TITLE: ANALYSIS OF LAM PATIENT TOMOGRAMS:
STEREOLOGY AND MODELING

AUTHOR(S): V. Faber, G. M. Wing, L. Winter, J. D. Zahrt

SUBMITTED TO: External Distribution

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive royalty-free license to publish or reproduce the published form of this contribution or to allow others to do so, for U.S. Government purposes.

The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Los Alamos

MASTER
Los Alamos National Laboratory
Los Alamos New Mexico 87545

Analysis of LAM Patient Tomograms: Stereology and Modeling

Vance Faber, G. Milton Wing, Larry Winter, and John D. Zahrt
Los Alamos National Laboratory
Computing, Information, and Communications Division
Los Alamos, New Mexico

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

FOREWORD

This is primarily an applications paper using the Abel integral equation to interpret the CT (computed tomography) images of the lungs of LAM (lymphangioleiomyomatosis) patients.

There is mathematics in this report, but it is hoped that the approach taken here is accessible to all who have finished calculus and had a semester of linear algebra.

There exists an earlier Los Alamos manuscript (LA-11016-MS) on, "The Abel Integral Equation," by G. Milton Wing and Vance Faber which guided the numerics of this report. We hope, in the near future, to issue another report on the Abel equation which focuses more on the analytic mathematics. For those new to the field of integral equations and wish a painless introduction the book by G. Milton Wing, "A Primer on Integral Equations of the First Kind," is highly recommended. Another recent reference for work with the Abel equation is Peter Linz's book, "Analytical and Numerical Methods for Volterra Equations" (and included references).

Finally, and most importantly, it is hoped that the material herein and the results reported here, in some small way will enable physicians to improved diagnosis and/or treatment of LAM patients.

Contents

1	Introduction	1
2	Conversion of Wicksell's Equation to Area Variables	2
3	The Adjoint	4
4	Analytic Solutions to the Adjoint Equation	6
5	Power Law Model	7
6	Exponential Model	21
7	Solution of Wicksell's Area Equation and Use of the Golberg Device	24
8	The Total Number of Cysts and Their Average Size	27
9	Conclusions	39
	References	41
	APPENDIX A: Derivation of Wicksell's Equation	42
	APPENDIX B: Calculation of the Number Spheres	45
	APPENDIX C: The 'Normal' Lung	47

1 Introduction

Lymphangioleiomyomatosis (LAM) is an insidious disease of young women which attacks the lungs and if untreated results in death. Treatment is radical. Currently a woman diagnosed as having LAM (usually by computed tomography (CT) of the lung) is run through an exhausting set of tests and exercises to determine the extent of the disease.

Image processing techniques have been able to quantify the tomograms by providing a histogram of the number of cyst intercepts of a given area. The images are digital, so the data are discrete, and the areas are in terms of square pixels. It is hoped that by analyzing the data and by recovering the probability density of the cysts themselves (as opposed to their intersection with the CT plane) we can provide a set of parameters that will aid in the early diagnosis of the disease and that will correlate well with the physiology.

The desire to invert the data, that is to determine the probability density of the cysts from the probability density of the circular intersections of the cysts with the tomographic plane, led to a re-examination of the Wicksell equation. Wicksell [1], in 1925, derived a generalized Abel integral equation relating the distribution of diameters of spheres (corpuscles, follicles, cysts) to the measured distribution of diameters of the circular contours obtained (assumed random) in a plane section through the spheres. His derivation assumes that the corpuscles are approximately spherical in form and that they are fairly uniformly distributed in the sample (see Appendix A). It turns out that the Wicksell equation is a specific case of the set of generalized Abel equations first introduced by Cormack [2] to solve the tomography problem.

That the two distributions (of the true diameters of the spheres and of the diameters of the circular projections of these spheres) are generally different is clear from the following two considerations:

- (1) The apparent diameters are as a rule smaller than the true diameters, and (2) a random section will contain a relatively greater number of actually large follicles¹, than of small ones, because the former more frequently will come within reach of the section plane. The two causes will, however, work in opposite directions, thus having a chance to balance each other. [1]

The last statement by Wicksell above implies that he was able to find an eigenfunction of his integral

¹sphere in our case

equation. Our interest in this will be apparent in Section 2.

In Wicksell's era, they measured diameters with a ruler even taking the average of multiple measurements in some cases. Other measurements of areas came from cutting out images on paper and weighing them. Today, image processing determines the number and area of these circular projections of the cysts by computer algorithms without any human intervention. So over the years the natural variable has changed from diameter to area.

The first three sections are preliminary, more mathematical, but meant to be self-contained, and not highly motivated. They may be skipped by the reader willing to take later results on faith.

In Section 2, we transform Wicksell's equation in diameter (a generalized Abel equation) into an equation in areas (the original Abel equation). In Section 3 we present the adjoint equation in several equivalent forms. Analytic solutions of the adjoint equation necessary for later use are developed in Section 4. Outside of the framework of integral equations we have done some modeling of the data. We find that, to a very high degree of reliability, the data follows a simple power law. The distribution of spheres likewise follows a simple power law. These considerations are given in Section 5. Then in Section 6 we explore the exponential model as an alternative. Section 7 solves the Wicksell equation and uses the Golberg device, via the adjoint, to solve a related problem. We finally get to the number of cysts and their average size in Section 8. Our conclusions are in Section 9.

We also present three appendices. For completeness we present the derivation of Wicksell's equation in Appendix A. We use his arguments with a slight change in symbols. Even our modified Wicksell equation give only a probability density of there being cysts of a certain size. In order to determine the number of cysts in the lung we need more machinery. The probability machinery necessary to estimate the number of cysts is developed in Appendix B. In Appendix C we present an analysis of our one example of a 'normal' lung. This data and its analysis were instrumental in the development of the double exponential model but otherwise plays no role in the analysis of LAM patient data.

2 Conversion of Wicksell's Equation to Area Variables

The Wicksell equation is

$$\hat{g}(x) = \frac{x}{y_0} \int_x^\infty dy \frac{\hat{f}(y)}{\sqrt{y^2 - x^2}} , \quad (2.1)$$

where x is the diameter of the circles in the plane, y is the diameter of the spheres, y_0 is the average diameter of the spheres, $\hat{g}(x)$ is the normalized probability density of the circles in the data, and $\hat{f}(y)$ is the normalized probability density of the spheres. As implied above

$$\int_0^\infty \hat{g}(x) dx = 1 , \quad (2.2)$$

and

$$\int_0^\infty \hat{f}(y) dy = 1 . \quad (2.3)$$

Now let

$$a = \pi x^2 / 4 , \quad (2.4a)$$

$$A = \pi y^2 / 4 , \quad (2.4b)$$

and using these and (2.1) gives

$$\hat{g}((4a/\pi)^{1/2}) = \frac{\sqrt{a}}{y_0} \int_a^\infty dA \frac{(\pi A)^{-1/2} \hat{f}((4A/\pi)^{1/2})}{\sqrt{A - a}} . \quad (2.5)$$

Since we desire to keep the structure of (2.2) and (2.3) we define

$$\hat{G}(a) = \hat{g}((4a/\pi)^{1/2}) (\pi a)^{-1/2} , \quad (2.6)$$

and

$$\hat{F}(A) = \hat{f}((4A/\pi)^{1/2}) (\pi A)^{-1/2} , \quad (2.7)$$

and thus rewrite (2.5) as

$$\hat{G}(a) = \frac{1}{\sqrt{\pi}y_0} \int_a^\infty da \frac{\hat{F}(a)}{\sqrt{A-a}} . \quad (2.8)$$

A further definition,

$$A_0 \equiv \frac{\pi}{4} y_0^2 , \quad (2.9)$$

allows a final rewrite of (2.8) as

$$\hat{G}(a) = \frac{1}{2\sqrt{A_0}} \int_a^\infty dA \frac{\hat{F}(A)}{\sqrt{A-a}} . \quad (2.10)$$

This is just Abel's equation which is well known. It perhaps should be noted here that $\hat{G}(a)$ is related to the half integral of $\hat{F}(A)$. This becomes more apparent in Section 4.

Wicksell was able to show that

$$\psi(y) = \frac{y}{\sigma^2} \exp(-y^2/2\sigma^2) , \quad (2.11)$$

where $\sigma = y_0\sqrt{2/\pi}$ is an eigenfunction of (2.1). By the transformation indicated in (2.4a,b) one sees then that

$$\hat{\psi}(A) = \frac{1}{\sigma} e^{-A/\sigma} , \quad (2.12)$$

where $\sigma = 4A_0/\pi$, is an eigenfunction of (2.10), namely

$$\psi(a) = \frac{1}{2\sqrt{A_0}} \int_a^\infty dA \frac{\psi(A)}{\sqrt{A-a}} . \quad (2.13)$$

3 The Adjoint

In the coming sections we will have occasion to need the adjoint equation to (2.10). This requirement stems from use in Section 7 of the Golberg device [3] which we sketch below.

Let us digress here and expound on this device. Suppose we have the integral equation

$$Kf = g \quad , \quad (3.1)$$

and we don't desire knowledge of f so much as we desire some moment of f

$$M = \int_0^\infty f(x)m(x)dx = (f, m) \quad . \quad (3.2)$$

Then let us solve

$$K^*\mu = m \quad , \quad (3.3)$$

so that

$$(f, m) = (f, K^*\mu) = (Kf, \mu) = (g, \mu) \quad . \quad (3.4)$$

In (3.4) we have used (3.3), the property of the adjoint and (3.1).

The classical *raison d'être* for the use of (3.4) and (3.3) is that $m(x)$ is presumed known without noise hence μ is known with more reliability than f . In our case K is not so ill-posed that we could not use (3.2) but our reasons for using (3.4) is that for an m of interest to us (see Section 5) (3.3) has an analytic solution. No computation algorithm for the solution of an integral equation is necessary.

The adjoint operator to Wicksell's integral operator (Cormack's $n = 1$) is [4]

$$\tilde{K}^* = \int_0^r \tilde{K}^*(r, x) \cdot dx \quad , \quad (3.5)$$

$$\tilde{K}^*(r, x) = \begin{cases} x/(r_0\sqrt{r^2 - x^2}) & 0 < x < r \\ 0 & \text{otherwise} \end{cases} \quad . \quad (3.6)$$

With (2.4a,b), (3.5) becomes

$$K^*(A, a) = \begin{cases} 1/(2\sqrt{A_0}\sqrt{A-a}) & 0 \leq a \leq A \\ 0 & \text{otherwise} \end{cases} \quad (3.7)$$

which is identical to the kernel of (2.10) but with a change in limits. It is useful to let

$$A = \alpha - t \quad , \quad a = \alpha - s$$

so that

$$K^*(\alpha - t, \alpha - s) = \begin{cases} 1/(2\sqrt{A_0}\sqrt{A-t}) & t \leq s \leq \alpha \\ 0 & \text{otherwise} \end{cases} \quad (3.8)$$

With $a = \alpha = 1$ (3.7) and (2.10) become equal in form and will become very useful.

4 Analytic Solutions to the Adjoint Equation

We have found the adjoint kernel (3.7) such that the adjoint equation to the area form of Wicksell's equation is

$$\frac{1}{2\sqrt{A_0}} \int_0^A \frac{\mu(a)}{\sqrt{A-a}} da = m(A) \quad (4.1)$$

We can speak of $m(A)$ being the Abel transform of $\mu(a)$. Let us see what $m(A)$ becomes if

$$\mu(a) = 2\sqrt{A_0}a^n \quad , \quad n \geq -\frac{1}{2} \quad (4.2)$$

The integral is related to the beta function or can be evaluated by Laplace transforms, but we choose a more elementary approach.

Let

$$a = A \sin^2 \theta \quad , \quad (4.3)$$

and substituting this into (4.1) gives

$$m(A) = A^{n+\frac{1}{2}} 2 \int_0^{\pi/2} \sin^{2n+1} \theta d\theta \quad , \quad n \geq -\frac{1}{2} \quad (4.4)$$

A table of definite integrals gives the integral and the result is

$$m(A) = A^{n+\frac{1}{2}} \frac{\sqrt{\pi} \Gamma(n+1)}{\Gamma(n+\frac{3}{2})} \quad (4.5)$$

where $\Gamma(z)$ is the gamma function with the following relevant properties

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad , \quad (4.6a)$$

$$\Gamma(1) = 1 \quad , \quad (4.6b)$$

$$\Gamma(n) = 1 \cdot 2 \cdot 3 \dots (n-1) = (n-1)! \quad , \quad n = 2, 3, 4 \dots \quad (4.6c)$$

$$\Gamma(z+1) = z\Gamma(z) \quad . \quad (4.6d)$$

Table 1 presents the transforms necessary for this work.

<p style="text-align: center;">Table 1. Transform Pairs for (4.1)</p>	
$m(A)$	$\mu(a)$
1	$2\sqrt{A_0} \frac{1}{\pi} \frac{1}{\sqrt{a}}$
$\sqrt{A}$	$2\sqrt{A_0} \frac{1}{2}$
A	$2\sqrt{A_0} \frac{2}{\pi} \sqrt{a}$
$A^{\frac{3}{2}}$	$2\sqrt{A_0} \frac{4}{3} a$

5 Power Law Model

Our discrete data is given as the number of cysts of a given area, $n(a)$ (a histogram). The areas are 'quantized,' being an integer number of pixels from the CT image. We will talk of bins of data, i.e., the j th bin has area $j \cdot a$, and has occupancy $n(a_j)$; a_1 is the area of one pixel.

Observing the data $n(a)$ (number of circles of a given area) it was decided to assume a model

$$n(a) = n_0 a^{-p} = g(a) \quad (5.1)$$

so that the slope of a log-log plot of $n(a)$ against a would give p and $n_0 = n(1)$. These graphs for 10 data sets are remarkably linear for the first 20 areas or so and more surprisingly the slopes are all nearly $-3/2$. The graphs are shown in Figures 1 to 10 and the slopes and intercepts ($\ln n_0$) are tabulated in Table 2.

About 90% of the data occurs in the first 20 bins

$$\sum_{i=1}^{20} n(a_i) \approx 0.90 \sum_{i=1}^N n(a_i) , \quad (5.2)$$

where N may be as high as 2500 which allows the necessary approximations below.

Assume that $f(A) = K/A^2$ so that

$$\frac{1}{2\sqrt{A_0}} \int_a^{A_M} \frac{K dA}{A^2 \sqrt{A-a}} = \frac{K}{2\sqrt{A_0}} \left\{ \frac{\sqrt{A-a}}{aA} + \frac{1}{2a} \frac{2}{\sqrt{a}} \tan^{-1} \sqrt{\frac{A-a}{a}} \right\}_a^{A_M} \approx \frac{K}{2\sqrt{A_0}} \cdot \frac{1}{a^{\frac{3}{2}}} \frac{\pi}{2} = \frac{n_0}{a^{\frac{3}{2}}} \quad (5.3)$$

where A_M is some maximal area such that $A_M \rightarrow \infty$, $\frac{1}{\sqrt{A_M}} \simeq 0$ and $\sqrt{A_M - a}/\sqrt{a} \simeq \infty$. Thus if we take $K = 4\sqrt{A_0}n_0/\pi$ we have the model solution to (2.10) with the model data²

$$\left. \begin{aligned} \hat{G}(a) &= \hat{n}(a) = \hat{n}_0 a^{-\frac{3}{2}} \\ \hat{F}(A) &\sim A^{-2} \end{aligned} \right\} a_1 < a, A < 20a_1 \quad (5.4a) \quad (5.4b)$$

Obviously the model does not apply for $a < a_1$ or for very large a , $a > 20a_1$. The latter limitations is graphically illustrated in Figure 11.

From (5.4b)

$$\hat{F}(A) = \frac{Aa_1}{A - a_1} \frac{1}{A^2} \quad (5.5)$$

where we have taken the upper limit of the normalization integral to be $\mathcal{A}$. If $\mathcal{A}$ is taken to be ∞ both the average values of area and volume are infinite—an undesirable consequence. $\mathcal{A}$ must then be chosen large but finite. Since $\mathcal{A}$ severely influences the values of $\langle a \rangle$ and $\langle v \rangle$ where $\langle g(a) \rangle = \int_{a_1}^{A_M} g(a) \hat{F}(a) da$ and as before $A_M \rightarrow \infty$ and $a_1 \rightarrow 0$. Since the model is only good up to $a_{20} = 20a_1$ we take $\mathcal{A} = 20a_1$.

²The models are only approximate but can be made more accurate than our data by choosing A_M large enough.

Figure 1: Data, linear least squares fit and models for patient 1238.

Figure 2: Data, linear least squares fit and models for patient 1257.

Figure 3: Data, linear least squares fit and models for patient 1306.

Figure 4: Data, linear least squares fit and models for patient 346.

Figure 5: Data, linear least squares fit and models for patient 485.

Figure 6: Data, linear least squares fit and models for patient 515.

Figure 7: Data, linear least squares fit and models for patient 520.

Figure 8: Data, linear least squares fit and models for patient 575.

Figure 9: Data, linear least squares fit and models for patient 83.

Figure 10: Data, linear least squares fit and models for patient 966.

Table 2.
Parameters of the Power Law Fit, Reduced
Chi-Square of Fit and Percent Data Used in Fit

<i>Medical Case</i>	$\ln n_0$	p	$\bar{\chi}^2$	P_d^*
1238	2.94	1.67	0.72	98
1257	4.49	1.41	1.31	82
1306	4.24	1.47	0.98	87
346	3.52	1.58	1.33	87
485	3.17	1.46	0.64	88
515	4.78	1.48	0.86	92
520	4.55	1.56	1.52	89
575	6.12	1.70	1.39	97
83	4.15	1.44	1.59	93
966	4.74	1.47	1.32	86
$< 1.52 > \pm 0.10$				

*in first 20 bins

From (3.2) we obtain the average values of d , a , and v

$$< d > = < \sqrt{\frac{4}{\pi}} a > = \sqrt{\frac{4}{\pi}} \frac{40}{19} \frac{\sqrt{20} - 1}{\sqrt{20}} \sqrt{a_1} = 1.84 \sqrt{a_1} \quad (5.6)$$

$$< a > = \frac{20}{19} \ln 20 a_1 = 3.15 a_1 \quad (5.7)$$

$$< v > = < \frac{8}{3\sqrt{\pi}} a^{\frac{3}{2}} > = \frac{8}{3\sqrt{\pi}} \frac{40}{19} (\sqrt{20} - 1) a_1^{\frac{3}{2}} = 11.00 a_1^{\frac{3}{2}} \quad (5.8)$$

Values of these quantities for the 10 patients are given in Table 3.

Table 3.
Average Values of Areas, Diameters and Volumes of Cysts

patient	$< a_{dc} >^*$	$< a >_G^*$	$< a > = (a, f)^*$	$< d >_{3/2}^*$	$< a >_{3/2}^*$	$< v >_{3/2}^*$
1238	1.44 mm ²	1.33 mm ²	0.67 mm ²	1.02 mm	1.10 mm ²	1.40 mm ³
1257	7.04	2.89	0.93	0.96	0.96	1.37
1306	4.78	2.15	0.84	0.93	0.90	1.35
346	5.15	2.41	0.87	0.99	1.03	1.38
485	6.15	2.34	0.80	0.99	1.03	1.65
515	4.96	2.24	1.05	1.02	1.10	2.00
520	5.29	2.02	0.76	0.96	0.96	1.63
575	1.89	1.84	0.97	1.09	1.25	1.95
83	2.18	1.86	0.93	0.99	1.03	1.15
966	5.03	2.32	0.83	0.93	0.90	1.21

*dc = data circles; G = analytic Golberg; (a, f) = from solution out to 50 pixels; $3/2 = 3/2$ model

Figure 11: Data for patient 1257, linear least squares fit (dash) and power law model (solid). Note the long tail of the data distribution (oscillating between 3, 2, 1, and $10^4 = 0$) which cannot be fit by simple models (10^4 was added to each $n(a_i)$ so that logarithms of zero could be approximated).

6 Exponential Model

The power law model gives physically meaningless results if extrapolated to $a = 0$. Furthermore, it has been discovered that healthy lungs have a large number of structures that have areas in a tomographic slice of only a few square pixels. Our data is thus contaminated by the presence of benign structures (we will occasionally also refer to these as cysts). The occupation numbers (for cysts alone) for the first few bins are certainly too large but by how much is uncertain. In the long run, because of individual variability, it even seems impossible to make a correction for this.

However, if we assume

$$n(a) = N_s^0 e^{-\tau_s a} + N_l^0 e^{-\tau_l a} , \quad (6.1)$$

where the subscript l refers to 'larger' and s refers to 'smaller,' we can make a tenuous identification of the s term to the benign structures and the l term to the LAM cysts. To determine the four parameters is the next task.

We first determine N_l^0 and τ_l by least squares using only the data in the 'larger' bins, i.e., a_3 through a_{20} . We then determine N_s^0 and τ_s by least squares using only the adjusted data

$$n_s(a) = n(a) - n_l(a) , \quad (6.2)$$

from the 'smaller' bins, i.e., a_1 through a_5 . We then combined these according to

$$n(a_i) = n_s(a_i) + n_l(a_i) , \quad i = 1, 2, 3, 4, 5 \quad (6.3a)$$

and

$$n(a_i) = n_l(a_i) , \quad i = 6, \dots, 20 . \quad (6.3b)$$

thus the smaller sized 'cysts' are of two kinds, benign and LAM, while the larger sized cysts are assumed to be only LAM cysts.

The parameter values and the reduced chi-square of the fit for the double exponential fit are given in Table 4. The reduced chi-square of the power law fit seems to be only slightly better. The double exponential

has more predictive power and, within the limits of the above assumptions, allows direct knowledge of the LAM cysts as opposed to the entire population of benign and LAM cysts.

Table 4.
Determined Parameters and Reduced Chi-Squares

case	n_0	p	$\bar{\chi}_p^2$	A_ℓ	t_ℓ	A_s	t_s	$\bar{\chi}_e^2$
1238	2.98	-1.67	0.72	2.86	-0.55	5.78	-3.30	0.86
1257	4.49	-1.40	1.31	4.14	-0.36	7.49	-3.80	1.67
1306	4.24	-1.48	0.98	4.16	-0.46	7.08	-3.83	0.99
346	3.50	-1.58	1.30	3.11	-0.39	6.36	-3.22	1.30
485	3.16	-1.42	0.62	2.73	-0.35	6.38	-3.95	0.71
515	4.77	-1.47	0.86	4.55	-0.41	7.55	-3.75	0.93
520	4.55	-1.56	1.52	4.35	-0.45	8.31	-5.16	1.47
575	6.12	-1.70	1.39	5.68	-0.42	8.85	-3.22	3.12
83	4.15	-1.44	1.59	3.94	-0.41	8.12	-5.68	2.89
966	4.74	-1.47	1.32	4.73	-0.47	7.54	-4.02	1.47

Having demonstrated that the double exponential fit is quite good (if somewhat subjective) we can glean some knowledge of the LAM cysts (l term) without the benign (s term) contamination.

The data fit (6.1) can be rewritten as

$$g(a) = \frac{N_s^0}{\tau_s} [\tau_s e^{-\tau_s a}] + \frac{N_l^0}{\tau_l} [\tau_l e^{-\tau_l a}] , \quad (6.4)$$

in terms of the (normalized) probability densities. Normalizing $g(a)$ we obtain

$$\hat{g}(a) = \frac{\eta_s \tau_s e^{-\tau_s a} + \eta_l \tau_l e^{-\tau_l a}}{\eta_s + \eta_l} , \quad (6.5)$$

where $\eta_i = N_i^0 / \tau_i$. Now

$$\begin{aligned} \hat{g}(a) &= \frac{1}{2\sqrt{A_0}} \int_a^\infty \frac{\hat{f}(A)}{\sqrt{A-a}} dA \\ &= \frac{\eta_s}{\eta_s + \eta_l} \frac{1}{2\sqrt{\alpha_s}} \int_a^\infty \frac{\hat{f}_s(A)}{\sqrt{A-a}} dA + \frac{\eta_l}{\eta_s + \eta_l} \frac{1}{2\sqrt{\alpha_l}} \int_a^\infty \frac{\hat{f}_l(A)}{\sqrt{A-s}} da \\ &= \frac{\eta_s}{\eta_s + \eta_l} \hat{g}_s(a) + \frac{\eta_l}{\eta_s + \eta_l} \hat{g}_l(a) . \end{aligned} \quad (6.6)$$

where the α can be thought of as an effective area or as in (6.10) below.

Comparing the second and third terms of (6.6) we see that

$$\hat{f}(A) = \frac{\eta_s}{\eta_s + \eta_l} \sqrt{\frac{A_0}{\alpha_s}} \hat{f}_s + \frac{\eta_l}{\eta_s + \eta_l} \sqrt{\frac{A_0}{\alpha_l}} \hat{f}_l(A) , \quad (6.7)$$

which requires that

$$\frac{n_s \sqrt{A_0/\alpha_s} + \eta_l \sqrt{A_0/\alpha_l}}{\eta_s + \eta_l} = 1 , \quad (6.8)$$

or

$$\sqrt{A_0} = \frac{(\eta_s + \eta_l) \sqrt{\alpha_s \alpha_l}}{\eta_s \sqrt{\alpha_l} + \eta_l \sqrt{\alpha_s}} . \quad (6.9)$$

this last expression gives the correct limits when $\eta_s = 0$, when $\eta_l = 0$, and when $\sqrt{\alpha_s} = \sqrt{\alpha_l}$. So while $\hat{g}(a)$ is not an eigenfunction, its decomposition into eigenfunctions is not only possible but also useful.

Since

$$\sqrt{\alpha_i} = \sqrt{\pi/4\tau_i} \quad (6.10)$$

we can write, in terms of the least square parameters,

$$\sqrt{A_0} = \sqrt{\frac{\pi}{4}} \frac{N_s^0 \tau_l + N_l^0 \tau_s}{\sqrt{\tau_s} N_s^0 \tau_l + \sqrt{\tau_l} N_l^0 \tau_s} . \quad (6.11)$$

For the exponential probability density the average values of the diameter, area, and volume are quite simple expressions. Since

$$d = (4a/\pi)^{\frac{1}{2}} , \quad (6.12)$$

and

$$V = \frac{4\pi}{3} \left(\frac{d}{2}\right)^3 = \frac{4}{3\sqrt{\pi}} a^{\frac{3}{2}} , \quad (6.13)$$

we obtain

$$\langle a \rangle = \tau \int_0^\infty e^{-\tau a} a da = \tau^{-1} \quad (6.14a)$$

$$\langle d \rangle = \tau^{-1/2} \quad (6.14b)$$

$$\langle v \rangle = \tau^{-\frac{3}{2}} \quad (6.14c)$$

Using (6.7), (6.9), and (6.10) with $x = d$ (diameter), $x^2 = a$ (maximal cross-sectional area) and $x^3 = V$ (volume) we have

$$\langle x^m \rangle = \frac{N_s^0 \sqrt{\tau_l} \tau_s^{-\frac{m}{2}} + N_l^0 \sqrt{\tau_s} \tau_l^{-\frac{m}{2}}}{N_s^0 \sqrt{\tau_l} + N_l^0 \sqrt{\tau_s}} \quad (6.15)$$

7 Solution of Wicksell's Area Equation and Use of the Golberg Device

If we accept the double exponential fit of the data and recognize the eigenfunction relationship, then we are done. There is no necessity to solve the integral equation (2.10). Historically, however, solving (2.10) was the first thing we did and only near the end of this phase did we stumble upon the double exponential fit. In this section we will thus discuss the computer code used to solve (2.10) and show our results.

It is crucial to keep in mind that the models are strictly valid only out to $a = 20a_1$ while the results reported in this section go out to $a = 50a_1$. The Golberg device uses *all* data, in some cases going out to $a = 2950a_1$ (!).

To solve

$$\hat{g}(a) = \frac{1}{2\sqrt{A_0}} \int_a^{\mathcal{A}} \frac{\hat{f}(A)}{\sqrt{A-a}} dA \quad , \quad (7.1)$$

we discretize the variables a and A and limit (for convenience of computer time and memory) both a and A to a maximum value of $50.5a_1$. The upper limit in (7.1) becomes $\mathcal{A} = 50.5a_1$.

Since $\sqrt{A_0}$ is unknown we incorporate it into $\hat{f}(A)$ as

$$\tilde{f}(A) = \hat{f}(A)/\sqrt{A_0} \quad (7.2)$$

When the solution $\tilde{f}(A_i)$ is determined it will NOT be normalized so we find

$$F = \sum_{i=1}^{50} \tilde{f}_i \quad (7.3)$$

and

$$\hat{f}_i = \tilde{f}_i / F \quad (7.4)$$

and

$$\sqrt{A_0} = \frac{1}{F} \quad (7.5)$$

We also assume that $\hat{g}(a)$ and $\tilde{f}(A)$ are constant over one bin width (the i th bin ranges from $e_i = (i - \frac{1}{2})a_1$ to $e_{i+1} = (i + \frac{1}{2})a_1$; $\Delta a = a_1$). Thus

$$\int_{e_i}^{e_{i+1}} g(a) da = g(a_i) \Delta a = g_i \cdot \Delta a \quad (7.6)$$

If we sum the g_i and divide each g_i by this sum we obtain $\hat{g}_i$,

$$\hat{g}_i = \frac{g_i}{G} = \int_{e_i}^{e_{i+1}} \tilde{g}(a) da \quad (7.7)$$

where $G = \sum_{i=1}^{50} g_i$.

Hence integrating both sides of (7.1) over da from e_i to e_{i+1} gives

$$\hat{g}_i = \frac{1}{\Delta a} \int_{e_i}^{e_{i+1}} da \int_{e_i}^{e_{51}} \frac{\tilde{f}(A)}{2\sqrt{A-a}} dA = \frac{1}{\Delta a} \int_{e_i}^{e_{i+1}} da \sum_{j=1}^{50} \int_{e_j}^{e_{d+1}} \frac{dA}{2\sqrt{A-a}} \tilde{f}(A_j) \quad (7.8)$$

($\tilde{f}(A)$ constant over one bin). Thus

$$\hat{g}_i = \sum_{j=1}^{50} K_{ij} \tilde{f}_j \quad (7.9)$$

where

$$\begin{aligned} K_{ij} &= \frac{1}{\Delta a} \int_{e_i}^{e_{i+1}} da \int_{e_j}^{e_{j+1}} dA \frac{1}{2\sqrt{A-a}} \quad , \quad j \geq i \\ K_{ij} &= 0 \quad \text{otherwise} \end{aligned} \quad (7.10)$$

The integrations are simple with some care being necessary for $j = i$. The integral equation (7.1) has thus been transformed into the matrix problem

$$\hat{g} = K\tilde{f} \quad (7.11)$$

where K is square (50×50) upper triangular.

We chose to solve (6.11) by singular value decomposition. The kernel matrix K has the unique decomposition

$$K = U\Sigma V^T \quad (7.12)$$

where U and V are orthogonal matrices and Σ is a diagonal matrix. (Software to do this decomposition is readily available.) Now if we write

$$\hat{g} = \sum_{i=1}^{50} \gamma_i u_i \quad (7.13)$$

where u_i is the i th column of U and is referred to as the i th *left* hand singular vector, then

$$\gamma_i = (u_i, \hat{g}) \quad (7.14)$$

is a known quantity.

Now formally write

$$\tilde{f} = \sum_{i=1}^{50} \phi_i v_i \quad (7.15)$$

where v_i is the i th column of V and is called the i th *right* hand singular vector. It can be demonstrated that

$$Kv_i = \sigma_i u_i \quad (7.16)$$

where $\sigma_i = \sum_{ii}$, so that

$$K\tilde{f} = K \sum \phi_i v_i = \sum \phi_i \sigma_i u_i = \sum \gamma_i u_i \quad (7.17)$$

Finally we find the ϕ_i as

$$\phi_i = \frac{\gamma_i}{\sigma_i} , \quad (7.18)$$

and the solution is accomplished

$$\tilde{f} = \sum_{i=1}^{50} \frac{\gamma_i}{\sigma_i} v_i . \quad (7.19)$$

$\hat{f}_i$ is then determined from (7.3) and (7.4).

The solutions for the ten patients are shown in Figures 12 to 21.

8 The Total Number of Cysts and Their Average Size

The relationship between the number of areas measured n and the number of cysts in the volume sampled N is simply

$$\langle N \rangle = n(1 - \rho)/\rho \quad (8.1)$$

where $\rho = \langle d \rangle / L$, $\langle d \rangle$ is the average cyst diameter and L is the separation between tomographic planes ($= 20mm$ for the data in this report). The variance in $\langle N \rangle$ is

$$\sigma_N^2 = n(1 - \rho)/\rho^2 = \langle N \rangle / \rho . \quad (8.2)$$

The derivation of these formulae is given in Appendix B.

$\langle d \rangle$ is a difficult number to come by. As mentioned before the models only extend to $20a_1$ and thus ignore the very large cysts: $\langle d \rangle$ determined from the models will be too small. The solution to the integral equation extends out to $50a_1$. $\langle d \rangle$ obtained from $\langle d, f \rangle$ will be larger but still too small. But we *can* use the Golberg device.

As shown at the beginning of Section 3

$$(m, \hat{f}) = (\mu, \hat{g}) , \quad (8.3)$$

Figure 12: Data and solution to (7.1) for patient 1238.

Figure 13: Data and solution to (7.1) for patient 1257.

Figure 14: Data and solution to (7.1) for patient 1306.

Figure 15: Data and solution to (7.1) for patient 346.

Figure 16: Data and solution to (7.1) for patient 485.

Figure 17: Data and solution to (7.1) for patient 515.

Figure 18: Data and solution to (7.1) for patient 520.

Figure 19: Data and solution to (7.1) for patient 575.

Figure 20: Data and solution to (7.1) for patient 83.

Figure 21: Data and solution to (7.1) for patient 966.

where

$$K^* \mu = m \quad . \quad (8.4)$$

Choosing

$$m = d = \sqrt{\frac{4}{\pi} a} \quad , \quad (8.5)$$

from Table 1 we find $\mu = 2\sqrt{A_0} \cdot \frac{1}{2}$. So

$$\langle d \rangle = \sqrt{\frac{4}{\pi}} (\sqrt{a}, \hat{f}) = \sqrt{\frac{4}{\pi}} \sqrt{A_0} (1, \hat{g}) \quad , \quad (8.6)$$

where the unknown $\sqrt{A_0}$ arises again. But since $(1, \hat{f}) = 1$, (7.3) can be rewritten as

$$\frac{(m, \hat{f})}{(1, \hat{f})} = \frac{(d, \hat{f})}{(1, \hat{f})} = \frac{\sqrt{\frac{4}{\pi}} (\sqrt{a}, \hat{g})}{(1, \hat{f})} = \frac{\sqrt{\frac{4}{\pi}} \sqrt{A_0} (1, \hat{g})}{2\sqrt{A_0} \frac{1}{\pi} (\frac{1}{\sqrt{a}}, \hat{g})} \quad (8.7)$$

so that

$$\langle d \rangle = \frac{\sqrt{\pi}}{(\frac{1}{\sqrt{a}, \hat{g}})} \quad . \quad (8.8)$$

This and similar expressions for $\langle a \rangle$ and $\langle v \rangle$ have been determined for all ten patients. This the most reliable value of the average diameter of the cysts that we have.

As an aside it should be mentioned that the Golberg device is usually employed as a noise reduction technique for when kernel matrices are singular or nearly so. This was not the case here. Rather here it enabled us to obtain a better estimate of $\langle d \rangle$ by including all data and not just a truncated set of it. This is turn was possible because of the availability of an analytic solution to the adjoint problem.

Average values of the area of the cysts as determined by Golberg device are presented in Table 3. We ontemplate all of the average values in Table 3 and recognize that the average area as obtained by the analytic Golberg device (AGD) is the best estimate. First note that the average area of the cysts (3rd column) is significantly smaller than the average area of the projected circles in the tomograms (2nd column). Many small cysts are missed by two tomographic slices 20 mm apart. Next note that the AGD numbers are 2-3

times larger than the values obtained from the solution (4th column) or a model (6th column). For semi-completeness the average values of the diameter and volumes, as obtained from the model, are included. The reversals occurring between $\langle d \rangle$ and $\langle a \rangle$ for patients 1306 and 966 in the model are the result of the model. They do not occur when using the power from the best least squares fit.

9 Conclusions

In the early stages of work on the CT images of 10 LAM patients the average area of the cyst projections were supplied as a diagnostic. We now can supply reliable average values of diameter, cross-sectional area and volume of the cysts themselves.

More than this however we can supply the distribution (normalized) of cyst sizes and the total number of cysts. All of the cyst information, as contrasted with information on cyst projections, comes from the study of the integral equation relating the two distributions.

Furthermore, we currently have two different models that describe the data about equally well. The models have fidelity only out to areas of about 20 square pixels. The power law model is simpler but does not allow knowledge of cysts below 1 square pixel and cannot discriminate between benign and LAM cysts. The double exponential model, while more complicated, gives physically reasonable results when extrapolated to zero area, allows the data analyst to discriminate (if even somewhat ambiguously) between benign and LAM cysts, and it relates to the eigenfunction of the integral equation. Whether the latter is of any consequence is doubtful but it is 'cute.'

If we graph N_l^0 against η_l for the 10 LAM patients we obtain the scattergram shown in Figure 22. We see that there are four clusters. Patients 1238 and 575 are in clusters by themselves. Patients 346 and 485 are in a cluster and the remaining six patients are in a cluster.

Patient 1238 is known to be a very mild case, patients 346 and 485 are also mild cases. Patient 966 has just received a lung transplant. Patient 575 was said to be near death if not already deceased.

Thus a whole suite of parameters are now available to the medical community in order to try and quantitatively relate the CT images of a LAM patient to her physiology.

Figure 22: Scattergrams obtained from plotting the amplitude against the exponent for the large cysts ($AL = N_{\ell}^0$).

References

1. Wicksell, S., "The Corpuscle Problem," *Biometrika*, **17**, 84-99 (1925).
2. Cormack, A., "Representation of a Function by the Line Integrals with some Radiological Applications," *J. App. Phys.* *I*, **34**, 2722-2727 (1963); *II*, **35**, 2908-2913 (1964).
3. Golberg, M., "A Method of Adjoints for Solving Some Ill-Posed Equations of the First Kind," *App. Math. Comp.*, **5**, 123-130 (1979).
4. Faber, V., G. Milton Wing, and John Zahrt, "Meta-Tomography," to be submitted.

APPENDIX A

Derivation of Wicksell's Equation

The derivation here follows Wicksell except for the following. Wicksell used the symbol r to represent the *diameters* of a sphere which we find exceedingly awkward. Using d for diameter leads to the differential dd which is equally abhorrent. As Wicksell used x to represent the diameter of a circle we have used y to represent the diameter of a sphere.

Let the probability density (relative frequency density) of a sphere having a diameter y be $f(y)$ and the probability density that a circle has diameter x be $g(x)$. That $f(y)$ and $g(x)$ are probability densities implies that

$$\int_0^{\infty} f(y)dy = 1 \quad \text{and} \quad \int_0^{\infty} g(x)dx = 1 \quad (\text{A.1})$$

Note that from the data $g(x)$ is known and that we wish to infer $f(y)$.

The number of spheres of diameter y to $y + dy$ coming within reach of an arbitrary plane P is

$$Nyf(y)dy$$

where N is the number density of spheres. Since

$$\int_0^{\infty} yf(y)dy = \langle y \rangle = y_0 \quad , \quad (\text{A.2})$$

where $\langle y \rangle$ or y_0 is the average diameter we can write

$$Nyf(y)dy = Ny_0f(y)dy \quad , \quad (\text{A.3})$$

where

$$f(y) = \frac{yf(y)}{y_0} \quad , \quad (\text{A.4})$$

is the probability density that a sphere of diameter y to $y + dy$ is cut by the plane P .

The probability that a sphere of diameter y , once cut by the plane P , has its center from z to $z + dz$ from the plane (see Figure A.1) is

Figure A.1: A plane P cutting a sphere a distance z from its center.

$$P(z, y) = \frac{dz}{y/2} \quad (\text{A.5})$$

Such a sphere would project in the plane P a circle of diameter x where

$$z^2 + (x/2)^2 = (y/2)^2 . \quad (\text{A.6})$$

The relative number of spheres of diameter y cut by the plane with centers at z to $z + dz$ is thus

$$\frac{2dz}{y} F(y) dy = 2dy \frac{F(y)}{y} \left| \frac{dz}{dx} \right| dx = 2dy \frac{f(y)}{y_0} \frac{1}{2} \frac{x}{\sqrt{y^2 - x^2}} dx , \quad (\text{A.7})$$

using (A.4) and (A.6).

Integrating over y (considering spheres of all diameters) gives the relative frequency of all spheres cut by the plane with centers at z to $z + dz$ and because of the constraint (A.6) this is the relative frequency of circles of diameter x to $x + dx$ in the plane P ,

$$g(x) dx = \frac{x}{y_0} \int_x^\infty \frac{f(y)}{\sqrt{y^2 - x^2}} dy dx , \quad (\text{A.8})$$

or

$$g(x) = \frac{x}{y_0} \int_x^\infty \frac{f(y) dy}{\sqrt{y^2 - x^2}} . \quad (\text{A.9})$$

This is Wicksell's equation.

APPENDIX B

Calculation of the Number of Spheres

Given a number, n , of measured circular areas in a plane, we wish to estimate the number, N , of spheres in a volume. Let D_i be the i^{th} measured diameter and let these be independent and identically distributed. Further let the n measurements be labeled hits (H) and let there be m misses (M) such that $N = n + m$. Consider the process

1. choose ball of diameter D_i ,
2. place it randomly into volume which contains the cutting plane,
3. do 1 and 2 N times, and
4. last ball must hit (otherwise we would have $n - 1$ hits).

The conditional probability of having N given the set $\{D_i\}$ and n hits is

$$P(N|n, \{D_i\}) = \binom{n+m-1}{m} \frac{D_1^{(H)}}{L} \dots \frac{D_n^{(H)}}{L} \left(1 - \frac{D_1^{(M)}}{L}\right) \dots \left(1 - \frac{D_m^{(M)}}{L}\right) \quad (\text{B.1})$$

where A is the area of the plane and $V = AL$ is the volume in which the spheres are placed.

So

$$P(N_1\{D_i\}|n) = P(N|n, \{D_i\})P(\{D_i\}) = P(N|n, \{D_i\})P(D_1^{(H)}) \dots P(D_m^{(M)}) \quad (\text{B.2})$$

and

$$\begin{aligned} P(N|n) &= \int P(N_1\{D_i\}|n) dD_1^{(H)} \dots dD_m^{(M)} \\ &= \binom{n+m-1}{m} \int \frac{D_1^{(H)}}{L} P(D_1^{(H)}) \dots \left(1 - \frac{D_m^{(M)}}{L}\right) P(D_m^{(M)}) dD_1^{(H)} \dots dD_m^{(M)} \\ &= \binom{n+m-1}{m} \left(\frac{\bar{D}}{L}\right)^n \left(1 - \frac{\bar{D}}{L}\right)^m, \end{aligned} \quad (\text{B.3})$$

where $\bar{D}$ is the average diameter. If

$$p = \bar{D}/L \ , \tag{B.4}$$

then

$$P(N|n) = \binom{n+m-1}{m} p^n (1-p)^m \ , \tag{B.5}$$

and this is the negative binomial distribution. The expectation value of $N = n + m$ and its variance are easily obtained [3] and the results are

$$E(N) = \bar{N} = n(1-p)/p \ , \tag{B.6}$$

$$var(N) = n(1-p)/p^2 \ . \tag{B.7}$$

APPENDIX C

The 'Normal' Lung

We include this brief section on the 'normal' lung since it was crucial in the seminal ideas of the double exponential model. But our 'normal' lung data is from a patient suffering from lung cancer. This may, in some people's eyes, make this data suspect and we would probably agree. For certain though, this data is from a patient who does not have LAM disease.

The CT image of a normal lung shows numerous 'cysts' of areas of a few square pixels. These are not cysts in the LAM sense but rather other biological structures such as blood vesicles. They are benign and have nothing to do with the LAM disease. Ideally they should **not** be included in any analysis of the data from LAM patients. As there is no appreciable density difference between these structures and LAM cysts, the CT image cannot be used to differentiate.

We also anticipate that individual variation will be so pronounced that a universal subtraction algorithm will not work. Our normal lung has more 'cysts' than some LAM patients have.

It seems that any distinction would therefore have to be made in the model. The distinction while real and important physiologically, is going to be somewhat subjective and ambiguous in any analysis of the data.

Figure C.1 shows the data and the fit to the data. Note especially that the $3/2$ power law (solid) does not do a good job of fitting the data but that an exponential fit is quite good. Notice also that by the sixth data point the number of these cysts is less than 1% of the first data point.

Figure C.2 is a striking display that indeed even the power fit (not $3/2$ power) does not belong to the same population as the fits from the LAM patients. Even our questionable normal lung is quite different from a LAM lung.

Turning to the two exponential fit we see in Figure C.3 that the parameters for the small cysts fit in fairly well with the parameters for the small cysts from LAM patients. This should be expected if the small cysts are benign and consist of vesicles, etc., that should be present in all lungs. Figure C.4 shows the large cyst parameters and here we see the dramatic differences in populations that we saw in Figure C.2.

These findings certainly add credence to our somewhat ad hoc separation of cysts into two populations and our use of exponentials to fit them. It would be more reassuring to investigate healthy lungs but this

data has not been available to us in the format required.

Figure C.1: Data linear least squares fit and models for a 'normal' lung.

Figure C.2: Power law parameters for the LAM data and a 'normal' lung.

Figure C.3: Exponential parameters of the small cysts for the LAM patients and a 'normal' lung.

Figure C.4: Exponential parameters of the large cysts for the LAM patients and a 'normal' lung.