Statistical Models for Anisotropic X-Ray Dark-field Tomography

Nathanael Schilling*, Matthias Wieczorek*, Tobias Lasser*
* Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany

***Abstract*—Anisotropic X-ray Dark-field Tomography (AXDT) is a recently introduced imaging modality to recover anisotropic scattering functions of volumetric objects from dark-field data acquired by an X-ray grating interferometer. Current reconstruction methods implicitly assume a Gaussian distribution of the noise in the negative logarithm of the data. In this paper we derive log-likelihood functions for use in statistical reconstruction methods, based on more accurate noise models of the measured X-ray grating interferometer data. We also discuss some properties of these functions relevant to optimization.**

I. Introduction

Anisotropic X-Ray Dark-Field Tomography (or in short: AXDT, [10]) is a recently introduced imaging modality, reconstructing scattering functions from dark-field data. This dark-field data, which relates to small angle X-ray scattering, is acquired using an X-ray grating interferometer [8], see Fig. 1 for a schematic overview.

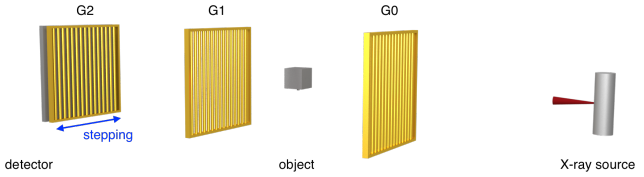

Fig. 1. Schematic overview of an X-ray grating interferometry setup.

As the dark-field signal changes when rotating the object around the X-ray beam direction, a full tomography of the directionally dependent dark-field signal requires reconstruction of anisotropic non-scalar quantities. In AXDT [10], this is realized using spherical functions, which are represented using spherical harmonics coefficients in a discretized vector η. Using the dark-field signals $\tilde{d}_j$, $j = 0, \dots, J-1$, as extracted from the raw X-ray grating interferometer measurements, the work presented in [10] reconstructs η from the $\tilde{d}_j$ via a simple least squares minimization approach, implicitly assuming a gaussian noise property of the dark-field signal.

In this work we use the known, correct noise properties of the dark-field signal and the underlying X-ray grating interferometer signal [3] to formulate log-likelihood functions suitable for statistical reconstruction of the scattering functions. We also derive the first two derivatives of the log-likelihood functions for use in gradient-based optimization algorithms. Finally, we discuss various implications for practical implementations, as well as the relationship to similar works such as [2], [9].

II. Preliminaries

Let $\tilde{s}_{n,j}$ denote the raw measurements of the X-ray grating interferometer, indexed by the stepping of grating G2 $n = 0, \dots, N-1$, and by $j = 0, \dots, J-1$ for each source-detector pair. Using the simplifying assumption that the only noise in the measurements comes from that inherent to photon-counting, it has been shown (c.f. [3]) that the corresponding random variables $S_{n,j}$ are independently Poisson distributed with mean

$$s_{n,j} = a_j + b_j \cos\left(\frac{2\pi n}{N} - \varphi_j\right).$$

Here a_j is proportional to the absorption and $\frac{b_j}{a_j}$ is proportional to the j-th dark-field coefficient denoted by d_j. Throughout this document, we use capital letters to refer to random variables, and lower case letters with a tilde (for example $\tilde{s}_{n,j}$) to refer to values derived directly from measured quantities.

With the discrete Fourier transform over n of the $s_{n,j}$,

$$c_{k,j} = \frac{1}{N} \sum_{n=0}^{N-1} \exp(-2\pi i k/N) s_{n,j},$$

we have

$$a_j = c_{0,j} \quad \text{and} \quad b_j = 2|c_{1,j}|.$$

In the same way, we obtain $\tilde{a}_j$ and $\tilde{b}_j$ from the discrete Fourier transform coefficients of the measurements $\tilde{s}_{n,j}$. Using this and by approximating the Poisson-distributions by Gaussian distributions, Chabior et al. [3] derive that for the corresponding random variables it holds that

$$A_j \sim \mathcal{N}\left(a_j, \frac{a_j}{N}\right) \quad \text{and} \quad \frac{1}{2}B_j \sim \mathcal{R}\left(\frac{b_j}{2}, \sqrt{\frac{a_j}{2N}}\right),$$

where $\mathcal{N}$ denotes the Gaussian distribution and $\mathcal{R}$ denotes the Rician Distribution.

The j-th dark-field coefficient is given by $d_j = \alpha_j^{-1} a_j^{-1} b_j$ with the corresponding random variable D_j, where α_j denotes the visibility of the reference or "flat-field" scan (without the object in the beam). Throughout this document we make the assumption that the flat-field values are known, which is an assumption commonly made for statistical reconstruction in the transmission tomography case (see for example [4]).

Finally, in order to keep the notation short and concise, we introduce subscript-less variants of the quantities just introduced to represent the full vectors for all $j = 0, \dots, J-1$. For example we have the random vector $D = (D_j)$ with $d = (d_j)$ and realizations $\tilde{d} = (\tilde{d}_j)$, with $j = 0, \dots, J-1$.

DOI: 10.12059/Fully3D.2017-11-3203019

III. Statistical Models

The direct reconstruction method of AXDT [10] maximizes the negative 2-norm of the residual, that is

$$f_0(\eta) = -\left\| \left(-\ln(\tilde{d}) \right) - \mathcal{B}\eta \right\|_2^2, \tag{1}$$

where $\ln$ is to be understood component-wise, $\mathcal{B}$ denotes the system-matrix for AXDT, and η is the vector of spherical harmonics coefficients to be reconstructed. This approach would be expected to perform well in a statistical sense if $\ln(D)$ were Gaussian, or at least close to Gaussian. However, if instead D were Gaussian (or close to a Gaussian distribution), due to Jensen's inequality and the convexity of the logarithm a bias would be expected.

In the following we introduce several statistical models, which are increasingly more realistic in the statistical assumptions of the underlying signal.

A. Model 1: Assuming Gaussian D

We first assume the D_j to be independent, Gaussian with equal variance σ^2. The log-likelihood is then

$$f_1(\eta) = -\left\| \exp(-\mathcal{B}\eta) - \tilde{d} \right\|_2^2,$$

leading to the optimization problem $\arg\max_\eta f_1(\eta)$. Using the AXDT forward model we have $d = \exp(-\mathcal{B}\eta)$, which allows us to write:

$$f_1(\eta) = -\mathbb{1}^T (d - \tilde{d})^2, \tag{2}$$

where $\mathbb{1}^T$ is defined as the row-vector $(1, \ldots, 1)$ used as a short-hand notation for the sum. Using the chain-rule, we receive the gradient and Hessian as

$$\begin{aligned} \nabla f_1(\eta) &= 2\mathcal{B}^T d(d - \tilde{d}) \\ H_{f_1} &= -2\mathcal{B}^T \operatorname{diag}\left(d(2d - \tilde{d})\right)\mathcal{B}. \end{aligned}$$

B. Model 2: Using the Rician Distribution

Using the definition of the dark-field coefficients, we have $b = \alpha a d$. Furthermore, given flat-field absorption values $\gamma = (\gamma_j)$, the Beer-Lambert Law states $a = \gamma \exp(-\mathcal{A}\mu)$, where $\mathcal{A}$ denotes the system matrix for X-ray CT and μ is the vector of absorption coefficients. For statistical reconstruction, we must therefore perform a joint optimization of the log-likelihood over both μ and η. Applying that $A_j \sim \mathcal{N}(a_j, \frac{a_j}{N})$ and $\frac{1}{2}B_j \sim \mathcal{R}(\frac{1}{2}b_j, \sqrt{\frac{a}{2N}})$ we obtain a log-likelihood function of:

$$f_2(\mu, \eta) = \mathbb{1}^T \Big(\ln\left(g_{a,\frac{a}{N}}(\tilde{a})\right) + \ln\left(r_{\frac{1}{2}b_j, \sqrt{\frac{a}{2N}}}(\tfrac{1}{2}\tilde{b_j})\right)\Big), \tag{3}$$

where g and r are the probability density functions (pdf) of the Gaussian and Rician distributions, respectively (to be understood component-wise).

We have that $\ln(g_{a,\frac{a}{N}}(\tilde{a})) = \frac{\ln(N) - \ln(a) - \ln(2\pi)}{2} - \frac{N(\tilde{a}-a)^2}{2a}$. We leave out the constant terms when forming the functional, as they do not affect the optimization problem.

1) Approximating Rician Distribution by a Gaussian: The pdf of the Rician-Distriution is quite complicated, but for large values of ν/σ it is very similar to a Gaussian distribution of mean $\sqrt{\nu^2 + \sigma^2}$ and variance σ^2 [5]. We further make the assumption that σ is small enough that the distribution has a mean of approximately ν. The term from the Rician distribution (ignoring additive constants) then becomes $\frac{-\ln(a)}{2} + \frac{2N\left(\frac{1}{2}\tilde{b} - \frac{1}{2}b\right)^2}{2a}$. Rearranging and adding the contribution from the Gaussian term yields:

$$f_{2G}(\mu, \eta) = \mathbb{1}^T \left(-\ln(a) - \frac{2N(\tilde{a} - a)^2 + N(\tilde{b} - a\alpha d)^2}{4a} \right), \tag{4}$$

with again component-wise operations. Applying the chain-rule gives the following gradient:

$$\begin{aligned} \nabla_\mu f_{2G}(\mu, \eta) &= \mathcal{A}^T \left(1 + \frac{N(2a^2 - 2\tilde{a}^2 + a^2\alpha^2 d^2 - \tilde{b}^2)}{4a} \right) \\ \nabla_\eta f_{2G}(\mu, \eta) &= \frac{1}{2}\mathcal{B}^T \left(\alpha d N (a\alpha d - \tilde{b}) \right) \end{aligned}$$

The Hessian has the following form:

$$H_{f_{2G}} = \begin{bmatrix} \mathcal{A}^T & 0 \\ 0 & \mathcal{B}^T \end{bmatrix} \begin{bmatrix} H^{(1,1)} & H^{(1,2)} \\ H^{(2,1)} & H^{(2,2)} \end{bmatrix} \begin{bmatrix} \mathcal{A} & 0 \\ 0 & \mathcal{B} \end{bmatrix}$$

where:

$$\begin{aligned} H^{(1,1)} &= \frac{-N}{2} \operatorname{diag}\left(\frac{2a^2 + 2\tilde{a}^2 + a^2\alpha^2 d^2 + \tilde{b}^2}{2a} \right) \\ H^{(1,2)} &= H^{(2,1)} = \frac{-N}{2} \operatorname{diag}\left(a\alpha^2 d^2 \right) \\ H^{(2,2)} &= \frac{-N}{2} \operatorname{diag}\left(d(2\alpha^2 a d - \alpha\tilde{b}) \right). \end{aligned}$$

2) Using the Rician Distribution Directly: The Rician distribution has a pdf of: $r_{\nu,\sigma}(x) = \frac{x}{\sigma^2} \exp\left(\frac{-(x^2+\nu^2)}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right)$, where I_k denotes the k-th modified Bessel Function of the first kind with $I_k(t) = \frac{1}{\pi}\int_0^\pi e^{t\cos x}\cos(kx)\, dx$ [7].

Discarding constants, we can write the Rician term of the log-likelihood function exactly as:

$$-\ln(a) - \frac{2N\left((\frac{1}{2}\tilde{b})^2 + (\frac{1}{2}b)^2\right)}{2a} + \ln I_0\left(\frac{N\tilde{b}b}{2a}\right)$$

Combining this with the term from the normal distribution and discarding constants yields

$$\begin{aligned} f_{2R}(\mu, \eta) = \mathbb{1}^T \Bigg(& \frac{-3\ln(a)}{2} - \frac{N(2\tilde{a}^2 + 2a^2 + \tilde{b}^2 + a^2\alpha^2 d^2)}{4a} \\ & + \ln I_0\left(\frac{\tilde{b}\alpha d N}{2}\right) \Bigg) \end{aligned} \tag{5}$$

For the gradient we have:

$$\nabla_\mu f_{2R}(\mu,\eta) = \mathcal{A}^T\left(\frac{3}{2} + \frac{N(2a^2 - 2\tilde{a}^2 + a^2\alpha^2d^2 - \tilde{b}^2)}{4a}\right)$$
$$\nabla_\eta f_{2R}(\mu,\eta) = \mathcal{B}^T\left(\frac{N}{2}a\alpha^2d^2 - \frac{\tilde{b}\alpha dN I_1\left(\frac{\tilde{b}\alpha dN}{2}\right)}{2I_0\left(\frac{\tilde{b}\alpha dN}{2}\right)}\right)$$

We note in passing that for the numerical calculation of the $\nabla_\eta f_{2R}$ it may be useful to use the bounds obtained by [1]. They prove that $\coth(x) - \frac{1}{x} < \frac{I_1(x)}{I_0(x)} < \tanh(x)$, from which it follows that $\frac{I_1(x)}{I_0(x)}$ quickly goes to 1 with $x \to \infty$.

For readability, we define the short notations $q(x) := \frac{I_1(x)}{I_0(x)}$ and $z = \frac{\tilde{b}\alpha dN}{2}$. Computing the Hessian in the format described above, we get:

$$H^{(1,1)} = \operatorname{diag}\left(\frac{-N(2a^2 + 2\tilde{a}^2 + a^2\alpha^2d^2 + \tilde{b}^2)}{4a}\right)$$
$$H^{(2,1)} = H^{(1,2)} = \frac{-N}{2}\operatorname{diag}\left(a\alpha^2d^2\right)$$
$$H^{(2,2)} = \operatorname{diag}\left(-Na\alpha^2d^2 + z^2(1 - q(z)^2)\right).$$

C. Model 3: Statistical Reconstruction directly from Measurements

If we directly use the fact that the $S_{n,j}$ have a Poisson distribution with mean $s_{n,j}$, then we must also include the φ_j as variables. In lieu of setting up a forward-model for them, we can simply add them as additional terms to be optimized for. The log-likelihood functional then has the form

$$f_3(\mu,\eta,\varphi) = \sum_{n=0}^{N-1} \mathbb{1}^T\left(-a + \tilde{s}_n \ln\left(a + a\alpha d\cos\left(\frac{2\pi n}{N} - \varphi\right)\right)\right) \tag{6}$$

To aid readability, we define $\psi_n = \cos\left(\frac{2\pi n}{N} - \varphi\right)$ and $\Gamma_n = \sin\left(\frac{2\pi n}{N} - \varphi\right)$. The gradient is then

$$\nabla_\mu f_3(\mu,\eta,\varphi) = \sum_{n=0}^{N-1} \mathcal{A}^T\left(a - \tilde{s}_n\right)$$
$$\nabla_\eta f_3(\mu,\eta,\varphi) = \sum_{n=0}^{N-1} \mathcal{B}^T\left(\frac{-\tilde{s}_n\alpha d\psi_n}{1 + \alpha d\psi_n}\right)$$
$$\nabla_\varphi f_3(\mu,\eta,\varphi) = \sum_{n=0}^{N-1} \frac{\tilde{s}_n\alpha d\Gamma_n}{1 + \alpha d\psi_n}$$

For the Hessian, we observe that it has the form:

$$H_{f_3} = \operatorname{diag}(\mathcal{A},\mathcal{B},\mathcal{I})^T \begin{bmatrix} H^{(1,1)} & 0 & 0 \\ 0 & H^{(2,2)} & H^{(2,3)} \\ 0 & H^{(3,2)} & H^{(3,3)} \end{bmatrix} \operatorname{diag}(\mathcal{A},\mathcal{B},\mathcal{I}),$$

where $\mathcal{I}$ denotes the identity matrix of size M. We further have

$$H^{(1,1)} = -N\operatorname{diag}(a)$$
$$H^{(2,2)} = \sum_{n=0}^{N-1}\operatorname{diag}\left(\frac{\tilde{s}_n\alpha d\psi_n}{(1 + \alpha d\psi_n)^2}\right)$$
$$H^{(2,3)} = H^{(3,2)} = \sum_{n=0}^{N-1}\operatorname{diag}\left(\frac{-\tilde{s}_n\alpha d\Gamma_n}{(1 + \alpha d\psi_n)^2}\right)$$
$$H^{(3,3)} = \sum_{n=0}^{N-1}\operatorname{diag}\left(\frac{-\tilde{s}_n\alpha d(\alpha d + \psi_n)}{(1 + \alpha d\psi_n)^2}\right)$$

IV. Discussion

As can be seen from Fig. 2, all likelihood functions derived in this document penalize small values much stronger than those of the least squares functional f_0 minimized by [10]. For the particular data example chosen for the plot (which was taken from a real grating interferometer measurement), the approximations derived in this document resemble the theoretical log-likelihood f_3 of the most accurate model 3, see also Fig. 3. Whilst this even seems to be true for the simplistic model 1 (f_1), we would nevertheless expect better results from the more advanced models, especially in reconstructions that would otherwise suffer from streaks of noise originating from high-variance readings. In these the variance depends on the detector value and the high variance from higher pixel values is taken into account by models 2 and 3. Of course our single data example may also not be representative.

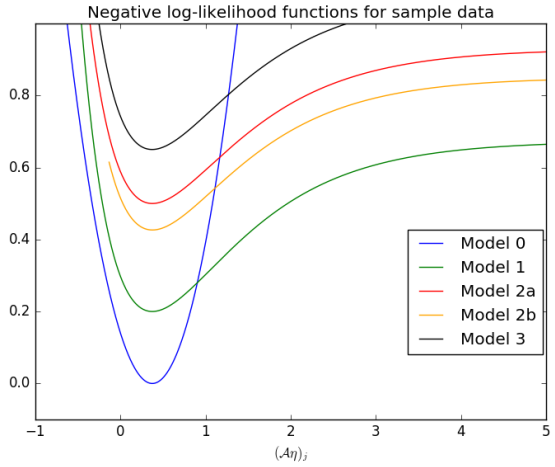

Fig. 2. Graphs of negative log-likelihood functions $f_0, f_1, f_{2G}, f_{2R}, f_3$ for example data $\tilde{s}_n$ with $a = \tilde{a}$ and $\varphi = -\arg(\tilde{c}_1)$. All graphs, except for the one for f_0, were scaled by the factor $\frac{1}{50}$ and translated vertically in order to fit in one plot nicely.

All of the log-likelihood functions derived are both non-quadratic (due to the exponential dependence on μ and η) and also asymmetric around the minimum. Physical considerations limit the values of a and b to be positive with $a > b$,

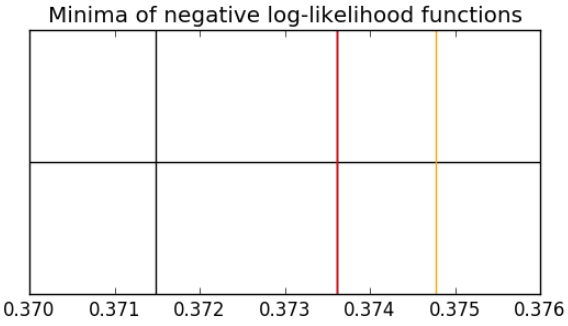

Fig. 3. Minima of the functions $f_0, f_1, f_{2G}, f_{2R}, f_3$ as plotted in Fig. 2. The black line (f_3) and the blue line (f_0) coincide with the red one (f_{2G}).

furthermore $0 < d \leq 1$ and $0 < \alpha \leq 1$. This should be taken into account when optimizing, and the analysis below depends on these assumptions.

For model 1 (f_1), it is clear from the Hessian that the log-likelihood function is (locally) strictly concave whenever $d > \frac{1}{2}\tilde{d}$ and $\mathcal{A}$ has trivial kernel. For model 2.1 (f_{2G}), Theorem 7.7.7 of [6] gives a neccessary and sufficient condition for positive-definiteness to be that $-H^{(1,1)}$ is positive definite and that $(-H^{(2,2)})-(-H^{(2,1)})(-H^{(1,1)})^{-1}(-H^{(2,1)})^T$ is positive definite. Writing this out and simplifying gives equivalently that $2b(2a^2 + 2\tilde{a}^2 + \tilde{b}^2) - \tilde{b}(2a^2 + b^2 + 2\tilde{a}^2 + \tilde{b}^2)$ is positive (in each component). This is the case when $b > \frac{3}{4}\tilde{b}$ and $a > b$. Hence a sufficient condition for local concavity (if $\mathcal{A}$ and $\mathcal{B}$ have trivial kernel) is given by $\frac{3}{4}\tilde{b} < b < a$. A practical challenge for using these functionals is the fact that the current formulations can produce very large values, creating numerical issues. For implementation purposes, we suggest scaling or translating the functionals in a way that preserves their properties but results in smaller values.

Similar statistical reconstruction methods have already been demonstrated to perform well on phantom images and images of mice by [9] and [2]. The methods there are most similar to those described in model 3 (f_3), in that they perform reconstruction from the raw detector readings $\tilde{s}_n$. However, both only calculate the mean scattering and do not attempt to reconstruct the anisotropic scattering as is done by AXDT.

In a sense our work in model 3 can be seen as a generalization of this; though it should also be noted that we do not reconstruct the differential phase contrast image as [9] and [2] do. A direction explored by [9], which we did not look into further yet, is reconstruction from only a subset of the values in each vector $(\tilde{s}_n)_{n=0,\dots,N_1}$ conventionally required for direct reconstruction. This has the potential to greatly reduce the number of required readings and shows promising results. We note that the equations from model 3 can easily be modified to take this into account, though some additional work is required to estimate the α_j and γ_j required by the formulae. In [9] and [2] a slightly different forward model is used to account for this, and it is possible that this can be modified to work with AXDT.

We also note that model 3 can be modified to reconstruct η only (instead of doing a joint reconstruction), or extended with a forward-model for the phase-contrast φ. Evaluations using real experimental data are currently in progress. A further aspect that could be looked into is the addition of a regularization term as was done by [2].

V. Conclusion

We have presented several log-likelihood functions, as well as their derivatives, for use in statistical reconstruction methods for X-ray grating interferometry data. We were also able to show (local) concavity for one of advanced models (f_{2G}). While the more accurate models are harder to optimize, we expect improved reconstruction results in practice.

Acknowledgement

This work was partially funded by the DFG Cluster of Excellence Munich–Centre for Advanced Photonics (MAP).

References

[1] Árpád Baricz. Bounds for modified bessel functions of the first and second kinds. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 53(03):575–599, 2010.

[2] Bernhard Brendel, Maximilian von Teuffenbach, Peter B. Noël, Franz Pfeiffer, and Thomas Koehler. Penalized maximum likelihood reconstruction for x-ray differential phase-contrast tomography. *Medical Physics*, 43:188–194, 2016.

[3] Michael Chabior, Tilman Donath, Christian David, Manfred Schuster, Christian Schroer, and Franz Pfeiffer. Signal-to-noise ratio in x ray dark-field imaging using a grating interferometer. *Journal of applied physics*, 110(5):053105, 2011.

[4] Jeffrey A Fessler. Statistical image reconstruction methods for transmission tomography. *Handbook of medical imaging*, 2:1–70, 2000.

[5] Hákon Gudbjartsson and Samuel Patz. The rician distribution of noisy mri data. *Magnetic resonance in medicine*, 34(6):910–914, 1995.

[6] Charles R Johnson and Roger A Horn. *Matrix analysis*. Cambridge University Press Textbooks, 2012.

[7] Venkatarama Krishnan and Kavitha Chandra. *Probability and random processes*. John Wiley & Sons, 2015.

[8] Franz Pfeiffer, Martin Bech, Oliver Bunk, Philipp Kraft, Eric F. Eikenberry, Christian Brönnimann, Christian Grünzweig, and Christian David. Hard-x-ray dark-field imaging using a grating interferometer. *Nature Materials*, 7(2):134–137, 2008.

[9] André Ritter, Gisela Anton, and Thomas Weber. Simultaneous maximum-likelihood reconstruction of absorption coefficient, refractive index and dark-field scattering coefficient in x-ray talbot-lau tomography. *PloS one*, 11(10):e0163016, 2016.

[10] Matthias Wieczorek, Florian Schaff, Franz Pfeiffer, and Tobias Lasser. Anisotropic x-ray dark-field tomography: A continuous model and its discretization. *Physical Review Letters*, 117(15):158101, 2016.