat{t}_1}{\partial d_i}\right) + \operatorname{tr}\left(\Sigma^{-1}\Sigma^{(1,1)} - \Sigma^{-1}\Sigma^{(0,1)}\Sigma^{-1}\Sigma^{(1,0)}\right)\left(\frac{\partial \hat{t}_2}{\partial d_i}\right) \\
&- 2e^{(1,0)T}\Sigma^{-1}\left[\delta_i - \left(e^{(1,0)}\left(\frac{\partial \hat{t}_1}{\partial d_i}\right) + e^{(0,1)}\left(\frac{\partial \hat{t}_2}{\partial d_i}\right)\right)\right]
\end{aligned}$$

Combining the results from taking the partial of the second equation in (12) with respect to d_i,

$$\begin{bmatrix} \frac{\partial \hat{t}_1}{\partial d_i} \\ \frac{\partial \hat{t}_2}{\partial d_i} \end{bmatrix} = \begin{bmatrix} S & R \\ R & T \end{bmatrix}^{-1} \begin{bmatrix} e^{(1,0)T}\Sigma^{-1}\delta_i \\ e^{(0,1)T}\Sigma^{-1}\delta_i \end{bmatrix},$$

where

$$\begin{bmatrix} S & R \\ R & T \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\Sigma^{(2,0)} - \Sigma^{-1}\Sigma^{(1,0)}\Sigma^{-1}\Sigma^{(1,0)}\right) + e^{(1,0)T}\Sigma^{-1}e^{(1,0)} & \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\Sigma^{(1,1)} - \Sigma^{-1}\Sigma^{(1,0)}\Sigma^{-1}\Sigma^{(0,1)}\right) + e^{(1,0)T}\Sigma^{-1}e^{(0,1)} \\ \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\Sigma^{(1,1)} - \Sigma^{-1}\Sigma^{(1,0)}\Sigma^{-1}\Sigma^{(0,1)}\right) + e^{(1,0)T}\Sigma^{-1}e^{(0,1)} & \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\Sigma^{(0,2)} - \Sigma^{-1}\Sigma^{(0,1)}\Sigma^{-1}\Sigma^{(0,1)}\right) + e^{(0,1)T}\Sigma^{-1}e^{(0,1)} \end{bmatrix}$$

Now we can compute the variances and covariance of the ML estimates. Here we compute the variance of estimate $\hat{t}_1$:

$$\begin{aligned}
\operatorname{Var}\left(\hat{t}_1\right) &\approx \sum_i \sum_j \left(\frac{\partial \hat{t}_1}{\partial d_i}\right)\left(\frac{\partial \hat{t}_1}{\partial d_j}\right)\operatorname{Cov}\left(d_i, d_j\right) \\
&= \left(\frac{1}{ST - R^2}\right)^2 \sum_i \sum_j \left(Te^{(1,0)T}\Sigma^{-1}\delta_i - Re^{(0,1)T}\Sigma^{-1}\delta_i\right)\left(Te^{(1,0)T}\Sigma^{-1}\delta_j - Re^{(0,1)T}\Sigma^{-1}\delta_j\right)\left(w_{(i,j)}^T\lambda\right) \\
&= \left(\frac{1}{ST - R^2}\right)^2 \left[T^2\left(e^{(1,0)T}\Sigma^{-1}e^{(1,0)}\right) - 2TR\left(e^{(1,0)T}\Sigma^{-1}e^{(0,1)}\right) + R^2\left(e^{(0,1)T}\Sigma^{-1}e^{(0,1)}\right)\right].
\end{aligned}$$