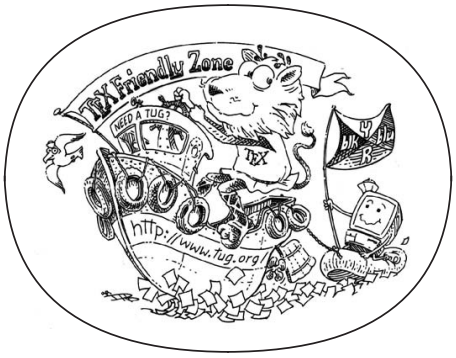
PROJECT THESIS IN NANOTECHNOLOGY AT NTNU

RUBEN SKJELSTAD DRAGLAND



Optimisation of Small Angle X-ray Scattering Tensor Tomography Gradient Descent Algorithm by Automatic Differentiation

December 2022 – version 1.0

Ruben Skjelstad Dragland: *Project Thesis in Nanotechnology at NTNU,* Optimisation of Small Angle X-ray Scattering Tensor Tomography Gradient Descent Algorithm by Automatic Differentiation, © December 2022

SUPERVISORS:
Dag Werner Breiby
Basab Chattopadhyay

Ohana means family.
Family means nobody gets left behind, or forgotten.

— Lilo & Stitch

Dedicated to the loving memory of Rudolf Miede.
1939 – 2005

ABSTRACT

Short summary of the contents. . . a great guide by Kent Beck how to write good abstracts can be found here:

https://plg.uwaterloo.ca/~migod/research/beckOOPSLA.html

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

Put your publications from the thesis here. The packages `multibib` or `bibtopic` etc. can be used to handle multiple different bibliographies in your document.

We have seen that computer programming is an art,
because it applies accumulated knowledge to the world,
because it requires skill and ingenuity, and especially
because it produces objects of beauty.

— Donald E. Knuth [6]

ACKNOWLEDGEMENTS

Put your acknowledgements here.

Many thanks to everybody who already sent me a postcard!

Regarding the typography and other help, many thanks go to Marco Kuhlmann, Philipp Lehman, Lothar Schlesier, Jim Young, Lorenzo Pantieri and Enrico Gregorio[1], Jörg Sommer, Joachim Köstler, Daniel Gottschlag, Denis Aydin, Paride Legovini, Steffen Prochnow, Nicolas Repp, Hinrich Harms, Roland Winkler, and the whole LaTeX-community for support, ideas and some great software.

Regarding LyX: The LyX port was initially done by *Nicholas Mariette* in March 2009 and continued by *Ivo Pletikosić* in 2011. Thank you very much for your work and the contributions to the original style.

1 Members of GuIT (Gruppo Italiano Utilizzatori di TeX e LaTeX)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

DRY Don't Repeat Yourself

API Application Programming Interface

UML Unified Modeling Language

FBP Filtered Back Projection

CT Computed Tomography

ML Maximum-Likelihood

GD Gradient Descent

CGD Conjugated Gradient Descent

AD Automatic Differentiation

SAXSTT Small Angle X-ray Scattering Tensor Tomography

SH Spherical Harmonics

Part I

INTRODUCTION

1 INTRODUCTION

Machine learning has grown to be a powerful tool in many disciplines within physics. This includes computed tomography...

Part II

REVIEW OF THE LITERATURE

You can put some informational part preamble text here. Illo principalmente su nos. Non message *occidental* anglo-romanic da. Debitas effortio simplificate sia se, auxiliar summarios da que, se avantiate publicationes via. Pan in terra summarios, capital interlingua se que. Al via multo esser specimen, campo responder que da. Le usate medical addresses pro, europa origine sanctificate nos se.

2 COMPUTED TOMOGRAPHY

2.1 X-RAYS

X-rays are electromagnetic waves with energy in the orders of keV. From Planck's Equation (1), this corresponds to nanometer wavelengths. The equation relates energy of a photon E to the angular frequency ω or wavelength λ of the corresponding electromagnetic wave as

$$E = \hbar\omega = 2\pi\hbar\frac{c}{\lambda}, \tag{1}$$

with $c \sim 2.99776 \times 10^8 m/s$ being the speed of light [2]. The other constant is the reduced Planck's constant $\hbar \sim 1.0543 \times 10^{-34} Js$.

Excitation, acceleration, and deceleration are the three most commonly utilised processes for producing X-rays. The first method is commonly referred to as "Characteristic X-ray radiation", which occurs when a highly energetic electron collides into a target atom. The accelerated electron transfers enough energy to eject an inner-shell electron from the atom. An outer electron may therefore occupy a lower-energy state. Due to conservation of energy, this process causes emission of a photon, as illustrated in Equation (2). As the atomic energy levels are discrete, this process is characterised by a spectrum of discrete X-ray emission lines.

$$E_{photon} = -\Delta E = -(E_f - E_i) \tag{2}$$

In addition to excitation, scattering events occur when electrons pass through an anode material. These events accelerate the electrons in a new direction, and X-rays known as "Bremsstrahluhng" are emitted.

The synchrotron is the last common form of X-ray production, and is also based upon the principle of "Bremsstrahluhng". Generally, charged particles are accelerated to very high energies, and magnets maintain their circular path. As moving objects in a circular path experience a centrifugal acceleration perpendicular to its directions, "Bremsstrahluhng" X-rays are emitted [4].

2.2 BEER-LAMBERT'S LAW

The intensity of X-rays attenuates upon interacting with matter. This is due to photoelectric absorption, elastic Rayleigh scattering, and

inelastic Compton scattering. The attenuation coefficient μ describes this attenuation in an inhomogeneous sample as

$$I(s) = I(0)\exp(-\int_0^s \mu(\nu)d\nu), \tag{3}$$

where s is the distance from the initial intensity to the end of the sample, effectively the thickness of the sample, and $I(0)$ is the initial intensity. Here, the spectral dependence, $\mu(E,\nu)$ is often neglected as it is unknown [3]. A simple manipulation of the expression gives the projection line integral

$$p(s) = -\ln(\frac{I(s)}{I(0)}) = \int_0^s \mu(\nu)d\nu. \tag{4}$$

2.3 RADON TRANSFORM

The projection line integral in Equation (4) may be viewed as a Radon transform of an object function $f(x,y)$ for a single orientation θ [10]. Confidence in this statement may be achieved by comparing Equation (4) with a single-angle Radon transform (5)

$$p_\theta(r) = \int_{-\infty}^{\infty} f(r,\nu)d\nu. \tag{5}$$

2.4 FOURIER SLICE THEOREM

The key in computed tomography is to determine the spatial dependency of the attenuation coefficient. By sampling many projections, meaning line integrals from different orientations and crossections, data necessary to reconstruct a three-dimensional image is collected. For a given crossection of the object $f(x,y)$, the detected intensity is plotted as a function of projection number and pixel number in what is called a sinogram. By utilising this sinogram and the Fourier slice theorem, the object $f(x,y)$ may be determined by other means than computing the full inverse Radon transform.

The Fourier slice theorem states that the full 2D Fourier transform $F(\omega_x,\omega_y)$ of an object $f(x,y)$ can be constructed from a series of 1D Fourier transforms $P(\omega)$ of projections $p(s)$ with different orientations [10].

2.5 FILTERED BACK PROJECTION

In short, the filtered back projection (FBP) algorithm reconstructs the object by forward and inverse Fourier transforms. Firstly, the sinogram of projections is mapped to frequency space in polar coordinates by subsequent 1D Fourier transforms, as shown in Equation (6):

$$P(\theta, \omega) = \int_{-\infty}^{\infty} p(\theta, r) e^{-2\pi i \omega r} dr. \tag{6}$$

With this the 2D Fourier transform $F(u, v)$ of the object $f(x, y)$ is found. The final step is an inverse 2D Fourier transform with a ramp-filter of $|\omega|$ to account for the radial distribution of points in polar coordinates. This filter is also the Jacobian of the area integration element in the polar Fourier space. Consequently, the object function can be expressed as

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} |\omega| P(\theta, \omega) e^{-2\pi i \omega (x \cos\theta - y \sin\theta)} d\omega d\theta. \tag{7}$$

3 MACHINE LEARNING OPTIMISATION

3.1 MAXIMUM LIKELIHOOD ESTIMATION

The maximum likelihood estimator is defined to be the set of parameters $\boldsymbol{\theta}_{\mathrm{ML}}$ that maximise the probability $P(\boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta})$. In other words, it chooses the parameters that produce the most probable estimation $\widehat{\boldsymbol{y}}$ of the true output $\boldsymbol{y}$ given the input $\boldsymbol{x}$. Equation (8) defines this mathematically [5].

$$\boldsymbol{\theta}_{\mathrm{ML}} = \operatorname{argmax}_{\boldsymbol{\theta}} P(\boldsymbol{y} \mid \boldsymbol{x}; \boldsymbol{\theta}) \tag{8}$$

ML estimation is an example of supervised learning, because the true output, called the targets, are known. In supervised learning, the estimation is evaluated by computing the error relative to the true output. The expression for the total error of the model is called the cost function $J(\boldsymbol{\theta})$, which is a sum of loss functions $\mathcal{L}(\boldsymbol{y}, \widehat{\boldsymbol{y}})$ representing the error of a single data point.

3.2 GRADIENT DESCENT

Gradient descent is an optimisation algorithm that updates the model's parameters based on the gradient of the cost function and the step size α, as shown in Equation (9) [8].

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \alpha \nabla_{\theta} J(\boldsymbol{\theta}) \tag{9}$$

The idea of this algorithm is that by following the gradient of the cost function, it is possible to find the global, or at least a sufficiently good local, minima in parameter space. In this way, the estimation will converge and minimise the error. Gradient descent in parameter space is visualised in Figure **??**

3.3 CONJUGATE GRADIENT DESCENT

The basic gradient descent algorithm is prone to require many iterations before converging. When solving computationally expensive optimisation tasks, one should therefore consider a method like Conjugate Gradient Descent (CGD). This method serves as a compromise between basic first order gradient descent and Newton's second order method. The latter uses the Hessian to converge in a small number of highly expensive iterations. The essential characteristics of the CGD

algorithm will be summarised in this section, while it is clearly explained by Jonathan Richard Shewchuk [9].

The first improvement over basic gradient descent is a line search to determine the optimal step size α in the direction of the calculated gradients. This is done by computing the cost function $J(\boldsymbol{\theta})$ for a range of step sizes α. At the minima, the gradient vector of the current and next iteration are orthogonal, which optimises the path of convergence. Intuitively, this reduces the number of gradient calculations required to reach convergence, as the algorithm always exhausts the potential of the current direction, and proceeds orthogonal to this direction.

However, with this procedure, which is commonly referred to as "Method of Steepest Descent", the algorithm often steps in the same direction multiple times. To prevent this, CGD only performs one step per basis vector in a set of orthogonal search directions. The orthogonal set is derived from Gram-Schmidt conjugations of the gradients. In order to optimise any continuous nonlinear function, the Polak-Ribiere formula (10) is used to determine the optimal Gram-Schmidt constant β, where $\boldsymbol{g}$ is the current gradient vector and $\boldsymbol{g}_{k-1}$ is the gradient vector of the previous iteration.

$$\beta_k = \max\left(\frac{\boldsymbol{g}_k^T(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}{\boldsymbol{g}_{k-1}^T\boldsymbol{g}_{k-1}}, 0\right) \tag{10}$$

Due to the max function, the algorithm will restart CGD with "Method of Steepest Descent" if the gradients are no longer decreasing, thus ensuring convergence. Note that this resets the orthogonal set of search directions, but the algorithm still converges in the order of $O(N)$ iterations, where N is the number of parameters. Consequently, CGD can be mathematically expressed as follows [9]:

$$\begin{aligned} \boldsymbol{d}_0 &= -\boldsymbol{g}_0 \\ \alpha_i &= \min_{\alpha}\left(J(\boldsymbol{\theta}) + \alpha \boldsymbol{d}_i\right) \\ \boldsymbol{\theta}_{i+1} &= \boldsymbol{\theta}_i + \alpha_i \boldsymbol{d}_i \\ \boldsymbol{d}_{i+1} &= \boldsymbol{g}_{i+1} + \beta_{i+1}\boldsymbol{d}_i. \end{aligned} \tag{11}$$

In Equation (11), $\boldsymbol{d}_i$ is the search direction, $\boldsymbol{g}_i$ is the gradient, and $\boldsymbol{\theta}_i$ is the current set of parameters. As already mentioned, $J(\boldsymbol{\theta})$ is the cost function, α_i is the optimal step size, which is determined by a line search, and β_{i+1} is the Gram-Schmidt constant derived from the Polak-Ribiere formula (10).

3.4 AUTOMATIC DIFFERENTIATION

Automatic differentiation (AD) is an algorithmic technique for computing the analytical gradients of a function using computational

graphs and the chain rule. It is important to contrast AD from numerical differentiation using finite differences, which cannot calculate the expression of the gradients analytically. Moreover, AD should not be confused with symbolic differentiation, which is a method for calculating the full symbolic gradient expression, like one would do by hand [1]. The superior routine for implementing AD is the "reverse mode", which consists of a forward pass and a backward pass. The forward pass is a function evaluation, where a computational graph, like Figure **??** illustrates, is constructed.

In regard to computer science, the computational graph can easily be constructed from an object-oriented operator overloading approach. To elaborate, the AD object inserts the operators from the function to a computational graph. Moreover, the rules of differentiation for the operators are pre-implemented. Therefore, it is required that the evaluated function only consists of operators that are supported by the AD object. With the function evaluation completed, the backward pass is initiated. Using the chain rule from Equation (12), the gradients with respect to the input variables are calculated given the output value. The chain rule,

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial y}\frac{\partial y}{\partial x}, \tag{12}$$

shows how the gradient of a functional is simplified to a product series of simple gradient expressions. Here, $w(y(x))$ is a functional depending on the function $y(x)$, which is a function of the input variable x. The backpropagation algorithm recursively applies the chain rule on the computational graph, eventually ending up with the gradients of the input variables [1].

In terms of advantages, AD can be a powerful tool for optimisation if the symbolic expression is difficult to derive, or ends up being an uncondensed expression. Furthermore, AD is versatile in deep learning tool boxes such as Pytorch and Tensorflow, because it allows everyone to implement their own custom neural network architectures or optimisation algorithms. However, AD is not without its disadvantages. As pointed out by Baydin, Pearlmutter, Radul, and Siskind [1], AD is not immune to floating point numbers, other numeric issues, and vanishing gradients for deep neural networks. Vanishing gradients are a consequence of the chain rule applied too many times, which will cause the gradients to approach zero.

4 SCATTERING OF X-RAYS

4.1 THE WAVE-PARTICLE DUALITY AND DIFFERENTIAL CROSSECTION

As mentioned in section 2.1, X-rays can be described as electromagnetic waves. Furthermore, X-rays are described as plane waves, or Transverse electromagnetic waves (TEM-waves), assuming they are free, coherent, and monochromatic. TEM-waves are characterised by a magnetic field H and an electric field E that are perpendicular to each other and to the direction of propagation, called the Poynting vector S. Nevertheless, the expression of the TEM wave in terms of time t and position $\boldsymbol{r}$ of the electric field E is

$$\boldsymbol{E}(\boldsymbol{r}, t) = \widehat{\boldsymbol{\epsilon}} E_0 \exp\left(-i\left(\omega t - \boldsymbol{k}\cdot\boldsymbol{r}\right)\right), \tag{13}$$

where $\widehat{\epsilon}$ is the unit vector in the direction of the electric field, E_0 is the amplitude of the electric field, ω is the angular frequency, and $\boldsymbol{k}$ is the wave vector. Therefore can scattering events between X-rays and electrons, for instance, be described by exertion of force between the electric field of the wave and the charge of the electron [7].

However, quantisation of photons, which is performed within the scope of quantum mechanics, results in X-rays being described as photons. This is a more useful description of X-rays in terms of scattering events, because it allows for particle-particle interactions between photons and electrons, with transfer of momentum, elastic scattering, but also energy in events called inelastic scattering [7].

The definition of the "differential scattering crossection" is also more intuitive from the perspective of quantisation. The differential scattering crossection is the number of photons scattered relative to number of incoming photons per unit solid angle and time. Mathematically, this is expressed in terms of scattering intensity I_s, the incoming flux Φ_0, and the differential solid angle $\Delta\Omega$:

$$\frac{dI_s}{d\Omega} = \frac{I_s}{\Phi_0 \Delta\Omega}. \tag{14}$$

4.2 CLASSICAL SCATTERING DESCRIPTION

In the classical description of scattering, the scattering vector $\boldsymbol{Q}$ and the atomic form factor $f(Q)$ are characteristic properties.

$\boldsymbol{Q}$ is linked to the phase shift of the scattering event, which can be understood from analysing equation 13. It is defined as

$$\boldsymbol{Q} = \boldsymbol{k} - \boldsymbol{k'}, \tag{15}$$

where $\boldsymbol{k}$ is the incoming wave vector and $\boldsymbol{k}'$ is the scattered wave vector.

The atomic form factor $f(Q)$ is a function of the scattering vector $\boldsymbol{Q}$, from Equation (15), and describes the scattering of X-rays by the electron density of the atom. Generally, it is a Fourier transform of the electron density distribution of the atom $\rho(\boldsymbol{r})$,

$$f(Q) = \int \rho(\boldsymbol{r}) \exp\left(i\boldsymbol{Q} \cdot \boldsymbol{r}\right) d\boldsymbol{r}. \tag{16}$$

As a result,

4.3 TIME DEPENDENT PERTUBATION THEORY

4.4 RECIPROCAL SPACE RELATED TO ELECTRON DENSITY

5

SMALL ANGLE X-RAY SCATTERING TENSOR TOMOGRAPHY

Part III

PROJECT WORK

6 DATA SETS

6.1 CARBON KNOT FROM SYNCHROTRON MEASUREMENT

6.2 CONSTRUCTED KNOWN MODEL

7 IMPLEMENTATION

8

CALCULATIONS

Part IV

PRESENTATION OF RESULTS

9 DISPROVAL OF DERIVED SYMBOLIC GRADIENTS

10

PERFORMED RECONSTRUCTIONS

11

COMPUTATION PERFORMANCE

Part V

DATA ANALYSIS

12 VALIDATION OF SAXSTT ALGORITHMS

13

COMPARISON OF SYMBOLIC AND AUTOMATIC GRADIENTS

14

IMPROVEMENTS OF COMPUTATIONAL PERFORMANCE

Part VI

APPENDIX

BIBLIOGRAPHY

[1] Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. "Automatic differentiation in machine learning: a survey." In: *Journal of Machine Learning Research* 18 (2018), pp. 1–43.

[2] Mikhail Arnol'dovich Blokhin. *The Physics of X-rays*. Vol. 4502. United States Atomic Energy Commission, Office of Technical Information, 1961, p. 12.

[3] Thorsten M Buzug. *Computed tomography: from photon statistics to modern cone-beam CT*. Soc Nuclear Med, 2009, pp. 15–46.

[4] T. Editors of Encyclopaedia. *synchrotron*. 2018.

[5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[6] Donald E. Knuth. "Computer Programming as an Art." In: *Communications of the ACM* 17.12 (1974), pp. 667–673.

[7] Des McMorrow and Jens Als-Nielsen. *Elements of modern X-ray physics*. John Wiley & Sons, 2011.

[8] Sebastian Ruder. "An overview of gradient descent optimization algorithms." In: *arXiv preprint arXiv:1609.04747* (2016).

[9] Jonathan Richard Shewchuk et al. *An introduction to the conjugate gradient method without the agonizing pain*. 1994.

[10] Gengsheng Lawrence Zeng. *Medical image reconstruction: a conceptual tutorial*. Springer, 2010, pp. 10–24.

DECLARATION

Put your declaration here.

Trondheim, December 2022

Ruben Skjelstad Dragland

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/