Deep learning based phase retrieval for X-ray phase contrast imaging

Kannara Mom

N°d'ordre NNT : 2023ISAL0087

THESE de DOCTORAT DE L'INSA LYON, membre de l'Université de Lyon

Ecole Doctorale N°160
Electronique, Electrotechnique, Automatique

Spécialité/ discipline de doctorat :
Traitement du signal et de l'image

Soutenue publiquement le 20/11/2023 par :

Kannara Mom

Deep learning based phase retrieval for X-ray phase contrast imaging

---

Devant le jury composé de :

Bousse, Alexandre Chargé de recherche, LaTIM, INSERM	Rapporteur
Soussen, Charles Professeur, CentralSupélec	Rapporteur
Denis, Loïc Professeur, Université de Saint-étienne	Examinateur
Desbat, Laurent Professeur, Université Grenoble Alpes	Examinateur
Pustelnik, Nelly Chargée de recherche, CNRS CRCN, ENS de Lyon	Examinatrice
Rolland Du Roscoat, Sabine Professeur, Université Grenoble Alpes	Examinatrice
Villanueva Perez, Pablo Associate senior lecturer, Lund university	Examinateur
Sixou, Bruno Maître de Conférences, INSA Lyon	Directeur de thèse
Langer, Max Chargé de recherche, CNRS, TIMC	Invité

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Département de la Formation par la Recherche et des Études Doctorales (**FEDORA**)

Bâtiment INSA direction, 1er étage
37, av. J. Capelle
69621 Villeurbanne Cédex
fedora@insa-lyon.fr

Référence : TH1022_Kannara MOM

L'INSA Lyon a mis en place une procédure de contrôle systématique via un outil de détection de similitudes (logiciel Compilatio). Après le dépôt du manuscrit de thèse, celui-ci est analysé par l'outil. Pour tout taux de similarité supérieur à 10%, le manuscrit est vérifié par l'équipe de FEDORA. Il s'agit notamment d'exclure les auto-citations, à condition qu'elles soient correctement référencées avec citation expresse dans le manuscrit.

Par ce document, il est attesté que ce manuscrit, dans la forme communiquée par la personne doctorante à l'INSA Lyon, satisfait aux exigences de l'Etablissement concernant le taux maximal de similitude admissible.

INSA LYON
Campus LyonTech La Doua
20, avenue Albert Einstein - 69621 Villeurbanne cedex - France
Tél. +33 [0]4 72 43 83 83 - Fax +33 [0]4 72 43 85 00
www.insa-lyon.fr

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **https://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage secretariat@edchimie-lyon.fr	**M. Stéphane DANIELE** C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne **directeur@edchimie-lyon.fr**
E.E.A.	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **https://edeea.universite-lyon.fr** Sec. : Stéphanie CAUVIN Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	**M. Philippe DELACHARTRE** INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 **philippe.delachartre@insa-lyon.fr**
E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	**Mme Sandrine CHARLES** Université Claude Bernard Lyon 1 UFR Biosciences Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69622 Villeurbanne CEDEX **sandrine.charles@univ-lyon1.fr**
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://ediss.universite-lyon.fr** Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	**Mme Sylvie RICARD-BLUM** Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (ICBMS) - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 **sylvie.ricard-blum@univ-lyon1.fr**
INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	**M. Hamamache KHEDDOUCI** Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 **hamamache.kheddouci@univ-lyon1.fr**
Matériaux	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 yann.de-ordenana@ec-lyon.fr	**M. Stéphane BENAYOUN** Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 **stephane.benayoun@ec-lyon.fr**
MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon mega@insa-lyon.fr	**M. Jocelyn BONJOUR** INSA Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69621 Villeurbanne CEDEX **jocelyn.bonjour@insa-lyon.fr**
ScSo	**ScSo*** **https://edsciencessociales.universite-lyon.fr** Sec. : Mélina FAVETON INSA : J.Y. TOUSSAINT Tél : 04.78.69.77.79 melina.faveton@univ-lyon2.fr	**M. Bruno MILLY** Université Lumière Lyon 2 86 Rue Pasteur 69365 Lyon CEDEX 07 **bruno.milly@univ-lyon2.fr**

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Contents

List of Figures

List of Tables

Introduction

In 1895, Wilhelm Conrad Röntgen made a groundbreaking discovery by detecting a previously unknown form of electromagnetic radiation resulting from electron-matter collisions. This radiation, now known as X-rays, had the unique property of being able to penetrate through the human body (Röntgen 1896). This important discovery led to the birth of X-ray radiography, revolutionizing the field of medical imaging. Recognizing its immense potential, Röntgen was awarded the Nobel Prize in 1901 for his pioneering work. In the early 1970s, the first Computed Tomography (CT) scans was produced by Godfrey Hounsfield and Allan Cormack, allowing to produce detailed cross-sectional images of the body using X-rays. For their work in developing the CT scanner, they were awarded the Nobel Prize in Physiology or Medicine in 1979.

In the field of biomedical imaging, X-ray CT has risen in prominence and is an increasingly used technique. Traditional CT imaging is based on the attenuation of X-rays as they pass through an object and its primary objective is to reveal the three-dimensional (3D) internal structures while providing quantitative information. X-ray microtomography (μCT), a high resolution form of CT form, extends its application to smaller lenght scales (Bonse 1999; Vásárhelyi et al. 2020).

X-rays are greatly attenuated by bones, metals and other hard materials, but less so in soft tissues and other light materials. Within the energy range of hard X-rays, i.e. energies exceeding 5 keV, conventional X-ray attenuation contrast imaging faces a well-known weaknesse of insufficient sensitivity. In this region, phase-based methods exhibit remarkable sensitivity, surpassing absorption-based techniques by up to three orders of magnitude (Momose and Fukuda 1995), which makes this imaging approach particularly appealing for the biomedical examination of soft tissues. The relatively recent innovation of X-ray phase imaging has certain requirements, such as a highly spatially coherent source and monochromaticity.

Consequently, most X-ray phase imaging techniques have been developed in synchrotron facilities, which are currently the main sources satisfying these conditions. The emergence of third-generation synchrotrons has conducted to new possibilities for X-ray imaging methods based on phase contrast imaging. Although this limits the accessibility of these techniques, other alternatives are being developed, such as coherent X-ray microscopes (S. Mayo et al. 2003; Eggl et al. 2016; Günther et al. 2020).

Different phase contrast techniques have been introduced, which can be roughly divided into interferometric (U. Bonse and Hart 1965; Momose, Takeda, et al. 1996; Takeda et al. 1995), crystal analyzer (Chapman et al. 1997; Davis et al. 1995), grating interferometry (David et al. 2002; Timm Weitkamp et al. 2005) and propagation based techniques (Snigirev et al. 1995; Peter Cloetens et al. 1996). These different imaging techniques have advantages and disadvantages that depend on the type of study, the experimental setup and the sample. Some studies, for example (Pagot et al. 2005), propose to compare some of these techniques.

One of the first setups for conducting phase contrast measurements was the Laue-Laue-Laue (LLL) X-ray interferometer, introduced by Bonse and Hart in 1965 (U. Bonse and Hart 1965). Phase contrast arises from the interference between the reference beam, which does not pass through the object, and the beam that has been transmitted through the object. Momose et al. (Momose, Takeda, et al. 1996) exhibited the considerable potential of phase contrast imaging for high-sensitivity visualization of soft tissues in their 1996 study. However, the requirements are very restrictive because of the mechanical stability of the optical system.

Grating interferometry, also referred to as Talbot interferometry, is a technique whose first experience was recently reported (David et al. 2002). This technique relies on the Fresnel

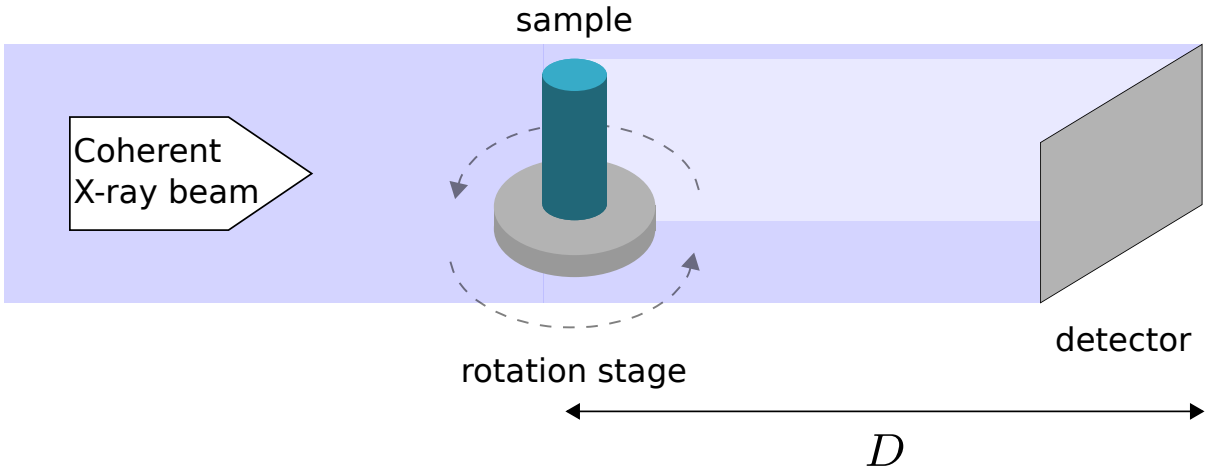


Figure 1.: Schematic of the propagation based phase tomography setup.

diffraction arising from periodic structures and relies on the Talbot effect. To achieve this, a spatially coherent beam is essential, or it can be rendered spatially coherent using a source grating (Timm Weitkamp et al. 2005). This approach exhibits reduced sensitivity to the precise positioning and stability of the experimental setup compared to interferometry.
Analyser based imaging (Chapman et al. 1997) involves the placement of an analyzer crystal between the sample and the image detector. This crystal analyzer functions as an angular filter, specifically identifying the X-rays that have traversed the sample and meet the Bragg law's criteria. Consequently, the detector records an image composed only of this restricted spectrum of X-rays, yielding a highly sensitive X-ray phase contrast imaging technique. The main drawbacks of this technique are the need for perfect crystals and high spatial stability. In addition, the photon flux is considerably reduced by the crystals, resulting in long acquisition times.

Propagation-based imaging is based on the observation that phase contrast is visible if the beam is allowed to propagate in free space after interacting with the object. This type of phase contrast image is often referred to as a diffraction pattern. The experimental setup is simple as it is essentially similar to standard X-ray radiography but with the added capability of moving the detector away from the object 4. When the detector is positioned close to the sample, at the contact plane, it records an absorption image of the object. By moving the detector downstream, phase contrast is achieved and a diffraction pattern is recorded (Figure 5). This technique, in conjunction with X-ray focusing optics, can attain very high spatial resolution in projection mode.

The image formation process can be well understood within the framework of Fresnel diffraction theory (Goodman 1996). This yields a quantitative relationship between the object and the recorded intensity. Since the phase of the beam is not directly captured in the measured intensity, this relationship can be used to reconstruct the phase shift introduced by the object, through a process known as *phase retrieval*. Since phase shift can be seen as proportional to the projection of the complex refractive index distribution in the object, it can be combined with tomography. This results in a reconstruction of the refractive index decrement, i.e. the real component of the complex refractive index (P. Cloetens, Ludwig, et al. 1999). The refractive index decrement is typically one to three orders of magnitude larger than the imaginary part, which corresponds to the absorption index. This characteristic is responsible for the enhanced sensitivity observed in phase imaging. This has found practical applications across various

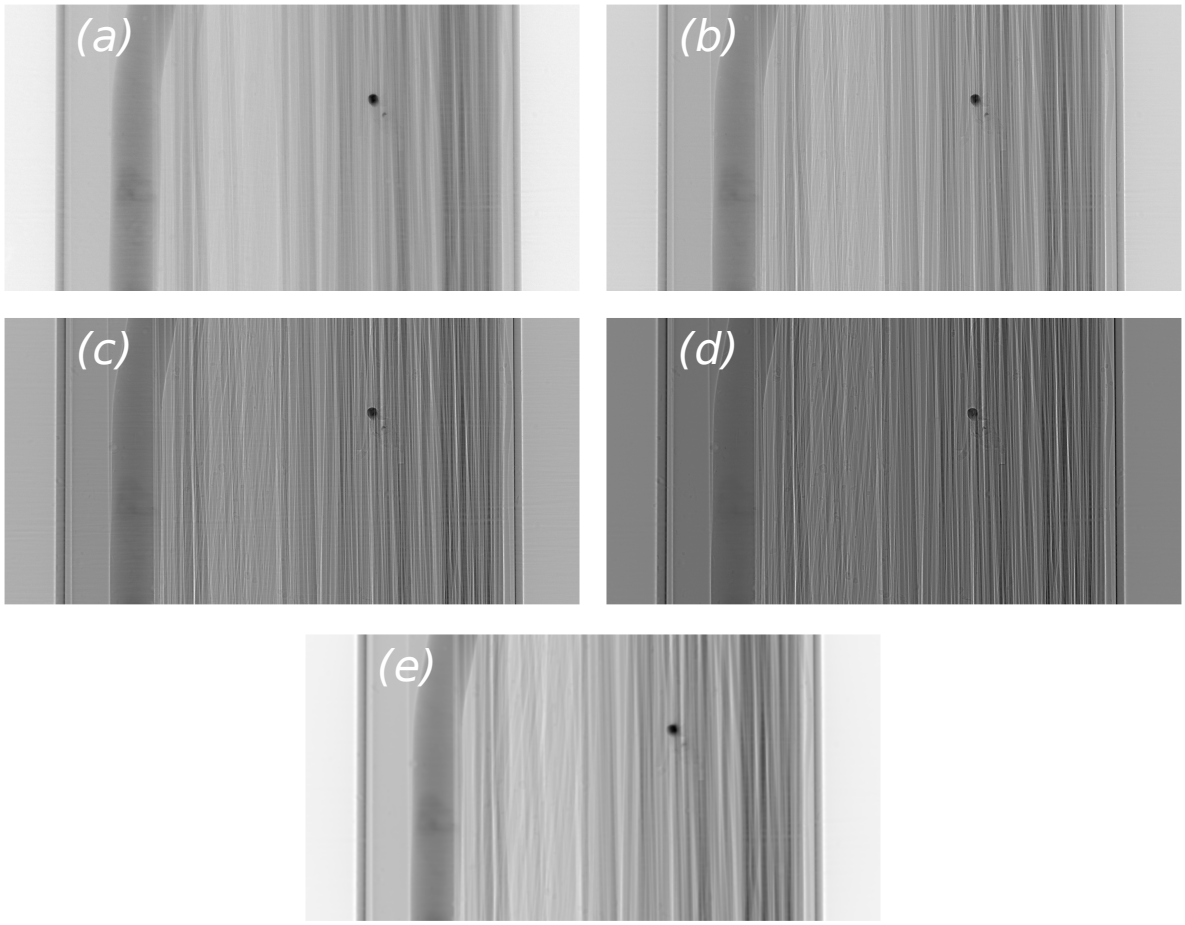


Figure 2.: Propagation phase contrast radiographs of a test object constructed from fibres of Al, Al_2O_3 , poly(ethylene terephthalate) (PET) and polypropylene (PP). It was imaged at 0.7 µm pixel size with an X-ray energy of 22.5 keV, using four propagation distances (*a*) $D = 2$ mm, (*b*) $D = 10$ mm, (*c*) $D = 20$ mm and (*d*) $D = 45$ mm. (*e*) Phase map retrieve using the TIE method (Paganin).

fields, including materials science (P. Cloetens, Pateyron-Salomé, et al. 1997; Sheridan C. Mayo, Stevenson, and Wilkins 2012) and biomedical imaging (Cancedda et al. 2007; Max Langer, Pacureanu, et al. 2012).

Several algorithms have been proposed for the phase retrieval problem, among them are algorithms that rely on the linearization of the Fresnel integral to obtain fast reconstructions. The *Contrast Transfer Function* (CTF) (J.-P. Guigay 1977; Peter Cloetens et al. 1996; Zabler et al. 2005) and the *Transport of Intensity Equation* (TIE) (Teague 1983; T. E. Gureyev and Nugent 1996; D. Paganin and Nugent 1998; D. Paganin, S. C. Mayo, et al. 2002; Paganin 2006) are such methods (Figures 5 and 6). These methods all have limitations and are only valid under some restrictive assumptions about the imaging conditions or the object. The reconstructions obtained from these are often used as initial estimates for iterative methods that refine the reconstruction (Gerchberg 1972; J. R. Fienup 1978; James R. Fienup 1982; Max Langer, Pacureanu, et al. 2012; Max Langer 2008). More recently, variational approaches, based on the minimization of a functional, make it possible to add different types of informations and constraints about the

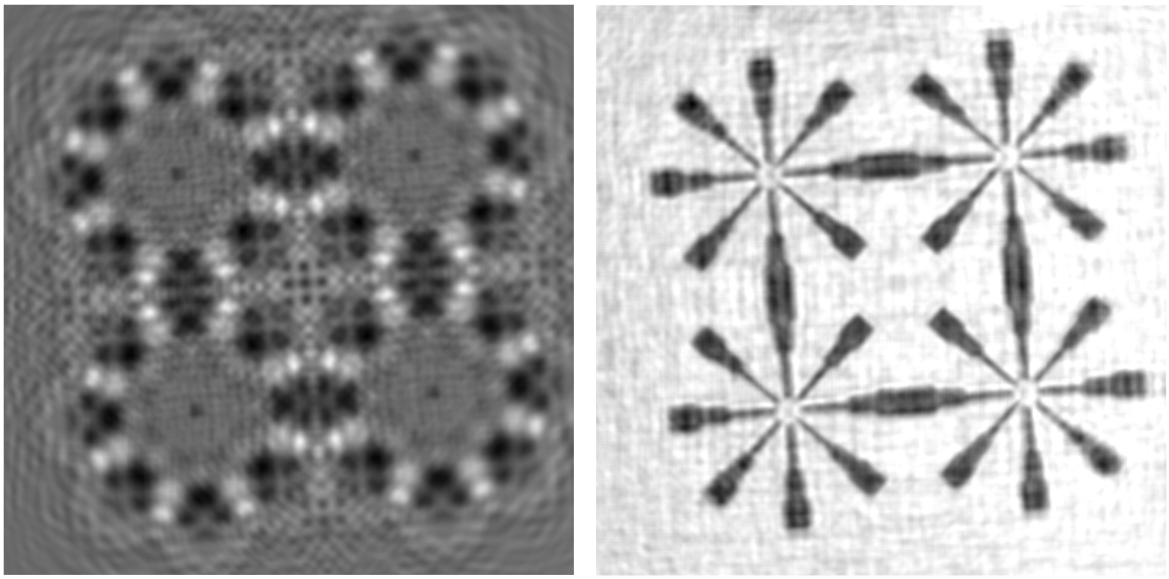


Figure 3.: (*left*) Propagation phase contrast images acquired at a distance $D = 10$ mm, imaged using X-rays at 13 keV energy and a pixel size equal to 6 nm. (*right*) Phase map retrieve using the CTF method, under homogeneous assumption.

solution (V. Davidoiu et al. 2011; Bruno Sixou et al. 2013; Maretzke, Bartels, et al. 2016). Despite these advances, there are still a number of obstacles limiting the use of these methods, such as the choice of a priori information or the computation time required.

In recent years, deep learning techniques has led to many advances in several image processing tasks. They outperform conventional methods in terms of noise reduction, segmentation and image recognition (LeCun, Y. Bengio, and Hinton 2015). Recent advances in learning methods have made it possible to develop new, efficient algorithms for inverse problems (Arridge et al. 2019). Retrieving the phase information from the recorded intensity is an example of nonlinear inverse problem. Although this remains a difficult problem, the direct problem associated with it, i.e. generating phase contrast images from a given complex refractive index, is relatively straightforward. This makes phase retrieval in X-ray phase contrast imaging a prime candidate for machine learning based inversion approaches.

Objectives

The main objective of this thesis is to propose and evaluate new algorithms based on data-driven approaches for the phase retrieval problem. These learning algorithms aim to overcome the limitations of classical approaches, such as the computation time, the choice of a priori information or restrictive assumptions on the forward model.

A key motivation of this work is to reduce the number of measurements by taking into account a single propagation distance. In particular, different neural network architectures have been studied and evaluated. A convolutional neural network for combining information at different scales has been proposed to take into account the action of the Fresnel propagator. Primal-dual approaches were introduced to regularize absorption and phase separately. The effectiveness of variational approaches has brought us to combine them with deep learning, using the so-called *unrolled* approaches.

The algorithms were evaluated using simulated data as well as experimental data acquired at NanoMAX (MAX IV, Lund, Sweden). The aim was also to make these algorithms as user-friendly as possible, so most of them have been integrated into the *PyPhase* package (Max Langer, Y. Zhang, et al. 2021), a Python package for X-ray phase imaging.

Summary

The first chapter serves as the theoretical foundation of this thesis. Within this chapter, we introduce the basics of linear inverse problems. We provide an in-depth explanation of the interactions between X-rays and matter, highlighting the advantages of employing phase imaging techniques. Additionally, we explain the contrast formation process, which constitute the *direct problem* of phase retrieval, covering both near-field (Fresnel diffraction) and far-field (Fraunhofer diffraction) scenarios. An overview is given of the main phase retrieval methods in the Fresnel domain. These include linear methods, methods based on alternating projections and variational approaches (linear and nonlinear). We give a brief introduction to deep learning and its use in image reconstruction tasks. A state-of-the-art of the different types of learning algorithms that have been proposed for phase recovery is given.

In chapter 2, we present an architecture of neural networks, the Mixed Scale Dense Network (MS-D Net). The MS-D Net combines dense connection and dilated convolution. It was able to reconstruct the phase and absorption of a heterogeneous object from a single intensity measurement.
This chapter was published in *Applied Optics* (Kannara Mom, Bruno Sixou, and Max Langer 2022).

In chapter 3, we introduce the Nonlinear Primal-Dual Hybrid Gradient (NL-PDHG), a primal-dual method that allows to regularize phase and absorption separately. A linearized version by the contrast transfer functions (CTF), the PDHG-CTF, is presented. We study the contribution of the nonlinearity information as well as the use of different regularizations for the images to be reconstructed.
Main results of this chapter were published in *Optics Letters* (Kannara Mom, Max Langer, and Bruno Sixou 2022b).

In chapter 4, we develop a new learned iterative method, the Deep Gauss-Newton (DGN). The DGN algorithm is obtained by unrolling a regularized Gauss-Newton scheme. A more efficient variant, the Deep Proximal Gauss-Newton (DPGN), is also proposed. These approaches combine the power of neural networks with the versality of classical iterative approaches.
The first part of this chapter on the DGN algorithm was published in *Optics Letters* (Kannara Mom, Max Langer, and Bruno Sixou 2023).

In chapter 5, a comparative study of different unrolling methods, based on gradient descent algorithm, primal-dual scheme, and Gauss-Newton type methods. These approaches are compared to their classical iterative counterpart. An empirical study of their stability is undertaken, as well as the effect of the number of unrolled iterations.
Part of this chapter was the subject of a conference paper (Kannara Mom, Max Langer, and Bruno Sixou 2022a).

In appendix, we show that the previous unrolling methods can be use for the basis material decomposition, another nonlinear inverse problem. Several ways of improving these approaches are suggested, such as the use of different steps and a regularized cost function.
This chapter was the subject of a conference paper (Kannara Mom, Lesaint, et al. 2023).

Finally, we explore potential avenues for future research interests.

Résumé étendu

Introduction

Dans le domaine de l'imagerie biomédicale, la tomographie par rayons X a gagné en importance et est de plus en plus utilisée. L'imagerie CT (Computed Tomography) ou tomodensitométrie (TDM) traditionnelle repose sur l'atténuation des rayons X lorsqu'ils traversent un objet, son objectif principal étant de révéler les structures internes tridimensionnelles tout en fournissant des données quantitatives. La microtomographie à rayons X (μCT), une forme de CT à haute résolution, étend son application à de plus petites échelles de longueur (Bonse 1999; Vásárhelyi et al. 2020).

À travers ses observations, Wilhelm Conrad Röntgen a découvert que les rayons X étaient fortement atténués par les os et les métaux, mais très peu absorbés par les tissus mous. Dans la gamme d'énergie des *rayons X durs*, i.e. des rayons X ayant des énergies photoniques élevées (supérieures à 5 keV), l'imagerie conventionnelle par contraste d'atténuation se heurte à une faiblesse bien connue, à savoir un manque de sensibilité. Dans cette région des rayons X durs, les méthodes basées sur la phase présentent une sensibilité remarquable et peuvent être trois ordres de grandeur plus sensibles que les techniques basées sur l'absorption (Momose and Fukuda 1995), ce qui rend cette approche d'imagerie particulièrement intéressante pour l'examen biomédical des tissus mous. Les avancées récentes de l'imagerie de phase par rayons X nécessitent cependant quelques conditions préalables, telles qu'une source hautement cohérente, la monochromaticité et un flux élevé. Par conséquent, la plupart des techniques d'imagerie de phase par rayons X ont été développées dans des installations synchrotrons, qui sont actuellement les principales sources remplissant ces conditions. L'émergence des synchrotrons de troisième génération a conduit à de nouvelles possibilités pour les méthodes d'imagerie par rayons X basées sur le contraste de phase. Bien que cela limite l'accessibilité à ces techniques, d'autres alternatives sont en cours de développement, telles que les microscopes à rayons X cohérents (S. Mayo et al. 2003).

Différentes techniques de contraste de phase ont été introduites, qui peuvent être grossièrement divisées en 4 groupes: l'interférométrie (U. Bonse and Hart 1965; Momose, Takeda, et al. 1996; Takeda et al. 1995), les analyseurs de cristaux (Chapman et al. 1997; Davis et al. 1995), l'interférométrie à réseau (David et al. 2002; Timm Weitkamp et al. 2005) et les techniques basées sur la propagation (Snigirev et al. 1995; Peter Cloetens et al. 1996). Ces différentes techniques d'imagerie présentent des avantages et des inconvénients qui dépendent du type d'étude et de l'échantillon en question. Certaines études, par exemple (Pagot et al. 2005), proposent de comparer certaines de ces techniques d'imagerie de phase.

L'interféromètre à rayons X Laue-Laue-Laue (LLL), introduite par Bonse et Hart en 1965 (U. Bonse and Hart 1965) a été l'une des premières installations permettant d'effectuer des mesures de contraste de phase. En 1996, Momose et al. (Momose, Takeda, et al. 1996) ont montré le potentiel considérable de l'imagerie par contraste de phase pour la visualisation des tissus mous avec une grande sensibilité. Cependant, les conditions expérimentales sont très restrictives en raison d'un besoin de stabilité mécanique du système optique. L'interférométrie à réseau, également appelée interférométrie de Talbot-Lau, est une technique dont la première expérience a été décrite récemment (David et al. 2002). Cette technique s'appuie sur la diffraction de Fresnel provenant de structures périodiques et sur l'effet Talbot. Cette approche est moins sensible au positionnement précis et à la stabilité du dispositif expérimental que demande l'interférométrie classique.

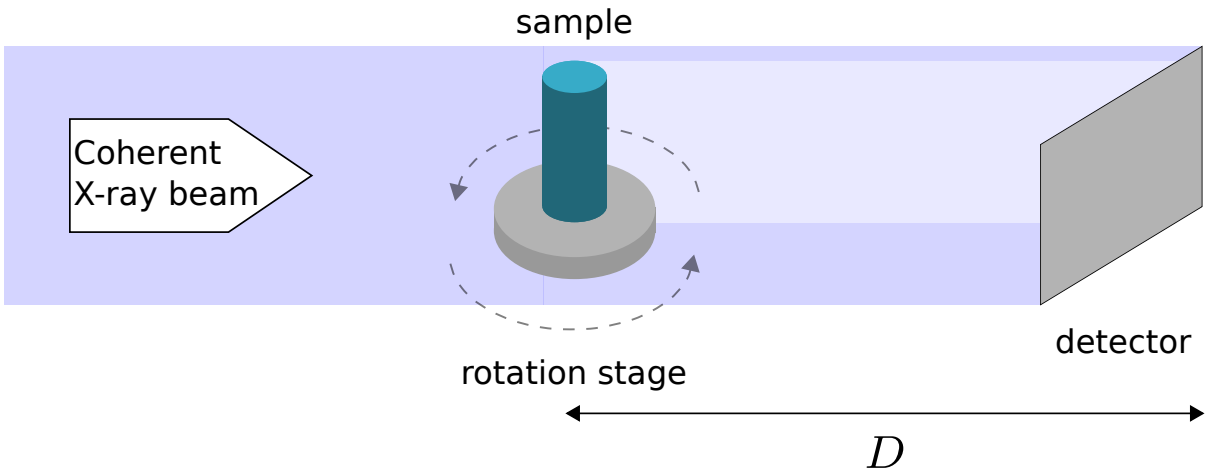


Figure 4.: Principe expérimental de la tomographie par contraste de phase basée sur la propagation.

Les techniques basées sur la propagation reposent sur le principe selon lequel le faisceau de rayons X peut se propager librement dans l'espace après avoir interagi avec l'objet, l'image enregistrée est un diagramme de diffraction (Snigirev et al. 1995; Peter Cloetens et al. 1996). Le dispositif expérimental est simple car il est similaire à celui de la radiographie standard, mais avec la possibilité supplémentaire d'éloigner le détecteur par rapport à l'objet (Figure 4). Lorsque le détecteur est placé près de l'échantillon, au niveau du plan de contact, une image d'absorption de l'objet est enregistrée. En éloignant le détecteur de l'objet, on obtient un contraste de phase et on enregistre un diagramme de diffraction (Figure 5). Comme cette configuration expérimentale ne nécessite aucun élément optique, contrairement aux techniques mentionnées précédemment, la résolution spatiale obtenue peut être très élevée, de l'ordre du micron, voire même plus fine.

Le processus de formation de l'image est bien décrit dans le cadre de la théorie de la diffraction de Fresnel (Goodman 1996). On obtient ainsi une relation quantitative entre l'objet et l'intensité enregistrée. Comme l'information de phase n'est pas directement capturée dans l'intensité mesurée, cette relation peut être utilisée pour calculer le déphasage introduit par l'objet, ce processus est connu sous le nom de *récupération de phase*. Le déphasage est proportionnel à une projection de la distribution de l'indice de réfraction complexe de l'objet. Par conséquent, la récupération de phase peut être combinée avec la tomographie, la reconstruction qui en résulte est la partie réelle de l'indice de réfraction complexe (P. Cloetens, Ludwig, et al. 1999). Cette composante réelle de l'indice de réfraction complexe est généralement plus grand d'un à trois ordres de grandeur que la partie imaginaire, qui correspond à l'indice d'absorption. Cette caractéristique est responsable de la sensibilité observée dans l'imagerie de phase. La tomographie par contraste de phase a trouvé des applications pratiques dans divers domaines, notamment la science des matériaux (P. Cloetens, Pateyron-Salomé, et al. 1997; Sheridan C. Mayo, Stevenson, and Wilkins 2012) et l'imagerie biomédicale (Cancedda et al. 2007; Max Langer, Pacureanu, et al. 2012).

Plusieurs algorithmes ont été proposés pour le problème de la récupération de phase, parmi lesquels on retrouve des algorithmes qui reposent sur la linéarisation de l'intégrale de Fresnel pour obtenir des reconstructions directes (analytiques). La *fonction de transfert du contraste* (CTF) (J.-P. Guigay 1977; Peter Cloetens et al. 1996; Zabler et al. 2005), *l'équation de transport d'intensité* (TIE) (Teague 1983; T. E. Gureyev and Nugent 1996; D. Paganin and Nugent 1998; D. Paganin, S. C. Mayo, et al. 2002; Paganin 2006) sont de telles méthodes (Figures 5 et 6). Ces algorithmes

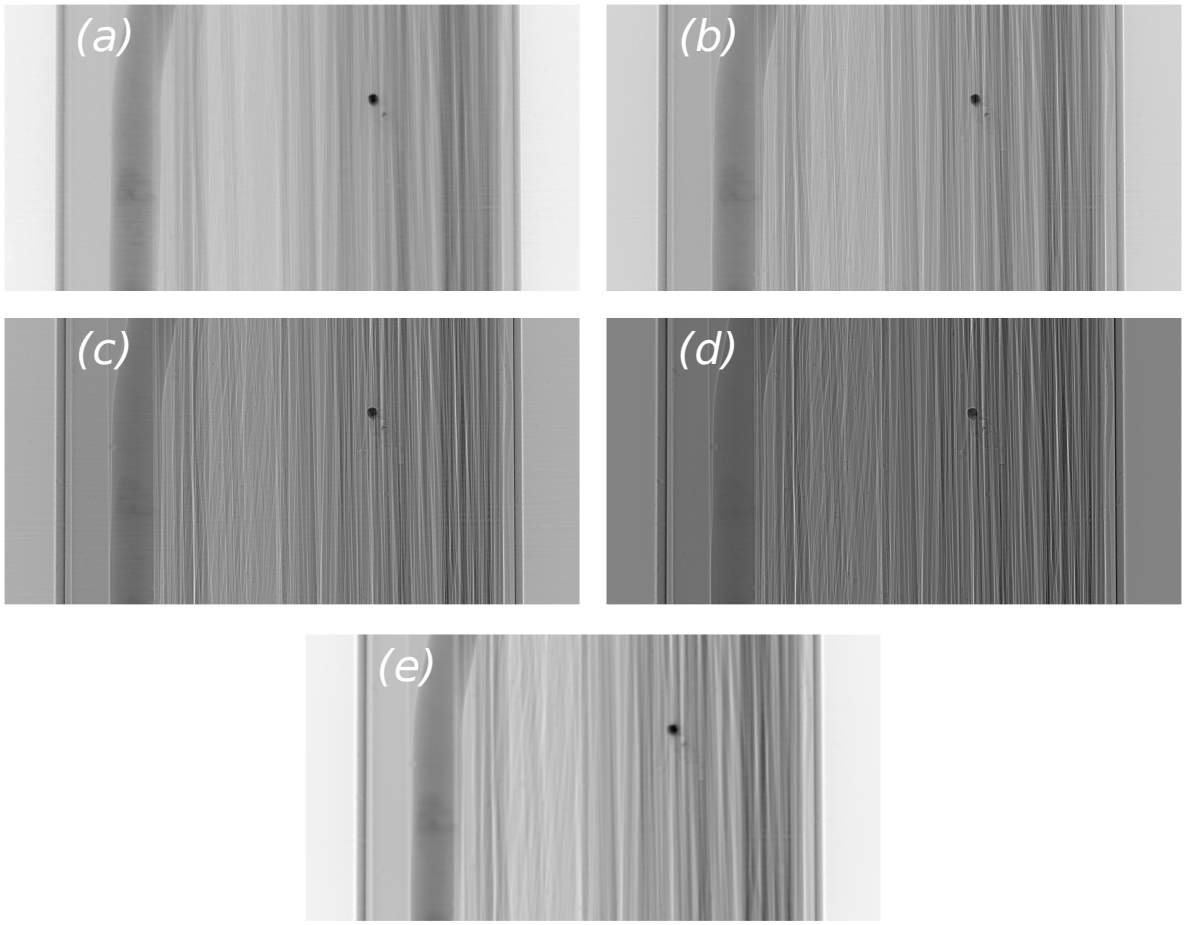


Figure 5.: Radiographies de contraste de phase d'un objet test constitué de fibres d'aluminium (Al), d'oxyde d'aluminium (Al_2O_3), de polyéthylène téréphtalate (PET) et de polypropylène (PP). Les images ont une taille de pixel de 0.7 µm, elles ont été obtenues avec une énergie de rayons X de 22.5 keV et en utilisant quatre distances de propagation (*a*) $D = 2$ mm, (*b*) $D = 10$ mm, (*c*) $D = 20$ mm et (*d*) $D = 45$ mm. (*e*) Image de phase récupérée avec la méthode TIE (Paganin).

présentent des limites et ne sont valables que sous certaines hypothèses restrictives concernant les conditions d'imagerie ou l'objet. Les reconstructions obtenues à partir de ces méthodes sont souvent utilisées comme estimations initiales pour des méthodes itératives qui ont pour but d'affiner la reconstruction (Gerchberg 1972; J. R. Fienup 1978; James R. Fienup 1982; Max Langer, Pacureanu, et al. 2012; Max Langer 2008). Plus récemment, les *approches variationnelles*, basées sur la minimisation d'une fonctionnelle, permettent d'ajouter différents types d'informations a priori et de contraintes sur la solution (V. Davidoiu et al. 2011; Bruno Sixou et al. 2013; Maretzke, Bartels, et al. 2016), à travers un terme de régularisation. Malgré tout, un certain nombre d'obstacles limitent encore l'utilisation de ces méthodes, comme le choix de la régularisation ou le temps de calcul nécessaire.

Ces dernières années, les techniques d'apprentissage profond ont permis de nombreuses avancées dans plusieurs tâches de traitement d'images. Elles surpassent les méthodes conventionnelles en termes de réduction du bruit, de segmentation et de reconnaissance d'images

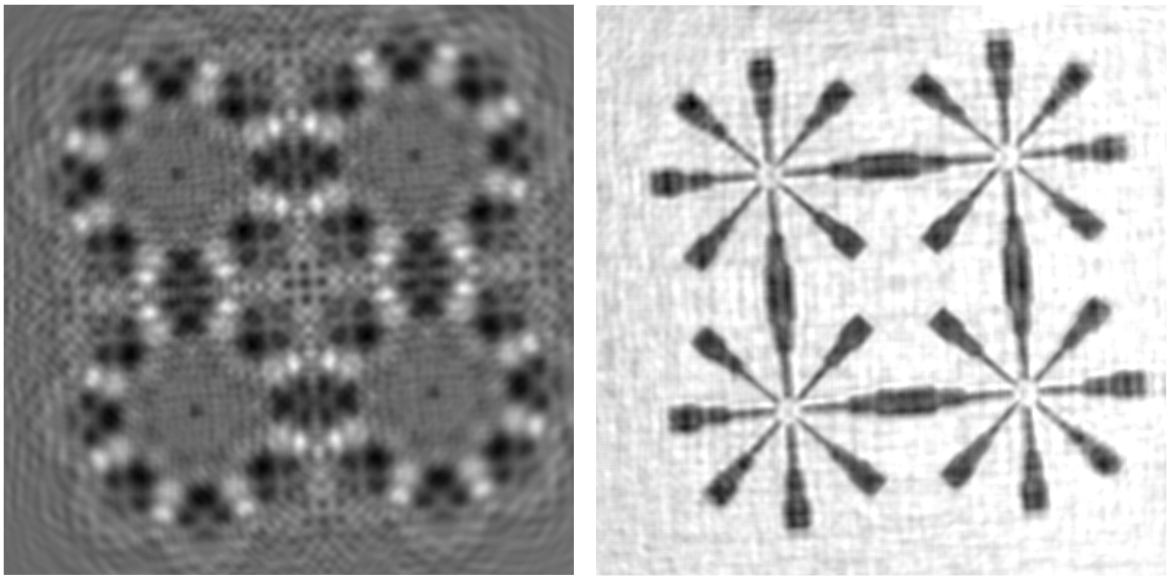


Figure 6.: Image de contraste de phase obtenue avec une distance de propagation $D = 10\,\text{mm}$, avec une énergie de rayons X de 13 keV et une taille de pixel égale à 6 nm (*gauche*). Récupération de la phase à l'aide de la méthode CTF, sous une hypothèse d'homogénéité (*droite*).

(LeCun, Y. Bengio, and Hinton 2015). Les progrès récents des méthodes d'apprentissage ont permis de développer de nouveaux algorithmes efficaces pour les problèmes inverses (Arridge et al. 2019). La récupération de l'information de phase à partir de l'intensité mesurée est un exemple de problème inverse non linéaire. Bien que ce problème reste difficile car mal posé, le problème direct qui lui est associé, à savoir la génération d'images de contraste de phase à partir d'un indice de réfraction complexe donné, est relativement simple. Le problème de récupération de phase est donc un candidat de choix pour les approches d'inversion basées sur l'apprentissage.

Objectifs

L'objectif principal de cette thèse est de proposer et d'évaluer de nouveaux algorithmes basés sur l'apprentissage pour le problème de récupération de phase. Ces algorithmes d'apprentissage visent à surmonter les limitations des approches classiques, telles que le temps de calcul, le choix des informations a priori ou les hypothèses restrictives sur les conditions d'imagerie.

L'une des principales motivations de ce travail était de réduire le nombre de mesures en prenant en compte seulement une seule distance de propagation. Nous avons également examiné les performances de nos méthodes en présence de bruit. Différentes architectures de réseaux neuronaux ont été étudiées et évaluées. Un réseau neuronal convolutionnel permettant de combiner des informations à différentes échelles a été proposé pour prendre en compte l'action du propagateur de Fresnel. Des approches primales-duales ont été introduites pour régulariser l'absorption et la phase séparément. L'efficacité des approches variationnelles nous a amenés à les combiner avec l'apprentissage profond, en utilisant les approches dites *unrolled*.

Nos algorithmes ont été évalués en utilisant des données simulées ainsi que des données expérimentales acquises à NanoMAX (MAX IV, Lund, Suède). L'objectif était également de

rendre ces algorithmes utilisables par un maximum de personnes, c'est pourquoi la plupart de ces algorithmes ont été intégrés dans le package *PyPhase* (Max Langer, Y. Zhang, et al. 2021), un package Python pour l'imagerie de phase à rayons X.

Organisation du manuscrit

Le premier chapitre sert de base théorique à cette thèse. Dans ce chapitre, nous introduisons les bases des problèmes inverses linéaires. Nous fournissons une explication approfondie des interactions entre les rayons X et la matière, en soulignant les avantages de l'utilisation des techniques d'imagerie de phase. En outre, nous expliquons le processus de formation de l'image de contraste, qui constitue le *problème direct* de la récupération de phase, couvrant à la fois la notion de champ proche (diffraction de Fresnel) et celle de champ lointain (diffraction de Fraunhofer). Une vue d'ensemble est donnée des principales méthodes pour la récupération de phase dans le domaine de Fresnel. Cela inclut les méthodes linéaires, les méthodes basées sur des projections alternées ainsi que les approches variationnelles (linéaires et non linéaires). Ensuite, nous donnons une brève introduction à l'apprentissage profond et son utilisation dans les tâches de reconstruction d'images. Un état de l'art non exhaustif des différents algorithmes d'apprentissage qui ont été proposés pour la récupération de phase est donné.

Dans le chapitre 2, nous présentons une architecture de réseau de neurones, appelé *Mixed scale dense network* (MS-D Net). Le réseau MS-D combine l'utilisation de convolutions dilatées et celle de connexions denses. Ce type de réseau a permis de reconstruire simultanément la phase et l'absorption d'un objet hétérogène à partir d'une seule mesure d'intensité, sans information a priori.
Les travaux présentés dans ce chapitre ont donné lieu à une publication dans le journal *Applied Optics* (Kannara Mom, Bruno Sixou, and Max Langer 2022).

Dans le chapitre 3, nous présentons une méthode primale-duale non linéaire, appelée *nonlinear primal-dual hybrid gradient* (NL-PDHG), qui permet de régulariser la phase et l'absorption séparément. Une version linéarisée par la fonction de transfert de contraste (CTF), nommée PDHG-CTF, est aussi présentée. Nous étudions la contribution de l'information de non linéarité ainsi que l'utilisation de différentes régularisations pour les images que nous souhaitons reconstruire.
Les principaux résultats de ce chapitre ont été publiés dans le journal *Optics Letters* (Kannara Mom, Max Langer, and Bruno Sixou 2022b).

Dans le chapitre 4, nous développons une nouvelle méthode itérative basée sur l'apprentissage, le Deep Gauss-Newton (DGN). L'algorithme DGN est obtenu en déroulant un schéma de Gauss-Newton régularisé. Une variante plus efficace, le Deep Proximal Gauss-Newton (DPGN), est également proposée. Ces approches combinent la puissance des réseaux neuronaux avec la polyvalence des approches itératives classiques.
La première partie de ce chapitre sur l'algorithme DGN a été publiée dans le journal *Optics Letters* (Kannara Mom, Max Langer, and Bruno Sixou 2023).

Dans le chapitre 5, une étude comparative de différentes méthodes déroulées, basées sur l'algorithme de descente de gradient, le schéma primal-dual et les méthodes de type Gauss-Newton. Ces approches sont comparées à leur équivalent itératif classique. Une étude empirique de leur stabilité est donnée, ainsi que l'effet du nombre d'itérations utilisé pour le déroulement. Une partie de ce chapitre a fait l'objet d'un article de conférence (Kannara Mom, Max Langer, and Bruno Sixou 2022a).

En annexe, nous montrons que les méthodes déroulées précédentes peuvent aussi être utilisées pour un autre problème inverse non linéaire, la décomposition en matériaux de base. Plusieurs façons d'améliorer ces approches sont suggérées, en particulier, l'utilisation de différents pas de

descente et une fonction de coût régularisée.
Ce chapitre a fait l'objet d'un article de conférence (Kannara Mom, Lesaint, et al. 2023).

Cadre théorique

Le principe de base de l'imagerie basée sur la propagation consiste à laisser le faisceau de rayons X se propager librement après avoir traversé l'objet avant d'atteindre le détecteur. Au cours de cette propagation sur une distance D, les rayons X qui traversent l'objet subissent un changement de phase dû aux différences d'indice de réfraction. Les mesures obtenues après la propagation sont des images de contraste de phase, également appelées diagrammes de diffraction de Fresnel.

Imagerie de contraste de phase basée sur la propagation

L'interaction d'un faisceau de rayons X avec un objet est décrit par son *indice de réfraction complexe* 3D qui est généralement exprimé de la manière suivante:

$$n(x, y, z) = 1 - \delta_r(x, y, z) + i\beta(x, y, z) \tag{0.1}$$

La partie réelle de l'indice de réfraction n est liée au décalage de phase φ induit par l'objet, tandis que la partie imaginaire est liée à l'absorption B de l'objet. Si l'on note les coodonnées du plan par $\mathbf{x} = (x, y)$ et que l'axe z suit la direction de propagation du faisceau de rayons X, on a les relations suivantes:

$$B(\mathbf{x}) = \frac{2\pi}{\lambda} \int \beta(\mathbf{x}, z)\mathrm{d}z \tag{0.2}$$

$$\varphi(\mathbf{x}) = -\frac{2\pi}{\lambda} \int \delta_r(\mathbf{x}, z)\mathrm{d}z \tag{0.3}$$

Le décalage de phase φ et l'absorption B peuvent alors être considérés comme les projections des indices δ_r et β, à un facteur multiplicatif près, qui dépend de la longueur d'onde λ. L'indice δ_r est appelé *décrément de l'indice de réfraction* et l'indice β est appelé *indice d'absorption*, ils dépendent tous les deux de la composition de l'objet ainsi que de l'énergie du faisceau. Il est à noter que, pour les tissus mous et au sein de la plage des rayons X durs, le décrément de l'indice de réfraction peut être jusqu'à trois ordres de grandeur plus grand que l'indice d'absorption, comme le montre les graphes de la Figure 7.

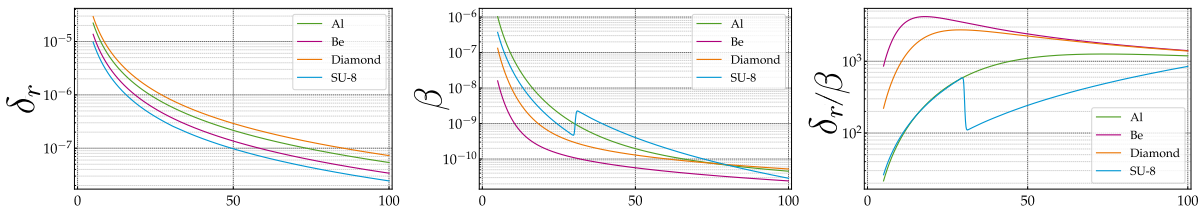


Figure 7.: Décrément de l'indice de réfraction δ_r (*gauche*), indice d'absorption β (*milieu*) et ratio δ_r/β pour l'aluminium (Al), le Béryllium (Be), le diamant et le SU-8 (Celestre 2021) (*droite*). L'axe des abscisses représente les énergies (en keV).

Considérons un objet éclairé par un faisceau de rayons incidents $u_{\mathrm{inc}}(\mathbf{x})$, monochromatique de longueur d'onde λ. Pour une propagation rectiligne du faisceau suivant la direction z (Figure 8), l'interaction du faisceau avec la matière peut être décrite par une fonction de transmittance

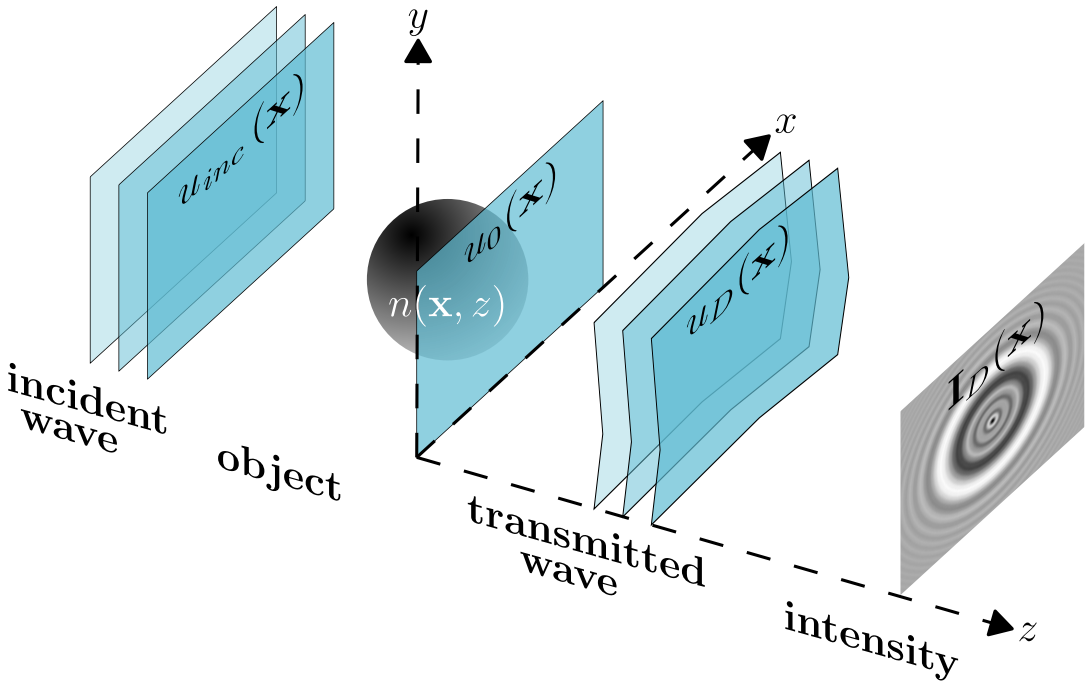


Figure 8.: Schéma du modèle physique de base de l'imagerie de contraste de phase par rayons X basée sur la propagation.

2D, notée T, qui s'écrit:

$$T(\mathbf{x}) = \exp[-B(\mathbf{x}) + i\varphi(\mathbf{x})] \tag{0.4}$$

L'*intensité* de l'onde à la sortie de l'objet ($D = 0$) est donnée par

$$\mathbf{I}_0(\mathbf{x}) = |T(\mathbf{x})u_{\text{inc}}(\mathbf{x})|^2 = \mathrm{e}^{-2B(\mathbf{x})}\mathbf{I}_{\text{inc}}(\mathbf{x}) \tag{0.5}$$

En particulier, si l'intensité incidente $\mathbf{I}_{\text{inc}}(\mathbf{x})$ est connue, on peut accéder directement à l'information de l'atténuation et par conséquent à l'absorption.

Dans le cadre de la théorie de la diffraction de Fresnel, la fonction d'onde complexe $u_D(\mathbf{x})$ à une distance D en aval de l'objet peut être décrite comme une convolution (Goodman 1996)

$$u_D(\mathbf{x}) = P_D(\mathbf{x}) * u_0(\mathbf{x}), \quad \text{avec} \quad P_D(\mathbf{x}) = \frac{1}{i\lambda D}\exp\left(i\frac{\pi}{\lambda D}|\mathbf{x}|^2\right) \tag{0.6}$$

où $P_D(\mathbf{x})$ est appelé *propagateur de Fresnel*. Dans la suite, nous ferons l'hypothèse que $u_{\text{inc}}(\mathbf{x}) = 1$, dans ce cas, l'intensité détectée à une distance D de l'objet peut s'écrire

$$\mathbf{I}_D(\mathbf{x}) = |u_D(\mathbf{x})|^2 = |P_D(\mathbf{x}) * T(\mathbf{x})|^2 \tag{0.7}$$

Problème direct et inverse

L'opérateur qui décrit la formation de l'image d'intensité à une distance de propagation D de l'objet est défini par l'équation (0.7), il est également connu sous le nom de *transformation de Fresnel*. Pour faire le lien entre les images que l'on souhaite reconstruire (B, φ) et l'information mesurée, on définit l'*opérateur direct*

$$\mathbf{F}_D(B, \varphi) = \left|P_D * \mathrm{e}^{-B+i\varphi}\right|^2 \tag{0.8}$$

Tableau 1.: Résumé des méthodes linéaires pour la récupération de phase.

Méthode	Validité	Distances de propagation
TIE	distance de propagation courte	2 (proches)
Paganin	objet homogène	1
CTF	absorption faible et phase variant peu	2
Mixed	objet variant peu	2

Il est évident que $\mathbf{F}_D$ est un opérateur *non linéaire* en raison de l'exponentielle et de l'opération du module au carré. Le *problème inverse* associé à ce modèle direct peut donc être exprimé de la façon suivante:
Pour un certain ensemble A, nous cherchons à reconstruire l'absorption et le déphasage $(B, \varphi) \in A$ à partir de la mesure d'intensité mesurée $\mathbf{I}_D^{\text{obs}}$, de sorte que

$$\mathbf{I}_D^{\text{obs}} \approx \mathbf{F}_D(B, \varphi)$$

Ce problème est *mal posé* au sens d'Hadamard, c'est-à-dire qu'il ne satisfait pas au moins une des conditions suivantes: 1) Une solution existe. 2) La solution est unique. 3) La solution dépend continûment des données.

Méthodes linéaires pour la récupération de phase

Pour simplifier le problème, on peut linéariser l'opérateur direct en se basant sur différents modèles de contraste. Il en existe plusieurs, les plus connus étant l'équation de transport d'intensité (TIE) (Teague 1983), la fonction de transfert du contraste (CTF) (D. Paganin, S. C. Mayo, et al. 2002) et le modèle mixte (J. P. Guigay et al. 2007). Chacun de ces modèles repose sur différentes hypothèses sur les conditions d'imagerie ou sur l'objet. Par exemple, le modèle CTF suppose que l'absorption est faible et que la phase varie peu, dans ce cas on obtient une relation linéaire dans le domaine de Fourier

$$\widehat{\mathbf{I}}_D(\mathbf{f}) \approx \delta(\mathbf{f}) - 2\cos\left(\pi\lambda D\,|\mathbf{f}|^2\right)\widehat{B}(\mathbf{f}) + 2\sin\left(\pi\lambda D\,|f|^2\right)\widehat{\varphi}(\mathbf{f})$$

Cette relation permet de définir un opérateur direct, qui cette fois-ci est linéaire et correspond au modèle CTF :

$$\mathbf{F}_D^{\text{CTF}}(B, \varphi) = -2\mathcal{F}^{-1}\left[\cos(\pi\lambda D\,|\mathbf{f}|^2)\widehat{B}(\mathbf{f}) - \sin(\pi\lambda D\,|f|^2)\widehat{\varphi}(\mathbf{f})\right]$$

En combinant les modèles linéaires avec une méthode des moindres carrés, on peut avoir une formule explicite pour la phase et l'absorption. Cependant, dans le cas où l'on a une seule mesure d'intensité, ces algorithmes ne fonctionnent qu'à condition d'avoir une hypothèse d'homogénéité sur l'objet, ce qui est assez restrictif. Un résumé des méthodes linéaires/analytiques peut être retrouvé dans le tableau 1 et on peut voir le domaine de validité en fonction des distances de propagations utilisées dans la Figure 9. Le grand avantage de ces méthodes analytiques est d'avoir un temps de reconstruction relativement court. Mais ces approches ne prennent pas en compte l'information non linéaire et ne peuvent pas inclure de contrainte de non-négativité, car cela reviendrait à résoudre un problème non linéaire.

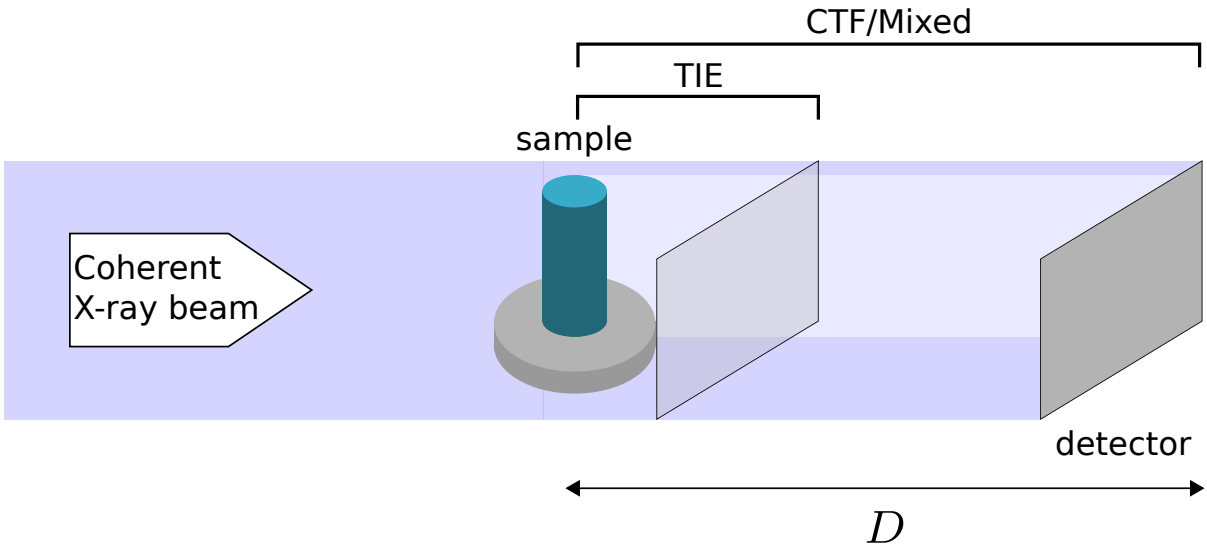


Figure 9.: Schéma du système de formation d'images. Le contraste de phase peut être enregistré à plusieurs distances, typiquement 2 pour la méthode TIE et 4 pour les méthodes CTF et mixte.

Méthodes itératives

Les algorithmes itératifs ne sont pas limités par ces contraintes. Parmi eux, on retrouve les méthodes qui consistent à projeter sur des contraintes (Gerchberg 1972), en alternant entre le domaine objet et le domaine de Fourier (Figure 10). Différentes méthodes se basant sur cette idée

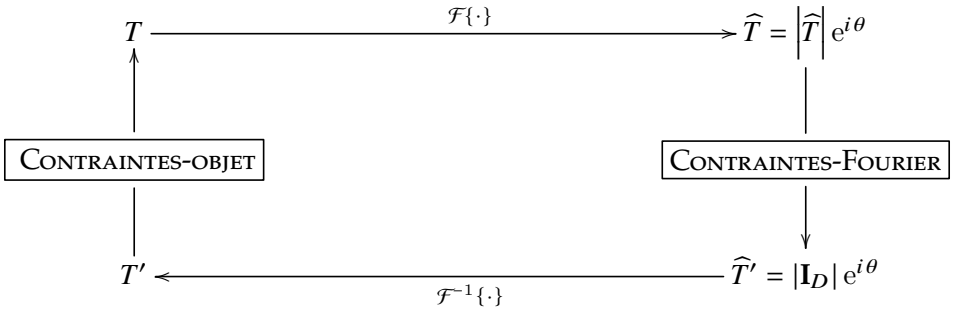


Figure 10.: Diagramme de l'algorithme *Error-Reduction*.

ont été proposées, en s'inspirant de l'optimisation convexe notamment (Bauschke, Combettes, and D. Russel Luke 2003; D Russel Luke 2005; Elser 2003).

Toujours dans les méthodes itératives, on retrouve les *approches variationnelles* qui reposent sur la reformulation du problème de reconstruction en un problème de minimisation

$$\underset{B,\varphi}{\operatorname{argmin}} \left\{ d\left[\mathbf{F}(B,\varphi), \mathbf{I}_D^{\text{obs}}\right] + \mathcal{R}(B,\varphi) \right\} \tag{0.9}$$

où $\mathbf{F} \in \left\{\mathbf{F}_D, \mathbf{F}_D^{\text{CTF}}\right\}$, d est un terme d'attache aux données et $\mathcal{R}$ est un terme de régularisation permettant d'inclure des informations *a priori* sur la solution recherchée. Lorsque le problème inverse est linéaire, alors on peut se ramener à une fonctionnelle convexe et résoudre le problème

variationnel à l'aide d'outil d'optimisation convexe classique. Cela inclut les méthodes de types descente de gradient, schéma forward-backward splitting, ou encore des méthodes primales-duales (Chambolle and Pock 2016a). Dans le cas où l'opérateur direct est non linéaire, d'autres méthodes ont été développées. On peut citer le gradient conjugué non linéaire ou encore la méthode de Gauss-Newton régularisée (Kaltenbacher, Andreas Neubauer, and Scherzer 2008). Il subsiste de nombreux freins à l'utilisation de ces méthodes en pratique. En effet, le choix d'une régularisation et celle du poids qui lui est associé est important dans la qualité de reconstruction et ce choix reste délicat. De plus, la complexité en temps peut être grande puisque ces méthodes demandent parfois des milliers d'itérations. Ces différents facteurs limitant justifient la recherche de solutions fondées sur l'apprentissage.

Les progrès récents des techniques d'apprentissage profond ont donné lieu à de nombreuses applications dans le domaine du traitement des images (LeCun, Y. Bengio, and Hinton 2015). En particulier, des approches innovantes ont vu le jour pour résoudre les problèmes inverses, y compris les tâches de reconstruction (Arridge et al. 2019). Bien qu'un bon nombre de ces approches se soient concentrées sur les problèmes inverses linéaires (Ongie et al. 2020), un nombre croissant de travaux portent sur des cas non linéaires (Hoop, Lassas, and Wong 2022), en particulier sur des problèmes tels que la récupération de phase (Deshpande et al. 2022).

Différentes stratégies ont été proposées pour utiliser l'apprentissage profond pour des problèmes de reconstruction. L'une des approches les plus simples consiste à entraîner un réseau de neurones à reconstruire directement à partir des mesures observées. Dans ce scénario, un réseau de neurones est entraîné pour apprendre (ou approcher) l'inverse de l'opérateur direct, prenant ainsi en charge l'ensemble de la tâche de reconstruction, on parle de *reconstruction directe*. Si une estimation initiale est accessible, par exemple par une méthode analytique, on peut utiliser l'apprentissage profond pour améliorer cette images. Ces méthodes sont appelées *post-traitement*.

Plus récemment, des méthodes proposent de traiter le problème de reconstruction comme un problème d'optimisation et de combiner des techniques d'optimisation traditionnelles avec l'apprentissage profond. C'est l'idée de l'approche *Plug-and-Play* (PnP), où l'on va itérer sur une estimée, en alternant entre : *(1)* une étape d'attache aux données, qui inclut la connaissance du modèle direct et *(2)* l'application d'un réseau qui a été entraîné en amont à effectuer une tâche spécifique (débruitage, déconvolution). Suivant la même idée, il y a les méthodes dites, d'*unrolling*. Ces méthodes partent d'un schéma d'optimisation et le déroulent en un nombre fini d'itérations, ce schéma déroulé est remplacé par un réseau de neurones profond où chaque itération consiste en l'application d'un réseau de neurones.

Les prochaines sections représentent les contributions apportées dans le domaine de récupération de phase.

Réseaux de neurones convolutionnels denses à échelle mixte pour l'imagerie de contraste de phase des rayons X

Dans ce chapitre, nous présentons une approche d'apprentissage supervisé pour extraire simultanément les informations de phase et d'absorption à partir des images de contraste de phase des rayons X. Dans ce contexte, l'ensemble du processus de reconstruction est appris durant l'entraînement d'un réseau neuronal profond afin d'établir une correspondance directe entre les mesures et le résultat souhaité. L'architecture du réseau joue un rôle crucial dans l'apprentissage profond pour la reconstruction directe.

Nous introduisons un réseau de neurones convolutionnels denses à échelle mixte, appelé *mixed scale dense network* (MS-D network). La conception de ce réseau s'appuie sur des *convolutions*

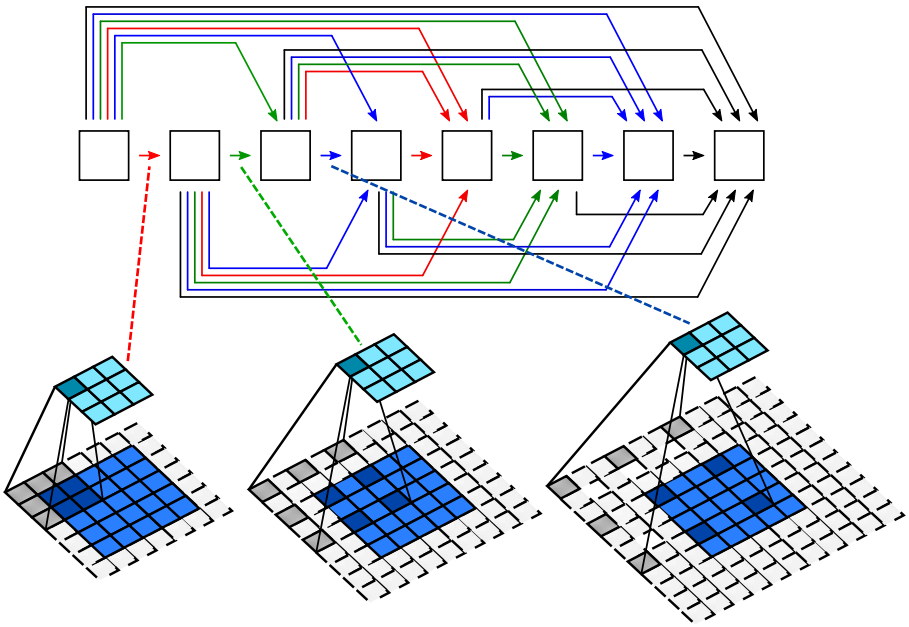


Figure 11.: Architecture d'un réseau MS-D utilisant $L = 7$ couches et un facteur de dilatation $l = 3$.

dilatées pour englober des caractéristiques à différentes échelles d'image et sur l'utilisation de *connexions denses* afin de combiner toutes les cartes caractéristiques entre elles. Par conséquent, il intègre rapidement les informations à longue portée dans les images et atteint une taille de champ réceptif plus importante sans sacrifier la résolution.

L'objectif de ce travail est de développer une approche d'apprentissage profond, dite *end-to-end*, pour la récupération de la phase et de l'absorption à l'aide des réseaux neuronaux MS-D. Nous comparons les reconstructions obtenues avec celles données par l'approche linéaire CTF ainsi qu'un autre réseau bien connu, le U-Net. Les réseaux ont été entraînés sur des données simulées d'objets constitués de combinaisons d'un ou plusieurs matériaux homogènes différents à plusieurs niveaux de signal sur bruit, impliquant une seule ou plusieurs distances de propagation.

Expériences

Génération des données simulées

Nous avons généré des images synthétiques de contraste de phase de rayons X. L'énergie des rayons X a été fixée à 13 keV pour une longueur d'onde correspondante $\lambda = 0.095$ nm, et la taille des pixels dans l'espace objet a été fixée à 6 nm. Nous avons créé des ensembles de données de projections d'objets 3D à partir de combinaisons aléatoires d'une à dix formes, constituées soit d'un matériau homogène (pour créer des objets homogènes), soit de trois matériaux différents (pour créer des objets hétérogènes): or, palladium et zinc. Les formes utilisées étaient des ellipsoïdes et des paraboloïdes avec des positions et des orientations aléatoires. Elles ont été obtenues à partir des projections tomographiques analytiques 2D des parties réelle et imaginaire de l'indice de réfraction, correspondant respectivement à la phase et à l'absorption, à partir de ces objets 3D.

La taille de l'image a été fixée à 2048×2048 pixels et le logiciel *TomoPhantom* a été utilisé pour la génération des objets et des projections. Les images de contraste de phase ont

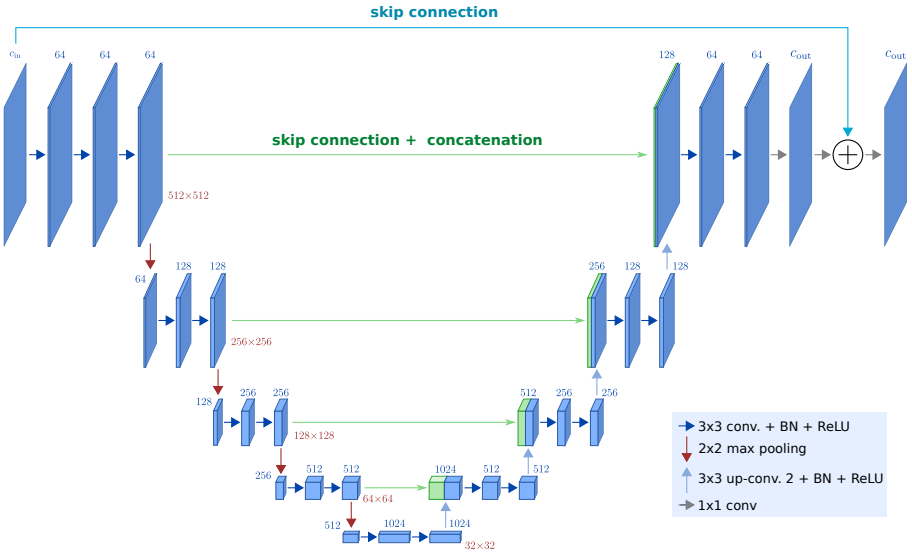


Figure 12.: Architecture du réseau U-Net.

été produites à partir des images de projection à différentes distances de propagation $D = [10.1, 15.5, 17.8, 19, 20.3]$ mm, et sous-échantillonnées à 512×512 pixels pour éviter les effets d'*aliasing* lors des calculs du diagramme de diffraction. Différents niveaux de bruit blanc gaussien ont été ajoutés aux ensembles de données afin d'obtenir un rapport signal/bruit crête à crête spécifique (PPSNR) à la plus grande distance. Le niveau de bruit est resté constant sur toutes les distances, imitant ainsi les conditions expérimentales typiques. Deux ensembles de données ont été générés : l'un avec des objets homogènes et l'autre avec des objets hétérogènes.

Chaque ensemble de données comprenait 12 000 paires de 5 images d'entrée (images de contraste de phase à différentes distances de propagation) et 2 images de sortie (absorption et phase). Pour chaque ensemble de données, 10 000 images ont été utilisées pour l'entraînement, 1 000 pour la validation et 1 000 pour l'évaluation. Une augmentation des données a été effectuée, utilisant des rotations aléatoires de 90 degrés ou des retournements, ce qui a permis d'obtenir un total de 20 000 images d'entraînement.

Différents entraînements

Pour les simulations, nous avons entraîné différents réseaux MS-D (et U-Net). Pour chacune des bases (homogène/hérogène), la sortie désirée, i.e. la vérité terrain, correspond à un couple (absorption,phase), et pour les données en entrées, nous avons considérés différents cas de figure :

1. une seule image d'intensité correspondant à $D = 10.1$ mm
2. plusieurs images d'intensité correspondant aux différentes distances de propagation $D = [10.1, 15.5, 17.8, 19, 20.3]$ mm
3. une estimation initiale (B_0, φ_0) donnée par l'adjoint de la dérivée de Fréchet de $\mathbf{F}_D$ au point $(0, 0)$

$$(B_0, \varphi_0) = \sum_{i=1}^{N_D} \left[\mathbf{F}'_{D_i}(0,0)\right]^* (\mathbf{I}^{\text{obs}}_{D_i}) \tag{0.10}$$

ce dernier cas correspond donc à une méthode post-traitement.

Ce qui fait un total de 6 réseaux MS-D entrainés. Dans l'ensemble, nous avons utilisé la même architecture de réseau MS-D, composée de $L = 100$ couches et d'un noyau de con-

Tableau 2.: Erreur quadratique moyenne normalisée et écart-type (en %) pour les 1 000 images de test, objets homogènes.

	#Distances	#Paramètres	NMSE (en %)	
			Absorption	Phase
CTFHomo	5	-	13.5 (3.92)	13.5 (3.92)
CTFHomo	1	-	21.5 (14.0)	21.5 (14.0)
U-Net	5	31.10^6	4.29 (4.82)	4.29 (4.82)
MS-D Net	5	49.10^3	**3.95 (4.41)**	**3.95 (4.41)**
MS-D Net	1	45.10^3	4.37 (5.50)	4.37 (5.50)

Tableau 3.: Erreur quadratique moyenne normalisée et écart-type (en %) pour les 1 000 images de test, objets hétérogènes.

	#Distances	#Paramètres	NMSE (en %)	
			Absorption	Phase
CTF	5	-	42.5 (19.8)	30.4 (8.99)
U-Net	5	31.10^6	11.1 (12.3)	7.65 (9.35)
MS-D Net	5	49.10^3	**7.67 (10.7)**	**5.33 (6.74)**
MS-D Net	1	45.10^3	11.9 (9.05)	7.76 (6.36)

volution dilaté de taille 3×3. Les facteurs de dilatation ont été sélectionnés dans la liste $[1, 2, \ldots, 10, 1, 2, \ldots, 10, 1, 2 \ldots]$. Les réseaux ont été entraînés à l'aide de l'optimiseur ADAM avec la norme l_2 entre les vérités terrains et les prédictions comme fonction de coût.

Résultats

Le réseau MS-D a été comparé au réseau U-Net ainsi qu'à la méthode analytique CTF (CTFhomo pour la version homogène). La métrique utilisée pour les données simulées est l'erreur quadratique moyenne normalisée (NMSE).

Données simulées

Les résultats obtenus sur les objets homogènes sont résumés dans le Tableau 2. Le réseau MS-D reconstruit correctement l'absorption et la phase comme étant identiques à un facteur multiplicatif près, correspondant au ratio $\frac{\delta_r}{\beta}$. Cela se traduit par une erreur normalisée identique pour les prédictions d'absorption et de phase. En revanche, le U-Net n'a pas été en mesure de reconstruire simultanément l'absorption et la phase. Il ne récupère que la phase tout en mettant l'absorption à zéro. C'est la raison pour laquelle nous n'avons entraîné le U-Net qu'à produire une seule image, la phase, et à considérer l'absorption comme proportionnelle à la phase.

Les résultats obtenus sur les objets hétérogènes sont résumés dans le Tableau 3. Le réseau MS-D est un peu moins performant sur les objets hétérogènes en raison de la diversité de l'ensemble de données, mais les résultats restent très bons. Nous constatons que le réseau récupère un peu mieux la phase que l'absorption.

Enfin, l'utilisation du réseau en post-traitement nous permet de prendre en compte les

Tableau 4.: Erreur quadratique moyenne normalisée et écart-type (en %) pour les 1 000 images de test, initialisation donnée par (0.10).

	Homogène		Hétérogène	
	Absorption	Phase	Absorption	Phase
U-Net	9.60 (13.85)	4.74 (8.85)	16.22 (12.33)	5.77 (7.94)
MS-D Net	**1.96 (3.05)**	**1.96 (3.05)**	**8.19 (6.68)**	**4.83 (5.17)**

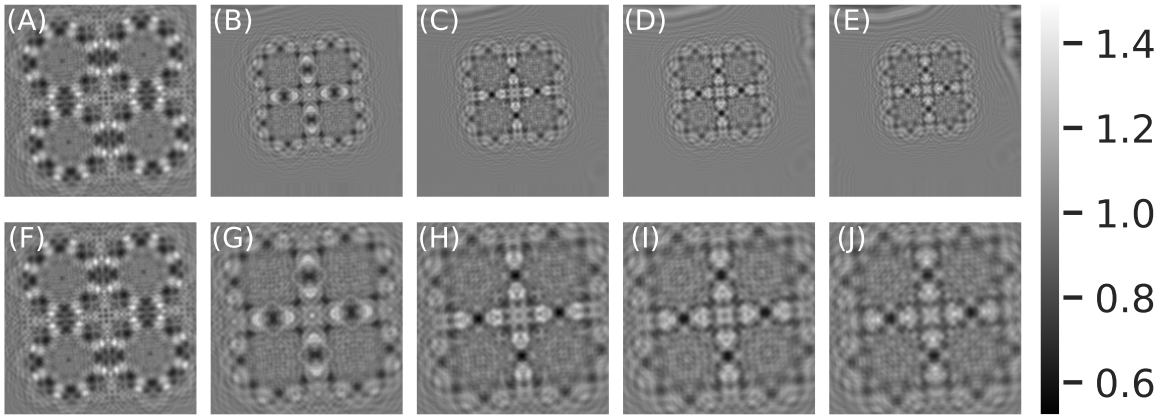


Figure 13.: (A)-(E) Images de contraste de phase acquises à des positions progressivement plus éloignées du foyer (et donc plus proches du détecteur). (F)-(J) Images de contraste de phase agrandies pour avoir la même taille de pixel (6 nm).

connaissances préalables de la physique du problème direct. Dans le Tableau 4, nous voyons que cette initialisation donnée par l'adjoint de la dérivée améliore la qualité de la reconstruction de la phase, mais est un peu moins performante pour l'absorption.

Données expérimentales

Nous avons appliqué les différentes méthodes sur les données expérimentales acquises sur la ligne de faisceau NanoMAX du synchrotron MAX IV (Lund, Suède). L'échantillon est placé à différentes positions par rapport aux positions du foyer et du détecteur pour différents niveaux de grossissement et, par conséquent, différentes distances de propagation effectives correspondant à $D = [10, 1, 15, 5, 17, 8, 19, 20, 3]$ mm. Les différents diagrammes de diffraction n'ont pas été directement utilisés comme données d'entrée, elles ont été agrandies afin d'avoir la même taille de pixel (Figure 13), qui a été mesurée à 6 nm. L'énergie des rayons X a été fixée à 13 keV.

Les réseaux entraînés sur objets hétérogènes donnent les reconstructions représentées dans la Figure 14. Nous constatons que la méthode CTF récupère bien la forme de l'objet mais laisse des artefacts au niveau des basses fréquences. D'autre part, le U-Net semble réduire ces artefacts mais récupère grossièrement les bords. Le réseau MS-D réduit les artefacts tout en reconstruisant bien l'objet, même lorsqu'une seule distance est donnée en entrée.

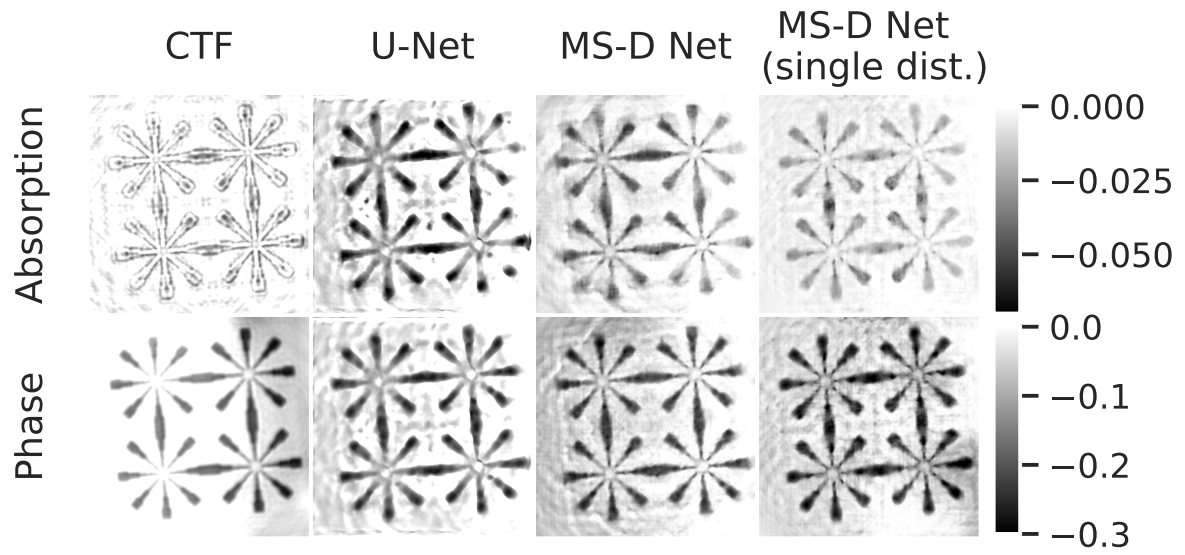


Figure 14.: Comparaison des différentes approches sur les données expérimentales lorsque l'entraînement se fait sur des objets hétérogènes.

Analyse et conclusion

Le réseau MS-D intègre efficacement les informations à courte et à longue portée, un facteur essentiel pour traiter les effets du propagateur de Fresnel. Les comparaisons qui ont été faites démontrent la nette supériorité des approches par apprentissage profond dans l'obtention de meilleurs résultats.

Le réseau MS-D récupère avec précision la phase et l'absorption, qui sont directement liées, pour les objets homogènes. Bien que ses performances soient un peu moins précises pour les objets hétérogènes, il réduit encore de manière significative les erreurs de reconstruction par rapport à l'approche linéaire. Il est important de noter qu'il y parvient sans information a priori entre la phase et l'absorption, en récupérant les deux à partir d'un seul diagramme de diffraction.

Dans l'ensemble, le réseau MS-D excelle dans la récupération de la phase et de l'absorption, surpassant à la fois la méthode linéaire et le U-Net, sur données expérimentales et simulées. Cependant, comme toutes les approches d'apprentissage, son efficacité dépend de la qualité et de la précision des données d'entraînement, qui sont influencées par divers paramètres physiques tels que l'énergie, les distances et la résolution. En outre, la mise en œuvre de la simulation peut introduire ses propres artefacts, y compris ceux provenant de l'échantillonnage et de la précision numérique.

Algorithme primal-dual non linéaire

Nous présentons ici un algorithme primal-dual non linéaire pour le problème de récupération de phase et d'absorption. Cet algorithme permet de régulariser les deux images séparément. Nous montrons que la prise en compte de la non linéarité dans la reconstruction améliore considérablement les résultats par rapport à une version linéarisée. En outre, le choix de régularisateurs distincts permet de prendre en compte les différentes contributions de l'atténuation et de la phase dans la formation de l'image de contraste. La méthode proposée (NL-PDHG) produit des

reconstructions avec moins d'artefacts et une erreur quadratique moyenne normalisée réduite par rapport à sa version linéarisée.

Méthodes

PDHG-CTF

Différentes approches ont été proposées pour récupérer simultanément la phase et l'absorption à partir d'une seule distance, mais aucune d'entre elles n'a proposé l'utilisation d'une régularisation telle que la variation totale (TV) tout en tenant compte de la non linéarité. La variation totale a été introduite pour le débruitage et la reconstruction d'images, et son utilisation a été étendue à plusieurs problèmes d'imagerie tels que la déconvolution, l'agrandissement, l'inpainting et la segmentation. Des travaux antérieurs sur la récupération de phase ont montré que la régularisation TV améliore la qualité des reconstructions de phases constantes par morceaux, mais ils n'ont traité que le cas linéaire. Pour résoudre le problème non linéaire avec la régularisation TV, nous proposons une approche basée sur une méthode primale-duale, nommée *primal-dual hybrid gradient* (PDHG), aussi connu sous le nom d'algorithme de Chambolle-Pock. Cet algorithme nous permet d'utiliser la régularisation TV ainsi que sa généralisation du second ordre (TGV^2). La méthode PDHG a été généralisée afin de prendre en compte les opérateurs non linéaires. Nous dérivons ensuite deux algorithmes, l'un basé sur la linéarisation CTF, que l'on nomme PDHG-CTF et l'autre, dans le cas non linéaire, appelé *nonlinear primal-dual hybrid gradient* (NL-PDHG).

Puisque la contribution de l'atténuation et de la phase à l'image de contraste de phase est différente, il peut être plus intéressant d'utiliser une régularisation différente pour B et φ. En choisissant les paramètres de pondération comme étant $\alpha, \beta, \nu > 0$, la formulation variationnelle revient alors à résoudre le problème de minimisation suivant :

$$\min_{\substack{B,\varphi \\ B>0,\varphi>0}} \left\{ \left\| \mathrm{F}_D^{\mathrm{CTF}}(B,\varphi) - \mathrm{I}_D^{\mathrm{obs}} \right\|_2^2 + \mathrm{TGV}^2_{(\alpha,\beta)}(B) + \nu \mathrm{TV}(\varphi) \right\} \tag{0.11}$$

où $\mathrm{I}_D^{\mathrm{obs}}$ est l'intensité mesurée à une distance de propagation égale à D. La variation totale est définie par

$$\mathrm{TV}(\varphi) = \left\| (\nabla\varphi)_x \right\|_1 + \left\| (\nabla\varphi)_y \right\|_1 \tag{0.12}$$

où ∇ est l'opérateur gradient discret, et la régularisation TGV^2 peut être exprimée de la manière suivante:

$$\mathrm{TGV}^2_{(\alpha,\beta)}(B) = \min_{\mathrm{v}} \left\{ \alpha \left\| \mathfrak{D}\mathrm{v} \right\|_1 + \beta \left\| \nabla B - \mathrm{v} \right\|_1 \right\} \tag{0.13}$$

où $\mathrm{v} = (v_1, v_2)$ est une variable auxiliaire et $\mathfrak{D}(\mathrm{v}) = \frac{\nabla v_1 + \nabla v_2}{2}$.

On peut voir comment le paramètre α oblige v à avoir un gradient parcimonieux et le paramètre β pénalise le gradient ∇B pour qu'il ne s'écarte que sur un ensemble parcimonieux de v. Le problème de minimisation (0.11) peut s'écrire sous la forme:

$$\min_{B,\varphi,\mathrm{v}} \left\{ \mathcal{H}\left[\mathcal{K}_{\mathrm{CTF}}(B,\varphi,\mathrm{v}) \right] + \mathcal{G}(B,\varphi,\mathrm{v}) \right\} \tag{0.14}$$

avec

$$\mathcal{K}_{\mathrm{CTF}}(B,\varphi,\mathrm{v}) = \left[\mathrm{F}_D^{\mathrm{CTF}}(B,\varphi), \mathfrak{D}(\mathrm{v}), \nabla B - \mathrm{v}, \nabla\varphi \right] \tag{0.15}$$

$$\mathcal{H}\left(h^1, h^2, h^3, h^4\right) = \left\| h^1 - \mathrm{I}_D^{\mathrm{obs}} \right\|_2^2 + \alpha \left\| h^2 \right\|_1 + \beta \left\| h^3 \right\|_1 + \nu \left\| h^4 \right\|_1 \tag{0.16}$$

$$\mathcal{G}(B, \varphi, \mathrm{v}) = \iota_+(B, \varphi) = \begin{cases} 0 & \text{si } B, \varphi > 0 \\ +\infty & \text{sinon} \end{cases} \tag{0.17}$$

Ici, $\mathcal{K}_{\mathrm{CTF}}$ est un opérateur linéaire et ι_+ est une fonction indicatrice qui contraint B et φ à être strictement positifs. En utilisant cette formulation, nous définissons l'algorithme PDHG-CTF (Alg. 1), qui itère sur le triplet $x_i = (B_i, \varphi_i, \mathrm{v}_i)$, où B_i et φ_i représentent l'absorption et le déphasage recherchés, et $\mathrm{v_i}$ est la variable auxiliaire qui apparaît dans la formule de TGV, à la i-ième itération. Ici, τ et σ sont les pas de descente dans l'espace primal et dual, respectivement, $\mathcal{K}^*$ désigne l'opérateur adjoint de $\mathcal{K}$, $\mathrm{prox}_{\tau\mathcal{G}}$ l'opérateur proximal de $\tau\mathcal{G}$ et $\mathcal{H}^*$ le conjugué de $\mathcal{H}$. Pour assurer la convergence de l'algorithme PDHG-CTF, il suffit de choisir des pas de descente

Algorithm 1 PDHG-CTF

Étant donnés :

- les pas de descente σ, τ tels que $\sigma\tau|||\mathcal{K}_{\mathrm{CTF}}|||^2 < 1$ et un paramètre de relaxation $\theta \in [0, 1]$
- les paramètres de régularisation α, β and ν
- l'initialisation $x_0 = \{B_0, \varphi_0, v_0\} \in X$ (primal) et $h_0 = [h_0^1, h_0^2, h_0^3, h_0^4] \in Y$ (dual)

for $k = 0, \ldots, N\mathrm{iter}$ **do** :

$h_{k+1} \leftarrow \mathrm{prox}_{\sigma\mathcal{H}^*}(h_k + \sigma\mathcal{K}_{\mathrm{CTF}}(\bar{x}_k))$

$x_{k+1} \leftarrow \mathrm{prox}_{\tau\mathcal{G}}(x_k - \tau\mathcal{K}_{\mathrm{CTF}}{}^*(h_{k+1}))$

$\bar{x}_{k+1} \leftarrow x_{k+1} + \theta(x_{k+1} - x_k)$

qui satisfont l'inégalité

$$\sigma\tau|||\mathcal{K}_{\mathrm{CTF}}|||^2 < 1$$

où $|||\mathcal{K}_{\mathrm{CTF}}||| = \sup_{\|x\| \leq 1} \{||\mathcal{K}_{\mathrm{CTF}} x||\}$ est la norme de l'opérateur $\mathcal{K}_{\mathrm{CTF}}$.

NL-PDHG

Conçu à l'origine pour les opérateurs linéaires, l'algorithme PDHG a été généralisé aux cas non linéaires. En remplaçant l'opérateur linéaire $\mathrm{F}_D^{\mathrm{CTF}}$ par l'opérateur non linéaire F_D dans le problème (0.11), nous obtenons un nouveau problème de minimisation que nous pouvons résoudre avec la méthode NL-PDHG. Le seul changement par rapport à PDHG-CTF est que l'opérateur $\mathcal{K}_{\mathrm{CTF}}$ est remplacé par l'opérateur non linéaire

$$\mathcal{K}_{\mathrm{NL}}(B, \varphi, \mathrm{v}) = [\mathrm{F}_D(B, \varphi), \mathfrak{D}(\mathrm{v}), \nabla B - \mathrm{v}, \nabla\varphi] \tag{0.18}$$

Dans ce cas $\mathcal{K}^*_{\mathrm{CTF}}$ doit être remplacé par $\left[\mathcal{K}'_{\mathrm{NL}}(x_i)\right]^*$ où $\mathcal{K}'_{\mathrm{NL}}(x_i)$ est la dérivée de Fréchet au point x_i, pour laquelle nous avons une formule explicite. Pour assurer la convergence de l'algorithme, le gradient de $\mathcal{K}_{\mathrm{NL}}$ doit être Lipschitz dans un voisinage d'une solution, l'estimation initiale doit être suffisamment proche d'une solution et les pas de descente doivent satisfaire les inégalités locales:

$$\sigma_k\tau_k \sup_{n=0,1,\ldots,k} \left\{|||\mathcal{K}'_{\mathrm{NL}}(x_n)|||^2\right\} < 1$$

pour tout k.

Expériences

À titre de comparaison, nous avons implémenté une descente de gradient projetée, régularisée avec une variation totale lisse (TV^ϵ), appelée GD-TV^ϵ, où ϵ est un facteur de lissage. Les paramètres des différentes méthodes ont été optimisés par balayage dans un large intervalle afin d'obtenir une forte diminution du terme d'attache aux données. Pour GD-TV^ϵ, le facteur de

Algorithm 2 NL-PDHG

Étant donnés :

- les pas de descente $\sigma_0, \tau_0 > 0$ et un paramètre de relaxation $\theta \in [0, 1]$
- les paramètres de régularisation α, β and ν
- l'initialisation $x_0 = \{B_0, \varphi_0, v_0\} \in X$ (primal) et $h_0 = [h_0^1, h_0^2, h_0^3, h_0^4] \in Y$ (dual)

for $k = 0, \ldots, N_{\text{iter}}$ **do** :

$h_{k+1} \leftarrow \text{prox}_{\sigma \mathcal{H}^*}\left(h_k + \sigma_k \mathcal{K}_{\text{NL}} \bar{x}_k\right)$

$x_{k+1} \leftarrow \text{prox}_{\tau \mathcal{G}}\left(x_k - \tau_k \left[\mathcal{K}'_{\text{NL}}(x_k)\right]^* h_{k+1}\right)$

$\bar{x}_{k+1} \leftarrow x_{k+1} + \theta\left(x_{k+1} - x_k\right)$

$\sigma_{k+1}, \tau_{k+1} \leftarrow \sigma_k, \tau_k$ tel que $\sigma_k \tau_k \sup_{n=0,1,\ldots,k} \left\{|||\mathcal{K}'_{\text{NL}}(x_n)|||^2\right\} < 1$

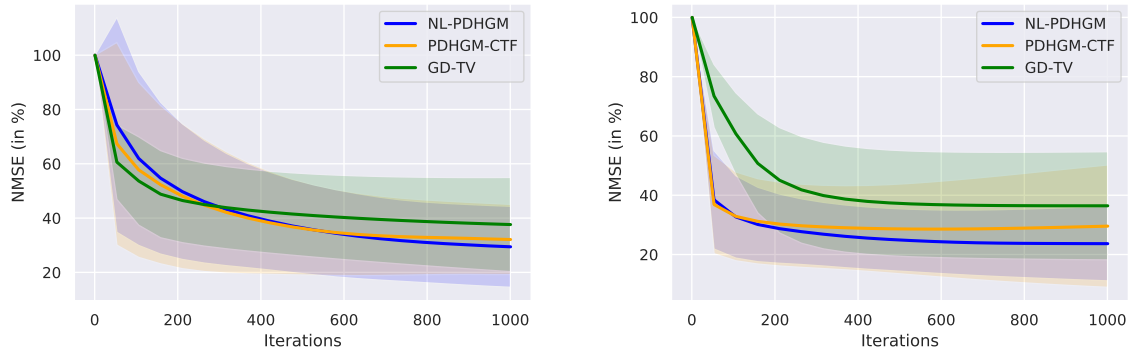


Figure 15.: Évolution de la NMSE moyenne (en %) pour les 1 000 images de test. Les zones transparentes correspondent à l'écart-type.

lissage a été fixé à 10^{-3}, avec des paramètres de pondération différents pour B et φ, ayant pour valeurs 10^{-1} et 10^{-3}, respectivement. Pour PDHG-CTF et NL-PDHG, nous avons utilisé le même ensemble de paramètres : $\alpha = 10^{-2}, \beta = 5 \times 10^{-3}, \nu = 10^{-2}$ et le paramètre de relaxation $\theta = 1$. Un critère d'arrêt basé sur l'écart de pseudo-dualité a été proposé (Valkonen 2014), mais son calcul est coûteux. Par conséquent, nous avons utilisé un nombre fixe d'itérations ($N_{\text{iter}} = 1\,000$) ce qui était suffisant pour atteindre la convergence (Fig. 15).

Et même si en théorie, il faut s'assurer que l'initialisation soit suffisamment proche d'une solution pour assurer la convergence, en pratique, l'algorithme converge en initialisant avec $(B_0, \varphi_0) = (0, 0)$. Dans la suite, l'algorithme est toujours initialisé à zéro. Pour GD-TV$^\epsilon$, la convergence n'a pas été analysée en détail, mais un pas de descente fixe suffisamment petit égal à 0.01 était suffisant pour obtenir la convergence en pratique. Pour PDHG-CTF, les pas de descente ont été fixés à $\sigma = \tau = 0,99||\mathcal{K}_{\text{CTF}}||^{-1}$, ce qui a permis d'assurer la convergence. Et pour NL-PDHG, les pas de descente ont été définis par $\sigma_i = \tau_i = 0.99 L_i$, où $L_i = \sup_{k=0,1,\ldots,i} \left\{|||\mathcal{K}'_{\text{NL}}(x_k)||^2\right\}$, et ils ont été mis à jour toutes les 50 itérations, afin de satisfaire les inégalités locales. En moyenne, il a fallu environ 104 s pour GD-TV$^\epsilon$ et 120 s pour les méthodes PDHG-CTF et NL-PDHG pour faire 1 000 itérations.

Tableau 5.: Moyennes des métriques NMSE, SSIM, PSNR et écart-type (en %) pour les 1 000 images de test en utilisant différentes stratégies de régularisation.

	Régularisation		**NMSE (en %)**		**SSIM (en %)**		**PSNR**	
	Absorption	Phase	Absorption	Phase	Absorption	Phase	Absorption	Phase
GD-TV$^\epsilon$	TV$^\epsilon$	TV$^\epsilon$	37.5 (17.4)	36.4 (18.2)	99.6 (0.440)	95.2 (6.95)	65.2 (9.43)	50.5 (11.1)
PDHG-CTF	TGV^2	TV	32.1 (12.9)	29.6 (20.9)	**99.8 (0.337)**	92.9 (7.87)	68.2 (9.10)	52.6 (8.19)
NL-PDHG	TGV^2	TV	**29.2 (14.8)**	**23.6 (12.6)**	**99.8 (0.237)**	**97.2 (3.12)**	**68.7. (8.63)**	**53.0 (6.40)**
NL-PDHG	TV	TV	41.3 (23.9)	25.6 (14.1)	99.7 (0.371)	94.8 (5.10)	65.0 (9.72)	52.8 (7.53)
NL-PDHG	TGV^2	TGV^2	32.4 (19.6)	29.3 (14.8)	**99.8 (0.249)**	92.3 (6.58)	66.8 (8.00)	51.4 (6.51)

Résultats et discussion

Pour l'évaluation des algorithmes, nous avons générés un ensemble d'images test (comme ce qui a été décrit plus haut). Un ensemble d'objets 3D a été généré en créant des combinaisons aléatoires d'un à dix ellipsoïdes ou paraboloïdes composés de trois matériaux différents : or (Au), palladium (Pd) et zinc (Zn). Des projections 2D de taille 512×512 pixels ont été générées, produisant des images de phase et d'absorption, qui ont ensuite été utilisées pour générer des images de contraste de phase. L'énergie des rayons X a été fixée à 13 keV pour une longueur d'onde de $\lambda = 0,095$ nm, la distance de propagation $D = 20.3$ mm et la taille des pixels à 12 nm. Pour mesurer quantitativement la qualité de la reconstruction et quantifier l'incertitude des reconstructions, nous avons généré un ensemble de données de test comprenant 1 000 images.

Afin d'évaluer la qualité de la reconstruction obtenue par les différentes méthodes, nous avons utiliser différentes métriques. L'erreur quadratique moyenne normalisée (NMSE), l'indice de similarité structurelle (SSIM) et le rapport signal sur bruit (PSNR). Les moyennes de ces métriques pour l'ensemble de données de test sont résumées dans le Tableau 5. Les approches primales-duales sont plus performantes que la descente de gradient, et globalement la méthode NL-PDHG permet d'obtenir les meilleures reconstructions. Nous pouvons voir que l'application de la même régularisation à B et φ donne de moins bonnes reconstructions, et que l'absorption est mieux récupérée avec une régularisation TGV^2 alors que la phase a de meilleurs résultats avec la régularisation TV. Cela peut s'expliquer par le fait que dans nos expériences, l'absorption récupérée a tendance à avoir des parties manquantes qui, lorsqu'elles sont reconstruites avec la régularisation TV, laissent des *effets d'escalier*. La régularisation TGV^2 pour B offre les meilleures performances, fournissant un bon compromis entre la régularité et des frontières nettes. Pour φ, on ne peut s'attendre à une meilleure reconstruction en prenant en compte le second ordre. De plus, si les objets considérés ont une atténuation suffisamment faible et une phase variant peu, i.e. qui satisfont les conditions de CTF, alors la méthode PDHG-CTF converge plus rapidement. Elle nécessite une centaine d'itérations pour obtenir la même qualité de reconstruction en moyenne que la version non linéaire, qui nécessite un millier d'itérations.

Sur données expérimentales (Figure 16), l'algorithme GD-TV$^\epsilon$ donne une reconstruction de moins bonne qualité que les autres méthodes. Des parties de l'objet sont manquantes, peut-être en raison de la projection sur les valeurs positives et de l'approximation de la variation totale. Les deux approches primales-duales semblent bien récupérer l'absorption, tout en réduisant les franges, et les phases récupérées sont presque exemptées d'artefacts. Il convient de noter qu'ici, l'objet n'est pas très absorbant puisqu'il est assez fin. Par conséquent, la méthode linéarisée donne également une très bonne reconstruction. Et comme l'objet est constant par morceaux, les régularisations de type TV et TGV sont bien adaptées.

Pour résumer, nous avons proposé d'utiliser différentes régularisations pour l'absorption et la

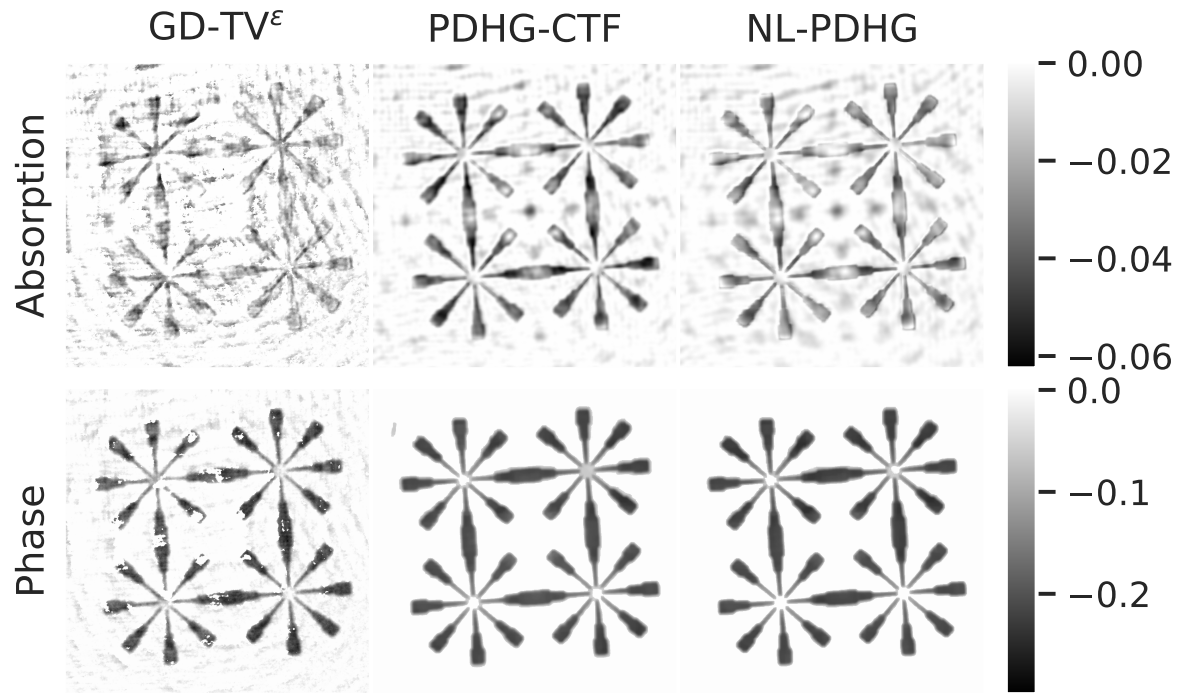


Figure 16.: Reconstructions obtenues sur donnée expérimentale.

phase. La sélection de régularisations qui tiennent compte de la manière dont l'atténuation et la phase contribuent à l'image de contraste de phase a permis d'améliorer les reconstructions. Nous avons observé la contribution significative de l'information non linéaire du problème, ce qui suggère que l'algorithme NL-PDHG pourrait être appliqué dans une grande variété de cas où les méthodes linéaires échoueraient. L'un des inconvénients de ces algorithmes est que trois paramètres de régularisation doivent être choisis en amont. Le choix de ces paramètres s'est toutefois avéré robuste lors de l'application à un large ensemble d'images. La méthode pourrait être utilisée pour des échantillons dont le rapport phase/absorption est plus élevé ou dont les structures sont plus complexes en ajustant les paramètres et en augmentant le nombre d'itérations. Cependant, des recherches plus approfondies devraient être menées pour des objets non parcimonieux, tels que ceux rencontrés dans l'imagerie des tissus mous biologiques. Le cas des rayons X provenant d'un environnement de laboratoire peut être étudié en considérant la divergence de Kullback-Leibler au lieu de la norme L^2. Une extension directe de cette approche serait d'appliquer les algorithmes proposés à la tomographie de contraste de phase, en particulier lorsqu'il n'y a pas d'hypothèse de multi-matériaux, en considérant la version 3D de la TV.

Deep Gauss-Newton pour la récupération de phase

Dans ce chapitre, nous proposons l'algorithme Deep Gauss-Newton (DGN), qui intègre la connaissance du modèle direct dans un réseau neuronal profond en déroulant un schéma de Gauss-Newton régularisé. DGN ne nécessite ni de sélection manuelle des paramètres, ni d'une bonne estimation initiale. Nous étendons également cette approche en créant l'architecture Deep Proximal Gauss-Newton (DPGN), qui permet d'améliorer la qualité de la reconstruction en termes de résolution et d'erreur tout en conservant un nombre de paramètres à peu près égal.

Déroulement d'un schéma de type Gauss-Newton

Pour simplifier, nous notons $f = (B, \varphi)$ le couple que nous voulons reconstruire. Pour ce faire, nous pouvons utiliser une approche variationnelle telle que la méthode itérative de Gauss-Newton régularisée (IRGN). La méthode IRGN consiste à résoudre, à chaque itération, le problème de minimisation

$$f_{k+1} = \underset{f}{\operatorname{argmin}} \left\{ \left\| \mathrm{F}_D(f_k) + \mathrm{F}'_D(f_k)\,(f - f_k) - \mathbf{I}_D^{\mathrm{obs}} \right\|_2^2 + \alpha_k \, ||f||_2^2 \right\} \tag{0.19}$$

où $\mathrm{F}'_D(f_k)$ désigne la dérivée de Fréchet de l'opérateur direct au point f_k, $\alpha_k > 0$ est un paramètre de régularisation à l'itération k, et $\mathbf{I}_D^{\mathrm{obs}}$ est l'intensité mesurée (bruitée). Ce problème de minimisation a toujours une unique solution qui est donnée par

$$f_{k+1} = f_k + \left[\mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k) + \alpha_k \mathrm{Id} \right]^{-1} \left\{ \mathrm{F}'_D(f_k)^* \left[\mathbf{I}_D^{\mathrm{obs}} - \mathrm{F}_D(f_k) \right] - \alpha_k f_k \right\} \tag{0.20}$$

où $\mathrm{F}'_D(f_k)^*$ est l'adjoint de l'opérateur linéaire $\mathrm{F}'_D(f_k)$ et Id est l'identité. Nous proposons ici d'apprendre une régularisation optimale en remplaçant le terme $\alpha_k f_k$ par un réseau de neurones convolutionnels (CNN) $\mathrm{G}_{\theta_k^g}$, dont les paramètres sont notés θ_k^g. Nous souhaitons aussi approcher l'inverse de l'opérateur

$$H(f_k) = \mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k) + \alpha_k \mathrm{Id}$$

avec un autre CNN $\mathrm{H}_{\theta_k^h}$ ayant pour paramètres θ_k^h, qui dépend de l'itéré actuelle f_k et de l'approximation de la Hessienne $\mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k)$. Le réseau $\mathrm{H}_{\theta_k^h}$ remplace alors l'approximation classique de l'inverse de la Hessienne (utilisée dans le schéma traditionnel de Gauss-Newton) par une approximation apprise potentiellement meilleure et plus rapide.

Si l'algorithme est arrêté après N itérations, on obtient

$$f_N = \left(\Gamma_{\theta_N}^{\mathrm{DGN}} \circ \cdots \circ \Gamma_{\theta_1}^{\mathrm{DGN}} \right) (f_0, \mathbf{I}_D^{\mathrm{obs}}) \tag{0.21}$$

où f_0 est l'estimation initiale, $\theta_k = \left(\theta_k^g, \theta_k^h \right)$ et chaque itération est donnée par

$$\Gamma_{\theta_k}^{\mathrm{DGN}}(f_k, \mathbf{I}_D^{\mathrm{obs}}) = f_k + \mathrm{H}_{\theta_k^h} \left[f_k, \mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k) \left\{ \mathrm{F}'_D(f_k)^* \left[\mathbf{I}_D^{\mathrm{obs}} - \mathrm{F}_D(f_k) \right] + \mathrm{G}_{\theta_k^g}(f_k) \right\} \right] \tag{0.22}$$

En déroulant ce schéma, nous pouvons considérer l'opérateur

$$\Lambda_\Theta^{\mathrm{DGN}} = \Gamma_{\theta_N}^{\mathrm{DGN}} \circ \cdots \circ \Gamma_{\theta_1}^{\mathrm{DGN}}$$

comme un réseau neuronal profond représentant N itérations avec $\Theta = (\theta_1, \ldots, \theta_N)$ les paramètres de ce réseau. On va considérer ici que la transformation est la même à chaque itération, c'est-à-dire que $\theta_k = \theta$ pour tout $k \in \{1, \ldots, N\}$ Dans ce cas, $\Lambda_\Theta^{\mathrm{DGN}}$ peut être considéré comme un réseau de neurones récurrent.

Plus récemment, une extension de la méthode de IRGN, la méthode de Gauss-Newton proximale, a été introduite et elle incorpore un terme proximal pour améliorer la convergence et gérer certains types de contraintes ou de régularisation. L'algorithme met à jour la solution de manière itérative en effectuant une étape de Gauss-Newton suivie d'une étape de régularisation

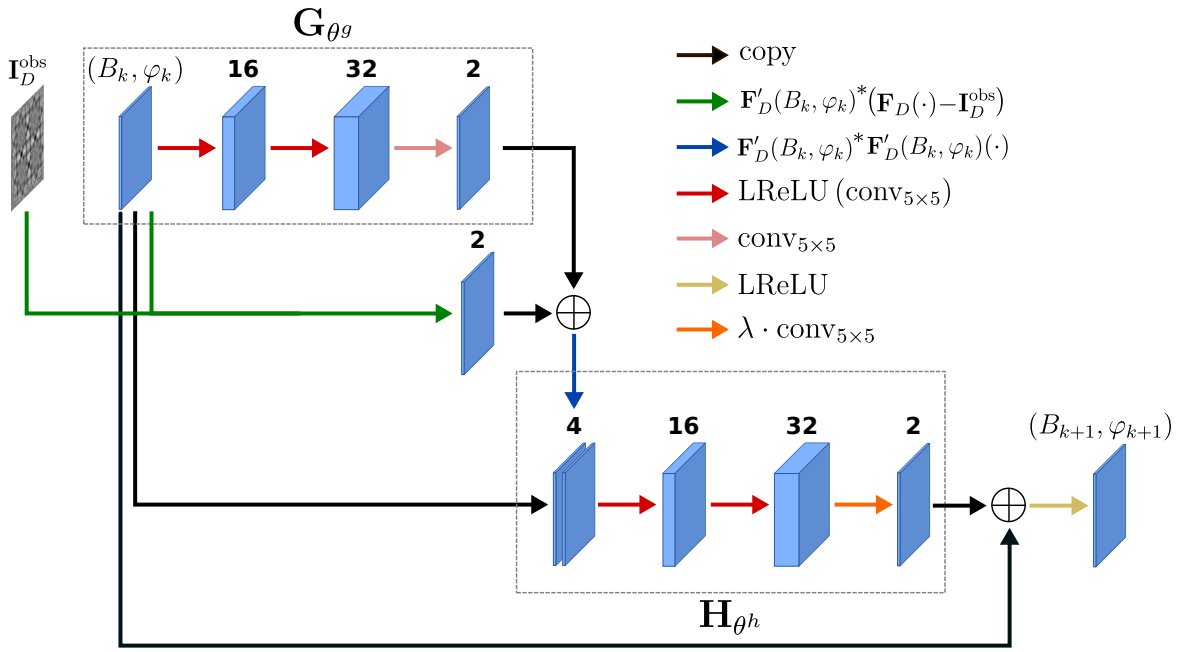


Figure 17.: Architecture du réseau $\Gamma_\theta^{\text{DGN}}$, représentant une itération de la méthode Deep Gauss-Newton.

proximale. Cela revient à ajouter un opérateur de proximité normalisé :

$$f_{k+1} = \text{prox}_J^{H(f_k)}\left(f_k + H(f_k)^{-1}\left\{\mathrm{F}'_D(f_k)^*\left[\mathbf{I}_D^{\text{obs}} - \mathrm{F}_D(f_k)\right] - \alpha_k f_k\right\}\right) \tag{0.23}$$

Ici $\text{prox}_J^{H(f_k)}$ est l'opérateur de proximité associé à une certaine fonction J et normalisé par $H(f_k)$:

$$\text{prox}_J^H(z) = \inf_x\left\{J(x) + \frac{1}{2}\|x - z\|_H^2\right\} \tag{0.24}$$

où la norme $\|\cdot\|_H$ est induite par le produit scalaire $\langle x|z\rangle_H = \langle x|Hz\rangle$. L'idée de la méthode Deep Proximal Gauss-Newton (DPGN) est exactement la même que DGN, à l'exception que l'on ajoute un autre CNN pour remplacer l'opérateur proximal normalise. Pour une comparaison équitable, l'architecture de DPGN diffère légèrement de celle du DGN afin d'avoir approximativement le même nombre de paramètres. Le réseau G_{θ^g} est simplifié et consiste simplement en 16 filtres convolutifs suivis de 2 filtres convolutifs. Le réseau H_{θ^h} est le même que dans DGN. Mais la différence majeure ici est que, après que la sortie du réseau H_{θ^h} soit ajoutée à l'itération courante, nous appliquons le réseau J_{θ^j}. Les architectures des réseaux $\Gamma_\theta^{\text{DGN}}$ et $\Gamma_\theta^{\text{DPGN}}$ utilisées pour chaque itération sont représentées sur les Figures 17 et 18, respectivement.

Résultats

Les données d'entraînement et de validation ont été générées dans les mêmes conditions que celles décrites précédemment pour l'algorithme NL-PDHG. L'ensemble de données se compose de 11 000 paires, dont 10 000 paires ont été utilisées pour l'entraînement et 1 000 pour la validation pendant l'entraînement. Les 1 000 données utilisées pour l'évaluation dans le chapitre précédent constituent notre ensemble de données de test.

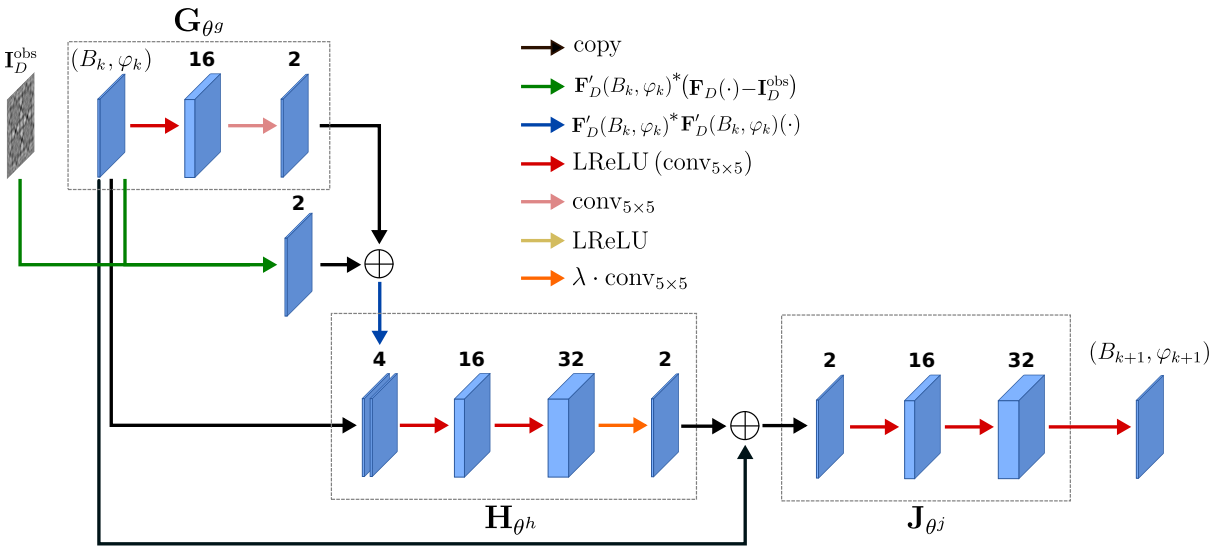


Figure 18.: Architecture du réseau $\Gamma_\theta^{\text{DPGN}}$, représentant une itération de la méthode Deep Proximal Gauss-Newton.

Entraînement

Nous comparons ces approches déroulées à la méthode IRGN classique ainsi qu'au réseau MS-D Net. Pour les méthodes DGN et DPGN, nous avons utilisé un nombre $N = 10$ d'itérations déroulées, ce qui signifie que la dérivée $\mathrm{F}'_D(f_k)$ et son adjoint $\mathrm{F}'_D(f_k)^*$ sont évalués 10 fois. Une bonne estimation initiale f_0 permet de réduire le temps d'entraînement mais ne donne pas de meilleurs résultats finaux. Par conséquent, pour des raisons de simplicité, nous avons initialisé par zéro, $f_0 = (0, 0)$.

Métriques d'évaluation

En plus de la NMSE, la résolution spatiale a été évaluée à l'aide de la FRC (*Fourier Ring Correlation*). La FRC consiste à comparer deux images en calculant la corrélation entre leurs transformées de Fourier, en particulier à l'intérieur des cercles dans le domaine des fréquences:

$$\mathrm{FRC}_{\{f,f_{\text{true}}\}}(R_i) = \frac{\sum_{r\in C(R_i)} \widehat{f^*}(r)\widehat{f_{\text{true}}}(r)}{\sqrt{\left(\sum_{r\in C(R_i)} \left|\widehat{f}(r)\right|^2\right)\left(\sum_{r\in C(R_i)} \left|\widehat{f_{\text{true}}}(r)\right|^2\right)}} \tag{0.25}$$

où R_i est le rayon du cercle $C(R_i)$ dans le domaine de Fourier dans lequel la corrélation est calculée, f^* est le conjugué de f et $\widehat{f}$ est la transformée de Fourier de f. Le calcul de la FRC génère un graphique appelé *courbe FRC*, qui représente les valeurs de corrélation en fonction de la fréquence spatiale (voir Figure 19). À partir de là, nous pouvons construire une métrique pour mesurer l'information que nous sommes capables de reconstruire à un certain niveau de fréquence. La *résolution* ρ de l'image reconstruite peut alors être définie comme suit

$$\rho(f, f_{\text{true}}) = \left(R_{\{\mathrm{FRC}_{\{f,f_{\text{true}}\}}(R)\leq\tau(R)\}}\right)^{-1} \tag{0.26}$$

Tableau 6.: Moyenne (écart-type) sur l'ensemble de données de test.

Méthode	NMSE (en %)		FRCM (en %)		Resolution (en nm)		Time (en s)
	Absorption	Phase	Absorption	Phase	Absorption	Phase	
IRGN	85.5 (40.7)	39.3 (15.0)	71.2 (9.95)	68.1 (5.45)	238 (136)	154 (43)	116
MS-D Net	13.6 (12.8)	10.6 (10.8)	48.8 (13.8)	47.8 (13.3)	102 (77.4)	98.5 (135)	**2.60**
DGN	12.1 (13.5)	4.61 (6.20)	35.7 (15.7)	23.0 (16.6)	72.2 (55.2)	62.3 (37.0)	5.88
DPRGN	**11.0 (12.3)**	**4.05 (5.25)**	**31.5 (13.9)**	**19.8 (15.8)**	**74.0 (63.4)**	**59.1 (28.3)**	6.31

où $R_{\{\mathrm{FRC}_{\{f,f_{\mathrm{true}}\}}(R)(R)\leq\tau(R)\}}$ est le rayon pour lequel le FRC est inférieur à un seuil τ. Pour le seuil, nous avons utilisé le seuil 2σ :

$$\tau(R) = \frac{2}{\sqrt{\frac{N_p(R)}{2}}} \tag{0.27}$$

avec R le rayon dans le domaine de Fourier et $N_p(R)$ le nombre de pixels contenus dans le cercle correspondant. À partir de la courbe FRC, nous calculons également la FRCM (*Fourier Ring Correlation Metric*), qui est la différence quadratique moyenne entre la FRC et l'unité sur toutes les fréquences spatiales :

$$\mathrm{FRCM}(f, f_{\mathrm{true}}) = \sum_i \left(1 - \mathrm{FRC}_{\{f,f_{\mathrm{true}}\}}(R_i)\right)^2 \tag{0.28}$$

Une petite FRCM implique une plus grande similarité dans le domaine de Fourier.

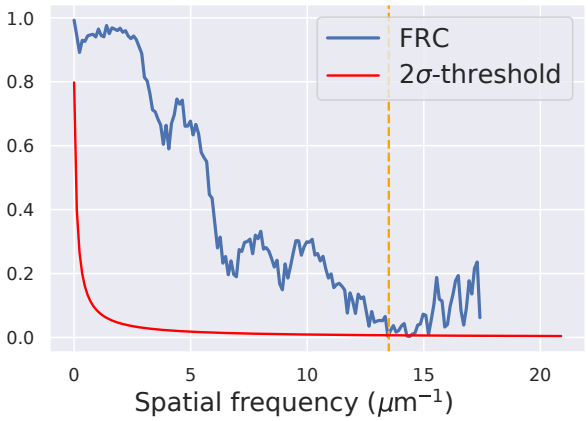
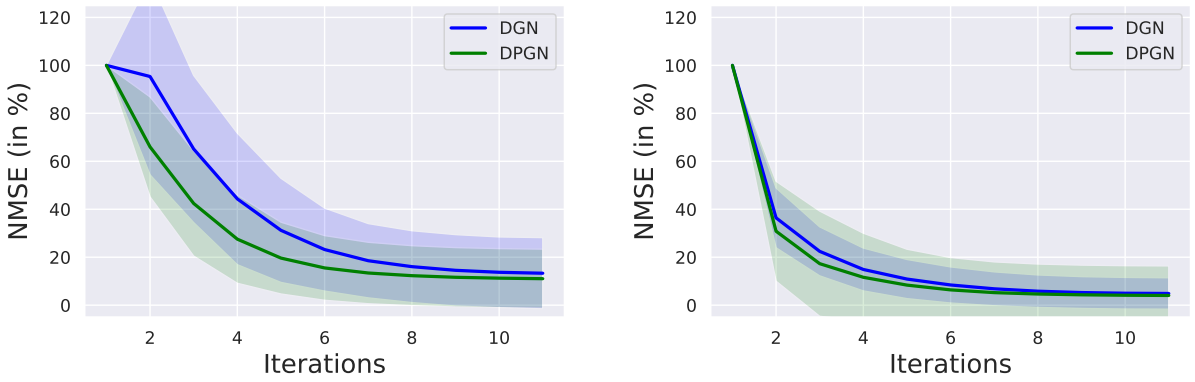


Figure 19.: Évaluation de la résolution à l'aide de la FRC. La résolution estimée par le critère du seuil 2σ est de 75 nm.

Données simulées

Les moyennes et les écart-types des métriques ont été calculés sur l'ensemble de données de test. Le temps de calcul moyen pour une reconstruction a été mesuré pour comparer le temps de reconstruction. Les résultats sont résumés dans le Tableau 6. En moyenne, les approches basées sur l'apprentissage profond surpassent la méthode IRGN en termes d'erreurs et de résolution. Pour la NMSE, les réseaux ont donné des résultats similaires pour l'absorption, mais DGN a obtenu de meilleurs résultats que MS-D Net pour la récupération de la phase. En



((a)) Absorption ((b)) Phase

Figure 20.: Évolution de la NMSE moyenne (%) des approches déroulées pour l'ensemble test. Les zones transparentes correspondent à l'écart-type.

outre, DGN a obtenu une meilleure résolution et une meilleure corrélation dans le domaine des fréquences. DPGN a encore amélioré les reconstructions obtenues par DGN en termes d'erreurs et de résolution. MS-D Net a été le plus rapide car il ne nécessite qu'une seule application du réseau contrairement aux autres méthodes qui consistent à itérer. Les approches déroulées ont été efficaces, 20 fois plus rapides que leur équivalent itératif standard, malgré la nécessité de multiples calculs de dérivée de l'opérateur direct. Les graphiques de la Figure 20 affichent l'évolution moyenne de la NMSE sur l'ensemble de données de test au fil des itérations pour DGN et DPGN. Ces approches semblent converger après 10 itérations, comme elles ont été entraînées à le faire. Pour la phase, la plus grande amélioration est obtenue dès la première étape, alors que l'absorption nécessite quelques itérations supplémentaires. Ces deux méthodes ont presque convergé après 5 itérations, mais les courbes continuent à décroître au delà.

Donnée expérimentale

Les reconstructions de la Figure 21 démontrent que DGN et DPGN fournissent des reconstructions de haute qualité avec un minimum d'artefacts visibles. Cependant, le réseau MS-D, bien que performant sur les données simulées, ne se généralise pas aussi bien aux données expérimentales en raison de sa stratégie d'apprentissage. Les méthodes déroulées prennent explicitement en compte le modèle physique, apprennent des statistiques de bruit à partir des données pour une régularisation adaptée, et tirent parti des propriétés de convergence des méthodes d'optimisation. En revanche, MS-D Net est entraîné à reconstruire directement à partir des mesures sans connaissance du modèle physique.

Conclusion

En intégrant des réseaux neuronaux convolutionnels dans un schéma de type Gauss-Newton régularisé, les méthodes DGN et DPGN permettent de surmonter certaines limites des approches itératives classiques tout en exploitant la puissance des réseaux neuronaux. Ces approches déroulées apprennent des termes de régularisation optimaux pour l'absorption et la phase, améliorant la qualité des reconstructions. L'incorporation de la connaissance du modèle direct dans un réseau simple améliore les reconstructions et permet une bonne généralisation aux données réelles.

Comparés à la méthode IRGN classique, ces méthodes améliorent considérablement les reconstructions et réduisent le temps de calcul. Le choix du schéma d'optimisation déroulé

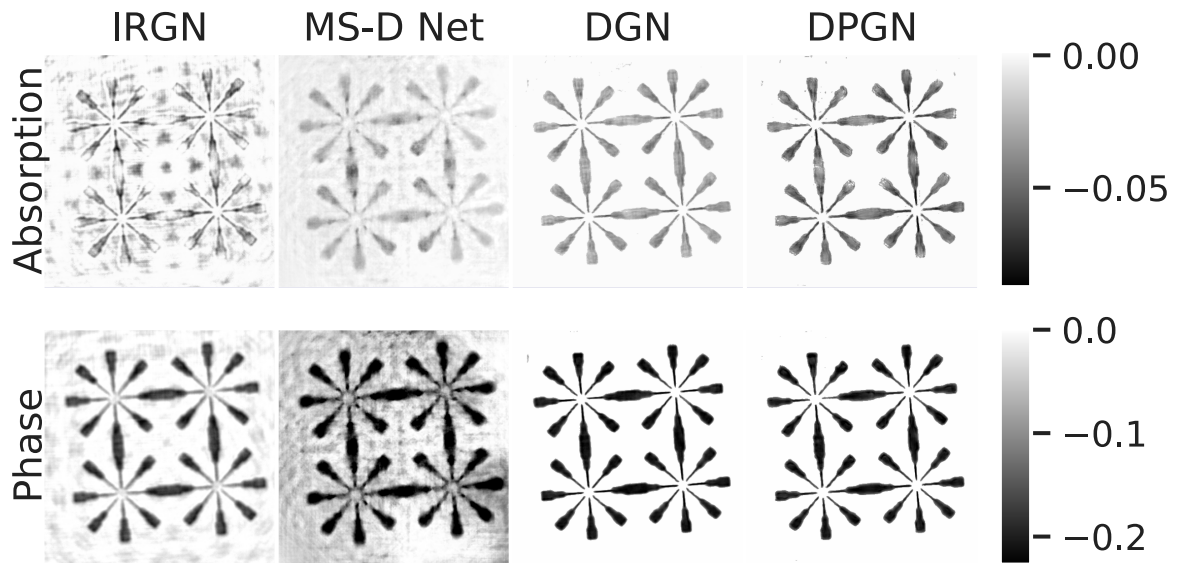


Figure 21.: Reconstructions obtenues sur donnée expérimentale

semble avoir un impact sur la qualité de la reconstruction : DPGN a permis d'améliorer les résultats tout en conservant le même nombre de paramètres que DGN.

Comparaison d'algorithmes déroulés pour la récupération de phase

Dans le chapitre précédent, nous avons présenté deux algorithmes basés sur l'idée de dérouler un algorithme itératif. Nous avons constaté que le schéma que nous déroulons influence la qualité de la reconstruction obtenue. Ici, nous approfondissons cette analyse en comparant les approches déroulées de différents schémas d'optimisation à leurs équivalents classiques.

Déroulement de solutions itératives

Deep Gradient Descent

De la même manière que l'on a procédé précédemment, nous pouvons dérouler une méthode de descente de gradient. A chaque itération, au lieu de choisir une régularisation et un pas de descente, nous proposons d'apprendre toute l'itération, comme fonction de l'itérée (B_k, φ_k) et du gradient du terme d'attache aux données

$$\mathbf{F}'_D(B_k, \varphi_k)^* \left(\mathbf{F}_D(B_k, \varphi_k) - \mathbf{I}_D^{\text{obs}} \right)$$

La k-ième itérée peut alors s'exprimer de la façon suivante:

$$(B_k, \varphi_k) = \Gamma_\theta^{\text{DGD}} \left\{ (B_{k-1}, \varphi_{k-1}), \nabla \mathbf{F}'_D(B_k, \varphi_k)^* \left(\mathbf{F}_D(B_k, \varphi_k) - \mathbf{I}_D^{\text{obs}} \right) \right\} \tag{0.29}$$

L'architecture qui a été utilisée pour la méthode Deep Gradient Descent (DGD) est inspirée de celle de (Hauptmann et al. 2018) et est affichée sur la Figure 22.

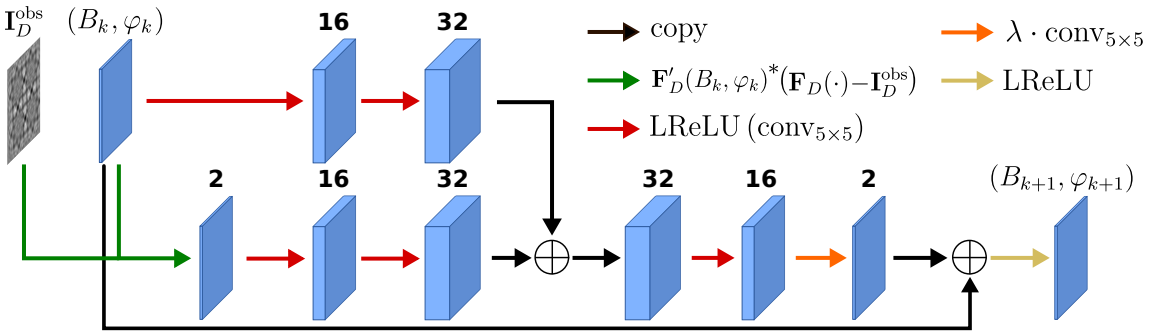


Figure 22.: Architecture du réseau $\Gamma_\theta^{\text{DGD}}$, représentant une itération de la méthode Deep Gradient Descent.

Deep Primal-Dual

Adler et Öktem (Adler:2018) ont proposé de dérouler un schéma primal-dual et d'apprendre les différentes étapes à travers deux réseaux de neurones. Un réseau P_{θ^p} pour la variable primale et un réseau D_{θ^d} pour la variable duale. L'idée est donc d'apprendre le processus itératif dans les deux espaces, celui des paramètres et celui des données. Une itération de la méthode Deep Primal-Dual (DPD) consiste à alterner entre ces deux réseaux, comme le montre la Figure 23.

Expériences

Nous avons comparé les différentes méthodes déroulées avec leurs homologues itératifs classiques. Les mêmes hyperparamètres que ceux qui ont été décrits pour les méthodes DGN et DPGN ont été utilisés. De même, les méthodes déroulées et itératives ont toutes été initialisées avec $(B_0, \varphi_0) = (0, 0)$.

Données simulées

La même base d'ensemble de données de test que celle du chapitre précédent a été utilisée ici. Les moyennes des métriques ont été calculées et les résultats sont résumés dans le Tableau 7. Toutes les approches déroulées ont une qualité de reconstruction similaire en moyenne, mais

	NMSE (en %)		FRCM (en %)		Résolution (en nm)		#Paramètres	Temps (en s)
Method	Absorption	Phase	Absorption	Phase	Absorption	Phase		
GD-TV$^\epsilon$	37.5 (17.4)	36.4 (18.2)	61.8 (12.2)	57.7 (13.2)	214 (101)	139 (78)	–	145
IRGN	85.5 (40.7)	39.3 (15.0)	71.2 (9.95)	68.1 (5.45)	238 (136)	154 (43)	–	116
NL-PDHG	29.2 (14.8)	23.6 (12.6)	58.4 (9.08)	50.7 (8.28)	146 (85.2)	99.7 (26.5)	–	147
DGD	13.2 (17.3)	4.74 (6.99)	37.6 (13.2)	23.8 (15.7)	82.2 (116)	64.3 (62.6)	41×10^3	**3.85**
DPD	12.5 (15.5)	4.48 (6.25)	39.2 (14.4)	24.3 (16.5)	107 (138)	75.5 (66.7)	31×10^3	4.63
DGN	12.1 (13.5)	4.61 (6.20)	35.7 (15.7)	23.0 (16.6)	76.8 (63.3)	63.0 (37.4)	31×10^3	5.88
DPGN	**11.0 (12.3)**	**4.05 (5.25)**	**31.5 (13.9)**	**19.8 (15.8)**	**74.0 (63.4)**	**59.1 (28.3)**	32×10^3	6.31

Tableau 7.: Comparaison des différentes approches déroulées et itératives.

DGD a la moins bonne NMSE parmi elles, bien qu'ayant le plus grand nombre de paramètres. Cela confirme que davantage de paramètres ne signifie pas nécessairement de meilleurs résultats. Ici, nous pouvons voir que le choix du schéma d'optimisation a un impact sur la qualité de la reconstruction. Le déroulement d'un schéma du premier ordre, tel que la descente de gradient, permet d'obtenir un temps de reconstruction plus court, mais il ne semble pas assez complexe

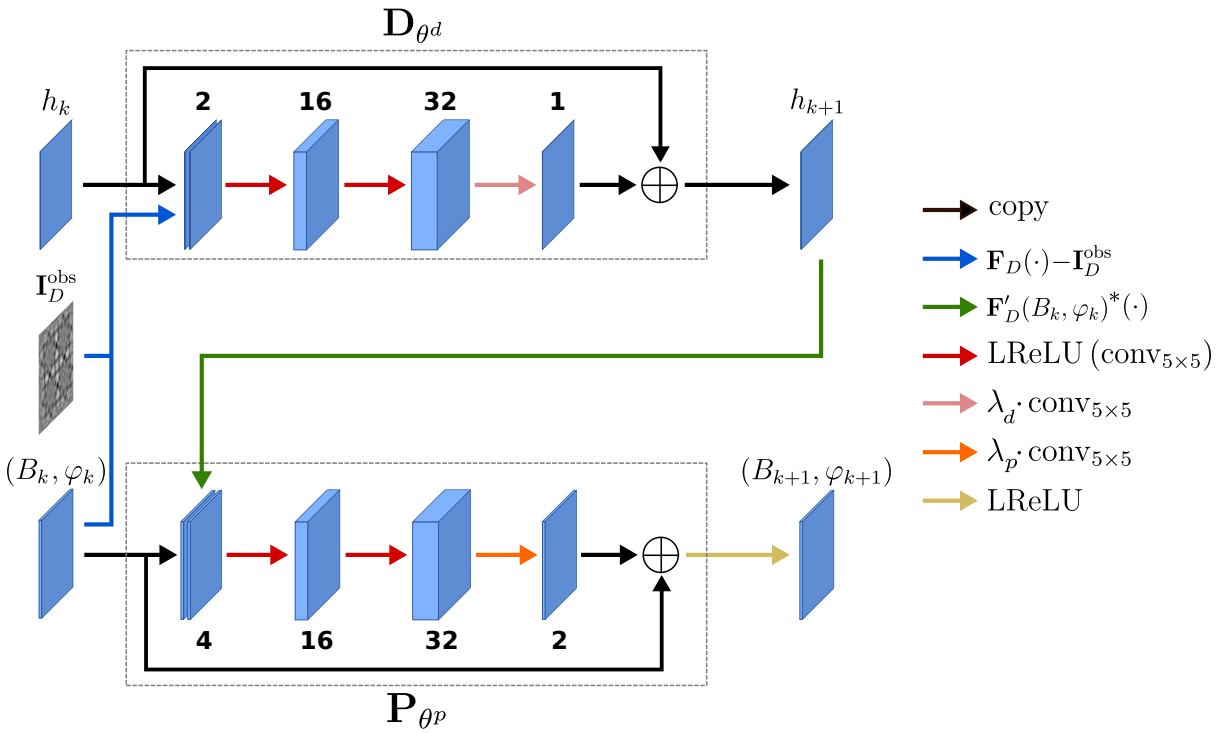


Figure 23.: Architecture du réseau $\Gamma_\theta^{\text{DPD}}$, représentant une itération de la méthode Deep Primal-Dual.

pour atteindre des résultats optimaux. L'apprentissage à la fois dans l'espace des paramètres du modèle et dans l'espace des données permet à la méthode DPD d'améliorer l'erreur normalisée, mais souffre d'une moins bonne qualité dans le domaine fréquentiel. Dans l'ensemble, ce sont les méthodes déroulées qui sont basées sur un schéma de type Gauss-Newton qui offrent les meilleures reconstructions, tant en termes d'erreur que de fréquence/résolution. Elles ont un temps de reconstruction légèrement plus long, car elles nécessitent davantage d'évaluations de l'opérateur à l'intérieur du réseau, mais restent tout de même rapides.

Données expérimentale

Dans la Figure 24, nous pouvons constater, une fois de plus, que les méthodes itératives classiques éprouvent des difficultés à récupérer l'absorption, à l'exception de la méthode NL-PDHG, dont la régularisation semble mieux adaptée. Les méthodes déroulées présentent des reconstructions très satisfaisantes, bien que des artefacts soient présents pour DGD. La méthode DPGN offre la meilleure reconstruction visuelle. En ce qui concerne la récupération de la phase, la Figure 25 montre que les méthodes déroulées donnent des résultats très similaires, et seule la méthode classique NL-PDHG parvient à obtenir des résultats comparables.

Discussion et perspectives

Nous avons illustré comment le choix du schéma de base avait un impact sur les reconstructions obtenues. Nous avons constaté qu'avec approximativement le même nombre de paramètres et la même stratégie d'entraînement, ces approches produisent des résultats similaires, même si certaines offrent une meilleure qualité à la fois visuelle et quantitative. Les réseaux exploitant des informations du second ordre, même avec une Hessienne approximative, comme DGN ou DPGN, donnent de meilleurs résultats de reconstruction. Globalement, les approches déroulées

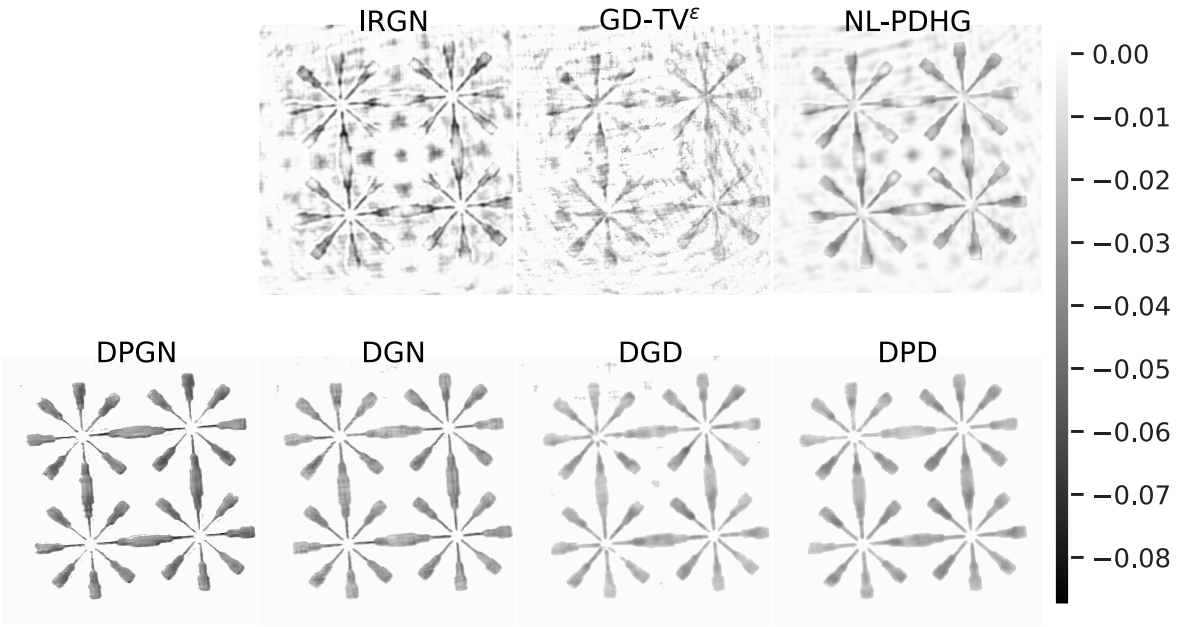


Figure 24.: Reconstructions de l'absorption pour les différentes méthodes.

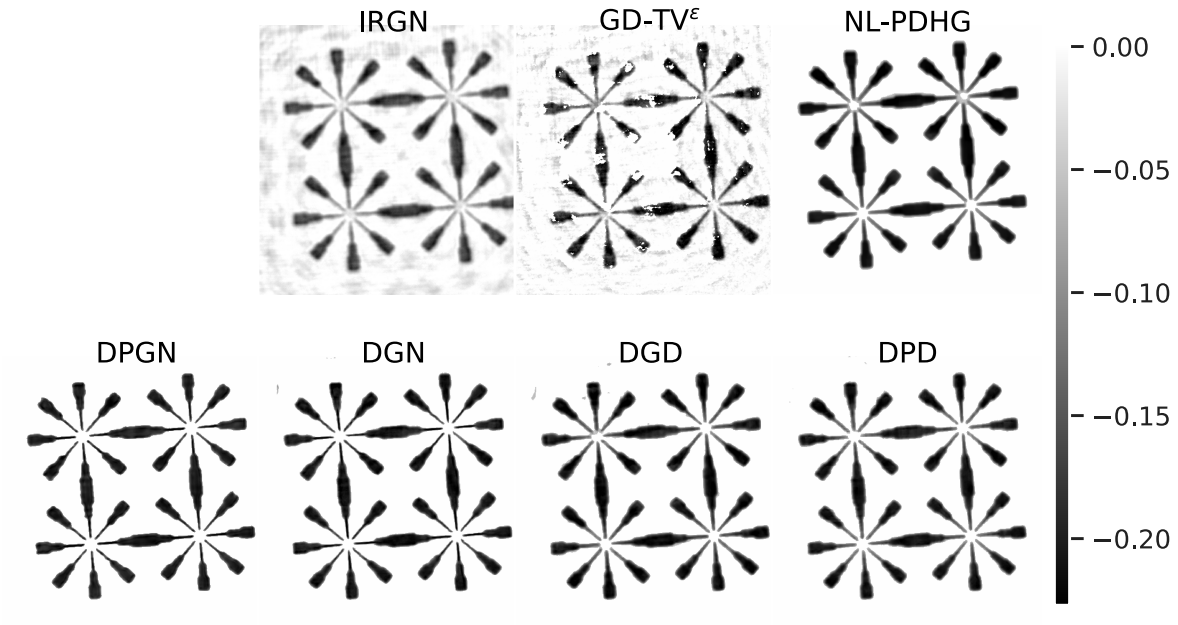


Figure 25.: Reconstructions de la phase pour les différentes méthodes.

sont plus efficaces que leurs homologues classiques, que ce soit sur des données simulées ou expérimentales. L'association de ces approches avec la tomographie pourrait nous fournir davantage d'informations sur leur comparaison.

Une autre approche intéressante pourrait être celle proposée par (Gilton, Ongie, and Willett 2021), où les auteurs ont suggéré de résoudre des problèmes inverses linéaires basés sur des modèles d'équilibre, ce qui correspond à un nombre infini d'itérations. Dans ce cas, les réseaux sont entraînés pour converger vers un point fixe. Une extension au cas non linéaire pourrait être intéressante pour l'approche DPGN, où le réseau $J_{\theta j}$ pourrait être entraîné d'une telle façon. Les propriétés de convergence des algorithmes déroulés que nous avons présentés ne sont pas encore claires. Nous pourrions essayer de montrer que le réseau $J_{\theta j}$ possède des propriétés similaires à un opérateur proximal, comme le font les auteurs de (Pesquet et al. 2021). Si ce n'est pas le cas, nous pourrions essayer d'imposer de telles propriétés. La difficulté ici réside dans le fait que le problème de récupération de phase est un problème non linéaire et que le pas de descente dépend de l'itéré actuel.

1 Theoretical background

Phase retrieval is a common problem in coherent imaging techniques.This is an inverse problem which consists in recovering phase information from intensity measurements. Since the majority of detectors only record intensity information, the phase information is lost, making its reconstruction an ill-posed problem.

In this chapter, an overview of the foundations of inverse problems and linear inverse problems is provided, and the basic theoretical concepts are reviewed. Next, we present the fundamental principles of propagation-based X-ray phase contrast imaging. The direct problem of phase retrieval is described by the Fresnel diffraction. Analytical methods can be obtained via linearization of the direct model, which yields fast filtering-based methods which are however limited by the type of linearization. Resolution of the non-linear problem usually require iterative methods which are more time consuming but avoid the limitations of analytical methods. Approaches include alternate projections onto constraints and variational methods, the latter in which the phase retrieval task is treated as a minimization problem. Finally, at the end of this chapter, we introduce the use of neural networks and present aspects of deep learning used for image reconstruction. The state of the art of machine learning-based approaches for reconstruction problems, and in particular for the phase recovery problem, is presented.

I Inverse problems

Following (Keller 1976), two problems are said to be *inverse* of one another if the formulation of each involves all or part of the solution of the other. This definition is somewhat arbitrary and gives a symmetrical role to the problems under consideration. A more practical definition is that an inverse problem means determining the *causes* knowing the *effects*. Solving an inverse problem therefore involves deducing the causes from what we call the *direct problem*, i.e. the effects. This definition highlights the common focus on studying direct problems, where causes lead to effects.

In practice, physical models can be described by a direct model, which consists of a map $\mathcal{A}$:

$$\mathcal{A} : \mathcal{P} \to \mathcal{D} \tag{1.1}$$

where $\mathcal{P}$ (model parameter space) and $\mathcal{D}$ (data space) are two Hilbert spaces The application $\mathcal{A}$ is generally a non-linear operator and is called the *forward operator*. Solving the direct problem refers to determining the effects or outputs that arise from a given set of known causes or input parameters within a specific system or model. Formally, this means determining $\mathcal{A}(p)$ knowing that we have a precise forward model $\mathcal{A}$ and parameters $p \in \mathcal{P}$, in other words, generating measurements from a know system. Conversely, solving the inverse problem means estimating the parameters $p \in \mathcal{P}$ given some data $d \in \mathcal{D}$. A practical difficulty of the study of inverse problems is that it often requires a good knowledge of the direct problem, manifesting through the utilization of a wide array of both physical and mathematical concepts. Inverse problems generally present other challenges, as they can have multiple solutions, and distinguishing

between them requires additional information. The concept of "same causes produce same effects" can be given mathematical meaning, ensuring well-posedness for direct problems, but inverse problems can result in various causes producing the same effects, posing the primary difficulty in their study. One of the main problem is that the inverse solution is very sensitive to noise in the output, in other words the inverse operator (or pseudo-inverse) is not continuous.

Hadamard introduced the notion of *well-posed* problem, namely one which satisfies three conditions:

1. A solution exists.
2. The solution is unique.
3. The solution depends continuously on the data.

From this, we can see that solving a well-posed inverse problem is easier. A problem that is not well-posed within the meaning of the definition above is said to be *ill-posed*. In reality, inverse problems are rarely well posed due to several reasons, including the inherent complexity and variability of real-world systems, limitations in data collection and measurement. Mathematically, the forward operators are often compact or smoothing operators, so the inverse is not continuous.

I.1 Solving linear inverse problems using least square method

If the forward operator $\mathcal{A}$ is linear we say that the inverse problem is linear. Assuming the problem is discrete or that we have an appropriate discretization, then the direct model can be written as

$$\mathbf{A}\mathbf{p} = \mathbf{d} \tag{1.2}$$

where $\mathbf{A}$ is a (possibly complex) matrix, and $\mathbf{d}$ and $\mathbf{p}$ are (possibly complex) vectors.

If the problem (1.2) is well-posed, then the matrix $\mathbf{A}^{-1}$ exists and the solution is given by $\mathbf{p} = \mathbf{A}^{-1}\mathbf{d}$. If the problem is ill-posed, for instance, the matrix $\mathbf{A}$ is not square or is singular, then what we can do is replace the problem (1.2) with a formulation as a least squares problem:

$$\underset{\mathbf{p}\in\mathcal{P}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{A}\mathbf{p} - \mathbf{d}\|_{\mathcal{D}}^2 \right\} \tag{1.3}$$

where $\|d\|_{\mathcal{D}} = \sqrt{\langle d|d\rangle_{\mathcal{D}}}$ is the norm induces by the scalar product. By calculating the gradient of the functional $\mathbf{p} \mapsto \frac{1}{2}\|\mathbf{A}\mathbf{p} - \mathbf{d}\|^2$, we obtain the *normal equation* associated with (1.3)

$$\mathbf{A}^*\mathbf{A}\mathbf{p} = \mathbf{A}^*\mathbf{d} \tag{1.4}$$

where $\mathbf{A}^*$ denotes the conjugate transpose of $\mathbf{A}$. The matrix $\mathbf{A}^*\mathbf{A}$ is square and is invertible as long as the columns of $\mathbf{A}$ are linearly independent (i.e. if the continuous operator $\mathcal{A}$ is injective), in this case $\mathbf{p}$ is given by

$$\mathbf{p} = \left(\mathbf{A}^*\mathbf{A}\right)^{-1}\mathbf{A}^*\mathbf{d} \tag{1.5}$$

Note that the normal equation can be rewritten as

$$\mathbf{A}^*\left(\mathbf{A}\mathbf{p} - \mathbf{d}\right) = 0 \tag{1.6}$$

which means that the residual $\mathbf{A}\mathbf{p} - \mathbf{d}$ is in the kernel of $\mathbf{A}^*$, i.e. that it is orthogonal to the image of $\mathbf{A}$ (since $\mathrm{Ker}(A^*) = \mathrm{Im}(A)^{\perp}$) The solution of the least squares problem (1.3) is such that $\mathbf{A}\mathbf{p}$ is the projection of the measure $\mathbf{d}$ on the image of $\mathbf{A}$ (Fig. 1.1).

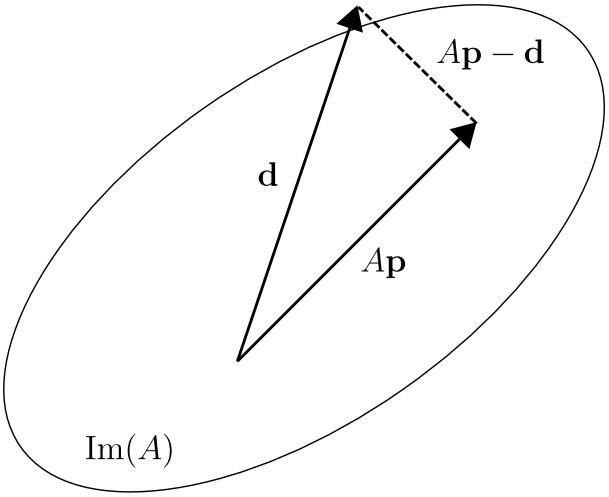


Figure 1.1.: Geometric illustration of the least squares.

I.2 Conditioning of the least squares problems

Matrix conditioning refers to the numerical behavior of a matrix with respect to its invertibility and sensitivity to changes in its elements. Suppose the matrix $\mathbf{A}$ is disturbed by an error $\delta\mathbf{A}$ and that our data $\mathbf{d}$ is disturbed by an error $\delta\mathbf{d}$. The solution to the perturbed problem can be denoted as $\mathbf{p} + \delta\mathbf{p}$, so that the perturbed system can be written as

$$(\mathbf{A} + \delta\mathbf{A})(\mathbf{p} + \delta\mathbf{p}) = (\mathbf{d} + \delta\mathbf{d}) \tag{1.7}$$

The *relative error* $\dfrac{\|\delta\mathbf{p}\|}{\|\mathbf{p}\|}$ can be quantified by the *condition number* of the matrix $\mathbf{A}$, $\mathrm{cond}(\mathbf{A})$, which is defined as

$$\mathrm{cond}(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\| = \max_{\mathbf{p}} \left\{\|\mathbf{A}\mathbf{p}\|\right\} \max_{\mathbf{d}} \left\{\|\mathbf{A}^{-1}\mathbf{d}\|\right\} \tag{1.8}$$

Using the L^2 norm defined by $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$, the condition number is then given by the eigenvalues of $\mathbf{A}$. If $\nu_{\min}$ and $\nu_{\max}$ denote the smallest and largest eigenvalues of $\mathbf{A}$, respectively, then

$$\mathrm{cond}(\mathbf{A}) = \frac{\nu_{\max}}{\nu_{\min}} \tag{1.9}$$

We see that if $\mathbf{A}$ is close to singular, i.e., if $\nu_{\min} \to 0$, the condition number goes to infinity.

The relative error is controlled by the condition number of the matrix $\mathbf{A}$ (see (Ciarlet 1989)):

$$\frac{\|\delta\mathbf{p}\|}{\|\mathbf{p}\|} \leq \frac{\mathrm{cond}(\mathbf{A})}{1 - \|\mathbf{A}\|\|\delta\mathbf{A}\|} \left(\frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|} + \frac{\|\delta\mathbf{d}\|}{\|\mathbf{d}\|}\right) \tag{1.10}$$

A matrix is considered *well-conditioned* if its condition number is close to 1. Eq. (1.10) shows that the condition number of a matrix measures how sensitive its output is to changes in the input. A small condition number indicates that the matrix is well-conditioned and small changes in the input lead to proportionally small changes in the output. An *ill-conditioned* matrix has a high condition number, which means that small changes in the input can lead to large changes in the output. Ill-conditioned matrices can cause problems in numerical computations, such as amplification of errors and loss of accuracy. They can also make it difficult to compute

an accurate or stable solution to systems of equations involving the matrix. Preconditioning techniques are often used to solve linear systems of equations in the field of inverse problems.

I.3 Regularization of linear inverse problems

Recall that the least squares method allows us to obtain the solution to the inverse problem by inverting the matrix $\mathbf{A}^*\mathbf{A}$ (1.5). However, this matrix can be ill-conditioned or even singular, making the solution sensitive to small variations. To overcome this problem, we can add *prior* information to enforce the solution to satisfy certain constraints by adding a *regularization* term to the functional (1.3). Regularizing the problem allows to add some knowledge about the solution, such as smoothness or sparsity in some domain, obtained with a sparsifying operator. A weighting parameter on the regularization term helps a balance between fitting the observed data and incorporating prior information or constraints.

I.3.1 Tikhonov regularization

One of the most well-known regularization techniques is Tikhonov regularization, also known as ridge regression. In order to introduce some prior knowledge about the solution, i.e., give preference to a solution with desirable properties, a *quadratic* regularization term can be included in the minimization:

$$\underset{\mathbf{p}\in\mathcal{P}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{A}\mathbf{p}-\mathbf{d}\|_{\mathcal{D}}^2 + \alpha\|\mathbf{L}(\mathbf{p}-\mathbf{p}_0)\|_{\mathcal{P}}^2 \right\} \tag{1.11}$$

where $\mathbf{p}_0 \in \mathcal{P}$ is an initial estimate of the solution and α is a *regularization parameter* which is used to control the trade-off between fitting the observed data and applying the regularization constraint. $\mathbf{L}$ is called the *Tikhonov matrix*, and is used to enforce desired properties in the solution. Usually, $\mathbf{L}$ is the identity or an approximation of the first or second order derivative operator (Tikhonov and Arsenin 1977). In the latter case, this leads to a preference for smooth solutions.

Again, by calculating the gradient of the functional $\mathbf{p} \mapsto \frac{1}{2}\|\mathbf{A}\mathbf{p}-\mathbf{d}\|^2 + \alpha\|\mathbf{L}(\mathbf{p}-\mathbf{p}_0)\|^2$, it can be shown that the solution of (1.11) is given by

$$\mathbf{p} = \left(\mathbf{A}^*\mathbf{A} + \alpha^2\mathbf{L}^*\mathbf{L}\right)^{-1}\left(\mathbf{A}^*\mathbf{d} + \alpha^2\mathbf{L}^*\mathbf{L}\mathbf{p}_0\right) \tag{1.12}$$

If $\mathbf{L}$ is the identity, the Tikhonov regularization will force the solution not to be too far from the initial guess, and if in addition there is no initial a priori solution, i.e. $\mathbf{p}_0 = 0$, the regularization will limit the energy in the solution. From (1.12), we see that whenever $\mathbf{A}$ is close to zero, the solution will be close to $\mathbf{p}_0$, and when $\mathbf{A}$ is singular, we will have $\mathbf{p} = \mathbf{p}_0$. Tikhonov regularization effectively mitigates overfitting by preventing the solution from closely following the noisy data, thus leading to more robust and stable solutions. For ill-posed inverse problems, the singular values of the operator tend to zero. Using a singular value decomposition of the operator $\mathbf{A}$, it can be seen that these singular values appear at the denominator of the solution leading to a divergence of the solution if there is no regularization. The regularization prevents this divergence. It can be noted that some regularization techniques use some filter functions to generalize the Tikhonov regularization.

I.3.2 Choice of the regularization parameter

Choosing the appropriate regularization parameter α in an inverse problem is a critical step to balance the trade-off between fitting the data well and preventing overfitting. We saw that regularization helps stabilize the solution. The choice of the regularization parameter depends

on the specific problem and the characteristics of the data, for example the noise level.

Some methods to automatically select the regularisation parameter have been proposed, for example the L-curve (Hansen 2000) method and the Morozov's discrepancy principle (Fromovitz 1984). The name L-curve comes from the characteristic shape of the curve when the fitting error is plotted against the regularization term on a logarithmic scale 1.2. The L-curve method provides a way to identify the point of maximum curvature where an optimal trade-off between fitting and regularization occurs.

The Morozov's discrepancy principle (Fromovitz 1984) suggests that you should choose a regularization parameter that results in a solution $\mathbf{p}$ such that the discrepancy between the model prediction $\mathbf{Ad}$ and the observed data $\mathbf{p}$ is comparable to the level of noise in the data (if known).

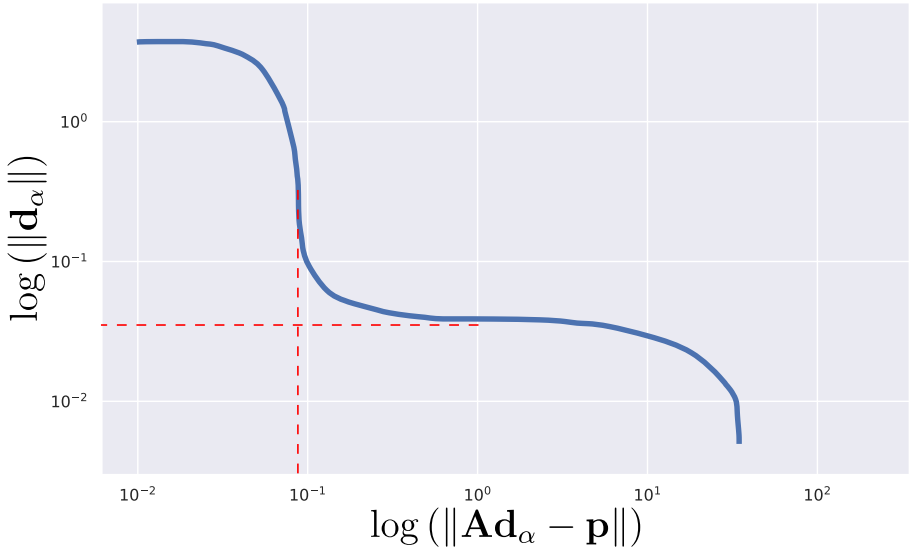


Figure 1.2.: A typical L-curve and its global corner for Tikhonov regularization method. For each value α, $\mathbf{p}_\alpha$ denotes the least-squares solution regularized by Tikhonov. The norm of the distance between the data and the model is reported on the horizontal axis, while the distance of $\mathbf{p}$ to $\mathbf{p}_\alpha$ is reported on the vertical axis. The L-curve selection criterion consists of locating the value which maximizes the curvature, that is the L-curve corner that separates the two regions: under-regularized on the left and over-regularized on the right.

II Propagation-based X-ray phase contrast imaging

Various phase sensitive imaging techniques have been proposed. In this thesis, we focus on propagation-based X-ray phase contrast imaging, also known as in-line phase contrast imaging. Assuming an incident beam is spatially coherent, phase contrast is established through the free-space propagation of X-rays between the sample and the detector, a diffraction pattern is then recorded. The image formation process, which describes the direct problem, can be understood within the framework of Fresnel diffraction theory.

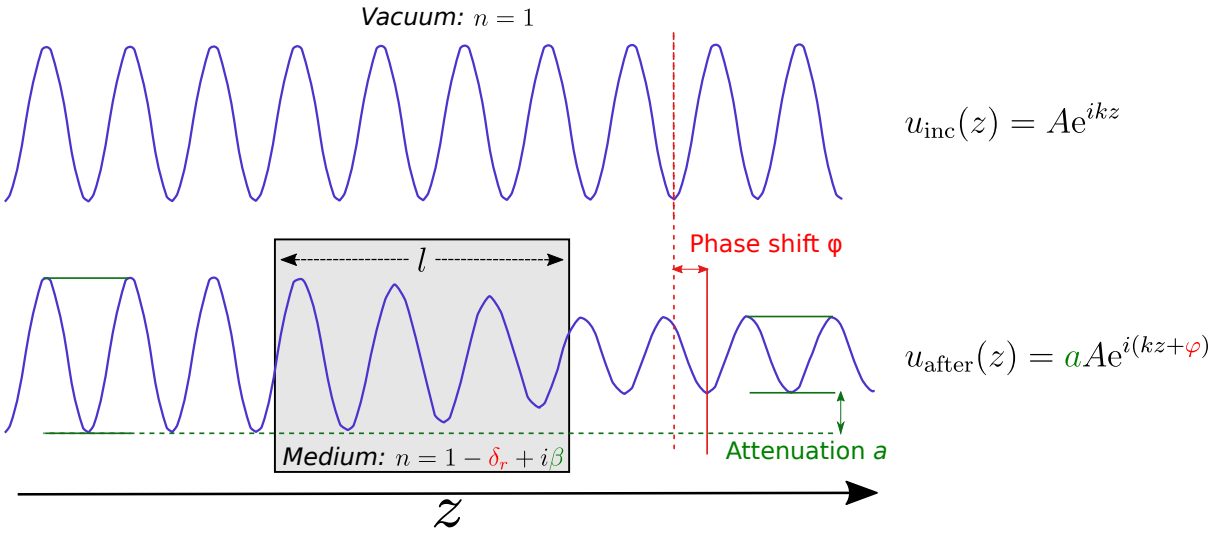


Figure 1.3.: Interaction of X-ray beam when passing through an object.

II.1 X-rays and their interactions with matter

When an X-ray beam passes through an object, it can undergo two distinct influences (Fig. 1.3). One possibility is absorption within the object, leading to an attenuation a in its amplitude. Moreover, the beam could experience a delay as it moves through the object, causing a phase shift φ. The phenomenon of an X-ray beam interacting with an object is related to its *complex refractive index* n, which is usually expressed as:

$$n = 1 - \delta_r + i\beta \tag{1.13}$$

where δ_r is the refractive index decrement and β is the absorption index. Both δ_r and β depend on the material as well as the X-ray wavelength λ. For the X-ray of wavelength λ, we define the *wavenumber* in vacuum

$$k = \frac{2\pi}{\lambda} \tag{1.14}$$

The refractive number is defined so that the k and the wavenumber in a material of refractive index n, denoted as k', are related by

$$k' = nk \tag{1.15}$$

If we denote

$$u_{\text{inc}}(z) = Ae^{ikz} \tag{1.16}$$

the wave propagating in the vacuum and u_{after} the wave that has passed through an object of length l, we have:

$$u_{\text{after}}(z) = aAe^{i(kz+\varphi)} \tag{1.17}$$

where the attenuation a and the phase shift φ are given by

$$a = e^{-k\beta s} \quad \text{and} \quad \varphi = -k\delta_r s \tag{1.18}$$

Thus, the absorption index β is related to attenuation, and the refractive index decrement δ_r is related to phase shift.

II.2 Attenuation contrast

Several types of interaction exist, of which only three are in the energy range of hard X-rays (10-100 keV) we are concerned with. These are photoelectric effect, Compton and Rayleigh scattering. The *cross section*, also known as the atomic attenuation, describes the probability of each interaction to occur. The type of interaction that contributes to attenuate an X-ray beam depends on the energy of the X-rays and the atomic number of the atoms encountered, so is the cross section.

II.2.1 Photoelectric effect

The incident X-ray photon ejects a bound atomic electron and is completely absorbed. The electron is ejected with an energy

$$E_{\text{electron}} = E_{\text{photon}} - E_{\text{binding}} \tag{1.19}$$

where E_{photon} is the energy of the incident photon and E_{binding} is the binding energy of the electron. The atom is thus ionized and releases a photon characteristic of the atom (secondary photon) in order to recover a stable state. The photoelectric cross section $\sigma_{\text{photoelectric}}$ depends on the energy of the X-ray E as well as the atomic number Z of the medium (Yi Wang 2007)

$$\sigma_{\text{photoelectric}} \propto \frac{Z^5}{E^{3.5}} \tag{1.20}$$

The relation (1.20) is important because it states that materials with a higher atomic number like iron ($Z = 26$) will absorb more photons than those with a smaller atomic number such as the aluminum ($Z = 13$). This will help to distinguish the two materials on the X-ray. The photoelectric effect is the one that causes the absorption part of the total attenuation.

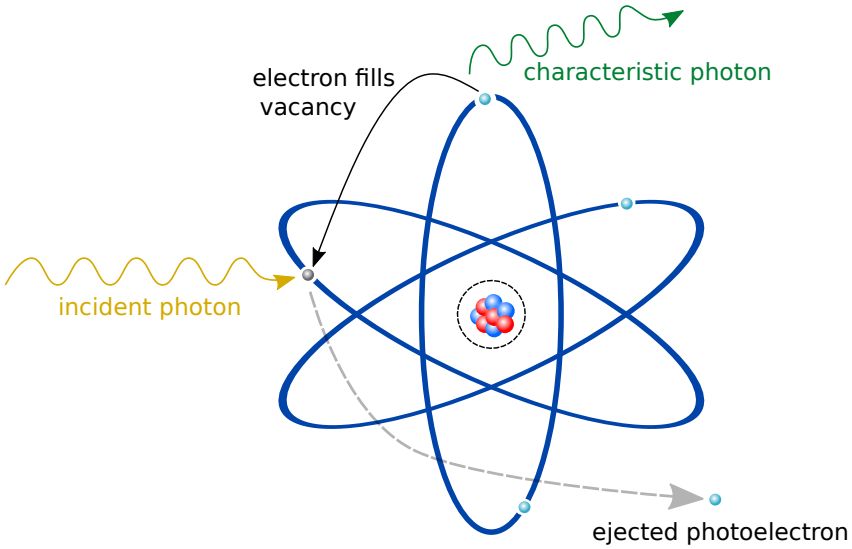


Figure 1.4.: Illustration of the photoelectric effect.

II.2.2 Compton scattering - Incoherent scattering

An incident photon of high energy collides with an electron of the atom but is not absorbed. The electron is ejected, and the incident photon, having lost a portion of its energy, is scattered in a random direction at an angle that depends on the initial energy and the lost energy. This is

also known as incoherent scattering, and it occurs in the same energy range as the photoelectric effect. The Compton scattering cross section depends on the atomic number and the energy

$$\sigma_{\text{Compton}} \propto \frac{Zm_e c}{E + m_e c^2} \tag{1.21}$$

where $m_e c^2 \approx 511$ keV denotes the electron rest mass energy. In the energy range we are concerned with, the relation (1.22) shows that the Compton scattering cross section is roughly proportional to the atomic number.

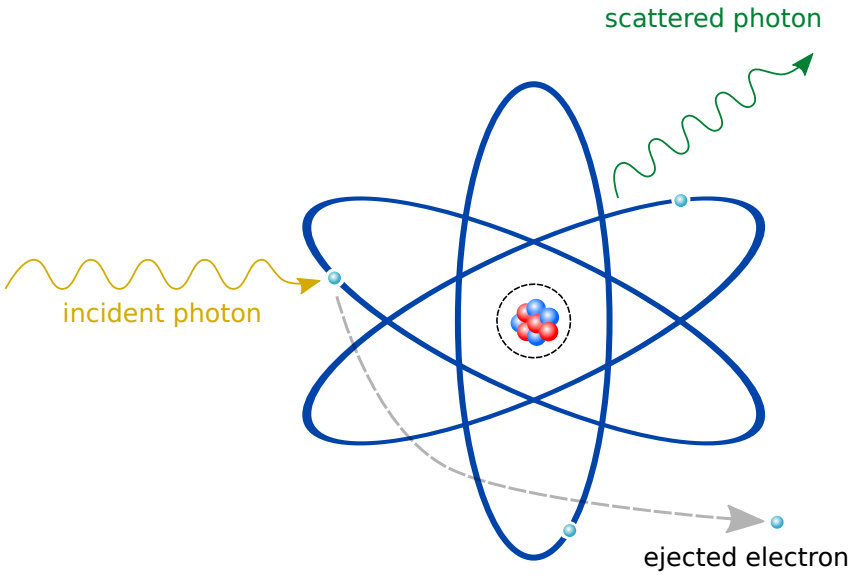


Figure 1.5.: Illustration of the Compton interaction.

II.2.3 Rayleigh scattering - Coherent scattering

The combined coherent summation of scatterings originating from all electrons within an atom results in the Rayleigh scattering effect, also known as elastic scattering. The photon excites the atom without colliding with an electron. When the atom returns to its stable state, it emits a new photon of the same energy, so the incident photon simply bounces off an atom, and is deviated. This interaction occurs at low energies. This effect is dependent on the photon energy E and the atomic number Z

$$\sigma_{\text{Rayleigh}} \propto \frac{Z^2}{E^2} \tag{1.22}$$

Rayleigh scattering effect has a small contribution of the total attenuation.

II.2.4 Linear attenuation coefficient

The *total attenuation coefficient* or *linear attenuation coefficient* μ is the measured quantity when attenuation projection is acquired, For a material with atomic mass A, we define

$$\mu = \frac{N_A \rho \sigma_{\text{total}}}{A} \tag{1.23}$$

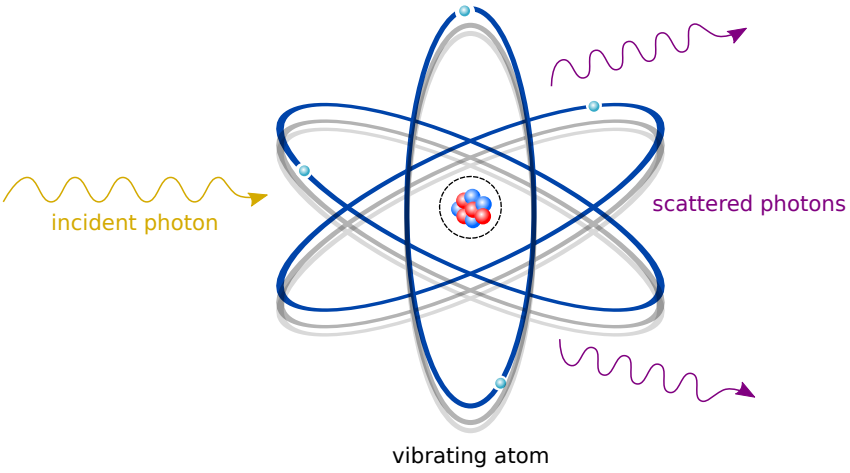


Figure 1.6.: Illustration of the Rayleigh scattering.

where N_A is Avogadro's number and ρ is the density of the material. It is based on the total attenuation cross section, composed of the three photons interactions we saw

$$\sigma_{\text{total}} = \sigma_{\text{photoelectric}} + \sigma_{\text{Compton}} + \sigma_{\text{Rayleigh}} \tag{1.24}$$

We can also write μ as follows

$$\mu = \frac{4\pi}{\lambda}\beta \tag{1.25}$$

so the linear attenuation coefficient is related to the attenuation index β of the material. From (1.23) and (1.25) we see the dependence of β on energy and atomic number

$$\beta \propto \frac{Z^5}{E^{4.5}} \tag{1.26}$$

II.2.5 Beer-Lambert Law

The Beer-Lambert law describes the attenuation of the photons in a medium. Let us consider a monochromatic incident beam composed of $N_0(E)$ photons at a certain energy E. This beam is passing through a homogeneous material of linear coefficient $\mu(z, E)$ along the z-direction. The number of photons transmitted is given by

$$N(z, E) = N_0(E)\mathrm{e}^{-\mu(z,E)l} \tag{1.27}$$

If the object is heterogeneous, we can generalized (1.27) by considering that the object is homogeneous over infinitesimal sections $\mathrm{d}z$, which translates into

$$N(z, E) = N_0(E)\mathrm{e}^{-\int \mu(z,E)\mathrm{d}z} \tag{1.28}$$

II.3 Refractive index decrement and phase contrast

The refractive index decrement δ_r describes how the wave propagates in the medium, following (Als-Nielsen and McMorrow 2011), it can be expressed as

$$\delta_r = \frac{r_c\lambda^2}{2\pi}\rho_{\mathrm{e}} \tag{1.29}$$

where r_c denotes the classical electron radius, λ the wavelenght and ρ_e is the electron density. Just like the aborption index (1.26), we see the dependence of δ_r on energy and atomic number

$$\delta_r \propto \frac{Z}{E^2} \tag{1.30}$$

In means that in the energy range of hard X-rays, the absorption index decreases significantly and the refractive index can be up to three orders of magnitude larger. The main advantage of phase-contrast imaging comes from the fact that δ_r is generally larger than β (Figure 1.7). It can also be shown that the refractive index is close to proportional to the mass density ρ and the wavelength squared λ^2 (Guinier 1994):

$$\delta_r \approx 1.3 \times 10^{-6} \rho \lambda^2 \tag{1.31}$$

This highlights another advantageous aspect of phase imaging techniques since the quantity δ_r is straightforward to interpret, it corresponds approximatively to the mass density.

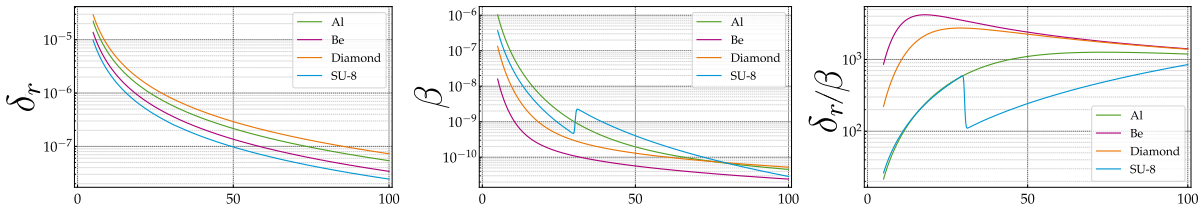


Figure 1.7.: (*left*) refractive index decrement δ_r, (*middle*) absorption index β and (*right*) ratio δ_r/β for aluminum (Al), beryllium (Be), diamond and SU-8 (Celestre 2021). The x-axis represents the energies (in keV).

II.4 Fresnel diffraction

The basic principle of propagation-based imaging (PBI), letting the beam propagate in free space after passing through the object, can be described by Fresnel diffraction. The effect of propagation induces a phase change of the beam through what is known as the *Fresnel propagator*. Therefore, the image recorded after propagation is often called a *Fresnel diffraction pattern*.

In the following, we consider an object that is completely described by its 3D complex refractive index

$$n(x, y, z) = 1 - \delta_r(x, y, z) + i\beta(x, y, z) \tag{1.32}$$

where $\delta_r(x, y, z)$ is the 3D refractive index decrement and $\beta(x, y, z)$ is the 3D absorption index of the spatial coordinates (x, y, z).

Let's consider an object illuminated with a monochromatic X-ray beam of wavelength λ. For straight-line propagation of the beam along the propagation direction z, this interaction can be described by a 2D complex *transmittance function* T of the coordinates $\mathbf{x} = (x, y)$:

$$T(\mathbf{x}) = \exp[-B(\mathbf{x}) + i\varphi(\mathbf{x})] = a(\mathbf{x}) \exp[i\varphi(\mathbf{x})] \tag{1.33}$$

$B(\mathbf{x})$ is the absorption and $\varphi(\mathbf{x})$ the phase shift induced by the object. The phase shift and the absorption are projections of the absorption and refraction index respectively defined with the following line integrals:

$$B(\mathbf{x}) = \frac{2\pi}{\lambda} \int \beta(\mathbf{x}, z) \mathrm{d}z \tag{1.34}$$

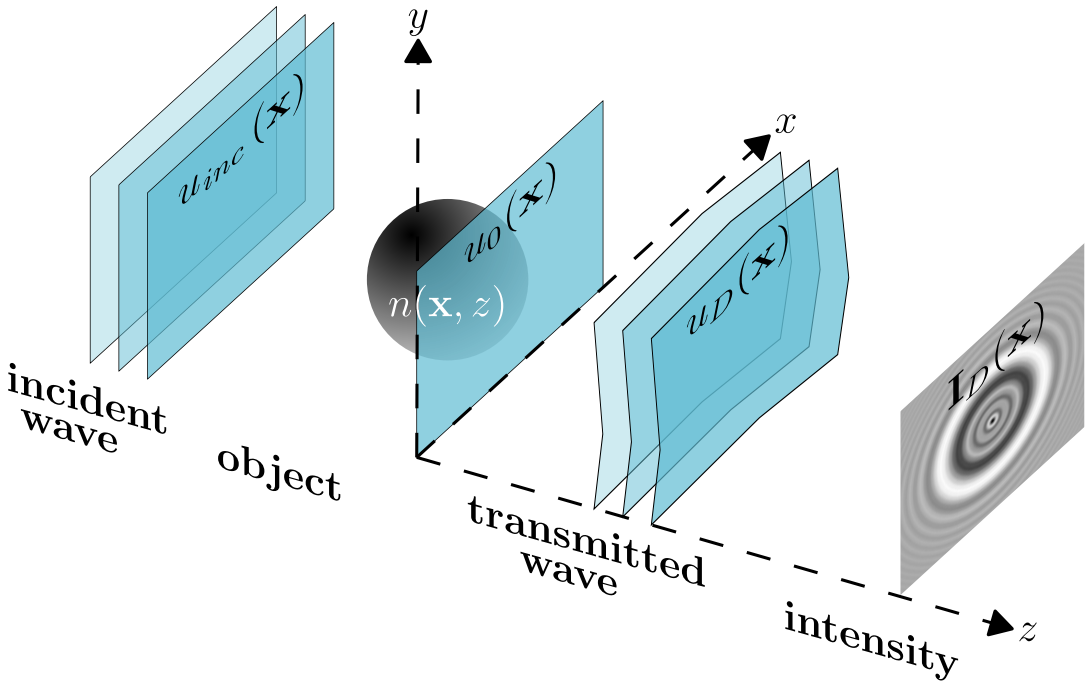


Figure 1.8.: Sketch of the basic physical model of propagation-based X-ray phase contrast imaging.

$$\varphi(\mathbf{x}) = -\frac{2\pi}{\lambda}\int \delta_r(\mathbf{x}, z)\mathrm{d}z \tag{1.35}$$

With this description, we have an expression for the wavefront $u_0(\mathbf{x})$ at the exit plane of the sample (corresponding to $D = 0$) as a function of the transmittance function $T(\mathbf{x})$ and the incident wave $u_{\text{inc}}(\mathbf{x})$

$$u_0(\mathbf{x}) = T(\mathbf{x})u_{\text{inc}}(\mathbf{x}) \tag{1.36}$$

The *intensity* of the wave at the exit plane of the sample is

$$\mathbf{I}_0(\mathbf{x}) = |T(\mathbf{x})u_{\text{inc}}(\mathbf{x})|^2 = \mathrm{e}^{-2B(\mathbf{x})}\mathbf{I}_{\text{inc}}(\mathbf{x}) \tag{1.37}$$

If the incident intensity $\mathbf{I}_{\text{inc}}(\mathbf{x})$ is known, the attenuation can be accessed directly from (1.37).

In the framework of the Fresnel diffraction theory, the complex wave function $u_D(\mathbf{x})$ at a distance D downstream of the sample can be described as a convolution (Goodman 1996)

$$u_D(\mathbf{x}) = P_D(\mathbf{x}) * u_0(\mathbf{x}), \quad \text{with} \quad P_D(\mathbf{x}) = \frac{1}{i\lambda D}\exp\left(i\frac{\pi}{\lambda D}|\mathbf{x}|^2\right) \tag{1.38}$$

where $P_D(\mathbf{x})$ is called the *Fresnel propagator*. So letting the wave propagate in free space can be seen as a multiplication in the Fourier domain

$$\widehat{u}_D(\mathbf{x}) = \widehat{P}_D(\mathbf{x}) * \widehat{u}_0(\mathbf{x}) \tag{1.39}$$

where the 2D Fourier transform is given by

$$\widehat{h}(\mathbf{f}) = \mathcal{F}h(\mathbf{f}) = \int_{\mathbb{R}} h(\mathbf{x})\mathrm{e}^{-2i\pi\mathbf{x}\cdot\mathbf{f}}\mathrm{d}\mathbf{x} \tag{1.40}$$

with $\mathbf{f}$ the frequency variable associate to $\mathbf{x}$. The Fourier transform of the Fresnel propagator can be written

$$\widehat{P}_D(\mathbf{f}) = \mathrm{e}^{-i\pi\lambda D|\mathbf{f}|^2} \tag{1.41}$$

In practice, the recorded images have to be corrected for $u_{\mathrm{inc}}(\mathbf{x})$ by flat field correction in order to obtain diffraction patterns that resemble those under a hypothetical illumination by plane waves. In the following, however, the assumption $u_{\mathrm{inc}}(\mathbf{x}) = 1$ is made. Assuming a unitary incident wavefront, which means $u_{\mathrm{inc}}(\mathbf{x}) = 1$, The intensity measured at a distance D from the sample can thus be written as

$$\mathbf{I}_D(\mathbf{x}) = |u_D(\mathbf{x})|^2 = |P_D(\mathbf{x}) * T(\mathbf{x})|^2 \tag{1.42}$$

In the Fourier domain, it has been shown that the intensity measurement can be expressed as (J.-P. Guigay 1977)

$$\widehat{\mathbf{I}}_D(\mathbf{f}) = \iint_{\mathbb{R}\times\mathbb{R}} T\left(\mathbf{x} - \frac{\lambda D\mathbf{f}}{2}\right) T^*\left(\mathbf{x} + \frac{\lambda D\mathbf{f}}{2}\right) \mathrm{e}^{-2i\pi\mathbf{x}\cdot\mathbf{f}} \mathrm{d}\mathbf{x} \tag{1.43}$$

From a theoretical perspective, this expression is valuable as it enables us to formulate linearized models (see IV).

II.4.1 Partial coherence

So far, we have assumed the presence of a perfectly coherent wave. We can classify coherence into two distinct types:

- *Longitudinal* (temporal) coherence, which is related to the source's monochromatic nature. A lack of spatial coherence can occur when the source emits photons with various frequencies, following a probability density distribution denoted as $p(\mathbf{f})$, which has a certain characteristic width. Photons of different frequencies do not exhibit coherent interactions.
- *Transverse* (spatial) coherence, which is associated with the physical dimensions of the source. Incomplete spatial coherence happens when the source region has a finite size, causing photons emitted from different positions to not interact coherently when reaching the detector.

Figure 1.9 summarizes the different types of coherence schematically.

II.4.2 Detector and phase contrast

We assumed so far that the intensity measured in the detector plane is defined by $\mathbf{I}_D = |u_D|^2$. In reality, various deviations from this ideal setup can occur, as detailed in the following.

Two-dimensional detectors are commonly employed to capture complete X-ray images in a single exposure. While effective for absorption radiography, they are less suitable for phase-sensitive imaging due to the contrast mechanism relying on interference, necessitating a broader incident beam. The resulting images typically comprise a substantial number of pixels, rendering scanning procedures prohibitively slow. Detector linearity, spatial invariance, and correction for non-linearities are essential considerations. The detector's detection quantum efficiency (DQE) plays a critical role in determining image quality and noise characteristics, it is given by (Rose 1946)

$$\mathrm{DQE} = \left(\frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}}\right)^2 \tag{1.44}$$

where $\mathrm{SNR}^2_{\mathrm{out}} = N_0(1 - a\mu)$ is the output signal-to-noise ratio and $\mathrm{SNR}^2_{\mathrm{out}} = N_0$ is the input

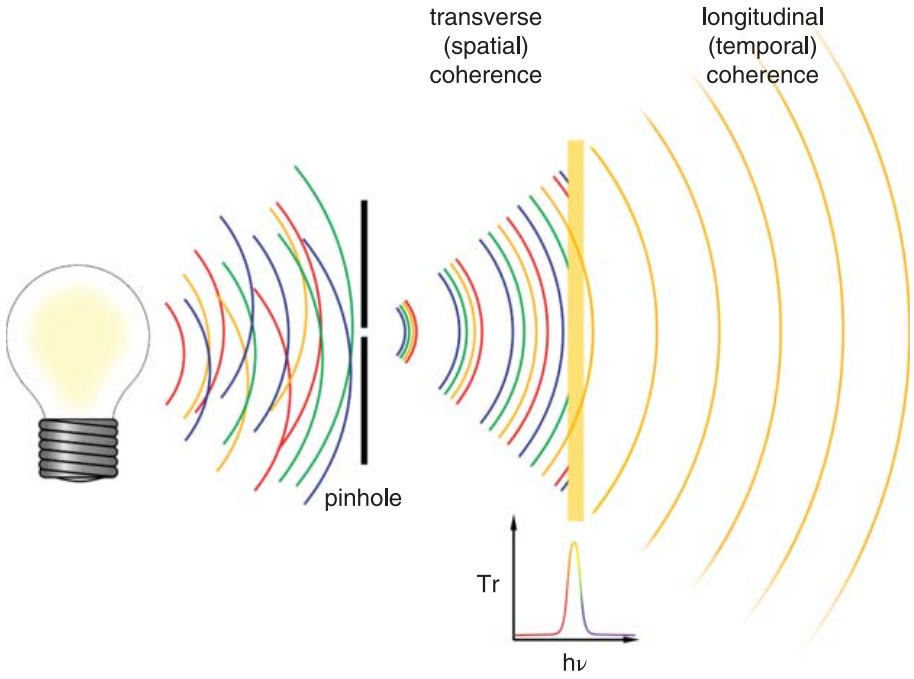


Figure 1.9.: Coherent radiation can be obtained from a broad, spatially extended source through a specific process. Initially, a pinhole is employed to function as a secondary, quasi-point source, allowing only a limited, spatially confined portion of the radiation to pass through. This results in an enhancement of transverse coherence because the product of the radiation's divergence and source size is significantly reduced, effectively reducing its emittance. Secondly, a filter, which could be a monochromator, is used to suppress all radiation except for a narrow bandwidth, much narrower than the original source's spectrum. At this stage, the radiation becomes both spatially and longitudinally coherent. Illustration from (Als-Nielsen and McMorrow 2011).

signal-to-noise ratio of the system. Here, a denotes the detector thickness, N_0 the photons, and μ the linear attenuation coefficient. In practical applications, DQE is typically not equal to 1 because of imperfections in the optical equipment.

For now, our focus is only on the *detector transfer function*, which explains how information spreads spatially across the detector. The *point spread function* (PSF) of the detector, denoted as $R(\mathbf{x})$, illustrates how the detector responds to a noiseless Dirac pulse $\delta D(\mathbf{x})$. In the real world, the recorded intensity $\mathbf{I}_D^{\text{obs}}(\mathbf{x})$ is derived in physical space by convolving the incident intensity with the point spread function.

$$\mathbf{I}_D^{obs}(\mathbf{x}) = R(\mathbf{x}) * \mathbf{I}_D(\mathbf{x}) \tag{1.45}$$

or equivalently by a multiplication in the Fourier domain

$$\widehat{\mathbf{I}}_D^{\text{obs}}(\mathbf{x}) = \widehat{R}(\mathbf{f})\widehat{\mathbf{I}}_D(\mathbf{f}) \tag{1.46}$$

where $\widehat{R}(\mathbf{f})$ is the detector transfer function, which is the Fourier transform of the point spread function.

II.5 Propagation regimes

II.5.1 Fresnel number

Recall that the wave $u_D(\mathbf{x})$ at a distance D from the sample is obtained by a convolution with the Fresnel propagator $P_D(\mathbf{x}) = \frac{1}{i\lambda D} \exp\left(i\frac{\pi}{\lambda D}|\mathbf{x}|^2\right)$. The quantity appearing in the exponential is related to the *Fresnel number*

$$\mathbf{f} = \frac{b^2}{\lambda D} \tag{1.47}$$

which is dimensionless. It is important to emphasize that the concept of a Fresnel number in an imaging setup doesn't stand alone. This is always associated with an implicit reference to a lateral lengthscale b which serves as the basis for defining unit length in the dimensionless coordinates being used.

The Fresnel number helps define various imaging regimes in X-ray phase contrast imaging:

1. *Contact regime* $\mathbf{f} \gg 1$: There are no observable effects related to Fresnel propagation. Consequently, the model described in (1.37) is applicable, there is an absence of phase contrast, we only get absorption information.
2. *Near-field regime* $\mathbf{f} \geq 1$: By propagating, we are sensitive to phase contrast. We observe edge enhancement and phase objects become detectable.
3. *Far-field regime* $\mathbf{f} \ll 1$: This is also known as *Fraunhofer diffraction*. The action of the Fresnel propagator is thus essentially given by a Fourier transform.

For clarity, Figure 1.10 illustrates intensities recorded for different propagation distances. On the intensity recorded at the exit plane (b) of the sample we only observe the pure-absorption objects (displayed in (a)). But as we move away from the object (c), pure-phase objects become visible, and phase information is built through propagation.

II.5.2 Far-field regime

The wave (1.38) can we rewritten under integral form

$$u_D(\mathbf{x}) = \frac{\exp(ikD)}{i\lambda D} \iint_\Sigma u_0(\mathbf{x}_0) \exp\left(i\frac{\pi}{\lambda D}\left[(x-x_0)^2 + (y-y_0)^2\right]\right) \mathrm{d}\mathbf{x}_0 \tag{1.48}$$

where $\mathbf{x}_0 = (x_0, y_0)$ and Σ is the support of the transmittance function. By developing the squares in the exponent, we obtained

$$u_D(\mathbf{x}) = \frac{\exp(ikD)}{i\lambda D} \exp\left(\frac{i\pi}{\lambda D}|\mathbf{x}|^2\right) \iint_\Sigma T(\mathbf{x}_0) \exp\left(i\frac{\pi}{\lambda D}|\mathbf{x}_0|^2\right) \exp\left[-i\frac{2\pi}{\lambda D}(\mathbf{x}_0 \cdot \mathbf{x})\right] \mathrm{d}\mathbf{x}_0 \tag{1.49}$$

The terms preceding the integrals, typically involving a phase component and a scaling factor, are commonly ignored or omitted. Following the previous section, if we assume the far-field regime, we get

$$\frac{\pi \max\left(|\mathbf{x}_0|^2\right)}{\lambda D} \ll 1 \tag{1.50}$$

This gives

$$u_D(\mathbf{x}) \propto \iint_\Sigma T(\mathbf{x}_0) \exp\left[-i\frac{2\pi}{\lambda D}(\mathbf{x}_0 \cdot \mathbf{x})\right] \mathrm{d}\mathbf{x}_0 \tag{1.51}$$

which is the Fourier transform of the transmittance function

$$u_D(\mathbf{x}) \propto \mathcal{F}T\left(\frac{\mathbf{x}}{\lambda D}\right) \tag{1.52}$$

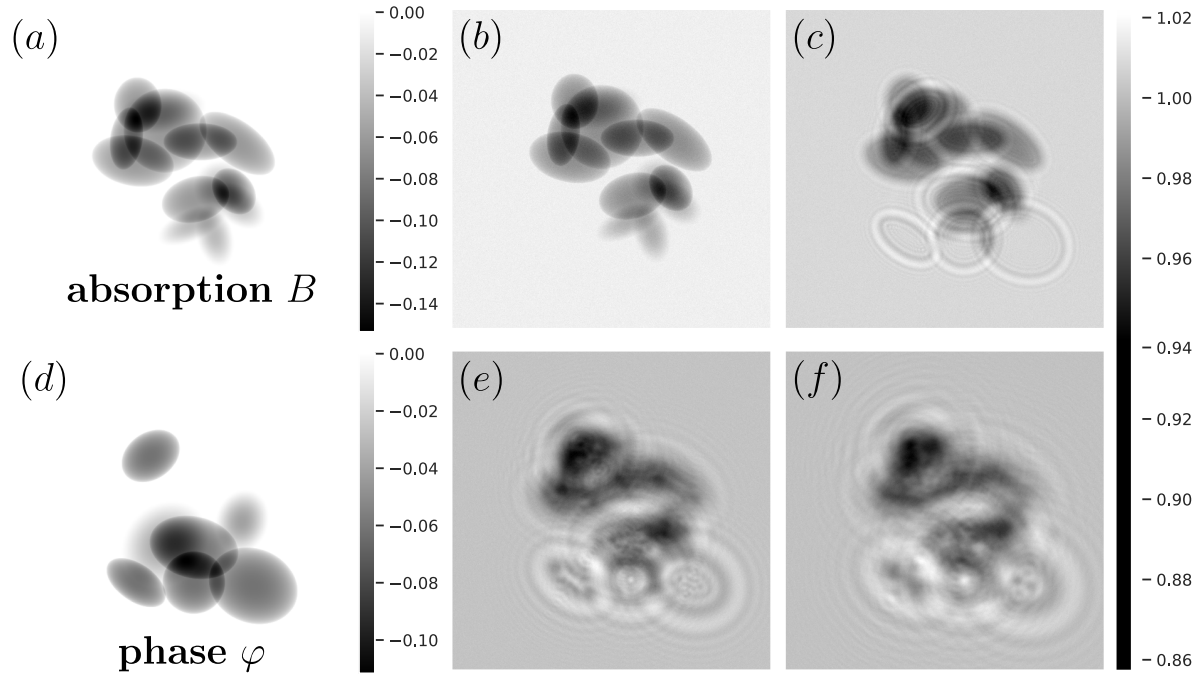


Figure 1.10.: Simulated diffraction patterns at different propagation distances for a combination of pure phase and pure absorption objects. The X-ray energy was set to 27 keV, the propagation distances $D = [0, 10, 50, 100]$ mm, and the pixel size is equal to 60 nm. (a), (d) assumed absorption and phase images. The simulated intensities displayed correspond to (b) $D = 0$ mm, (c) $D = 10$ mm, (e) $D = 50$ mm and (f) $D = 100$ mm.

Ultimately, measuring the wave's intensity in the diffraction plane yields:

$$\mathbf{I}_D(\mathbf{x}) = \left|\mathcal{F}T\left(\frac{\mathbf{x}}{\lambda D}\right)\right|^2 \tag{1.53}$$

i.e. the recorded intensity is the squared modulus of the Fourier transform of the transmittance function.

III Direct problem and forward operators

III.1 Nonlinear forward model

The forward operator of propagation-based imaging is defined by (1.42), it is also known as the *Fresnel transform* at a propagation distance D. It can be written as

$$\mathbf{F}_D(B, \varphi) = \left|P_D * \mathrm{e}^{-B+i\varphi}\right|^2 \tag{1.54}$$

Obviously, $\mathbf{F}_D$ is a *nonlinear* operator because of the exponential and the squared modulus operation. The inverse problem associated with this forward model can thus be expressed as follows:

Definition III.1 — General inverse problem. For some set A, we aim to reconstruct the

absorption and phase shift $(B, \varphi) \in A$ from the intensity measurement $\mathbf{I}_D^{\text{obs}}$ such that

$$\mathbf{I}_D^{\text{obs}} \approx \mathbf{F}_D(B, \varphi) \tag{1.55}$$

The set A of admissible images depends on available a priori knowledge on the images B and φ. As all physical problems, real world X-ray phase contrast imaging experiments never provide exact data (1.54), due to noise or other effects, that is why the sign $\approx$ is used in definition III.1. Solving the inverse problem (1.55) amounts to estimate the phase shift (resp. absorption) from the intensity $\mathbf{I}_D^{\text{obs}}$, or diffraction pattern, this process is called phase retrieval (resp. absorption retrieval).

III.2 Multi-distance formulation of the inverse problem

To gather more comprehensive details about the unknown images B and φ, it is customary to capture numerous diffraction patterns at varying distances $D = \{D_1, \ldots, D_{N_D}\}$ between the sample and detector. Configurations involving more than a single intensity can be represented by combining any of the aforementioned forward mappings into a vector-valued operator in a certain manner: let's denote by $\mathbf{I}_D = \left\{\mathbf{I}_{D_1}, \ldots, \mathbf{I}_{D_{N_D}}\right\}$ the intensities recorded, the forward operator which links the desired couple (B, φ) to the measurements, is defined by

$$\mathbf{F}_D = \begin{pmatrix} \mathbf{F}_{D_1} \\ \vdots \\ \mathbf{F}_{D_{N_D}} \end{pmatrix} : (B, \varphi) \mapsto \begin{pmatrix} \mathbf{F}_{D_1}(B, \varphi) \\ \vdots \\ \mathbf{F}_{D_{N_D}}(B, \varphi) \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{D_1} \\ \vdots \\ \mathbf{I}_{D_{N_D}} \end{pmatrix} \tag{1.56}$$

As we shall see, the acquisition of measures at different distances gives different diffraction patterns, which in turn provides frequency information. For the CTF-linearized model, this will be particularly useful.

III.3 Contrast Transfer Function linearized model

The Contrast Transfer Function (CTF) model is based on an assumption of weak absorption and *slowly varying phase shift*, characterized by sufficiently small phase-gradients $|\nabla\varphi|$:

$$B(\mathbf{x}) \ll 1, \quad |\varphi(\mathbf{x}) - \varphi(\mathbf{x} + \lambda D\mathbf{f})| \ll 1 \tag{1.57}$$

The forward model is linearized by Taylor expanding the transmittance function (1.33) to the first order

$$T(\mathbf{x}) \approx 1 - B(\mathbf{x}) + i\varphi(\mathbf{x}) \tag{1.58}$$

Substituting into (1.43) and again keeping only first order terms gives (Paganin 2006; P. Cloetens, Ludwig, et al. 1999):

$$\widehat{\mathbf{I}}_D(\mathbf{f}) \approx \delta(\mathbf{f}) - 2\cos\left(\pi\lambda D\,|\mathbf{f}|^2\right)\widehat{B}(\mathbf{f}) + 2\sin\left(\pi\lambda D\,|f|^2\right)\widehat{\varphi}(\mathbf{f}) \tag{1.59}$$

$\delta(\mathbf{f})$ is the unit impulse function, $\widehat{B}(\mathbf{f})$ is the Fourier transform of the absorption and $\widehat{\varphi}(\mathbf{f})$ is the Fourier transform of the phase. This approximation states that the Fourier transform of the linearized contrast $\left(\widehat{\mathbf{I}}_D(\mathbf{f}) - \delta(\mathbf{f})\right)$ is a superposition of the Fourier transform of phase and absorption image, modulated by the *contrast factors*, $c_0(\mathbf{f}) = \cos(\pi\lambda D\,|\mathbf{f}|^2)$ and $s_0(\mathbf{f}) =$

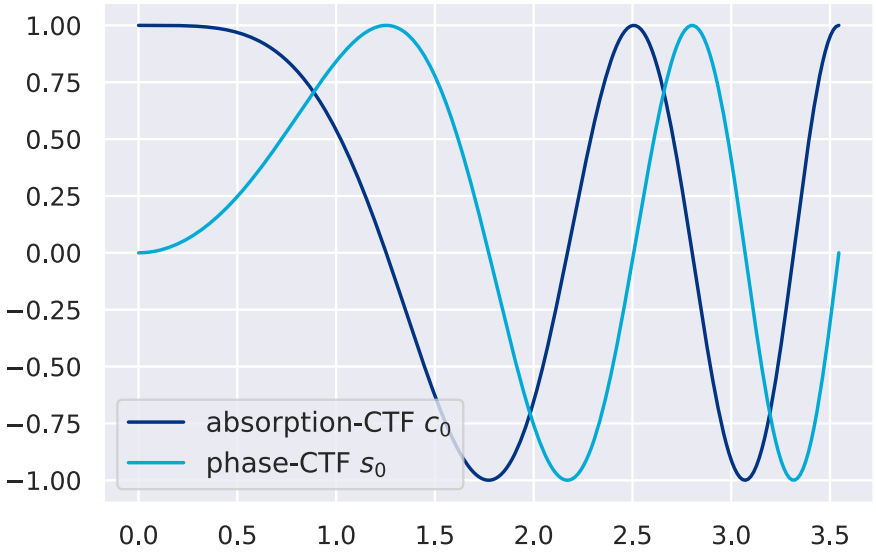


Figure 1.11.: Plot of the absorption and phase contrast transfer functions factors c_0 and s_0.

$\sin(\pi\lambda D\,|\mathbf{f}|^2)$, respectively:

$$\mathcal{F}\left(\mathbf{I}_D - 1\right)(\mathbf{f}) \approx -2c_0(\mathbf{f})\mathcal{F}(B)(\mathbf{f}) + 2s_0(\mathbf{f})\mathcal{F}(\varphi)(\mathbf{f}) \tag{1.60}$$

This result implies that attributing a specific characteristic in the data to either B or φ is generally challenging. Additionally, the points where the oscillatory functions c_0 and s_0, displayed in figure 1.11, intersect zero, correspond to Fourier components $\mathcal{F}(B)(\mathbf{f})$ or $\mathcal{F}(\varphi)(\mathbf{f})$ that can not be recovered. These components are inadequately represented in the intensity contrast $\mathbf{I}_D - 1$, making their recovery a challenging task. Notably, we observe that the second-order zero of the phase-CTF s_0 occurs at $\mathbf{f} = 0$, while for the absorption-CTF we have $c_0(0) = 1$. Consequently, coarse characteristics in the phase image φ induce low contrast.

This linear relationship in the Fourier domain between the intensity and (B, φ) allows us to define a new forward operator

$$\mathbf{F}_D^{\mathrm{CTF}}(B, \varphi) = -2\mathcal{F}^{-1}\left[\cos(\pi\lambda D\,|\mathbf{f}|^2)\widehat{B}(\mathbf{f}) - \sin(\pi\lambda D\,|f|^2)\widehat{\varphi}(\mathbf{f})\right] \tag{1.61}$$

Equation (1.61) will be referred to as the CTF-linearized forward model.

III.4 Homogeneity and pure-phase object constraints

III.4.1 Homogeneous object

Assuming we are working with a homogeneous object, which means that the refractive index decrement δ_r and the absorption β index are proportional, in other words, the ratio $c_{\beta/\delta_r} = \dfrac{\beta}{\delta_r t}$ is constant and $\beta = c_{\beta/\delta_r}\delta_r$. The *homogeneity* or *single-material constraint* can be imposed as a constraint, by defining $\varsigma = \arctan(c_{\beta/\delta_r})$, we can write the complex number

$$\varphi + iB = \mathrm{e}^{i\varsigma}\Psi \tag{1.62}$$

where Ψ is a real-valued function.

III.4.2 Pure-phase object

This definition of homogeneity includes the particular case of *pure-phase* or *non-absorbing object* when $\varsigma = 0$, i.e. $\Psi = \varphi$ and $B = 0$. This assumption can be useful for objects made of low-absorbent materials. Note that in the general case, since phase is a few orders of magnitude more sensitive than absorption, we will have $\beta \ll \delta_r$, so that $\mu \approx c_{\beta/\delta_r} \ll 1$ will be close to zero.

III.4.3 Consequences for the forward operators

Homogeneity and pure-phase objects are widely used in practice, so it is worth defining the equivalents of the forward operators in this particular case. The main advantage of imposing such constraints is that, instead of retrieving two unknowns B and φ, there is only one image Ψ (1.62) to retrieve. We can therefore rewrite the direct problem in terms of Ψ so that the constraint is incorporated directly into the model. For the general nonlinear operator (1.54), the forward operator becomes

$$\mathbf{F}_{D,\varsigma}(\Psi) = \left| P_D * \exp\left(-i\mathrm{e}^{i\varsigma}\Psi\right)\right|^2 \tag{1.63}$$

The CTF-linearized forward model (1.61) can be express as follows:

$$\mathbf{F}^{\mathrm{CTF}}_{D,\varsigma}(\Psi) = -2\mathcal{F}^{-1}\left(s_\varsigma \cdot \mathcal{F}(\Psi)\right), \quad \text{with} \quad s_\varsigma(\mathbf{f}) = \sin\left(\pi\lambda D\,|\mathbf{f}|^2 + \varsigma\right) \tag{1.64}$$

The new CTF-factors s_ς have analogous oscillatory behavior as the contrast-transfer-functions c_0 and s_0.

III.5 Properties of the forward operator

In this section, we focus on the properties of the direct operators, which will be useful when we will come to phase retrieval methods. These properties are discussed in detail in the thesis (Maretzke n.d.), and we will briefly discuss them here.

III.5.1 Domains of definition and constraints

In order to have a complete description of the mathematical model, we have to specify suitable model parameter space $\mathcal{P}$, data space $\mathcal{D}$, and admissible space A for the forward operator $\mathbf{F} : A \subset \mathcal{P} \to \mathcal{D}$.

Model parameter space

We will consider the absorption and the phase shift that describe the object belong to the Hilbert space of square-integrable functions on $\mathbb{R}^2$, i.e. $(B, \varphi) \in A \subset L^2(\mathbb{R}^2, \mathbb{R}) \times {}^2(\mathbb{R}^2, \mathbb{R})$ where

$$L^2(\mathbb{R}^2, \mathbb{R}) := \left\{ f : \mathbb{R}^2 \to \mathbb{R}, \quad \text{such that} \quad \|f\|_{L^2}^2 = \int_{\mathbb{R}^2} |f(\mathbf{x})|^2 \,\mathrm{d}\mathbf{x} < \infty \right\} \tag{1.65}$$

Data space

For all the forward maps $\mathbf{F} \in \left\{\mathbf{F}_D, \mathbf{F}_{D,\varsigma}, \mathbf{F}^{\mathrm{CTF}}_D, \mathbf{F}^{\mathrm{CTF}}_{D,\varsigma}\right\}$, we will see that the natural data-space for a single-distance measurement is $\mathcal{D} = L^2(\mathbb{R}^2, \mathbb{R})$. In case we have several propagation distances $D = \{D_i\}_{i=1}^{N_D}$, the data-space becomes $\mathcal{D} = \bigotimes_{i=1}^{N_D} L^2(\mathbb{R}^2, \mathbb{R})$.

Homogeneity constraints

In case of homogeneous object, or pure phase object, we saw that this amounts to retrieve a single real-valued function so the admissible set becomes $A = L^2(\mathbb{R}^2, \mathbb{R})$.

Support constraints

In practical applications, we often have prior information about the approximate dimensions of the imaged sample. This a priori knowledge can be modeled by a bounded object *support*:

$$\operatorname{supp}(f) = \overline{\left\{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) \neq 0\right\}} \subset \Omega \tag{1.66}$$

where f is the absorption or the phase shift, $\Omega \subset \mathbb{R}^2$ is some bounded support-domain and the overbar denotes the set-closure. *Support constraints* can be enforced by limiting the range of admissible objects to

$$A = L^2(\Omega, \mathbb{R}) = \left\{f \in L^2(\mathbb{R}^2, \mathbb{R}) : \operatorname{supp}(f) \subset \Omega\right\} \tag{1.67}$$

III.5.2 Continuity and differentiability

In the following, we will see that the different forward maps defined earlier are both well-defined and continuous within the model parameter and data spaces. Additionally, we will illustrate that these maps are differentiable, which is important for the development and analysis of algorithms aimed at approximating the inverse of these maps.

We will just mention the properties we are interested in and will not give any proof of them, which can be found in (Maretzke n.d.; Paganin 2006).

Properties of the Fresnel propagator

The regularity properties of the forward maps are mostly derived from those of the Fresnel propagator. Let us consider the direct nonlinear operator $\mathbf{F}_D$, which can be rewritten in the following form

$$\mathbf{F}_D(B, \varphi) = \left|\mathbb{P}_D\left(\mathrm{e}^{-B+i\varphi}\right)\right|^2 \tag{1.68}$$

where $\mathbb{P}_D$ represents the action of the Fresnel propagator and is given by

$$\mathbb{P}_D(z) = P_D * z = \mathcal{F}^{-1}\left(\mathcal{F}(P_D)\mathcal{F}(z)\right) = \mathcal{F}^{-1}\left(\mathrm{e}^{-i\pi\lambda D|\mathbf{f}|^2}\mathcal{F}(z)\right) \tag{1.69}$$

The properties of the Fresnel transform is thus given by those of

$$\begin{array}{rccl} \mathbb{P}_D : & L^2(\mathbb{R}^2, \mathbb{C}) & \to & L^2(\mathbb{R}^2, \mathbb{C}) \\ & g & \mapsto & \mathcal{F}^{-1}\left(\mathfrak{m}_D \cdot \mathcal{F}(g)\right), \quad \text{with} \quad \mathfrak{m}_D(\mathbf{f}) = \mathrm{e}^{-i\pi\lambda D|\mathbf{f}|^2} \end{array} \tag{1.70}$$

This operator is a bounded linear operator, which has the following properties (D. Paganin, S. C. Mayo, et al. 2002)

Theorem III.1

- **Unitary**: The operator $\mathbb{P}_D$ preserves the L^2-norm

$$\|\mathbb{P}_D(g)\|_2 = \|g\|_2, \quad \forall g \in L^2(\mathbb{R}^2, \mathbb{C}) \tag{1.71}$$

The inverse is given by the adjoint operator

$$\mathbb{P}_D^{-1}(g) = \mathbb{P}_D^*(g) = \mathcal{F}\left(\mathfrak{m}_D^{-1} \cdot \mathcal{F}^{-1}(g)\right) \tag{1.72}$$

In particular we have

$$\mathbb{P}_D \mathbb{P}_D^* = \mathbb{P}_D^* \mathbb{P}_D = \mathrm{Id}_{L^2(\mathbb{R}^2,\mathbb{C})} \tag{1.73}$$

- **Translation invariance**:

$$\mathbb{P}_D T_a = T_a \mathbb{P}_D, \quad \forall a \in \mathbb{R}^2 \quad \text{and} \quad T_a : g \mapsto g(\cdot + a) \tag{1.74}$$

- **Rotational invariance**: invariant under orthogonal coordinate transforms, i.e. for $A \in \mathbb{R}^{2\times 2}$ such that $AA^* = A^*A = \mathrm{Id}_{\mathbb{R}^2}$, we have

$$\mathbb{P}_D R = R \mathbb{P}_D, \quad \forall R : g \mapsto g\,[A(\cdot)] \tag{1.75}$$

Boundedness of CTF-linearized forward operator

A consequence of the linearity of the Fresnel propagator and its unitary property is the boundedness of linearized operators $\mathbf{F}_D^{\mathrm{CTF}}$ and $\mathbf{F}_{D,\mathcal{S}}^{\mathrm{CTF}}$.

Theorem III.2 — Boundedness of $\mathbf{F}_D^{\mathrm{CTF}}$ and $\mathbf{F}_{D,\mathcal{S}}^{\mathrm{CTF}}$. We have the following properties:

- The CTF-operator $\mathbf{F}_D^{\mathrm{CTF}} : L^2(\mathbb{R}^2,\mathbb{R}) \times L^2(\mathbb{R}^2,\mathbb{R}) \to L^2(\mathbb{R}^2,\mathbb{R})$ is bilinear and continuous with $\|\mathbf{F}_D^{\mathrm{CTF}}\| = 2$.
- The CTF-homogeneous-operator $\mathbf{F}_{D,\mathcal{S}}^{\mathrm{CTF}} : L^2(\mathbb{R}^2,\mathbb{R}) \to L^2(\mathbb{R}^2,\mathbb{R})$ is bilinear and continuous with $\|\mathbf{F}_{D,\mathcal{S}}^{\mathrm{CTF}}\| = 2$.

Well-definedness and continuity of the nonlinear forward map

It can be shown that the forward intensity operator $\mathbf{F}_D$ can be decomposed in simple operators ($\exp(\cdot)$ and $|.|^2$) and require some intermediate work on general L^p-spaces (Bruno Sixou et al. 2013). By combining the properties for the different sub-operators, one can prove the well-definedness and continuity of $\mathbf{F}_D$

Theorem III.3 — Well-definedness and continuity of $\mathbf{F}_D$. Let's Ω be a bounded subset of $\mathbb{R}^2$, the nonlinear operator

$$\begin{array}{rccc} \mathbf{F}_D : & L^2(\Omega,\mathbb{R}) \times L^2(\Omega,\mathbb{R}) & \to & L^2(\mathbb{R}^2,\mathbb{R}) \\ & (B,\varphi) & \mapsto & \left| P_D * \mathrm{e}^{-B+i\varphi} \right| \end{array} \tag{1.76}$$

is well-defined and continuous.

Fréchet-differentiability

The (Fréchet-)differentiability of $\mathbf{F}_D$ is crucial as it provides the mathematical basis for nonlinear image reconstruction via gradient descent, or Newton type methods (see VII)

Definition III.2 — Fréchet-differentiability. Let $\mathcal{X}$ and $\mathcal{Y}$ be two $\mathbb{R}$-vector spaces, $\mathcal{U} \subset X$ an open subset, $F : X \to Y$ and $a \in \mathcal{U}$. The map F is said to be *(Fréchet-)differentiable* at the point a if there exist a neighborhood $\mathcal{V} \in \mathcal{U}$ and a bounded linear operator $\mathcal{L} : \mathcal{X} \to \mathcal{Y}$ such that

$$F(a+h) \underset{h\to 0}{=} F(a) + \mathcal{L}(h) + o(\|h\|_X), \quad \forall a + h \in \mathcal{V} \tag{1.77}$$

where $\mathcal{L}$ is called the *differential* of F at a, we will use the notation $\mathcal{L} = F'[a]$ to show its dependence at point a.

Several mathematical subtleties need to be addressed to effectively describe the nonlinear

forward model of propagation-based imaging as a differential operator on L^2-spaces. The statement of the theorem that follows does not take into account all the technical details (which are detailed in (Maretzke n.d.)), and will be stated in a simplified way that will not affect the results of the next sections.

Theorem III.4 — Fréchet-differentiability of $\mathbf{F}_D$. Let Ω be a bounded subset of $\mathbb{R}^2$. The nonlinear forward operator $\mathbf{F}_D$ defined in (1.76) is Fréchet-differentiable. For $(B, \varphi) \in L^2(\Omega, \mathbb{R}) \times L^2(\Omega, \mathbb{R})$, the derivative is given by:

$$\begin{array}{rccc} \mathbf{F}'_D(B,\varphi): & L^2(\Omega,\mathbb{R}) \times L^2(\Omega,\mathbb{R}) & \to & L^2(\mathbb{R}^2) \\ & (f,g) & \mapsto & 2\mathrm{Re}\left\{\left(P_D * \left[(-f+ig)\mathrm{e}^{-B+i\varphi}\right]\right)\left(P_D * \mathrm{e}^{-B+i\varphi}\right)\right\} \end{array} \tag{1.78}$$

We will see later that what appears most often in nonlinear algorithms is not the derivative of the operator, but rather the adjoint of its derivative, namely $\mathbf{F}'_D(B,\varphi)^*$. Recall that the *adjoint* of a linear operator $A : \mathcal{X} \to \mathcal{Y}$ is the linear operator $A^* : \mathcal{Y} \to \mathcal{X}$ satisfaisant the rule

$$\langle Ax | y \rangle_{\mathcal{Y}} = \langle x | A^* y \rangle_{\mathcal{X}}, \quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}$$

where $\langle \cdot | \cdot \rangle_{\mathcal{X}}$ and $\langle \cdot | \cdot \rangle_{\mathcal{Y}}$ denotes the scalar product on the spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. Combining this definition with Theorem III.4, one can compute the explicit form of $\mathbf{F}'_D(B,\varphi)^*$.

Theorem III.5 — Adjoint of the Fréchet-differential of $\mathbf{F}_D$. The adjoint operator associated to the differential of $\mathbf{F}_D$ (1.78) is given by:

$$\begin{array}{rccc} \mathbf{F}'_D(B,\varphi)^*: & L^2(\mathbb{R}^2) & \to & L^2(\Omega,\mathbb{R}) \times L^2(\Omega,\mathbb{R}) \\ & h & \mapsto & 2\mathrm{Re}(v_1, v_2) \end{array} \tag{1.79}$$

where

$$v_1 = -\left\{P_D * \left[u\left(\overline{P_D} * \mathrm{e}^{-B+i\varphi}\right)\right]\right\} \mathrm{e}^{-B-i\varphi} \tag{1.80}$$

$$v_2 = \left\{P_D * \left[u\left(\overline{P_D} * \mathrm{e}^{-B+i\varphi}\right)\right]\right\} i\mathrm{e}^{-B-i\varphi} \tag{1.81}$$

IV Linear phase retrieval methods

Numerous X-ray techniques for phase retrieval have been suggested in the literature. Among these are methods that solve a linearized version of the problem, allowing direct reconstruction. One such method is based on the CTF-linearization (1.59), applicable when several propagation distances are available. Another approach is the Transport of Intensity Equation (TIE), which is valid when distances are small. Additionally, a hybrid strategy based on CTF has been proposed, transitioning to CTF with the assumption of weak absorption and to TIE for short distances.

IV.1 Contrast Transfer Function (CTF)

Since the contrast transfers functions factors c_0 and s_0 in (1.60) have zero crossings, several distances have to be used in order to cover as much of the Fourier domain as possible. By choosing multiple distances, typically three or four distances, zeros can be avoided. Then, a linear least squares optimization problem is considered, taking the different distances into

account, with the minimization of the sum

$$\min \left\{ \sum_{k=1}^{N_D} \left| \mathcal{F}\left[\mathbf{F}_{D_k}^{\mathrm{CTF}}(B,\varphi) \right](\mathbf{f}) - \widehat{\mathbf{I}}_{D_k}^{\mathrm{obs}}(\mathbf{f}) \right|^2 \right\} \tag{1.82}$$

where $\mathbf{I}_{D_1}^{\mathrm{obs}}, \cdots, \mathbf{I}_{D_{N_D}}^{\mathrm{obs}}$ denotes the recorded intensities. The problem (1.82) can be solved simultaneously for $\widehat{B}(\mathbf{f})$ and $\widehat{\varphi}(\mathbf{f})$ (Zabler et al. 2005). We then have two retrieval formulas for both absorption and phase

$$\widehat{B}(f) = \frac{1}{2\Delta + \alpha}[A \sum_{k=1}^{N_D} \widehat{\mathbf{I}}_{D_k}^{\mathrm{obs}}(\mathbf{f}) \sin(\pi\lambda D \left|\mathbf{f}\right|^2) - B \sum_{k=1}^{N_D} \widehat{\mathbf{I}}_{D_k}^{\mathrm{obs}}(\mathbf{f}) \cos(\pi\lambda D \left|\mathbf{f}\right|^2)] \tag{1.83}$$

$$\widehat{\varphi}(f) = \frac{1}{2\Delta + \alpha}[C \sum_{k=1}^{N_D} \widehat{\mathbf{I}}_{D_k}^{\mathrm{obs}}(\mathbf{f}) \sin(\pi\lambda D_k \left|\mathbf{f}\right|^2) - A \sum_{k=1}^{N_D} \widehat{\mathbf{I}}_{D_k}^{\mathrm{obs}}(\mathbf{f}) \cos(\pi\lambda D_k \left|\mathbf{f}\right|^2)] \tag{1.84}$$

where α is a Tikhonov regularization parameter and the coefficients are given by

$$A = \sum_{k=1}^{N_D} \sin(\pi\lambda D_k \left|\mathbf{f}\right|^2) \cos(\pi\lambda D \left|\mathbf{f}\right|^2) \quad B = \textstyle\sum_{k=1}^{N_D} \sin^2(\pi\lambda D_k \left|\mathbf{f}\right|^2) \tag{1.85}$$

$$C = \sum_{k=1}^{N_D} \cos^2(\pi\lambda D_k \left|\mathbf{f}\right|^2) \qquad \Delta = BC - A^2 \tag{1.86}$$

The parameter α is chosen to minimize the standard deviation outside the object.

IV.1.1 CTF homogeneous object

More recently, a single distance CTF has been introduced under homogeneous assumption (Turner et al. 2004; B. Yu et al. 2018). Considering a sample with a sufficiently homogeneous composition whose $c_{\delta_r/\beta} = \dfrac{\delta_r}{\beta}$ ratio is known. The phase can be retrieved using a single diffraction pattern, where the formula is given as:

$$\hat{\varphi}(\mathbf{f}) = \frac{1}{2} c_{\delta_r/\beta} \left[\frac{\widehat{\mathbf{I}}_D^{\mathrm{obs}}(\mathbf{f}) - \delta(\mathbf{f})}{\cos(D\pi\lambda \left|\mathbf{f}\right|^2) + c_{\delta_r/\beta} \sin(D\pi\lambda \left|\mathbf{f}\right|^2) + \alpha} \right] \tag{1.87}$$

IV.2 Transport of Intensity Equation (TIE)

The TIE method is based on the linearization of the transmittance function (1.33) with respect to the propagation distance D by Taylor expansion. If we express the first order term of the transmittance function

$$T\left(\mathbf{x} \pm \frac{\lambda D \mathbf{f}}{2}\right) \approx T(\mathbf{x}) \pm \frac{\lambda D \mathbf{f}}{2} \cdot \nabla T(\mathbf{x}) \tag{1.88}$$

Substituting this into (1.42), the intensity can be rewritten as

$$\mathbf{I}_D(\mathbf{x}) = \mathbf{I}_0(\mathbf{x}) - \frac{\lambda D}{2\pi} \nabla \left[\mathbf{I}_0(\mathbf{x}) \nabla \varphi(\mathbf{x}) \right] \tag{1.89}$$

This expression is only valid for *short* propagation distances D due to the Taylor expansion. The difference

$$\frac{\mathbf{I}_D(\mathbf{x}) - \mathbf{I}_0(\mathbf{x})}{D} = -\frac{\lambda}{2\pi}\nabla\left[\mathbf{I}_0(\mathbf{x})\nabla\varphi(\mathbf{x})\right] \tag{1.90}$$

can thus be used to approximate the partial derivative in the propagation direction z

$$\nabla\left[\mathbf{I}_0(\mathbf{x})\nabla\varphi(\mathbf{x})\right] = -\frac{2\pi}{\lambda}\cdot\frac{\partial}{\partial z}\mathbf{I}_0(\mathbf{x}) \tag{1.91}$$

This equation is known as the transport of intensity equation (Teague 1982).

Various strategies have been proposed to address the TIE, with solutions encompassing tackling the corresponding two-dimensional linear partial differential equation (Teague 1983; T. E. Gureyev and Nugent 1996; Allen and Oxley 2001) or employing Fourier-based techniques (D. Paganin and Nugent 1998; D. Paganin, S. C. Mayo, et al. 2002; Paganin 2006). The attention herein is directed towards the Fourier solution devised by Paganin and Nugent (D. Paganin and Nugent 1998; Paganin 2006), a methodology that shares implementation similarities with the other showcased phase retrieval techniques. By approximating $\mathbf{I}_0(\mathbf{x})\nabla\varphi(\mathbf{x})$ with the gradient of the scalar potential, an expression for the phase map can be obtained

$$\varphi(\mathbf{x}) = -\frac{2\pi}{\lambda}\nabla_\alpha^{-2}\left(\nabla\cdot\left\{\frac{1}{\mathbf{I}_0(\mathbf{x})}\nabla\left[\nabla_\alpha^{-2}\frac{\partial}{\partial z}\mathbf{I}_0(\mathbf{x})\right]\right\}\right) \tag{1.92}$$

where ∇_α^{-2} is the inverse Laplacian operator which can be implemented in the Fourier domain using

$$\nabla_\alpha^{-2} = -\frac{1}{4\pi^2}\mathcal{F}^{-1}\left(\frac{1}{|\mathbf{f}|^2+\alpha}\right)\mathcal{F} \tag{1.93}$$

The regularization term α is introduced in order to handle the singularity at $\mathbf{f}=0$ as well as the low frequencies where the Laplacian has small values. This approach requires the acquisition of two intensity measurements, at the exit plane of the sample for $\mathbf{I}_0(\mathbf{x})$ and at a short distance D from the sample for $\mathbf{I}_D(\mathbf{x})$.

IV.2.1 TIE homogeneous object

A single-distance phase retrieval method was developed by Paganin (D. Paganin, S. C. Mayo, et al. 2002) in the case of a homogeneous object having a constant ratio $c_{\delta_r/\beta} = \frac{\delta_r}{\beta}$. The phase reconstruction formula can be written as:

$$\widehat{\varphi}(\mathbf{f}) = \frac{1}{2}c_{\delta_r/\beta}\cdot\log\left(\frac{\mathcal{F}\left[\frac{\mathbf{I}_D}{\mathbf{I}_0}\right](\mathbf{f})}{1+\lambda D\pi c_{\delta_r/\beta}|\mathbf{f}|^2}\right) \tag{1.94}$$

Due to its straightforward formulation, (1.94) is also applied to scenarios involving inhomogeneous objects (T. Weitkamp et al. 2011). This application aims to acquire a non-quantitative phase map, which in turn simplifies the process of segmenting and visualizing the internal structures within the examined object.

In (B. Yu et al. 2018), they extend the Paganin's method (1.94) in case we have multiple

distances $\{D_k\}_{k=1}^{N_D}$

$$\widehat{\varphi}(\mathbf{f}) = \frac{1}{2} c_{\delta_r/\beta} \cdot \log \left(\frac{\frac{1}{N_D} \sum_{k=1}^{N_D} \left(1 + \lambda D_k \pi c_{\delta_r/\beta} |\mathbf{f}|^2\right) \cdot \mathcal{F}\left[\frac{\mathbf{I}_{D_k}}{\mathbf{I}_0}\right](\mathbf{f})}{\frac{1}{N_D} \sum_{k=1}^{N_D} \left(1 + \lambda D_k \pi c_{\delta_r/\beta} |\mathbf{f}|^2\right)^2} \right) \tag{1.95}$$

Recently, the application of the homogeneous object approximation has been expanded to encompass scenarios involving two or more homogeneous materials (Beltran et al. 2010). This approach, operating under the premise that the thickness of the enclosing material changes gradually, necessitates information about the overall projected thickness of the object.

IV.3 Mixed approach

As we have seen, the CTF and TIE methods are based on different assumptions, so they have their own strengths and weaknesses. The CTF is only valid for objects with weak absorption and use multiple distances. The TIE, on the other hand, requires images in two planes, but is restricted to short propagation distances where the contrast is weak. The desirability of merging these two approaches has led to attempts aimed at broadening the applicability of these approximations. Several techniques have been suggested with the intention of achieving this combination (T. Gureyev et al. 2004; Meng, H. Liu, and X. Wu 2007).

It has been shown that when $D \to 0$, the TIE and the CTF models do not yield the same expression (J. P. Guigay et al. 2007). To address this, a *mixed approach* has been introduced that connects the CTF and TIE, alleviating the impact of weak absorption assumptions in the CTF and the need for short propagation distances in the TIE. By Taylor expanding the phase term in (1.43), and retaining only the first term, the Fourier transform of the intensity yields

$$\widehat{\mathbf{I}}_D(\mathbf{f}) = \iint_{\mathbb{R}^2} \mathrm{e}^{-2i\pi\mathbf{x}\cdot\mathbf{f}} a\left(\mathbf{x} - \frac{\lambda D\mathbf{f}}{2}\right) a\left(\mathbf{x} + \frac{\lambda D\mathbf{f}}{2}\right) \left[1 + i\varphi\left(\mathbf{x} - \frac{\lambda D\mathbf{f}}{2}\right) - i\varphi\left(\mathbf{x} - \frac{\lambda D\mathbf{f}}{2}\right)\right] \mathrm{d}\mathbf{x} \tag{1.96}$$

With suitable variable change, and assuming that the absorption is slowly varying

$$|B(\mathbf{x} + \lambda D\mathbf{f}) - B(\mathbf{x} - \lambda D\mathbf{f})| \ll 1 \tag{1.97}$$

the equation (1.96) becomes

$$\begin{aligned} \widehat{\mathbf{I}}_D(\mathbf{f}) = \widehat{\mathbf{I}}_D^{\{\varphi=0\}}(\mathbf{f}) + 2\sin\left(\pi\lambda D |\mathbf{f}|^2\right) \iint_{\mathbb{R}^2} \mathrm{e}^{-2i\pi\mathbf{x}\cdot\mathbf{f}} \varphi(\mathbf{x}) a^2(\mathbf{x}) \mathrm{d}\mathbf{x} \\ +2\cos\left(\pi\lambda D |\mathbf{f}|^2\right) \lambda D\mathbf{f} \iint_{\mathbb{R}^2} \mathrm{e}^{-2i\pi\mathbf{x}\cdot\mathbf{f}} \varphi(\mathbf{x}) a(\mathbf{x}) \nabla a(\mathbf{x}) \mathrm{d}\mathbf{x} \end{aligned} \tag{1.98}$$

Here, $\widehat{\mathbf{I}}_D^{\{\varphi=0\}}(\mathbf{f})$ denotes the Fourier transform of the intensity when the phase shift is zero. We recognize in (1.98) that the integrals that appear are Fourier transforms, so this is rewritten in a simpler way:

$$\widehat{\mathbf{I}}_D(\mathbf{f}) = \widehat{\mathbf{I}}_D^{\{\varphi=0\}}(\mathbf{f}) + 2\sin\left(\pi\lambda D |\mathbf{f}|^2\right) \mathcal{F}[\mathbf{I}_0\varphi](\mathbf{f}) + \frac{\lambda D}{2\pi} \cos\left(\pi\lambda D |\mathbf{f}|^2\right) \mathcal{F}[\nabla \cdot (\nabla \mathbf{I}_0 \varphi)](\mathbf{f}) \tag{1.99}$$

From this, we can observe that this generalizes both the CTF and TIE. Indeed, when the distance D tends to 0, (1.99) becomes the TIE equation (1.91). And if the absorption is assumed to

Tableau 1.1.: Summary of the phase retrieval methods.

Method	Validity	Propagation distances
TIE	short propagation distance	2 (close)
Paganin	homogeneous object	1
CTF	Weak absorption and slowly varying phase	2
Mixed	Slowly varying object	2

be weak, i.e. when $\mathbf{I}_0(\mathbf{x}) \to 0$, we get the CTF solution (1.84). Equation (1.99) can be solved iteratively when regarding $\mathcal{F}[\mathbf{I}_0\varphi](\mathbf{f})$ as the unknown. Several propagation distances can be considered through a linear least squares method, the minimization problem is written as

$$\varphi_{n+1} = \underset{\varphi}{\operatorname{argmin}} \left\{ \sum_{k=1}^{N_D} \left| A_{D_k}(\mathbf{f}) \mathcal{F}[\mathbf{I}_0\varphi](\mathbf{f}) + \widehat{\mathbf{I}}_{D_k}^{\{\varphi=0\}}(\mathbf{f}) + \Delta_{D_k}(\mathbf{f}, \varphi_n) - \widehat{\mathbf{I}}_{D_k}(\mathbf{f}) \right| \right\} \tag{1.100}$$

where

$$A_D(\mathbf{f}) = 2\sin\left(\pi\lambda D |\mathbf{f}|^2\right) \tag{1.101}$$

$$\Delta_D(\mathbf{f}, \varphi_n) = \frac{\lambda D}{2\pi} \cos\left(\pi\lambda D |\mathbf{f}|^2\right) \mathcal{F}[\nabla \cdot (\varphi_n \nabla \mathbf{I}_0)](\mathbf{f}) \tag{1.102}$$

and φ_n is the phase at iteration n. The problem (1.100) has an explicit formula given by

$$\varphi_{n+1} = \frac{\sum_{k=1}^{N_D} A_{D_k}(\mathbf{f}) \left[\widehat{\mathbf{I}}_{D_k}(\mathbf{f}) - \widehat{\mathbf{I}}_{D_k}^{\{\varphi=0\}}(\mathbf{f}) - \Delta_{D_k}(\mathbf{f}, \varphi_n) \right]}{\sum_{k=1}^{N_D} \left| A_{D_k}^2(\mathbf{f}) \right|^2 + \alpha} \tag{1.103}$$

where α is a regularization parameter and the first iteration is assumed to be $\varphi_0 = 0$. Several regularization methods have been studied to solve this linear inverse problem, including classical Tikhonov regularization (Max Langer, Peter Cloetens, J.-P. Guigay, et al. 2008), wavelet shrinkage (Max Langer, Peter Cloetens, and Françoise Peyrin 2009) and phase-absorption duality prior (M. Langer, P. Cloetens, and F. Peyrin 2010). The mixed algorithm has been expanded to encompass multi-material objects, incorporating certain a priori knowledge. This enhancement involves employing regularization schemes, which begin with a tomographic reconstruction of attenuation. Subsequently, the regularization functional incorporates the relationship between the attenuation index β and the refractive index δ_r (Max Langer, Pacureanu, et al. 2012). In (Max Langer, Peter Cloetens, J.-P. Guigay, et al. 2008), a comparative analysis of the CTF, the TIE, and the mixed approach was conducted for scenarios involving mixed absorption and phase objects. The mixed approach exhibited higher accuracy and noise resilience compared to the other two methods. In contrast, the TIE showed better accuracy when no noise was involved. A summary of the methods and their validity is given in Tab. 1.1.

V Projection-based methods

Direct inversion methods are based on the linearization of the forward model and are only valid under some assumptions on the imaging conditions, either on the object or on the propagation distance. Such analytical approaches do not take into account the nonlinear information and

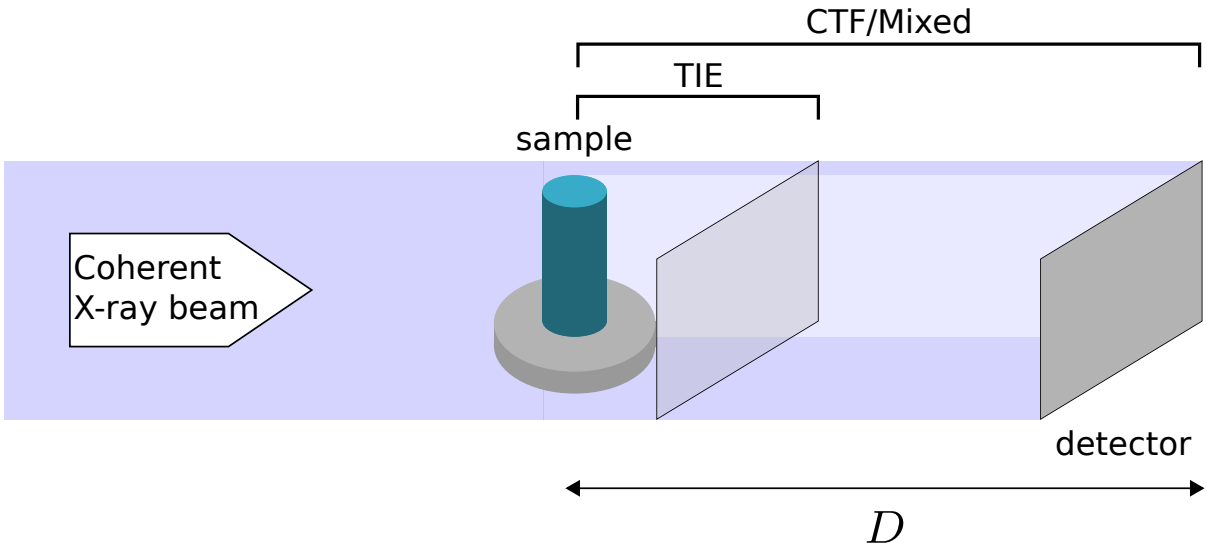


Figure 1.12.: Schematic of the imaging system. Phase contrast can be recorded at several distances, typically 2 using the TIE and 4 using the CTF and mixed approaches.

cannot include a non-negativity constraint, as this would be equivalent to solving a non-linear equation. Iterative algorithms are not limited by these constraints, among these are methods which we refer to as *alternating-projection* (Gerchberg 1972; James R. Fienup 1982) which are commonly applied to the classical phase retrieval problem of recovering an image f from the magnitude of its Fourier transform $|\mathcal{F}(f)|^2$. These challenges manifest in X-ray coherent diffractive imaging (CDI), which serves as the far-field imaging counterpart to X-ray phase contrast imaging (see II.5). This algorithmic approach can be easily modified for application in the near-field scenario of Fresnel diffraction.

V.1 Error-reduction algorithm

The *Gerchberg–Saxton* (GS) algorithm (Gerchberg 1972) is an iterative technique used for phase retrieval and image reconstruction. It addresses the problem of determining the phase from two intensity measurements. The GS algorithm iterates between the spatial and Fourier domains, using amplitude constraints to update the phase in the spatial domain. This interplay of constraints gradually refines the phase estimate to be consistent with the measured amplitudes.

The fundamental structure of this algorithm is known as the error reduction (ER) algorithm (J. R. Fienup 1978; James R. Fienup 1982). The aim of the ER algorithm is to estimate the object function $T(\mathbf{x})$, it consists in four steps:

1. Fourier transform of the current estimate $T_k(\mathbf{x})$, yielding $\widehat{T}_k(\mathbf{f})$
2. Replace the modulus of $\widehat{T}_k(\mathbf{f})$ with the measured Fourier modulus. This yields $\widehat{T}'_k(\mathbf{f})$, an estimate of the Fourier transform $\widehat{T}(\mathbf{f})$
3. Inverse Fourier transform to get $T'_k(\mathbf{x})$
4. Replace the modulus of $\widehat{T}_k(\mathbf{f})$ with the measured object modulus to obtain $T_{k+1}(\mathbf{x})$, a new estimate of the object transmission function

Suppose we have measured the intensity at the object's exit $\mathbf{I}_0(\mathbf{x})$ and the intensity $\mathbf{I}_D(\mathbf{x})$ in the diffraction plane, the ER algorithm is summarized in Algo. 3. It just comes down to convert between the two domains (Object-Fourier), ensuring that the constraints are met in one domain before transitioning back to the other.

Algorithm 3 Error-Reduction algorithm

for $k = 0, \ldots$ **do** :

1. $\widehat{T}_k(\mathbf{f}) = \mathcal{F}T_k(\mathbf{f}) = \left|\widehat{T}_k(\mathbf{f})\right| \exp\left[i\theta_k(\mathbf{f})\right]$
2. $\widehat{T}'_k(\mathbf{f}) = \sqrt{\mathbf{I}_D(\mathbf{f})} \exp\left[i\theta_k(\mathbf{f})\right]$
3. $T'_k(\mathbf{x}) = \mathcal{F}^{-1}\widehat{T}'_k(\mathbf{f}) = \left|T'_k(\mathbf{x})\right| \exp\left[i\varphi_k(\mathbf{x})\right]$
4. $T_{k+1}(\mathbf{x}) = \sqrt{\mathbf{I}_0(\mathbf{x})} \exp\left[i\varphi_k(\mathbf{x})\right]$

In the case where the contact object measurement $\mathbf{I}_0(\mathbf{x})$ is not available and we only have a single intensity measurement $\mathbf{I}_D(\mathbf{x})$, the first three steps of Algo. 3 remain the same and the last step is replaced by

$$T_{k+1}(\mathbf{x}) = \begin{cases} T'_k(\mathbf{x}), & \text{if } \mathbf{x} \notin \gamma \\ 0, & \text{if } \mathbf{x} \in \gamma \end{cases} \tag{1.104}$$

where C represents the set of points at which $T'_k(\mathbf{x})$ violates the constraints of the object domain. Such constraints may include non-negativity, real valuedness and object support.

▪ **Example 1.1** If we want the support of the object we are looking for to be contained in a bounded set $\Omega \subset \mathbb{R}^2$, and if we assume non-negativity of the object, the set C can be written as

$$\gamma = \left\{\mathbf{x} \in \mathbb{R}^2 : T'_k(\mathbf{x}) < 0 \text{ or } T'_k(\mathbf{x}) \cdot \mathbb{1}_\Omega(\mathbf{x}) = 0\right\} \tag{1.105}$$

where $\mathbb{1}_A$ is the characteristic function of the set A defined by

$$\mathbb{1}_A(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in A \\ 0, & \text{if } \mathbf{x} \notin A \end{cases} \tag{1.106}$$

The typical initiation of the algorithm involves employing an array of random numbers for the initial value of $T_0(\mathbf{x})$. But in practice, initialization is often performed using a linear method. Figure 1.13 illustrates a block diagram of the ER algorithm. This algorithm can be understood through various perspectives, for instance, it can be seen as a steepest descent gradient search (James R. Fienup 1982).

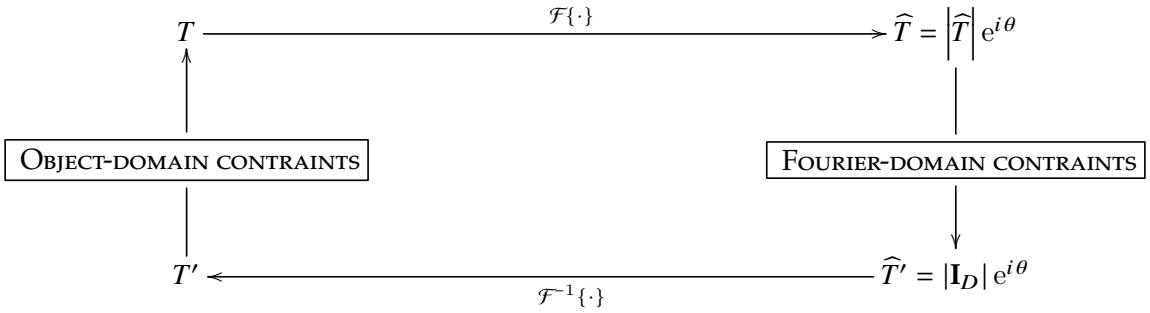


Figure 1.13.: Diagram of the error-reduction algorithm

V.1.1 Error-reduction as projections onto sets

Another point of view is given by (Bauschke, Combettes, and D. Russell Luke 2002) where the authors propose to make the connection with classical convex optimization algorithms. The set

of complex images that satisfy the Fourier domain constraint (Step 1-3 of algorithm 3) is

$$M = \left\{ T \in L^2(\mathbb{R}^2, \mathbb{C}) : \widehat{T}(\mathbf{f}) = \sqrt{\mathbf{I}_D(\mathbf{f})} \right\} \tag{1.107}$$

The ER algorithm with a single diffraction pattern (1.104) can thus be rewritten as

$$T_{k+1}(\mathbf{x}) = \begin{cases} P_M\left(T_k\right)(\mathbf{x}), & \text{if } \mathbf{x} \notin \gamma \\ 0, & \text{if } \mathbf{x} \in \gamma \end{cases} \tag{1.108}$$

where P_M is the projection operator (see (Bauschke, Combettes, and D. Russell Luke 2002) for more details). Let's assume that the object domain constraints are reduced to a support constraint Ω, (1.108) simplifies to

$$T_{k+1}(\mathbf{x}) = P_S P_M\left(T_k\right)(\mathbf{x}) \tag{1.109}$$

where $P_S(T) = T \cdot \mathbb{1}_\Omega$ is the projection on the support. In particular, the ER algorithm can be seen as a nonconvex version of projections onto convex sets (POCS) algorithm (Levi and Stark 1984).

V.2 Hybrid input-output algorithm

Numerous adaptations of the ER algorithm have been made for phase retrieval using a single diffraction pattern. The most effective approach among these is the *hybrid input-output* (HIO) algorithm, as introduced by Fienup (James R. Fienup 1982). The HIO algorithm consists of changing the last step of the ER algorithm, this can be written as

$$T_{k+1}(\mathbf{x}) = \begin{cases} P_M\left(T_k\right)(\mathbf{x}), & \text{if } \mathbf{x} \notin \gamma \\ T_k(\mathbf{x}) - \beta P_M\left(T_k\right)(\mathbf{x}), & \text{if } \mathbf{x} \in \gamma \end{cases} \tag{1.110}$$

where β is a constant. In contrast to the error-reduction algorithm, the *input* function $T_k(\mathbf{x})$ no longer needs to be perceived as the present optimal approximation of the object $T(\mathbf{x})$. Instead, it can be regarded as the driving factor for the forthcoming output, denoted as $T'_k(\mathbf{x}) = P_M\left(T_k\right)(\mathbf{x})$ The introduction of the constant β allows to gently nudges the solution toward the constraint where it is not satisfied.

In (Bauschke, Combettes, and D. Russel Luke 2003), they show that the form of the constraint set γ is crucial to the implementation and the study of HIO algorithm. The simplest case to consider is when there is only one support constraint.

Proposition V.1 — HIO with support constraint only. If the object domain constraint set incorporates only a support constraint Ω, then (1.110) is equivalent to

$$T_{k+1}(\mathbf{x}) = \frac{1}{2}\left[R_S\left(R_M + (\beta - 1)P_M\right) + \mathrm{Id} + (1-\beta)P_M\right](\mathbf{x}) \tag{1.111}$$

where R_S and R_M are reflection operators defined by

$$R_S = 2P_S - \mathrm{Id} \quad \text{and} \quad R_M = 2P_M - \mathrm{Id} \tag{1.112}$$

If we choose $\beta = 1$ then we end up with

$$T_{k+1}(\mathbf{x}) = \frac{1}{2}\left[R_S R_M + \mathrm{Id}\right](\mathbf{x}) \tag{1.113}$$

which is equivalent to the nonconvex version of the Douglas-Rachford projection algorithm (Bauschke, Combettes, and D. Russell Luke 2002; Aragón Artacho, Campoy, and Tam 2020).

If we now add a positivity constraint, we find the γ set (1.105) defined in the example 1.1.

$$T_{k+1}(\mathbf{x}) = \begin{cases} P_M\left(T_k\right)(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega \text{ and } P_M\left(T_k\right)(\mathbf{x}) \geq 0 \\ T_k(\mathbf{x}) - \beta P_M\left(T_k\right)(\mathbf{x}), & \text{otherwise} \end{cases} \tag{1.114}$$

Unlike the above V.1 here we can not rewrite the algorithm using reflection operations. This is because the projection P_{S^+} on the set of object constraints

$$S^+ = \left\{\mathbf{x} \in \mathbb{R}^2 : P_M\left(T_k\right)(\mathbf{x}) \geq 0 \text{ and } \left(P_M\left(T_k\right) \cdot \mathbb{1}_{\complement\Omega}\right)(\mathbf{x}) = 0\right\} \tag{1.115}$$

is a nonlinear operator (where $\complement\Omega$ denotes the complementary set of Ω).

V.3 Hybrid projection reflection algorithm

When there is only one support constraint, the HIO algorithm can be seen as a non-convex version of a classical optimization algorithm. Motivated to make this connection in the case of a non-negativity constraint, the authors (Bauschke, Combettes, and D. Russel Luke 2003) proposed a new algorithm called *hybrid projection reflection* (HPR), where the iterates are given by

$$T_{k+1}(\mathbf{x}) = \begin{cases} P_M\left(T_k\right)(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega \text{ and } R_M(T_k)(\mathbf{x}) \geq (1-\beta)P_M(T_k)(\mathbf{x}) \\ T_k(\mathbf{x}) - \beta P_M\left(T_k\right)(\mathbf{x}), & \text{otherwise} \end{cases} \tag{1.116}$$

The main difference with HIO is that the positivity condition is relaxed and now depends on the reflection R_M and the constant β. Now, we have a result analogous to proposition V.1, but this time in the case where the constraints are given by S^+ (1.115).

Proposition V.2 — HPR with support and nonnegativity constraints. If the object domain constraint set is given by S^+, then the HPR algorithm (1.116) is equivalent to

$$T_{k+1}(\mathbf{x}) = \frac{1}{2}\left[R_{S^+}\left(R_M + (\beta-1)P_M\right) + \mathrm{Id} + (1-\beta)P_M\right](\mathbf{x}) \tag{1.117}$$

Following (Bauschke, Combettes, and D. Russel Luke 2003), this new algorithm is not only useful to make the connection with convex optimization algorithms, but also to improve the quality of the reconstruction obtained. They showed that the HIO algorithm achieves its lowest error value at a faster pace compared to the HPR algorithm, but the HPR algorithm ultimately results in a reduction of the error metric beyond that obtained by the HIO algorithm.

V.4 Relaxed averaged alternating reflection

The principles of convex optimization offer a framework to formulate and understand analogous approaches for the nonconvex challenges encountered in phase retrieval problems. Approaches like the *relaxed averaged alternating reflection* (RAAR) have been suggested within the context of convex scenarios (D Russel Luke 2005). This new algorithm is inspired by the HPR algorithm, that we have seen involved a single-parameter relaxation derived from the widely recognized Douglas–Rachford algorithm as applied to phase retrieval. The RAAR algorithm is given by the

following updates

$$T_{k+1}(\mathbf{x}) = \begin{cases} P_M\left(T_k\right)(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega \text{ and } R_M(T_k)(\mathbf{x}) \geq 0 \\ \beta_k T_k(\mathbf{x}) - (1-2\beta_k) P_M\left(T_k\right)(\mathbf{x}), & \text{otherwise} \end{cases} \tag{1.118}$$

The update rule in RAAR algorithm (1.118) depends on the pointwise sign of the reflector $R_M(T_k)(\mathbf{x})$ whereas the update rule for Fienup's HIO algorithm (1.110) depends on the pointwise sign of the projector $P_M(T_k)(\mathbf{x})$. Moreover, instead of considering a single constant β, the authors propose to consider a sequence of steps $\{\beta_k\}_{k\in\mathbb{N}}$ which can be chosen automatically (given the initial value β_0) following the rule

$$\beta_{k+1} = \beta_0 + (1-\beta_0)\left(1 - \exp\left[-\left(\frac{n}{7}\right)^3\right]\right) \tag{1.119}$$

As with other algorithms, the RAAR updates can be rewritten as follows:

Proposition V.3 The RAAR algorithm (1.118) is equivalent to

$$T_{k+1}(\mathbf{x}) = \left[\frac{\beta_k}{2}\left(R_{S^+}R_M + \mathrm{Id}\right) + (1-\beta_k)P_M\right](\mathbf{x}) \tag{1.120}$$

Note that when $\beta = 1$, then we find the HPR algorithm, but when $\beta \neq 1$, these two algorithms are substantially different (D Russel Luke 2005). This intuitive framework proposed in the RAAR algorithm can be effectively analyzed mathematically and provides a straightforward approach for selecting a relaxation parameter, which further enhances the performance of the algorithm.

V.5 Difference map

Considering two arbitrary projections P_1 and P_2, we define the *difference map* of these projections (Elser 2003) by

$$\begin{array}{rccc} D: & L^2(\mathbb{R}^2,\mathbb{C}) & \to & L^2(\mathbb{R}^2,\mathbb{C}) \\ & T & \mapsto & T + \beta\Delta(T) \end{array} \tag{1.121}$$

where β is a non-zero real parameter and

$$\Delta = P_1 \circ f_2 - P_2 \circ f_1 \tag{1.122}$$

is the difference of the two projection operators, each composed with a map $f_i : L^2(\mathbb{R}^2,\mathbb{C}) \to L^2(\mathbb{R}^2,\mathbb{C})$. The specific structure of the maps f_i is of lesser importance compared to the overall behavior of the difference map. However, the typical selection is

$$f_i(T) = (1+\eta_i)P_i(T) - \eta_i T \tag{1.123}$$

which can be perceived as defining a parameterized line between T and $P_i(T)$ in the solution space, where T is regarded as a point, and the parameter η_i plays a defining role. Interestingly, the most effective option for the parameters η_i is to assign $\eta_1 = -\beta^{-1}$ and $\eta_2 = \beta^{-1}$, resulting in the simplified one-parameter configuration of the difference map. Consequently, a single iteration of the algorithm can be expressed as follows:

$$T_{k+1} = D(T_k) = T_k + \beta\left(P_1\left[(1+\beta^{-1})P_2(T_k) - \beta^{-1}T_k\right] - P_2\left[(1-\beta^{-1})P_1(T_k) - \beta^{-1}T_k\right]\right) \tag{1.124}$$

If T^* is a fixed point of D, it is characterized by $\Delta(T^*) = 0$, which is equivalent to

$$(P_1 \circ f_2)(T^*) = T_{1\cap 2} = (P_2 \circ f_1)(T^*) \tag{1.125}$$

Note that T^* is not necessarily a solution, but $T_{1\cap 2}$ is and must lie in the intersection of the corresponding constraint subspaces. The progress of the iterations T_k can be tracked by computing the norm of the difference

$$\varepsilon_k = \|\Delta(T_k)\|_2 \tag{1.126}$$

We observe that when $T_k \to 0$, the metric error ε_k tends towards zero. The Fienup's HIO algorithm can be seen as an instance of the difference map algorithm when P_1 is the Fourier modulus projection P_M, P_2 the support projection P_S and $\beta = 1$.

VI Linear inverse problem and convex optimization

In the previous section, we saw that projection-based methods not only take nonlinear information into account, but also allow us to add some a priori information about the sought object. However, these methods, although related to convex optimization, are not guaranteed to converge (Bauschke, Combettes, and D. Russell Luke 2002; Bauschke, Combettes, and D. Russel Luke 2003; D Russel Luke 2005), and even if this is the case, can be slow since they often require a high number of iterations to achieve accurate results. Projection-based methods do not naturally incorporate prior knowledge such as the regularity of the solution. The variational formulation of the problem enables the integration of prior knowledge about the solution, ensuring improved accuracy and stability. *Variational approaches* allows a flexible inclusion of priors such as Tikhonov (Tikhonov and Arsenin 1977), sparsity regularization (Ramlau and Teschke 2006) or the Total Variation (TV) (Rudin, Osher, and Fatemi 1992). Rooted in a solid mathematical framework, variational methods encompass various optimization techniques, enhancing adaptability. In this section, we present classical methods for functional minimization. Convex functionals are a significant special case, as they tend to possess global minima.

VI.1 Variational approach

Consider an ill-posed inverse problem where the forward operator is given by $\mathbf{F} : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ (model parameter space) and $\mathcal{Y}$ (data space) are Hilbert spaces.

Noisy data

Considering that in practice, the data $y \in \mathcal{Y}$ are rarely available with perfect precision, we denote the observed perturbed data as y^δ and make the assumption that these noisy measurements satisfy the following conditions:

$$\|y^\delta - y\|_{\mathcal{Y}} \leq \delta$$

Given y^δ we want to find $x^{\delta *} \in \mathcal{X}$ such that we have $F(x_*^\delta) \approx y^\delta$. The *variational approach* then consists in formulating this problem as an optimization problem

$$x_*^\delta \in \underset{x\in\mathcal{X}}{\operatorname{argmin}} \left\{\mathcal{J}(x, y^\delta)\right\} \tag{1.127}$$

The functional $\mathcal{J} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is given by

$$\mathcal{J}(x, y^\delta) = d\left[\mathbf{F}(x), y^\delta\right] + \mathcal{R}(x) \tag{1.128}$$

where $d : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is a data fidelity term and $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is a regularization term. The function d is not necessarily a distance and will ensure the alignment between the solution x and the measure y^δ. The regularization term enforces the solution x to satisfy some prior information. The data fidelity term incorporates a priori information about the statistical properties of the noise in the observed data y^δ.

▪ **Example 1.2 — Gaussian noise.** We consider the following model

$$y = \mathbf{F}(x) + \varepsilon \tag{1.129}$$

with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ In a bayesian framework, The probability distribution of y knowing x, the likelihood, is the probability of ε which is

$$\mathbb{P}(y|x) = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{F}(x) - y\|^2\right) \tag{1.130}$$

With the bayesian formalism, the posterior $\mathbb{P}(x|y)$ is given by:

$$\mathbb{P}(x|y) = \mathbb{P}(y|x)\mathbb{P}(x) \tag{1.131}$$

A solution to our inverse problem can be obtained by solving

$$x_{\text{MAP}} = \underset{x}{\text{argmax}} \{\mathbb{P}(x|y)\} \tag{1.132}$$

which is called the *maximum a posteriori* (MAP). The MAP estimate is equivalent to minimizing the following problem:

$$x_{\text{MAP}} = \underset{x}{\text{argmin}} \{-\log \mathbb{P}(y|x) - \log \mathbb{P}(x)\} \tag{1.133}$$

In case of Gaussian noise, this is equivalent to solve

$$\underset{x}{\text{argmin}} \left\{-\log \mathbb{P}(x) + \frac{1}{2\sigma^2}\|F(x) - y\|^2\right\} \tag{1.134}$$

If we choose a Gaussian a priori for x, with zero mean and variance equal to 1, we obtain the Tikhonov regularization.

▪ **Example 1.3 — Poisson noise.** Now, consider Poisson noise on the data and an uniform prior. The measurements y can be characterized as random variables following a Poisson distribution of mean $y \sim \mathcal{P}(\lambda \mathbf{F}(x))$. The probability of x knowing y is

$$\mathbb{P}(x|y) = \prod_i \frac{[\mathbf{F}(x)]_i^{y_i}}{y_i!} \exp\left(-[\mathbf{F}(x)]_i\right) \tag{1.135}$$

so maximizing the (log-)likelihood amounts to minimizing the sum of the Kullback-Leibler (KL) divergence and an a priori regularization term $\mathcal{R}(x)$, where the KL divergence is given by:

$$\text{KL}\left[\mathbf{F}(x), y\right] = \sum_i [\mathbf{F}(x)]_i - y_i + y_i \log(y_i) - y_i \log\left([\mathbf{F}(x)]_i\right) \tag{1.136}$$

if we consider only the terms depending on x.

Convergent regularization method

At this point, several issues have to be considered. The first one is the convexity of the data term and of the regularization term. The second one is the differentiability of these terms. Different approaches can be found depending on these properties. We can define the notion of convergence of a regularization method for an ill-posed problem (Kaltenbacher, Andreas Neubauer, and Scherzer 2008).

Definition VI.1 Let $\{y_k\}_k$ and $\{x_k\}_k$ be sequences such that $\lim_{k\to\infty} y_k = y^\delta$ and x_k is solution to the regularization method

$$x_k = \underset{x}{\operatorname{argmin}} \{d\left[\mathbf{F}(x), y_k\right] + \mathcal{R}(x)\}$$

This method is *convergent* if $\lim_{k\to\infty} x_k = x_*^\delta$.

Various notion of convergence can be studied depending on the topologies considered on the spaces. We provide a non-exhaustive overview of fundamental concepts in convex optimization algorithms and various approaches employed to seek an approximate solution to x_*^δ.

VI.2 Notions of convexity

We revisit foundational concepts of convexity (Nesterov 2014; Bauschke and Combettes 2011; Boyd and Vandenberghe 2004) and introduce the notations that will be used. As before, $\mathcal{X}$ and $\mathcal{Y}$ are presumed to be Hilbert (or even Euclidean) spaces equipped with a norm denoted as $\|\cdot\| = \sqrt{\langle\cdot|\cdot\rangle}$. While the forthcoming outcomes are typically conceived in finite dimensions, many are dimension-independent and often hold within a Hilbert space context.

VI.2.1 Convex functions

Let's consider an extended real-valued function $f : \mathcal{X} \to (-\infty, +\infty]$.

Definition VI.2 — Convex function. The function f is said to be *convex* if and only if its *epigraph*

$$\operatorname{epi}(f) := \{(x, \lambda) \in \mathcal{X} \times \mathbb{R} : f(x) \leq \lambda\}$$

is a convex set, that is, for all $(x, y) \in \operatorname{epi}(f) \times \operatorname{epi}(f)$

$$f\left(tx + (1-t)y\right) \leq tf(x) + (1-r)f(y), \quad \forall t \in [0,1] \tag{1.137}$$

The function f is said to be *strictly convex* if the inequality (1.137) is strict whenever $x \neq y$ and $0 < t < 1$.

Definition VI.3 — Proper function. The function f is *proper* if it is not identically $+\infty$, i.e. if the *domain* of f

$$\operatorname{dom}(f) := \{x \in \mathcal{X} \,|\, f(x) < +\infty\}$$

is not an empty set.

Definition VI.4 — Lower semi-continuous function. The function f is *lower semi-continuous* (l.s.c.) if, for all $x \in \mathcal{X}$ and all sequences $\{x_n\}_{n\in\mathbb{N}}$ such that $x_n \to x$ we have

$$f(x) \leq \liminf_{n\to\infty} f(x_n)$$

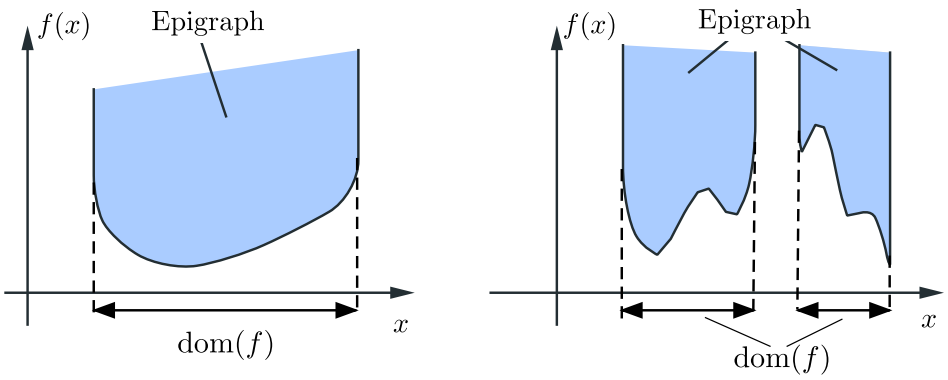


Figure 1.14.: (Left) Convex function and (Right) nonconvex function.

Definition VI.5 — Class $\Gamma_0(\mathcal{X})$. The set of functions $f : \mathcal{X} \to (-\infty, +\infty]$ that are convex, proper and lower semi-continuous is denoted $\Gamma_0(\mathcal{X})$.

An important example of function of $\Gamma_0(\mathcal{X})$ is the following

Definition VI.6 — Indicator function of a set. Let $C \subset \mathcal{X}$ be a set. The *characteristic function* or *indicator function* of the set C is denoted ι_C and is defined by

$$\forall x \in \mathcal{X}, \quad \iota_C : x \mapsto \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases} \tag{1.138}$$

The indicator function of C is convex, l.s.c. and proper if and only if C is convex, closed and non-empty. Minimizing such functions will provide us with a straightforward means to incorporate convex constraints into the formulation of our problems.

Definition VI.7 — Subdifferential. Given a function $f \in \Gamma_0(\mathcal{X})$, the *subdifferential* of f, denoted by ∂f, is defined as

$$\begin{array}{rccl} \partial f : & \mathcal{X} & \to & 2^{\mathcal{X}} \\ & x & \mapsto & \{p \in \mathcal{X} \,|\forall y \in \mathcal{X}, \quad f(x) + \langle y - x | p \rangle \leq f(y)\} \end{array} \tag{1.139}$$

where $2^{\mathcal{X}}$ is the power set of $\mathcal{X}$, i.e. the set of all subsets of $\mathcal{X}$. A vector in $\partial f(x)$ is a *subgradient* of f at x.

From this definition we are able to extend Fermat's rule to non-smooth convex functions, it can be expressed as follows:

$$0 \in \partial f(x^*) \Longleftrightarrow x^* \in \underset{x \in \mathcal{X}}{\mathrm{argmin}} \{f(x)\} \tag{1.140}$$

The function f is said to be *strongly convex* or *μ-convex* if for $x, y \in \mathcal{X}$ and $p \in \partial f(x)$, we have

$$f(x) + \langle p | y - x \rangle + \frac{\mu}{2}\|y - x\|^2 \leq f(y) \tag{1.141}$$

Equivalently, f is strongly convex if $x \mapsto f(x) - \frac{\mu}{2}\|x\|^2$ is convex, in particular f is strictly convex.

VI.2.2 Conjugate function and maximal monotone operators

To an arbitrary function f, one can associate the *Legendre-Fenchel conjugate* (Rockafellar 1974)

$$f^*(y) = \sup_{x \in \mathcal{X}} \{\langle y|x\rangle - f(x)\} \tag{1.142}$$

This supremum arises from the combination of linear and continuous functions, thus f^* is convex and lower semi-continuous. From this definition we observe that the *biconjugate* f^{**} is always below f, i.e. $f^{**} \leq f$. Actually f^{**} can be seen as the largest convex and lower semi-continuous function below f. In addition, if f is convex and l.s.c, the *Fenchel-Moreau theorem* states that the equality holds $f^{**} = f$.

Simple calculations yield the following equivalence, known as the *Legendre-Fenchel identity*

$$y \in \partial f(x) \Longleftrightarrow x \in \partial f^*(y) \Longleftrightarrow f(x) + f^*(y) = \langle y|x\rangle \tag{1.143}$$

From this identity, we can see that the subdifferential of a convex function is a *monotone operator*, i.e. it satisfies the following inequality

$$\langle p - q|x - y\rangle \geq 0, \quad \forall (x, y) \in \mathcal{X}^2, p \in \partial f(x), q \in \partial f(y) \tag{1.144}$$

Actually, we can show that the subdifferential of a function of class $\Gamma_0(\mathcal{X})$ is a special case of *maximal monotone operators*, which are multivalued operators $A : \mathcal{X} \to 2^{\mathcal{X}}$ that satisfy the monotonicity property

$$\langle p - q|x - y\rangle \geq 0, \quad \forall (x, p), (y, q) \in \mathrm{Graph}(A) \tag{1.145}$$

and whose graph $\mathrm{Graph}(A) = \{(x, p) : p \in Ap\} \subset \mathcal{X} \times \mathcal{X}$ is maximal (with respect to inclusion) in the class of graphs of operators which satisfy (1.145). In other words, A is a maximal monotone operator if it is monotone and if there is no monotone operator $B : \mathcal{X} \to 2^{\mathcal{X}}$ other than A such that $\mathrm{Graph}(A)$ is included in $\mathrm{Graph}(B)$.

VI.2.3 Proximity operator and resolvent

An increasingly significant notion in recent advancements within optimization is the concept of a *proximity operator* or *proximal map* that was introduced by Moreau (Moreau 1965). For a function $f \in \Gamma_0(\mathcal{X})$ it is defined as

$$\begin{array}{rccc} \mathrm{prox}_f : & \mathcal{X} & \to & \mathcal{X} \\ & x & \mapsto & \underset{y \in \mathcal{X}}{\mathrm{argmin}} \left\{ f(y) + \frac{1}{2}\|y - x\|^2 \right\} \end{array} \tag{1.146}$$

For any $x \in \mathcal{X}$, $\mathrm{prox}_f(x)$ can be understood as the outcome of a regularization-based minimization of f in a neighborhood of x It's important to note that since the function $x \mapsto f(y) + \frac{1}{2}\|y - x\|^2$ is strongly convex, the minimization to be performed to compute $\mathrm{prox}_f(x)$ always yields a single, unique solution. Once again, an important example is when f is the indicator function of a nonempty closed convex set $C \subset \mathcal{X}$, the proximity operator reduces to the projection P_C onto the set C

$$\forall x \in \mathcal{X}, \quad \mathrm{prox}_{\iota_C}(x) = P_C(x) = \underset{y \in C}{\mathrm{argmin}} \{\|y - x\|\}$$

This illustrates that proximity operators can be regarded as expansions of projections onto convex sets. The proximity operator shares several properties with projections, notably its

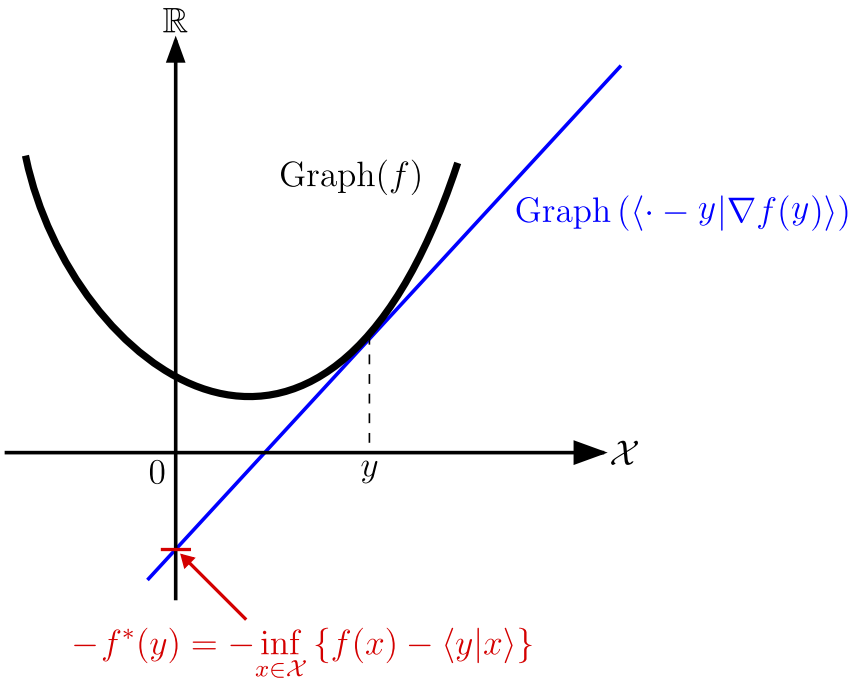


Figure 1.15.: The graph of a function $f \in \Gamma_0(\mathcal{X})$ is shown. The graph of the affine function $m_y : x \mapsto \langle x - y | \nabla f(y) \rangle$ is displayed in blue, it satisfies $m_x \leq f$ and it coincides with f at y. The value $f^*(y)$ can be understood as the intersection of the graph of m_y and the vertical axis.

nonexpansiveness (i.e. 1-Lipschitz) and its monotonicity (1.145). This nonexpansiveness can be seen as a broader form of the strict contraction property which is used in Banach–Picard fixed-point theorem. This characteristic enhances the effectiveness of the proximity operator in ensuring the convergence of fixed-point algorithms that rely on it.

In general, for any multivalued operators $A : \mathcal{X} \to 2^{\mathcal{X}}$, we can define the *resolvent* of A

$$J_A = (\mathrm{Id} + A)^{-1} \tag{1.147}$$

An important result by Minty characterizes the resolvent of a monotone maximal operator (Minty 1962). It states that A is maximal monotone if and only if its resolvent J_A is well defined and single-valued. Note that in the case of the proximity operator, we have the equivalences

$$x^* = \mathrm{prox}_{\tau f}(x) \iff 0 \in \partial f(x^*) + \frac{x^* - x}{\tau} \iff x^* = (\mathrm{Id} + \tau \partial f)^{-1}(x) \tag{1.148}$$

The proximity operator is thus the resolvent of the subdifferential of a function f, and if $f \in \Gamma_0(\mathcal{X})$, according to Minty theorem, we find that the proximity operator is well-defined and single-valued (because ∂f is maximal monotone). This allows us to make the link between the proximity operator of a function f and that of its conjugate f^*, given by *Moreau's identity*

$$x = \mathrm{prox}_{\tau f}(x) + \tau \mathrm{prox}_{\frac{1}{\tau} f^*}\left(\frac{x}{\tau}\right) \tag{1.149}$$

This shows that if we know how to compute prox_f, then we know how to compute prox_{f^*} (and vice versa). For a deeper exploration of proximity operators and their connections with

monotone operator theory, refer to (Bauschke and Combettes 2011; Combettes and Pesquet 2021). The proximity operator of a function sometimes has an explicit formulation, in which case we speak of a *closed form*, but generally, the functions we are interested in do not have one.

▪ **Example 1.4** Let us consider Ω an open subset of $\mathbb{R}^2$ and $u \in L^1(\Omega, \mathbb{R})$. We consider the functional $f(u) = \mathrm{TV}(u)$ the Total Variation of u (see §II.1 for more details), if u has a gradient $\nabla u \in L^1(\Omega, \mathbb{R}^2)$, this reduces to the integral $\mathrm{TV}(u) = \int_\Omega |\nabla u(x)| \, \mathrm{d}x$. It can be shown (Chambolle 2004) that $f^*(v) = \iota_K(v)$ where K is the closure of the set

$$\left\{ \mathrm{div}\xi \,\middle|\, \xi \in C_c^1\left(\Omega, \mathbb{R}^2\right), |\xi(x)| \leq 1, \ \forall x \in \Omega \right\}$$

We thus have $\mathrm{prox}_{f^*}(v) = P_K(v)$ which has no explicit formulation, in particular, using Moreau's identity, we can conclude that $\mathrm{prox}_f(u)$ has no closed form. Nevertheless, $P_K(v)$ can be approximated using an iterative method, known as Chambolle's algorithm (Chambolle 2004).

VI.3 Gradient methods

We begin by introducing the first group of methods: first-order gradient descent techniques. While they might appear basic, these methods remain crucial for solving straightforward imaging problems and are fundamental optimization techniques in machine learning. It offers a straightforward and intuitive approach to minimizing complex cost or loss functions.

Additionally, first-order method can be accelerated through various simple strategies, for instance by adding momentum. Building upon Nesterov's work (Nesterov 1983), in order to avoid oscillations around the minimum, slight modification of the descent algorithms can be done using over-relaxation techniques. In particular for deep learning optimization methods like the Adam method, acceleration techniques are widely used now (see §VIII.1.3). We will begin with the most basic approach and subsequently explore ways to enhance its performance.

VI.3.1 Gradient descent

We will start by introducing gradient descent methods and see that they are efficient and suitable for simple cases.

Let's consider we need to find a minimizer of a function $f \in \Gamma_0(\mathcal{X})$, that is, to solve the minimization problem

$$\min_{x \in \mathcal{X}} \{f(x)\} \tag{1.150}$$

Assuming the differentiability of f, the most direct method for solving the problem is implementing a gradient descent scheme with a constant step size $\tau > 0$ (algorithm 5). To guarantee

Algorithm 4 Gradient descent with fixed step

Choose $x_0 \in \mathcal{X}$

for $k \geq 0$ **do** :

$x_{k+1} \leftarrow x_k - \tau \nabla f(x_k)$

that the algorithm will converge to a minimum, the f function must be sufficiently smooth, i.e. the gradient ∇f should be Lipschitz with some constant L. Moreover, the step τ must not be too large if one wants to avoid that the method oscillates around the minimum, in fact, one must have $0 < \tau L < 2$ to ensure convergence.

Landweber iteration

An important algorithm for the resolution of linear inverse problems is the gradient descent applied to the data fidelity term, the minimization is thus rewritten as

$$\min_{x \in \mathcal{X}} \left\{ \|\mathbf{F}(x) - y\|_{\mathcal{Y}}^2 \right\} \tag{1.151}$$

The resulting method for solving (1.151) is known as the *Landweber iteration* (Landweber 1951)

$$x_{k+1} = x_k - \tau \mathbf{F}^* \left(\mathbf{F}(x_k) - y \right) \tag{1.152}$$

To ensure the convergence, the step τ must satisfy

$$0 < \tau < \frac{2}{\|\mathbf{F}^* \mathbf{F}\|} \tag{1.153}$$

where

$$\|\mathbf{F}^* \mathbf{F}\| = \sup_{x \in \mathcal{X}, \|x\|_{\mathcal{X}} \leq 1} \{ \|\mathbf{F}^* \mathbf{F}(x)\|_{\mathcal{X}} \}$$

VI.3.2 Sub-gradient descent

An alternative approach to execute a gradient-based scheme for a non-smooth convex objective f involves employing subgradient descent. In this method, the objective is to minimize the energy by moving in the direction of a subgradient chosen arbitrarily. This is used in practice

Algorithm 5 Sub-gradient descent method

Choose $x_0 \in \mathcal{X}$, $h_k > 0$ with $\sum_k h_k = +\infty$ and $\sum_k (h_k)^2 < +\infty$

for $k \geq 0$ **do** :

$$x_{k+1} \leftarrow x_k - h_k \frac{p_k}{\|p_k\|}, \quad \text{with } p_k \in \partial f(x_k)$$

when the subdifferential ∂f is easy to compute. A convergence condition for this method is that the function f should be L-Lipschitz near the optimal solution x^*.

VI.3.3 Implicit gradient descent and the proximal-point algorithm

When the gradient of the objective function ∇f is not Lipschitz, the gradient descent no longer satisfies the convergence conditions. Indeed the previous analysis becomes challenging, as it is difficult to predict the behavior of both the function f and its gradient at the new point. One way around this problem is to consider an *implicit gradient descent*, where the iteration defined in the algorithm 5 is substituted with a modified scheme.

$$x_{k+1} = x_k - \tau \nabla f(x_{k+1}) \tag{1.154}$$

If such an x_{k+1} exists, it is a critical point of the function

$$x \mapsto f(x) + \frac{\|x - x_k\|_{\mathcal{X}}^2}{2\tau} \tag{1.155}$$

thus it can be rewritten as $x_{k+1} = \operatorname{prox}_{\tau f}(x_k)$. And if $f \in \Gamma_0(\mathcal{X})$, then this critical point is the unique minimizer of this functional. Note that we have not assumed that f is smooth, we can in

fact work with the *Moreau-Yoshida regularization* of f defined by

$$f_\tau(y) = \min_{x \in \mathcal{X}} \left\{ f(x) + \frac{\|x - y\|_{\mathcal{X}}^2}{2\tau} \right\} \tag{1.156}$$

where $\tau > 0$. The function $f_\tau(y)$ is differentiable and its gradient $\nabla f_\tau(y)$ is $\frac{1}{\tau}$-Lipschitz which is given by

$$\nabla f_\tau(y) = \frac{y - \operatorname{prox}_{\tau f}(y)}{\tau} \tag{1.157}$$

Now the implicit gradient descent on f (1.154) can be seen as an explicit gradient descent on f_τ

$$x_{k+1} = \operatorname{prox}_{\tau f}(x_k) = (\mathrm{Id} + \tau \partial f)^{-1}(x_k) = x_k - \tau \nabla f_\tau(x_k) \tag{1.158}$$

Thus the convergence theory for the implicit algorithm is a straightforward consequence of the convergence theory for explicit gradient descent with a fixed step. Additionally, it is established that slightly larger steps can be employed in this context.

We can generalize equation (1.158) by replacing the subdifferential operator ∂f by any monotone operator A. This results in what is known as the *proximal-point algorithm* (PPA) whose general form can be expressed as

$$x_{k+1} = (\mathrm{Id} + \tau A)^{-1}(x_k) \tag{1.159}$$

It can be shown, under certain conditions, that the sequence defined by (1.159) will converge to a zero of A (Martinet 1970).

VI.3.4 Forward–backward splitting

We assume that the variational formulation (1.128) can be written in the general form

$$\min_{x \in \mathcal{X}} \{ f(x) + g(x) \} \tag{1.160}$$

where f is a convex function smooth enough (i.e. with Lipschitz gradient) and $g \in \Gamma_0(\mathcal{X})$ whose proximity operator is easy to compute. The core idea of the forward-backward (FB) splitting scheme (see algorithm 6) is to begin with an explicit step of descent (forward step) in the smooth function f followed by an implicit step of descent (backward), using its proximity operator, in g. Algorithm 6 can also be written under the following form

Algorithm 6 Forward-backward method with fixed step

Choose $x_0 \in \mathcal{X}$

for $k \geq 0$ **do** :

$x_{k+\frac{1}{2}} \leftarrow x_k - \tau \nabla f(x_k)$

$x_{k+1} \leftarrow \operatorname{prox}_{\tau g}\left(x_{k+\frac{1}{2}}\right)$

$$x_{k+1} = T_\tau x_k := \operatorname{prox}_{\tau g}(x_{k+1} - \tau \nabla f(x_k)) \tag{1.161}$$

This rewriting allows us to see the FB splitting scheme as a fixed-point algorithm. Indeed, the Fermat's rule (1.140) states that a critical point x^* of (1.160) must satisfy $0 \in \nabla f(x^*) + \partial g(x^*)$, or equivalently, $x^* = (\mathrm{Id} + \tau \partial g)^{-1}(x^* - \tau \nabla f(x^*))$, thus it should be a fixed-point of the operator T_τ.

VI.4 Saddle-point methods

Saddle point methods can be used to solve optimization problems with constraints, where the constraints can be taken into account with dual variables. They often involve the minimization of a convex-concave function, where one part is convex with respect to one set of variables and concave with respect to another set of variables. The goal is to find a solution that satisfies both the primal and dual conditions simultaneously. These methods typically alternate between updates in the primal and dual variables.

VI.4.1 Fenchel–Rockafellar duality

Let's consider the following minimization problem

$$\min_{x\in\mathcal{X}} \{f(Kx) + g(x)\} \tag{1.162}$$

where f and g are functions belonging to $\Gamma_0(\mathcal{X})$ and $\Gamma_0(\mathcal{Y})$, respectively, and $K : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator. According to the Fenchel-Moreau theorem VI.2.2, we have $f = f^{**}$ and we can show (Ekeland and Témam 1999) that the *primal* problem (1.162) is equivalent to

$$\max_{y\in\mathcal{Y}} \{-f^*(y) - g^*(-K^*y)\} \tag{1.163}$$

where K^* denotes the adjoint operator of K. The problem (1.163) is the (Fenchel-Rockafellar) *dual problem*. Problems (1.162) and (1.163) are related to the *saddle point problem*

$$\max_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \{\langle y|Kx\rangle - f^*(y) + g(x)\} \tag{1.164}$$

VI.4.2 Primal-dual methods

The basic idea is to use the FB splitting scheme in both primal and dual space to solve (1.164). As proposed by (Arrow 1958), this amounts to alternate between a proximal descent in the primal variable x and a proximal ascent in the dual variable y

$$x_{k+1} = \mathrm{prox}_{\tau g}(x_k - \tau K^* y_{k+1}) \tag{1.165}$$

$$y_{k+1} = \mathrm{prox}_{\tau f^*}(y_k + \sigma K x_k) \tag{1.166}$$

where τ and σ are gradient descent steps. Let us introduce the operator

$$T\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} \partial g(x) \\ \partial f^*(y) \end{pmatrix} + \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \tag{1.167}$$

which is maximal monotone, being the sum of two maximal monotone operators, and let M be the matrix operator

$$M := \begin{pmatrix} \frac{1}{\tau}\mathrm{Id} & -K^* \\ K & \frac{1}{\sigma}\mathrm{Id} \end{pmatrix} \tag{1.168}$$

Using these operators, iterations (1.165) and (1.166) can be written as

$$0 \in M\begin{pmatrix} x_{k+1} - x_k \\ y_{k+1} - y_k \end{pmatrix} + T\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} \tag{1.169}$$

thus this primal-dual algorithm can simply be considered as a proximal-point algorithm. Note that the matrix M is positive definite if $\tau\sigma\|K\|^2 < 1$. By varying the steps σ and τ, the algorithm can be accelerated (M. Zhu and T. F. Chan 2008). Proofs of convergence were also proposed in (Esser, X. Zhang, and T. Chan 2010).

A primal-dual algorithm widely used in practice is the primal-dual hybrid gradient (PDHG) method (Esser, X. Zhang, and T. Chan 2010), also known as the Chambolle-Pock algorithm (Chambolle and Pock 2011). In the PDHG method, the symmetry between dual and primal variables to reverse the order of iterations, and in addition, they add an over-relaxation step (see algorithm 7). The convergence conditions and accelerations of this method are studied in detail

Algorithm 7 Primal-dual hybrid gradient

Choose an initial pair $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, steps τ, σ and an over-relaxation parameter $\theta \in [0, 1]$

for $k \geq 0$ **do** :

$y_{k+1} \leftarrow \text{prox}_{\tau f^*}(y_k + \sigma K \overline{x}_k)$

$x_{k+1} \leftarrow \text{prox}_{\tau g}(x_k - \tau K^* y_{k+1})$

$\overline{x}_{k+1} \leftarrow x_{k+1} + \theta(x_{k+1} - x_k)$

in (Chambolle and Pock 2011).

VI.4.3 Alternating directions method of multipliers

The alternating direction method of multipliers (ADMM) can be understood as a primal-dual algorithm. This technique falls under the category of augmented Lagrangian methods, as one approach to deriving this algorithm involves seeking a saddle point of an augmented form of the traditional Lagrange function (Boyd, Parikh, et al. 2011). Classically, the ADMM algorithm aims at tackling problems under constraints

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax + By = b \end{array} \tag{1.170}$$

or equivalently

$$\min_{Ax+By=b} \{f(x) + g(y)\} \tag{1.171}$$

We get (1.162) if $b = 0$, $A = \text{Id}$ and $B = -K$. The idea is to introduce the *augmented Lagrangian*

$$L_\rho(x, y, z) = f(x) + g(y) + z^\top (Ax + By - b) + \frac{\rho}{2}\|Ax + By - b\|^2 \tag{1.172}$$

where $\rho > 0$ is a parameter. The constraint is handled by introducing the *Lagrange multiplier* z. The inclusion of the final quadratic term will not alter the saddle-point value and it significantly enhances iteration stability. This enhancement is typically achieved by rendering the minimization problem in x and y (with z fixed) solvable.

VII Iterative methods for nonlinear inverse problems

In this section, we treat ill-posed inverse problems given by a nonlinear forward operator $\mathbf{F} : \mathcal{X} \to \mathcal{Y}$. The variational formulation of the problem presented in §VI.1 can no longer be solved by convex optimization tools. Here we focus on the generalization of some of the approaches studied but we also introduce other methods. For simplicity's sake, we will assume

Algorithm 8 Alternating directions method of multipliers

Choose $y_0, z_0 \in \mathcal{Y}$ and $\rho > 0$

for $k \geq 0$ **do** :

$$x_{k+1} \leftarrow \underset{x \in \mathcal{X}}{\operatorname{argmin}} \left\{ L_\rho \left((x, y_k, z_k) \right) \right\}$$

$$y_{k+1} \leftarrow \underset{y \in \mathcal{Y}}{\operatorname{argmin}} \left\{ L_\rho \left((x_{k+1}, y, z_k) \right) \right\}$$

$$z_{k+1} \leftarrow z_k + \rho \left(A x_{k+1} + B y_{k+1} - b \right)$$

a Tikhonov-type regularization, so we aim to minimize:

$$\mathcal{J}(x, y^\delta) = \|\mathbf{F}(x) - y^\delta\|_{\mathcal{Y}}^2 + \alpha_k \|x - x_0\|_{\mathcal{X}}^2 \tag{1.173}$$

given an a priori estimate x_0. Since the proof of convergence of nonlinear methods is often technical, they will not be detailed here.

VII.1 Nonlinear Landweber algorithm

Recall that when dealing with a linear problem represented as $Ax = y^\delta$, where A is a linear operator operating on Hilbert spaces, iterative methods often involve the conversion of the normal equation into an equivalent fixed point equation:

$$x = x - A^* \left(Ax - y^\delta \right) \tag{1.174}$$

using that $A^* \left(Ax - y^\delta \right)$ is the direction of the gradient of the quadratic functional $\|Ax - y^\delta\|$. For nonlinear problems, the appropriate fixed point equation is given by

$$x = \varphi(x) := x - \mathbf{F}'(x)^* \left(\mathbf{F}(x) - y^\delta \right) \tag{1.175}$$

where $\mathbf{F}'(x)$ denotes the differential of F at x, assuming that the operator F is differentiable. The *nonlinear Landweber* iteration is thus defined by

$$x_{k+1} = \varphi\left(x_k\right) = x_k - \tau_k \mathbf{F}'(x_k)^* \left(\mathbf{F}(x_k) - y^\delta \right) \tag{1.176}$$

In contrast to the linear case, to ensure convergence, the τ_k step update depends on the current iteration, details about the convergence and convergence rates of this algorithm can be found in (Martin Hanke, Andreas Neubauer, and Scherzer 1995). Note that when $\mathbf{F}$ is linear, we get the Landweber iteration (1.152). In order to minimize the functional (1.173), the following modification of Landweber iteration was proposed (Scherzer 1998)

$$x_{k+1} = x_k - \tau_k \left[\mathbf{F}'(x_k)^* \left(\mathbf{F}(x_k) - y^\delta \right) + \alpha_k \left(x_k - x_0 \right) \right] \tag{1.177}$$

The additional term $\alpha_k \left(x_k - x_0 \right)$ compared to the classical Landweber iteration is obtained by the Tikhonov regularization term.

VII.2 Nonlinear conjugate gradient descent

The number of iterations can be reduced if instead of considering the direction of the negative gradient one uses an iteration procedure for a search direction that differs from the gradient

direction. A *nonlinear conjugate gradient* method can be written

$$x_{k+1} = x_k + \tau_k d_k \tag{1.178}$$

where τ_k is the step size and the directions d_k are obtained by

$$d_k = \begin{cases} -g_0 & \text{if } k = 0 \\ -g_k + \beta_k d_k & \text{if } k \geq 0 \end{cases} \tag{1.179}$$

Here β_k denotes the conjugate gradient update parameter and $g_k = \mathbf{F}'(x_k)^* \left(\mathbf{F}(x_k) - y^\delta\right) + \alpha_k (x_k - x_0)$ is the gradient direction.

Originally, Hestenes and Stiefel (Hestenes and Stiefel 1952) proposed an update of β_k in the case of symmetric, positive-definite linear systems. Later, different CG methods were introduced corresponding to different choices for the scalar β_k. The formula of Fletcher and Reeves

$$\beta_k^{\mathrm{FR}} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \tag{1.180}$$

is usually considered as the first nonlinear CG algorithm (Fletcher and Reeves 1964). Another nonlinear CG update was introduced by Polak and Ribière (Polak and Ribiere 1969) and by Polyak (Polyak 1969) and corresponds to

$$\beta_k^{\mathrm{PRP}} = \frac{g_{k+1}^\top (g_{k+1} - g_k)}{\|g_k\|^2} \tag{1.181}$$

The selection of the CG update parameter strategy often requires experimentation and problem-specific considerations.

VII.3 Iteratively Regularized Gauss-Newton method

The key idea of any Newton type method consists in repeatedly linearizing the forward operator F around some approximate solution x_k, and then solving the new linearized problem

$$\mathbf{F}(x_k) + \mathbf{F}'(x_k)(x_{k+1} - x_k) = y^\delta \tag{1.182}$$

for x_{k+1}. If we consider a Tikhonov-type regularization, then the approximate solution x_{k+1} is defined by

$$x_{k+1} \in \operatorname*{argmin}_{x \in \mathcal{X}} \left\{ \|\mathbf{F}(x_k) + \mathbf{F}'(x_k)(x - x_k) - y^\delta\|_{\mathcal{Y}}^2 + \alpha_k \|x - x_0\|_{\mathcal{X}} \right\} \tag{1.183}$$

It can be shown that (1.183) has always a unique solution (Engl, M. Hanke, and A. Neubauer 2000), thus one can define the iteratively regularized Gauss-Newton (IRGN) algorithm by

$$x_{k+1} = x_k + (\mathbf{F}'(x_k)^* \mathbf{F}'(x_k) + \alpha_k \mathrm{Id})^{-1} \left[\mathbf{F}'(x_k)^* (y^\delta - \mathbf{F}(x_k)) + \alpha_k (x_0 - x_k)\right] \tag{1.184}$$

In (1.184), the inverse of the selfadjoint positive-definite operator $\mathbf{F}'(x_k)^* \mathbf{F}'(x_k) + \alpha_k \mathrm{Id}$ has to be computed, which is bounded according to the estimate

$$\left\| (\mathbf{F}'(x_k)^* \mathbf{F}'(x_k) + \alpha_k \mathrm{Id})^{-1} \right\|_{\mathcal{X}} \leq \frac{1}{\alpha_k} \tag{1.185}$$

Therefore, the iterate x_{k+1} depends continuously on the measure y^δ so that the impact of data errors is regularized. Since the IRGN relies on iteratively updated linearizations of the imaging

operator, it is most effectively applied to problems characterized by moderate nonlinearity. It is possible to demonstrate the convergence of this method, provided there exist bounds on the nonlinearity of $\mathbf{F}$ and appropriately selected values for α_k and the initial guess x_0 (Kaltenbacher, Andreas Neubauer, and Scherzer 2008).

VIII Phase retrieval and deep learning

Recent developments in deep learning methods have led to numerous applications in the field of image processing (LeCun, Y. Bengio, and Hinton 2015). In particular, new methods have been proposed to solve inverse problems (Arridge et al. 2019) and this includes reconstruction problems (H.-M. Zhang and Dong 2020). Many of these approaches have focused on linear inverse problems (Ongie et al. 2020), but more and more advances are being made for the nonlinear case (Hoop, Lassas, and Wong 2022) and in particular for problems such as phase retrieval (Deshpande et al. 2022). In this section, we give a short introduction to deep learning with a focus on convolutional neural networks. This is followed by an overview of the application of neural networks to reconstruction problems. Finally, we will look at some recently proposed learning methods for the phase retrieval problem.

VIII.1 Foundations of deep learning

Deep learning can be seen as a subset of *machine learning* techniques, it focuses on algorithms inspired by the structure of the human brain. These networks learn through example data (training data), and the more diverse data they encounter, the more adept they become at recognizing various features and making generalizations. In this section, we provide a concise overview of the fundamental principles of machine learning and deep learning. We will take a quick look at the theory, and focus on the concepts we will be using for the rest of the thesis. For a comprehensive treatment of deep learning theory, readers are encouraged to consult (Goodfellow, Yoshua Bengio, and Courville 2016) and for practical insights into implementing deep learning, they can refer to (Chollet 2017; Geron 2019).

VIII.1.1 Machine learning

Machine learning is a science that aims to equip an algorithm with the ability to learn and adapt from a wide range of data types, including sound, images, statistical data and text. Instead of relying on fixed rules or hard-coded instructions, machine learning algorithms ingest data, extract information and features, and use this new knowledge to make *predictions*. This process typically involves constructing a *prediction model* and generally consists of two main phases. The first phase involves estimating/adjusting a model from the data (this is the *learning* or *training* phase). This step allows for matching/imitating a phenomenon given basic information (labels, rewards). The second phase corresponds to the *testing* phase, where the estimated model is applied to new data that was not used during the first phase. This serves to assess the model's consistency with new data and estimate its ability to generalize to new situations. Therefore, a learning algorithm aims to find a model that achieves the best performance/prediction scores on a test dataset among a chosen class of applications (model class) that are predetermined. This class could be the class of classifiers in the case of classification problems or the class of polynomials in a polynomial regression problem. In the context of deep learning, the class of interest is that of neural networks (see section VIII.1.2).

We can roughly distinguish between two types of learning.

Supervised learning

In *supervised learning*, the training data you feed to the algorithm includes the desired output, called *labels*. In this framework, each training data consists of a pair (x_i, y_i) where x_i is the input data annotated by the desired output y_i (*ground truth*). The aim of supervised learning is to find a model f_θ, parametrized by the parameters θ, capable of best matching inputs x to outputs y, based on the training set $\{(x_i, y_i)\}_{i=1}^{N}$, but also generalizing efficiently on new data. This can generally be summed up in the following steps:

1. *Training*: search for optimum parameters θ^* in order to minimize the error (*cost function*) between the predictions $f_\theta(x_i)$ and the ground truth y_i for the pair (x_i, y_i).
2. *Validation*: models are often dependent on *hyperparameters*, which are parameters of a learning algorithm and not a parameter of the model. As such, it is not affected by the learning algorithm itself; it must be set prior to training and remains constant during training. The validation phase consists of testing f_θ on data that has not been used for training, in order to adjust the hyperparameters. This phase prevents the model from *over-fitting*., i.e. to avoid that the hyperparameters are just fitted for the training data and not to new data.
3. *Testing*: After determining the hyperparameters (validation) and optimal parameters θ^* (training), we evaluate how well the model f_{θ^*} performs on new data that was not part of the training or validation process. Through testing, it becomes possible to assess the model's capacity to apply its learned knowledge to new data.

This is the approach that interests us in this manuscript and that we will be developing later in the neural network context.

Unsupervised learning

In *unsupervised learning*, the algorithm is trained on a dataset without explicit supervision or labeled targets In other words, the algorithm is not provided with predefined labels or categories for the input data. Instead, its objective is to learn, without a supervisor, to extract classes or groups with common characteristics. Unsupervised learning is particularly useful when you have large datasets and want to uncover hidden patterns or structures within the data. In this sense, unsupervised learning algorithms can perform more complex tasks than supervised learning but can also be highly unpredictable. Since the data is not labeled, the unsupervised algorithm cannot guarantee a success score with high certainty.

VIII.1.2 Deep learning

Deep learning algorithms are a special case of machine learning algorithms whose predictive model is an artificial neural network (ANN), which we will refer to as a neural network, or simply a network. The neural networks have garnered significant attention in recent times, they have been extensively studied and used in recent years. The initial network structure was conceived by McCulloch and Pitts, drawing inspiration from a simplified model of biological neural networks in the brain (McCulloch and Pitts 1943). They introduced a simplified computational model that demonstrated how biological neurons in animal brains could collaborate to execute intricate computations using propositional logic. This marked the creation of the first ANN architecture. Since then, various types of architectures have been proposed to tackle different machine learning problems. Presently, the widespread use of these structures is largely motivated by the fact that (deep) neural networks are highly effective in processing various types of data (images, audio, videos, text, ...).

Artificial neural network

A neural network refers to a *parameterized function* f_θ whose structure is built upon a sequence of linear and nonlinear operations. In practice, the parameters vector (or learning parameters) θ is adjusted to optimize an empirical performance score, the cost function. This parameter tuning phase corresponds to the training step (or learning procedure) of the neural network, which is central in machine learning.

Neuron The (artificial) neuron serves as the fundamental building block of neural networks. Neurons are inspired by biological neurons found in the human brain but are simplified mathematical models that process and transmit information. Each artificial neuron performs two main operations:

- *Linear combination*: The neuron takes a set of input values, often represented as a vector, and multiplies each input by a corresponding weight. These weighted inputs are then summed together along with a bias term (a constant).
- *Activation function*: The result of the linear combination is then passed through an activation function. This function introduces nonlinearity to the neuron, allowing it to model complex relationships in the data. Common activation functions include the sigmoid function, hyperbolic tangent (tanh), and rectified linear unit (ReLU). The choice of activation function depends on the specific neural network architecture and the problem being solved.

Mathematically, a neuron is a function f whose output $y \in \mathbb{R}$ can be represented as follows

$$y = f_\theta(x) = \rho\left(\langle x|w\rangle + b\right) = \rho\left(\sum_{i=1}^{d} x_i w_i + b\right)$$

where $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ is the inputs, $w = (w_1, \dots, w_d) \in \mathbb{R}^d$ represents the corresponding weights for each input, b is a bias term and $\rho : \mathbb{R} \to \mathbb{R}$ is an activation function. Here, the parameters to be optimized are weights and bias, i.e. $\theta = \{w, b\}$.

A neuron is only capable of solving very basic problems, such as linearly separable classification or linear regression. For more complex problems, a neuron can be used as the foundational building block for complex models by forming connections between multiple neurons. These models are what we refer to as neural networks.

Multi Layer perceptron A neural network is constructed by connecting multiple neurons in such a way that the output of one neuron serves as an input to another. The general structure of a neural network can be seen as a sequence of several *layers of neurons* (sets of neurons), where typically each layer takes its inputs from the outputs of the previous layer. It is worth noting that the output of a neuron can also be the input for a neuron in the same layer or in previous layers, in such cases, it is referred to as a *recurrent structure*. Here we are interested in *feedforward* neural network structures where a neuron only takes inputs from previous layers.

The *multilayer perceptron* (MLP) introduced by (Rosenblatt 1958), is a network with such a structure. In this model, each neuron in a layer takes its inputs from the outputs of all neurons in the previous layer, so the layers are said to be *fully connected* to each other. Mathematically, an MLP with L layers and learnable parameters θ is a function

$$\begin{array}{rccc} f_\theta : & \mathbb{R}^{N_0} & \to & \mathbb{R}^{N_L} \\ & x & \mapsto & x_L \end{array}$$

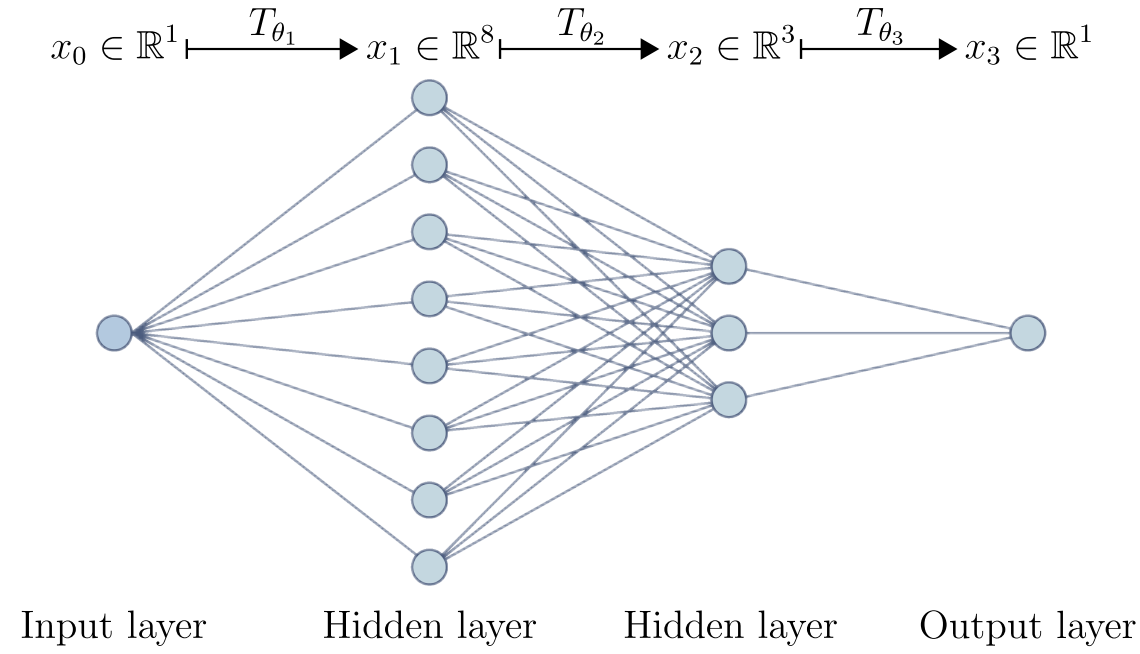


Figure 1.16.: Graph representation of a neural network with two *hidden layers* with respectively 8 neurons and 3 neurons, and an output layer with 1 neuron.

It can be written as a composition of operators, which correspond to layers

$$f_\theta = T_{\theta_L} \circ \cdots \circ T_{\theta_1},$$

where, for every $k \in \{1, \ldots, L\}$, $\theta_k = \{W_k, b_k\}$ represents the learnable parameters of the k-th layer T_{θ_k}. For $k \in \{1, \ldots, L\}$, k-th layer is defined as

$$\begin{array}{rccc} T_{\theta_k} : & \mathbb{R}^{N_{k-1}} & \to & \mathbb{R}^{N_k} \\ & x & \mapsto & \rho_k\left(W_k x + b_k\right) \end{array}$$

where $\rho_k : \mathbb{R}^{N_k} \to \mathbb{R}^{N_k}$ is a nonlinear activation operator, $W_k : \mathbb{R}^{N_{k-1}} \to \mathbb{R}^{N_k}$ is a linear operator and $b_k \in \mathbb{R}^{N_k}$ is a bias parameter vector. At each step, W_k can be seen as a matrix whose number of rows corresponds to the number N_k of neurons in the k-th layer and the number of columns is the number N_{k-1} of neurons in the $(k-1)$-th layer. Layers other than the input and output layers are called *hidden layers* (see Fig. 1.16). The learnable parameters of the neural network $\mathbf{f}_\theta$ are $\theta = \{\theta_1, \ldots, \theta_L\}$. These are the network parameters that are adjusted during the learning phase. These parameters are also called *weights* and biases of the neural network.

Activation function The activation function is not a learning parameter, but can rather be seen as a hyperparameter in the sense that it is chosen before the learning phase. The choice of activation function depends on the problem being studied. In practice, we require it to be nonlinear and differentiable almost everywhere. Indeed, for structures more complex than a neuron, nonlinearity can be used to approximate very complicated functions. Differentiability is also an important feature that enables gradient-based optimization algorithms to be used for adjusting learning parameters. A large number of activation functions have been introduced,

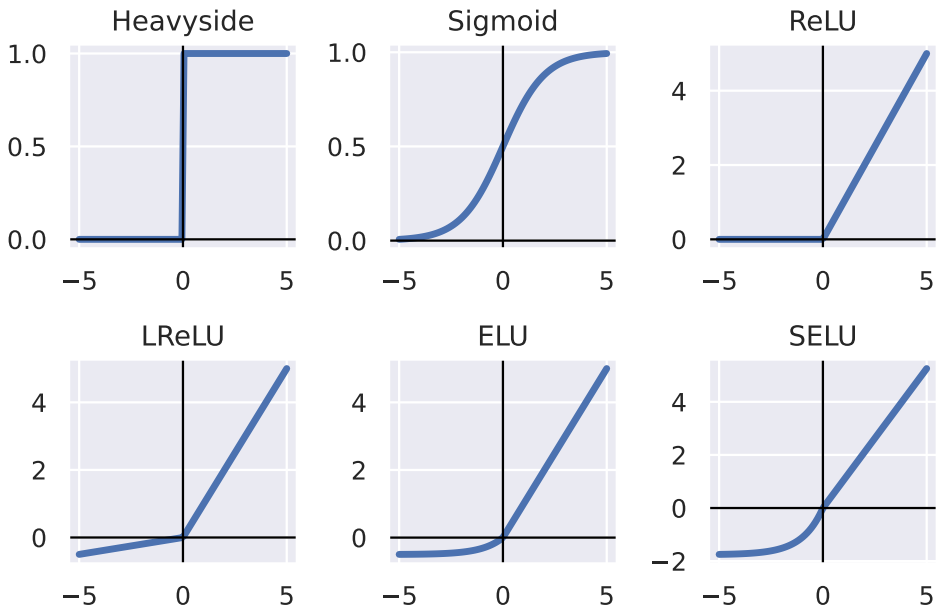


Figure 1.17.: Graphical representation of heavyside, sigmoid and ReLU (and variants) activation functions.

but, in practice, only a certain number of functions are used:

- The *Heaviside* function $\rho = \mathbb{1}_{\mathbb{R}_+}$ was introduced by McCulloch and Pitts (McCulloch and Pitts 1943) to demonstrate that simple logic functions like NOT, AND, and OR can be realized as a neural network. However, its derivative is zero almost everywhere, so this activation function is not suitable for modern learning.
- The *sigmoid* or *logistic* function $\rho(x) = \frac{1}{1+\exp(x)}$ is found in problems like logistic regression (where outputs are in $[0, 1]$) or in binary classification, interpreting each network output as a probability of belonging to a class. Another often-used variant is the hyperbolic tangent function.
- The *Rectified Linear Unit* (ReLU) function $\rho(x) = \mathrm{ReLU}(x) = \max(x, 0)$ is one of the most widely used functions today. It is not differentiable at 0, but in practice, it is relatively rare to have values close to 0. However, the gradient of ReLU is zero for negative values, so no information is retrieved for negative inputs. This can happen during training, then the neuron only outputs 0 on any input value, this is known as the *dying ReLU* problem. In such cases, it is advisable to use variants of ReLU, such as the Leaky ReLU

$$\mathrm{LReLU}_\alpha(x) = \max(x, 0) + \alpha \min(x, 0),$$

to include negative values. This activation depends on the hyperparameter α, which can become a learnable parameter, that is the idea of the parametric ReLU (PReLU) activation. Other more regularized variants have also been proposed, such as the Exponential Linear Unit (ELU) function (Clevert, Unterthiner, and Hochreiter 2016) or the Scaled Exponential Linear Units (SELU) function (Klambauer et al. 2017).

VIII.1.3 Training neural networks

In this section, we provide a more detailed description of how the learning parameters of a neural network are precisely adjusted.

Database and loss minimization

The learning process involves estimating the parameter vector θ of a neural network f_θ from a dataset $\{(x_i, y_i)\}_{i=1}^N$ (training data) using an optimization algorithm, also referred to as a learning algorithm in this context. To determine whether f_θ is a good predictive model or not, we have to quantify the error between the prediction $f_\theta(x_i,)$ and the label y_i. One way to address this question is by introducing an *cost* or *loss* function l, that quantifies the discrepancy between the prediction $f_\theta(x_i)$, for input x_i and the label y_i. The selection of the loss function is specific to the problem under consideration.

From the statistical learning point of view, the data points $\{(x_i, y_i)\}_{i=1}^N$ can be regarded as realizations of a joint probability distribution μ. The pairs (x_i, y_i) represent the elements for which we have direct access and can make measurements. The measure μ encompasses all possibilities, of which we only have partial knowledge through the pairs (x_i, y_i). In other words, our supervised learning problem can be described as an optimization problem involving the parameter θ:

$$\underset{\theta\in\Theta}{\operatorname{argmin}} \left\{\mathcal{L}(\theta) := \mathbb{E}_{(x,y)\sim\mu}\left[l\left(f_\theta(x_i), y_i\right)\right]\right\} \tag{1.186}$$

where Θ is a predetermined set of parameter vectors, representing the prior assumption about the network structure used during the learning process. The function $\mathcal{L}$ is often referred to as the *generalization error* as it quantifies the deviation for all possible (x_i, y_i) pairs under the measure μ.

In practice, since we do not have direct access to the joint measure μ, the parameter θ is optimized based on the *empirical error*

$$\mathcal{L}_N(\theta) = \sum_{i=1}^{N} l\left(f_\theta(x_i), y_i\right) \tag{1.187}$$

The training dataset $\{(x_i, y_i)\}_{i=1}^N$ should cover the entire range of possible configurations (defined by μ) so that the network obtained after minimization can generalize beyond the training data. As with variational approaches, a priori information can be added to the learned model by penalizing the cost function:

$$\mathcal{L}_N(\theta) = \sum_{i=1}^{N} l\left(f_\theta(x_i), y_i\right) + \lambda\mathcal{R}(\theta) \tag{1.188}$$

The first term corresponds to a data fitting penalty on $\{(x_i, y_i)\}_{i=1}^N$, while the term $\mathcal{R}(\theta)$ represents a regularization penalty on the learning parameters θ. Commonly used regularizations include the L^1 norm (for sparsity on learning parameters) and the L^2 norm (often used to make the problem strictly convex and accelerate learning), they often allow to avoid problems such as over-fitting. More advanced regularizations give properties related to the resolvent of a maximal monotone operator (Pesquet et al. 2021).

Optimization

To obtain the optimal network weights, we therefore need to minimize the (empirical) cost function. When dealing with an arbitrarily large database (when N is very large), it becomes impractical to use optimization algorithms whose complexity at each iteration depends on N. To address this challenge, modified gradient methods are often employed. The general idea behind these methods is to reduce the computational cost when direct computation of $\nabla_\theta\mathcal{L}_N$ becomes impractical.

The most used algorithm in this category is the Mini Batch Gradient Descent (MBGD). At each iteration, instead of computing $\nabla_\theta \mathcal{L}_N$ directly, it computes only

$$\sum_{k=1}^{b} \nabla_\theta l\left(f_\theta(x_i), y_i\right)$$

where the set $\{i_1, \ldots, i_b\}$ is randomly drawn from $\{1, \ldots, N\}$ and b is the *batch size*. When the batch size is set to $b = 1$, we obtain the Stochastic Gradient Descent (SGD). Starting from an initialization θ_0, the MBGD scheme is defined as follows:

$$\theta_{k+1} = \theta_k - \frac{\tau_k}{b} \sum_{k=1}^{b} \nabla_\theta l\left(f_\theta(x_i), y_i\right) \tag{1.189}$$

Here, the training set $\{(x_i, y_i)\}_{i=1}^N$ is divided into batches (of size b), and (1.189) is repeated for each batch until the entire database has been processed. This constitutes what is known as an *epoch*.

The selection of the gradient step τ_k, known as the *learning rate*, is crucial to ensure that the algorithm converges to a minimum. The conventional gradient descent can be quite slow, and faster optimizers have been introduced to accelerate the convergence speed. To do this and to avoid oscillations around the minimum, one option is to add *momentum*. This is essentially a slight modification of the descent algorithm, where the movement in the parameter space is averaged over multiple previous time steps. This method was initially proposed by Nesterov (Nesterov 1983) in 1983. More recently, the ADAM method (derived from ADAptive Moment estimation) have been introduced and is arguably the most widely used and efficient method for learning algorithms today. It is based on an adaptive learning rate that provides accurate estimations of the first and second moments of the gradient at each iteration (Kingma and Ba 2017). It shares similarities with the Nesterov method while ensuring better control.

Learning rate scheduling

Choosing an appropriate learning rate is important. If set too high, the training may diverge, and if too low, training will eventually converge to the optimal solution, but this will be a time-consuming process. It is possible to achieve better results than with a fixed learning rate. By initiating training with a high learning rate and subsequently decreasing it when progress slows down, you can reach a good solution more efficiently than with a constant learning rate. Various strategies exist for adjusting the learning rate during training. It can also be advantageous to commence with a low learning rate, increase it temporarily, and then decrease it again. These strategies are referred to as *learning rate scheduling*. While the optimizer focuses on the mechanics of parameter updates, learning rate scheduling deals specifically with controlling the size and adaptability of the learning rate throughout the training process.

Here are some of the most commonly employed learning schedules:

- **Step Decay**: The learning rate is reduced by a fixed factor after a predefined number of epochs or when a certain condition is met.
- **Exponential Decay**: The learning rate decreases exponentially over time, which can be beneficial for gradually slowing down the learning process.
- **Piecewise Constant Decay**: The learning rate is kept constant for a certain number of epochs and is then reduced abruptly.
- **Cosine Annealing**: The learning rate follows a cosine function, decreasing gradually and then increasing periodically. This method can help models escape local minima and find

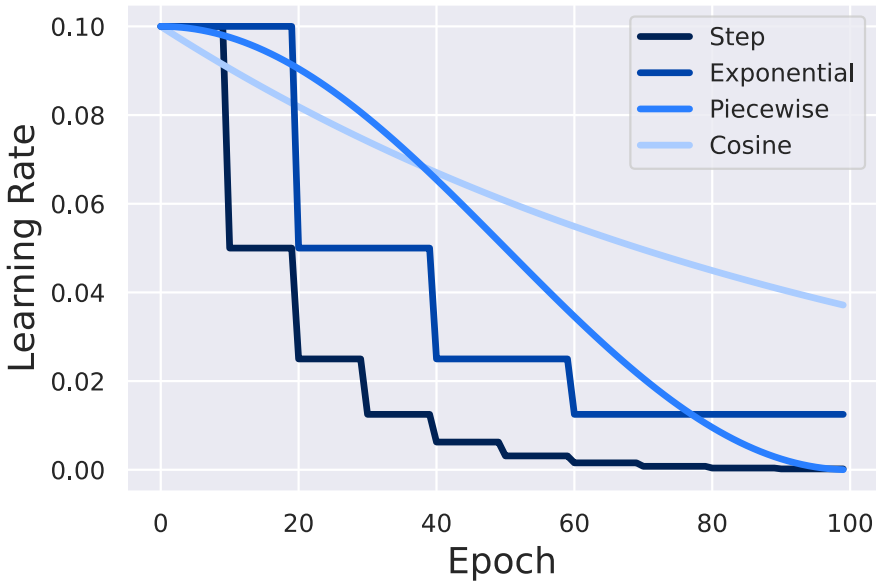


Figure 1.18.: Evolution of learning rate over 100 epochs for the step, exponential, piecewise constant and cosine annealing decay. Initial learning rate was set to $\tau_0 = 0.1$.

better solutions.

The choice of a specific learning rate schedule depends on the problem, the architecture, and the dataset. It often involves experimentation and fine-tuning to find the schedule that works best for a particular task.

VIII.2 Convolutional neural networks

During a time when fully connected neural networks struggled with training and generalization, a distinct type of deep feedforward network emerged as a more trainable and better generalizing alternative. This network, known as the Convolutional Network (ConvNet) (Lecun et al. 1998), not only demonstrated remarkable practical successes but also gained significant traction within the computer vision community. Since then, Convolutional Neural Networks (CNNs) have revolutionized the field of deep learning and have become the go-to architecture for a wide range of tasks (LeCun, Y. Bengio, and Hinton 2015). Their success lies in their ability to effectively capture spatial hierarchies and local patterns in data. CNNs employ convolutional layers with shared weights and local receptive fields, allowing them to efficiently recognize features within data, such as edges, textures, and shapes while being robust to variations in position and scale. This property makes CNNs exceptionally well-suited for tasks like image recognition, object detection, and image segmentation. Moreover, the hierarchical feature extraction in CNNs enables them to learn abstract representations from raw input, making them versatile in various domains, including computer vision and natural language processing.

VIII.2.1 Convolutional layers

A *convolutional layer* is the building block in CNN. As opposed to a fully connected layer, the neurons in the first convolutional layer do not establish connections with every individual pixel in the input image. Instead, they connect only with pixels residing within their receptive fields. Subsequently, each neuron in the following convolutional layer forms connections exclusively with neurons located within a small rectangular region in the preceding layer. This architectural

design enables the network to focus on small-scale, fundamental features within the initial hidden layer and then aggregate them into more comprehensive, high-level features in the next hidden layer. This hierarchical structure mirrors the organization of features in real-world images, underscoring why CNNs excel in image recognition tasks.

Formally, if we define an image as a set of pixels $x \in \mathbb{R}^{m\times n\times c}$ with m rows, n columns, and c channels and we denote the image corresponding to a single channel j of x as x^j. Like the MLP, we can define a CNN with L layers and learnable parameters θ as a function

$$\begin{array}{rccc} f_\theta : & \mathbb{R}^{m_0\times n_0\times c_0} & \to & \mathbb{R}^{m_L\times n_L\times c_L} \\ & x & \mapsto & x_L \end{array}$$

Every layer i generates an output image $x_i \in \mathbb{R}^{m_i\times n_i\times c_i}$, referred to as a *feature map*, by utilizing the output from the preceding layer x_{i-1} as its input. It is important to note that the dimensions of the layer's output x_i are not constrained to match those of the input x_{i-1}. Initially, the input image $x = x_0$ serves as the first layer, and the final layer yields the output image $y = x_L$.

Each individual layer is composed of a series of operations. In a typical layer architecture, the process begins by convolving each channel of the input feature map with distinct filters. Subsequently, the convolved images are pixel-wise summed, followed by the addition of a constant value (bias), to the resulting image. Lastly, like MLP, a non-linear operation is applied to each pixel in the image. These operations can be iteratively applied with varying filters and biases to create multiple channels within the output feature map. Consequently, the output x_i^j of a single channel j in such a convolutional layer can be expressed as follows:

$$x_i^j = \rho\left(g_{ij}(x_{i-1}) + b_{ij}\right) \tag{1.190}$$

where $\rho : \mathbb{R}^{m_i\times n_i} \to \mathbb{R}^{m_i\times n_i}$ is an activation function, $b_{ij} \in \mathbb{R}$ is the bias and

$$g_{ij} : \mathbb{R}^{m_{i-1}\times n_{i-1}\times c_{i-1}} \to \mathbb{R}^{m_i\times n_i}$$

convolves each channel of the input feature map with a different filter and sums the resulting images pixel by pixel:

$$g_{ij}\left(x_{i-1}\right) = \sum_{k=0}^{c_i-1} W_{ijk} * x_{i-1}^k$$

Here, $*$ denotes the discrete correlation operation or simply a discrete 2D convolution. The 2D convolution of an image I of size $n \times m$ by a kernel K of size $p \times q$ is defined as follows:

$$(K *_l I)(i,j) = \sum_s \sum_t I(s,t)K(s+i,t+j) \tag{1.191}$$

this will output an image of size $(n-p+1)\times(m-q+1)$.

Stride

The *stride* represents the spatial step size between two filter applications and influences the size of the image obtained after convolution. Stride can be used to reduce the image size while increasing the amount of information extracted from the input. It also contributes to making the convolutional layer more efficient by avoiding redundant computations.

Padding

Padding involves increasing the size of the image. It can be used to maintain the image size after convolution or to impose fixed dimensions on an image. Padding also helps reduce aliasing effects. When the added pixels have a value of zero, it's referred to as *zero padding*. Alternatively, one can add mirrored-value pixels if the intention is to avoid the introduction of zeros at the image's edges.

VIII.2.2 Pooling layers

A *pooling layer*, often referred to as subsampling or simply pooling, serves to condense information by merging similar features into a single representation. Pooling is typically done independently for each feature map of the input. Two types of pooling are often used for CNNs. The *max pooling* consists of selecting the maximum value from a small region (usually a 2x2 or 3x3 window) of the input feature map and retains only that value, discarding the others. In *average pooling*, we compute the average value within a small region of the input feature map. Pooling layers help reduce the computational complexity of the network and gradually increase the receptive field of neurons in deeper layers. These layers are crucial for hierarchical feature extraction.

VIII.2.3 The U-Net

In recent years, many convolutional neural network architectures have been introduced and analyzed. Among them, the U-net architecture stands out as arguably the most widely recognized and utilized. The U-Net architecture was originally designed to solve segmentation problems (Ronneberger, Fischer, and Brox 2015). Since then, it has been successfully used in image reconstruction as a post-processing tool of direct reconstruction in computed tomography (Jin et al. 2017). The U-Net is based on:

1. multilevel decomposition by dyadic scale decomposition based on max pooling, so that the effective filter size in the middle layers is larger than that of the early and late layers
2. multichannel filtering, such that there are multiple feature maps at each layer.

More precisely, the U-Net architecture consists of a downscaling and an upscaling part that gives it the U-shaped network structure. The downscaling follows the typical architecture of a convolutional neural network. It consists of the repeated application of convolutions with 3×3 filters, each followed by a ReLU activation function, a batch normalization layer and then a 2×2 max pooling operation with stride 2 for downsampling. At each downsampling step, the number of feature channels is doubled. On the other side, the upscaling part consists of an upsampling of the feature map with a 3×3 up-convolution that halves the number of feature channels and concatenation with the correspondingly cropped feature map from the downscaling path, from which we apply two convolutions with 3×3 filters, each followed by ReLU and batch normalization. At the final layer, a 1×1 convolution is used to learn a linear combination of all feature maps to reach the desired output. The two main parameters that influence performance are the number of downscaling (and subsequent upscaling) operations and the number of channels per feature map.

VIII.3 Deep learning for image reconstruction

Here, we focus on deep learning algorithms for solving inverse problems, and more specifically for image reconstruction. Different methods and architectures based on convolutional networks have been proposed for reconstruction problems. What we propose here is an (non-exhaustive) overview of the different categories of algorithms that have been introduced recently.

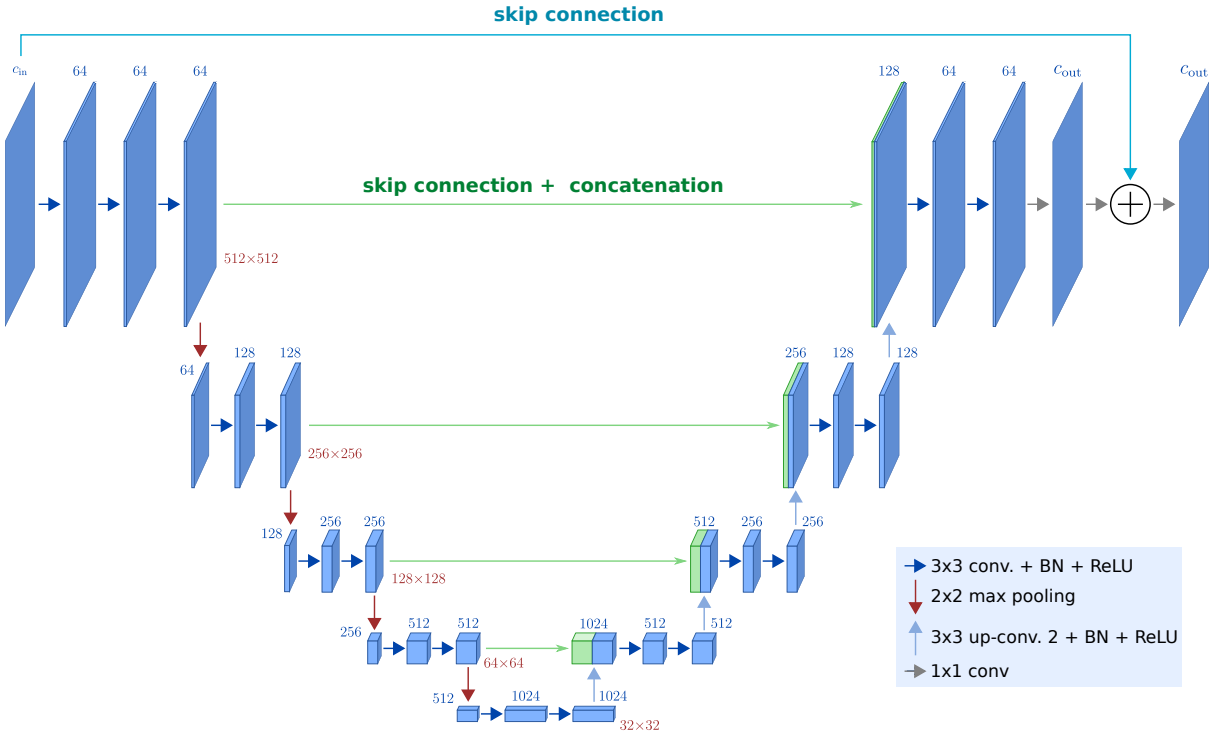


Figure 1.19.: Architecture of the U-Net. The network takes c_{in} image channels of size 512×512 as input and output c_{out} image channels of size 512×512. Adapted from (Jin et al. 2017)

We consider inverse (discretized) problems

$$y = \mathbf{F}(x) + \varepsilon \tag{1.192}$$

$y \in \mathbb{R}^M$ are the measurements, $x \in \mathbb{R}^N$ is the original data/image we aim to recover from the measurements, and $\varepsilon \in \mathbb{R}^M$ represents a vector of noise. Here $\mathbf{F} : \mathbb{R}^N \to \mathbb{R}^M$ is the (possibly nonlinear) forward operator that describes the direct problem.

VIII.3.1 Direct reconstruction

One of the simplest ways to use deep learning for reconstruction problems is to train the network to reconstruct directly from the measurements. In this case, the neural network f_θ is trained to learn (or rather approximate) the inverse operator $\mathbf{F}^{-1}$, so we leave the entire reconstruction task to the neural network.

Such an example is given by the Automated Transform by Manifold Approximation (AUTOMAP) (B. Zhu et al. 2017), where they reformulate image reconstruction as a supervised learning task driven by data, enabling the development of a mapping between the sensor and image domains. AUTOMAP was originally designed for Magnetic Resonance Imaging (MRI) reconstruction task, and its use has been extended to other types of imaging such as Computerized Tomography (CT) reconstruction (C. Liu et al. 2020). Its architecture includes two fully connected layers at the beginning, limiting its use to images of modest dimensions. Building upon this idea, the iRadonMap network (He, Yongbo Wang, and Ma 2020), whose architecture is composed of a first part completely connected, followed by a second part composed of CNN. This network, by being trained on generic images from ImageNet (Deng et al. 2009), was able to

generalize on medical data during the test phase

In the field of ultrasound elastography (S. Wu et al. 2018) proposed an end-to-end CNN to reconstruct directly from radio frequency (RF) data. They used a two-stage deep neural network, more specifically two cascaded CNNs, to first estimate motion after soft tissue compression and then compute the strain field. So they were able to reconstruct displacements and strains directly from the measurements.

Using deep learning for direct image reconstruction offers several advantages. Compared to traditional methods, this usually results in much better quality reconstructions and they can be significantly faster than iterative reconstruction. Moreover, that is an end-to-end learning process, meaning it directly maps raw measurement data to the reconstructed image so that simplifies the reconstruction pipeline. However, this approach has its own limitations. This generally requires large amounts of training data to generalize well, which is not always possible, as in the case of medical imaging. Further, this type of approach often requires complex architectures with many parameters, which results in a high demand for computational resources, but also in a risk of over-fitting. Finally, a lack of interpretability, makes it challenging to understand why a specific reconstruction was produced by the network.

Another important consideration for these methods is the total absence of model knowledge within the neural network. Without any information regarding the forward model in this context, concerns about robustness emerge.

VIII.3.2 Post-processing reconstruction

In this approach, image reconstruction is carried out using a simple inversion step, and a learned-based post-processing is used to remove artifacts and noise. Several methods have been proposed in this context, wherein a rapid and simple direct reconstruction algorithm is used to generate initially low-quality and flawed images. Subsequently, a CNN is trained to eliminate these artifacts.

In CT, for instance, instead of working in the projection domain, we can use an analytical method to recover an initial estimate to work in the image domain. Networks are trained in the image domain, which has proved to be more efficient than working directly on sinograms. One of the first methods for CT reconstruction post-processing is the FBPConvNet (Jin et al. 2017). From the measurements, an initial reconstruction is obtained using the filtered back projection (FBP), which is then used to feed a CNN. The architecture they used is the U-Net (the one depicted in Figure 1.19, with $c_{\text{in}} = c_{\text{out}} = 1$). Here the knowledge of the model is implicitly includes since the FBP incorporates the understanding of the physics underlying the inverse problem and additionally serves as an initial point for the CNN. A similar approach to FBPConvNet is taken by RED-CNN (H. Chen et al. 2017), which uses an encoder-decoder instead.

Inspired by FBPConvNet, authors of (Sandino et al. 2017) applied this idea to MRI imaging, using a U-Net type network. The raw Fourier data measurement is converted into the image domain, which is a more suitable space to leverage spatial structure and it offers a starting point for the CNN, as it eliminates the need for the network to learn the Fourier transform to reconstruct images. Although they were able to speed up the reconstruction time, they realized that their network struggled to perform well on cases substantially different from the training data. This indicates that the network learns spatial structure priors from the training data and applies them during inference on new data, rather than acquiring a deeper understanding of the MRI reconstruction process.

In photoacoustic tomography (PAT), a comparable approach has been proposed (Antholzer, Haltmeier, and Schwab 2018). The PAT filtered backprojection algorithm for the first layer, this

was then used to feed a U-Net architecture, which corresponds to the remaining layers. They have demonstrated the ability to achieve reconstruction quality comparable to state-of-the-art iterative algorithms with reduced computation time since it was comparable to a FBP.

The main concern with post-processing reconstruction methods is the potential loss of information from the raw data due to the initial reconstruction. Assessing the neural networks' capability to recover the solution becomes challenging when a portion of information from the measurements is lost during the initial reconstruction step.

VIII.3.3 Plug-and-Play priors

Introduced by Venkatakrishnan et al. (Venkatakrishnan, Bouman, and Wohlberg 2013), the key idea behind the Plug-and-Play (PnP) framework is to treat the reconstruction problem as an optimization task that combines traditional optimization techniques with modern denoising algorithms, including machine learning-based methods. We saw that solving the inverse problem (1.192) can be rewritten in its variational formulation

$$\underset{x\in\mathbb{R}^N}{\operatorname{argmin}} \left\{ \mathcal{J}(x,y) = \|\mathbf{F}(x) - y\|_2^2 + \mathcal{R}(x) \right\} \tag{1.193}$$

Instead of directly solving this optimization problem, the idea of PnP approach is to decompose it into two sub-problems: data fidelity and denoising.

Plug-and-Play: a practical example

For example, assuming that the forward operator $\mathbf{F}$ and the regularization are sufficiently smooth, one can solve (1.193) using the forward-backward splitting method 6. The k-th iteration of our FB scheme is then written:

$$\begin{cases} x_0 \text{ is given,} \\ x_{k-\frac{1}{2}} = x_{k-1} - \tau \mathbf{F}'\left(x_{k-1}\right)^* \left[\mathbf{F}\left(x_{k-1}\right) - y\right] \\ x_k = \operatorname{prox}_{\tau\mathcal{R}}\left(x_{k-\frac{1}{2}}\right), \quad \text{for } k = 1, \dots \end{cases} \tag{1.194}$$

We saw that the first step corresponds to a gradient descent step with respect to the data fidelity term, or as we will refer to it, a *data consistency step*. To simplify notations, we can introduce the $\mathrm{DC}[y] : \mathbb{R}^N \to \mathbb{R}^N$ function, which consists of this consistency step with respect to the measurement data y

$$\mathrm{DC}[y](x) := x - \tau \mathbf{F}'\left(x\right)^* \left[\mathbf{F}\left(x\right) - y\right] \tag{1.195}$$

Once we did that, we can think of this intermediate estimate $x_{k-\frac{1}{2}} = \mathrm{DC}(x_{k-1}, y)$ as a noisy version that has not been regularized. The second step consists in applying the proximal operator of the regularization, which is essentially a *denoising step* on the estimate $x_{k+\frac{1}{2}}$.

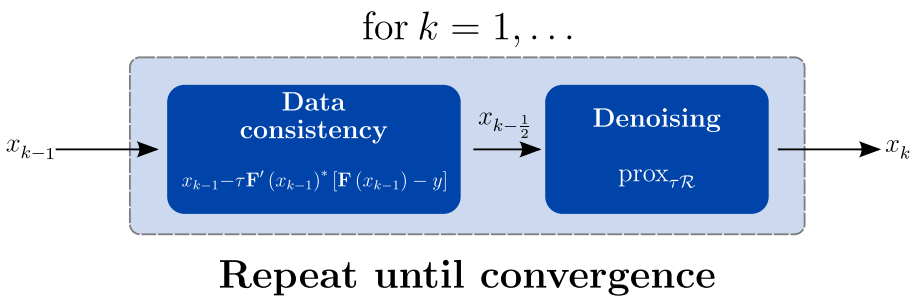


Figure 1.20.: Classical forward-backward splitting algorithm.

The idea behind the PnP approach is to replace this explicit step with a *plugged-in* denoising method. This includes non-local means (NLM) (Buades, Coll, and Morel 2011), Block-matching and 3D filtering (BM3D) (Dabov et al. 2007), and CNN as a denoiser. Here, we will only focus on using neural networks as denoisers. This is called plug-and-play because this network should be trained aside. Once trained, we just have to plug it inside an optimization scheme. If the denoising network is denoted by $\Gamma_\theta : \mathbb{R}^N \to \mathbb{R}^N$ the PnP-FBS method is given by

$$\begin{cases} x_0 \text{ known,} \\ x_k = (\Gamma_\theta \circ \mathrm{DC}[y])\,(x_{k-1}), \quad \text{for } k = 1, \dots \end{cases} \tag{1.196}$$

Iterations (1.196) are repeated until convergence or a criterion is reached. Unlike the previous approaches discussed in VIII.3.1 and VIII.3.2, these methods are not trained in an end-to-end manner. Instead, the denoising module is trained independently, without being tied to the specific inverse problem represented by the forward model $\mathbf{F}$. As explained in (Ongie et al. 2020), separating the training components from $\mathbf{F}$ leads to a flexible reconstruction pipeline that does not necessitate retraining for each new forward model. However, it may require a significantly larger number of training samples to achieve reconstruction accuracy comparable to methods trained end-to-end for a particular $\mathbf{F}$.

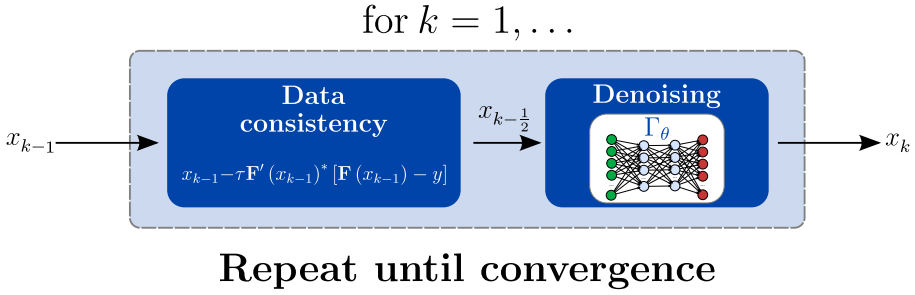


Figure 1.21.: PnP-forward-backward splitting algorithm.

Many PnP methods have been proposed, based on different optimization schemes. The first PnP method was the PnP-ADMM (Venkatakrishnan, Bouman, and Wohlberg 2013). Since then, several optimization methods such as the primal-dual algorithm (Meinhardt et al. 2017), Newton method (Buzzard et al. 2018) and stochastic FBS (Y. Sun, Wohlberg, and Kamilov 2019) have been used in a PnP framework.

Plug-and-Play: Theory

In recent years, PnP methods have received a great deal of attention, and have led to a major theoretical breakthrough. The convergence of PnP algorithms follows from the convergence of classical iterative algorithms. Proofs of convergence have been established for linear inverse problems, as long as the descent step τ is small enough and the denoiser network Γ_θ satisfies certain properties. For example, in (Ryu et al. 2019), the authors prove the convergence of the PnP-FBS and PnP-ADMM algorithms by forcing Γ_θ to satisfy a *Lipschitz condition*, namely

$$\| (\Gamma_\theta - \mathrm{Id})\,(x) - (\Gamma_\theta - \mathrm{Id})\,(y) \|^2 \leq \varepsilon^2 \|x - y\| \tag{1.197}$$

for all $x, y \in \mathbb{R}^N$ and for some $\varepsilon \geq 0$. This condition is achieved by using a technique based on the *spectral normalization* (Miyato et al. 2018). Another way of imposing properties on the Γ_θ network is proposed by (Pesquet et al. 2021). The authors propose to penalize the cost function so that the network acts like the resolvent of a maximally monotone operator. This is done by

constraining the network operator

$$Q_\theta = 2\Gamma_\theta - \mathrm{Id}$$

to be nonexpansive.

To our knowledge, no such proof of convergence has been proposed for nonlinear inverse problems. In this case, the convergence of classical algorithms requires the descent step to be variable (since it depends on the current iteration). This means, for example, having a different network for each iteration, or taking a small enough step size to prevent the algorithm from diverging. However, we do have convergence results towards nearly-optimal average points in practice (Y. Sun, Xu, et al. 2019; Kamilov, Mansour, and Wohlberg 2017).

VIII.3.4 Unrolling iterative methods

Using the variational formulation, we have written the reconstruction problem as follows: for each measurement y we want to find $\operatorname{argmin}_x \{\mathcal{J}(x,y)\}$ where $\mathcal{J}$ is given by (1.193). We can therefore define the function $\mathbf{J}$ which is parametrized by y and gives us a minimum of the functional $\mathcal{J}(x,y)$, in other words:

$$\begin{array}{rccc} \mathbf{J}: & \mathbb{R}^M & \to & \mathbb{R}^N \\ & y & \mapsto & \operatorname{argmin}_x \{\mathcal{J}(x,y)\} \end{array} \tag{1.198}$$

The core concept of unrolling involves training a deep neural network Λ_θ with learning parameters θ to effectively approximate the operator $\mathbf{J}$. This network is implicitly established through an iterative procedure.

Let us go back to the FBS scheme defined above (1.196) and let us suppose that we stop the algorithm after N iterations, which gives us

$$\begin{cases} x_0 \text{ is given,} \\ x_k = \operatorname{prox}_{\tau\mathcal{R}}\left(\mathrm{DC}[y]\left(x_{k-1}\right)\right), \quad \text{for } k = 1,\dots,N \end{cases} \tag{1.199}$$

Once the algorithm has been unrolled over N iterations, the idea is, at each iteration $k \in \{1,\dots,N\}$, to replace the proximity operator with a network Γ_{θ_k}:

$$\begin{cases} x_0 \text{ is given,} \\ x_k = \Gamma_{\theta_k}\left(\mathrm{DC}[y]\left(x_{k-1}\right)\right), \quad \text{for } k = 1,\dots,N \end{cases} \tag{1.200}$$

This allows us to express the N-th iterate by

$$x_N = \left(\Gamma_{\theta_N} \circ \mathrm{DC}[y]\right) \circ \cdots \circ \left(\Gamma_{\theta_1} \circ \mathrm{DC}[y]\right)(x_0) \tag{1.201}$$

Now, we can consider the function $\Lambda_\theta : \mathbb{R}^N \to \mathbb{R}^N$

$$\Lambda_\theta = \left(\Gamma_{\theta_N} \circ \mathrm{DC}[y]\right) \circ \cdots \circ \left(\Gamma_{\theta_1} \circ \mathrm{DC}[y]\right) \tag{1.202}$$

as a feedforward neural network representing N iterations, where $\theta = \{\theta_1,\dots,\theta_N\}$ are the learning parameters. The number of iterations N is a hyperparameter since it has to be chosen in advance. The network Λ_θ incorporates explicitly the model knowledge through the operators $\mathbf{F}'(x)^*$ and $\mathbf{F}$ inside the DC function, which can be seen as layers with fixed parameters. Now, Λ_θ can be trained in an end-to-end manner, so unlike PnP approaches, the network will be dependent on the forward model. But this is also what will enable the network to learn a regularization more adapted to the data. In addition, unrolling methods are often fast because

they require a fixed (often small) number of iterations N, whereas PnP approaches are just as time-consuming as conventional iterative methods.

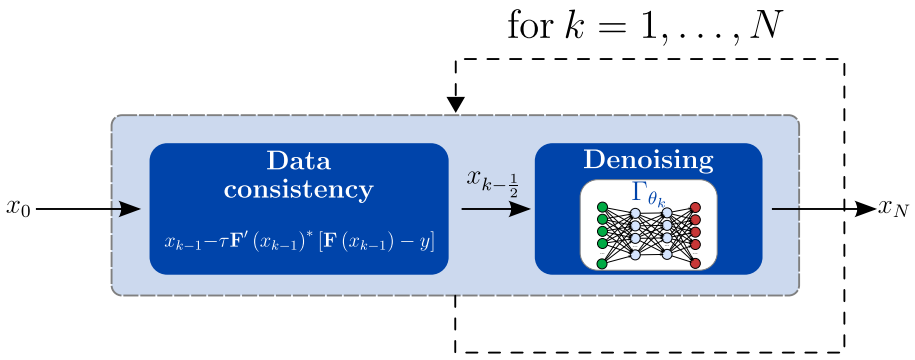


Figure 1.22.: Deep unrolled forward-backward splitting algorithm.

In the review article by Monga et al. (Monga, Y. Li, and Eldar 2021), we can see that unrolling methods have been getting a lot of attention lately. These approaches have been used for a wide range of applications and are based on different optimization schemes: gradient descent algorithms (Hauptmann et al. 2018), primal-dual schemes (Jonas Adler and Ozan Öktem 2018), and ADMM (Yan Yang et al. 2020). These approaches have outperformed state-of-the-art methods in a number of problems. However, one drawback is the memory required for training, since the network has to be copied N times, the memory needed for computing the backpropagation updates increases linearly with the number of unrolled iterations.

VIII.3.5 Deep Equilibrium Models

Deep equilibrium model (DEM) has been introduced by Bai et al. (S. Bai, Kolter, and Koltun 2019) for modeling sequential data. Inspired by the observation that the hidden layers in numerous pre-existing deep sequence models tend to converge to fixed points, they introduce the DEM approach, which directly identifies these equilibrium points through root-finding techniques. Inspired by this work, (Gilton, Ongie, and Willett 2021) proposed to apply this approach to inverse problems.

Let us consider a N layers network with learnable parameters θ, y the input data and x_k the output of the k-th hidden layer, we can write:

$$x_k = T_\theta^{(k)}(x_{k-1}, y), \quad \text{for } k = 1, \ldots, N \tag{1.203}$$

where k is the layer index and $T_\theta^{(k)}$ represents the action of the k-th layer (linear operation followed by activation). In case of *weight tying*, i.e. if the transformation is the same at each layer $T_\theta^{(1)} = \cdots = T_\theta^{(L)} = T_\theta$, this can be rewritten as the recursion

$$x_k = T_\theta(x_{k-1}, y), \quad \text{for } k = 1, \ldots, N \tag{1.204}$$

So, what happens when the number of layers L goes to infinity? Due to the increasing memory demands, that increase linearly with N, directly unrolling a sequence generated by the successive application of $T_\theta(\cdot, y)$ becomes generally infeasible for very large N. However, it is possible to model the sequence limit by employing a fixed point equation. In such cases, assuming that the limit $x_\infty = \lim_{k \to +\infty} x_k$ exists, then this is a fixed point of the operator $T_\theta(\cdot, y)$.

Let us see how to design an iteration map $T_\theta(\cdot, y)$ for the FBS algorithm so that the fixed-point

x_∞ satisfies

$$x_\infty = T_\theta\left(x_\infty, y\right) \tag{1.205}$$

In the same way as the PnP and unrolling approaches, we consider replacing the proximity operator with a CNN $\Gamma_\theta : \mathbb{R}^N \to \mathbb{R}^N$ which gives the map

$$T_\theta\left(x, y\right) = \Gamma_\theta\left(\mathrm{DC}[y](x)\right) \tag{1.206}$$

defining a DEM.

A way to obtain the fixed-points of $T_\theta\left(\cdot, y\right)$ is the Banach–Picard fixed-point theorem which consists in the iterations (1.204). Following this idea, the learning parameters θ must be adjusted so that applying $T_\theta\left(\cdot, y\right)$ a sufficient number of times produces a fixed point. This means that during both training and inference, the DEM model requires calculating a fixed point of the map $T_\theta\left(\cdot, y\right)$ (see (Gilton, Ongie, and Willett 2021; Heaton et al. 2021) for more details). In order to satisfy the hypothesis of the Banach-Picard fixed-point theorem, the iteration map $T_\theta\left(\cdot, y\right)$ should be *contractive*

$$\exists 0 < L < 1, \quad \forall x_1, x_2 \in \mathbb{R}^N \quad \|T_\theta\left(x_1, y\right) - T_\theta\left(x_2, y\right)\| \leq L\|x_1 - x_2\| \tag{1.207}$$

This is true if the regularization network Γ_θ satisfies some Lipschitz condition:

$$\exists 0 < \varepsilon, \quad \forall x_1, x_2 \in \mathbb{R}^N \quad \|\left(\Gamma_\theta - \mathrm{Id}\right)\left(x_1\right) - \left(\Gamma_\theta - \mathrm{Id}\right)\left(x_2\right)\| \leq \varepsilon\|x_1 - x_2\| \tag{1.208}$$

which is equivalent to assume that $\left(\Gamma_\theta - \mathrm{Id}\right)$ is ε-Lipschitz. This is done by using spectral normalization (as in (Ryu et al. 2019)).

In (Gilton, Ongie, and Willett 2021), they used this approach based on different optimization schemes (gradient descent, FBS, and ADMM) and apply them on several inverse problems: image deblurring, compressed sensing, and accelerated MRI reconstruction. They showed that their approach produced better results than PnP, RED and unrolling approaches.

One obvious advantage of DEM compared to unrolling is that it does not require the choice of a fixed number of iterations N and subsequently, it reduces the memory requirement because only a single iteration is trained. This is not without consequence, since each forward and backward pass requires the computation of a fixed-point, and this takes much longer to train.

VIII.4 Deep learning for phase retrieval

We have seen several ways in which neural networks can be used in reconstruction problems. Most of these approaches were introduced for linear inverse problems, and some of them are not necessarily suited to the nonlinear case. Here, we present a non-exhaustive review of deep learning methods for phase retrieval.

VIII.4.1 Direct reconstruction

Different architectures of convolutional neural networks have been proposed to recover the phase shift from diffraction patterns. The networks are trained in an end-to-end manner to reconstruct directly from the measurements, without explicit knowledge of the underlying physics or processes.

Rivenson et al. (Rivenson et al. 2017) proposed an approach based on deep learning that introduces a completely novel framework for conducting holographic imaging, enabling the rapid removal of spatial artifacts associated with twin-image and self-interference issues. The use of a multi-scale convolution neural network (Figure 1.23) has shown that we could reconstruct the amplitude and phase images of the unknown complex objects from a single distance diffraction

pattern and without a priori information. Unlike classic auto-encoders such as the U-Net, which contains millions of parameters, by combining short and long range information, we could reduce the number of parameters to learn while having a good quality of reconstruction. It has been shown that using a shallow U-Net could reduce the training time as well as the number of parameters while maintaining good performance (S. Z. Li et al. 2022).

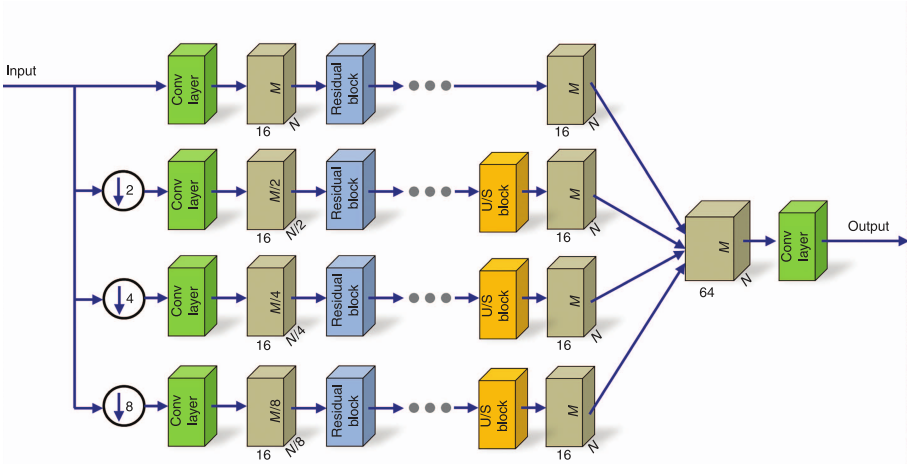


Figure 1.23.: The deep neural network consists of convolutional layers, residual blocks, and upsampling blocks, allowing for the simultaneous and efficient processing of complex-valued input images across multiple scales in parallel (Rivenson et al. 2017).

Authors of (Y. Wu et al. 2022) added a phase contrast refinement module before performing phase retrieval in order to stabilize and generalize the reconstruction. More recently, other approaches have proposed to use Multilayer Perception block to process the input measurement data, and to couple this with a residual attention mechanism (Ye, L.-W. Wang, and D. P. K. Lun 2022) or encoder-decoder network (Gugel and Dekel 2022) for the reconstruction part.

The main advantage of direct reconstruction is that, after being trained, the inference of these different networks is very fast. However, this approach often requires large training sets as well as long training time. As these methods only take the observation measurements as input, the reconstruction quality is limited by the accuracy of the training data, thus such end-to-end models may struggle to make accurate predictions outside the range of training data, in particular for experimental data.

VIII.4.2 Physics-informed neural networks

Physics-informed methods have been introduced to bridge the gap between data-driven approaches and prior knowledge in solving inverse problems. They combine the power of data-driven techniques with the richness of physics-based models to enable more accurate, robust, and interpretable solutions. The Physics-informed neural network (PINN) is a framework that combines physics-based modeling with deep learning techniques. Incorporating knowledge of the underlying physics in the models can be done explicitly or implicitly and in different ways.

Several physics-informed methods for phase retrieval have been inspired by the HIO algorithm (see section V.2) in order to add physical constraints in their network. For example, (Işil, F. S. Oktem, and Koç 2019) suggested an iterative method that utilizes modified U-nets and HIO method alternatively. They use HIO as initialization, then iteratively update between U-Net

and HIO, and finally, they refine the reconstruction with another U-Net. But such iterative physics-driven approaches are usually computationally demanding and cannot be end-to-end trained, which can affect the overall performance. In (Ye, L.-W. Wang, and D. P.-K. Lun 2022), they proposed a *Physics-Driven Phase Retrieval Network* (PPRNet) with an architecture inspired by U-Net. They added in the decoder part a *Physics-driven Unwinding Block* which integrates the physical information. Zhang et al. (Y. Zhang et al. 2021) introduced the *PhaseGAN* network which integrate the Fresnel propagator into the *CycleGAN* (J.-Y. Zhu, Park, Isola, and Alexei A Efros 2017a). This unsupervised approach allowed to enhance the phase reconstructions by using physical knowledge of image formation.

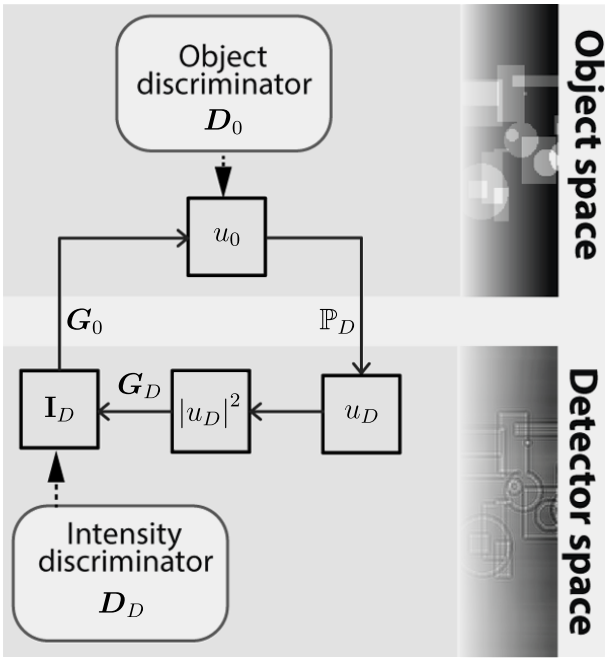


Figure 1.24.: Schematic approach of PhaseGAN (Y. Zhang et al. 2021). Each of the GANs is decomposed in their generator $\boldsymbol{G}$ and discriminator $\boldsymbol{D}$. The generators used in PhaseGAN are U-Net and the discriminators are PatchGAN discriminators (J.-Y. Zhu, Park, Isola, and Alexei A. Efros 2017b). $\boldsymbol{G}_0$ is the phase-reconstruction generator, which takes a single intensity measure and produces the phase and amplitude of the complex-object wave field u_0. $\boldsymbol{D}_0$ is the discriminator of the phase reconstruction. The object wavefield u_0 is then propagated using the fresnel propagator $\mathbb{P}_D$ to the detector plane $u_D = \mathbb{P}_D u_0$, and the intensity in the detector plane is computed $|u_D|^2$. $\mathbf{I}_D$ denotes the measured detector intensity.

Another way to take into account the physics of the image formation is by adapting the loss with the forward operator, which helps to regularize the learning process and ensure that the estimated phases adhere to the underlying physics. The loss then consists in minimizing the error between the measured intensity and the intensity obtained from the reconstruction, this has been done in both supervised (Manekar et al. 2020) and unsupervised (Xiang et al. 2022) ways.

There are several reasons why physics-informed methods have gained prominence. By incorporating prior knowledge about the physical properties of the imaging system, the PINN framework enhances the accuracy and reliability of the phase retrieval process. The advantage

of the PINN approach lies in its ability to leverage the physics of the imaging process, which can significantly improve the accuracy and robustness of the reconstructions. The network learns to generate physically plausible phase estimates, even in the presence of noise and limited data. However, it is important to note that the success of the PINN approach relies on accurate modeling of the underlying physics and the availability of representative training data.

VIII.4.3 Regularization by denoising

The Regularization by Denoising (RED) (Romano, Elad, and Milanfar 2017) is a particular case of the PnP framework where the regularizer has an explicit form

$$\mathcal{R}(x) = \frac{\alpha}{2} x^\top \left[x - \mathcal{D}(x)\right] \tag{1.209}$$

where $\alpha > 0$ is a regularization parameter and $\mathcal{D}$ is an arbitrary image denoising engine such as NLM, BM3D or a CNN denoiser (see section VIII.3.3). In this context, we are reduced to solving the problem

$$\underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \|\mathbf{F}(x) - y\|_2^2 + x^\top \left[x - \mathcal{D}(x)\right] \right\} \tag{1.210}$$

Authors of (Metzler et al. 2018) introduced the *prDeep* method, which consists of integrating a CNN denoiser Γ_θ such as the denoising convolutional neural network (DnCNN) (K. Zhang et al. 2017), into the RED framework, this amounts to solve

$$\underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ \|\mathbf{F}(x) - y\|_2^2 + x^\top \left[x - \mathcal{D}(x)\right] \right\} \tag{1.211}$$

They use the fast adaptive shrinkage/thresholding algorithm (FASTA) (Goldstein, Studer, and Baraniuk 2014) to solve (1.211), which is simply the forward-backward splitting algorithm with adaptive step sizes for acceleration. They showed that in practice, they had the convergence of the FASTA algorithm. Although the images used were only natural ones, the results were promising. So, this approach was then taken up in the propagation-based setting, using the CTF linearized model (C. Bai et al. 2019), the CTF-Deep phase retrieval method was proposed (Figure 1.25). Recovery of the phase was possible using a single diffraction pattern, under homogeneous object assumption. On experimental data, they showed that their method was robust even in the presence of high noise.

The advantage of such methods is their flexibility and adaptability. Different denoisers or regularization modules can be plugged into the iterative algorithm, allowing for customization based on the specific characteristics of the problem at hand. The fact that the denoiser is trained independently of the model allows the application of such a method in other experimental configurations without having to retrain. However, it is worth noting that this approach may not achieve comparable reconstruction accuracy to an end-to-end trained method specifically designed for a particular forward model. Moreover, the required time is greater than with direct methods, since a certain number of iterations are needed to reach convergence.

VIII.4.4 Deep unrolling approach

Recently, a few works on the deep unrolled network to solve phase retrieval problems by unrolling the physical model-based approach into a network fashion have been proposed. In (Naimipour, Khobahi, and Soltanalian 2020), they addressed the phase retrieval problem and the sparse phase retrieval problem by developing a hybrid architecture that combines model-based and data-driven elements, referred to as Unfolded Phase Retrieval (UPR). This architecture was then employed to search for the optimal parameters defined within the iterative optimization

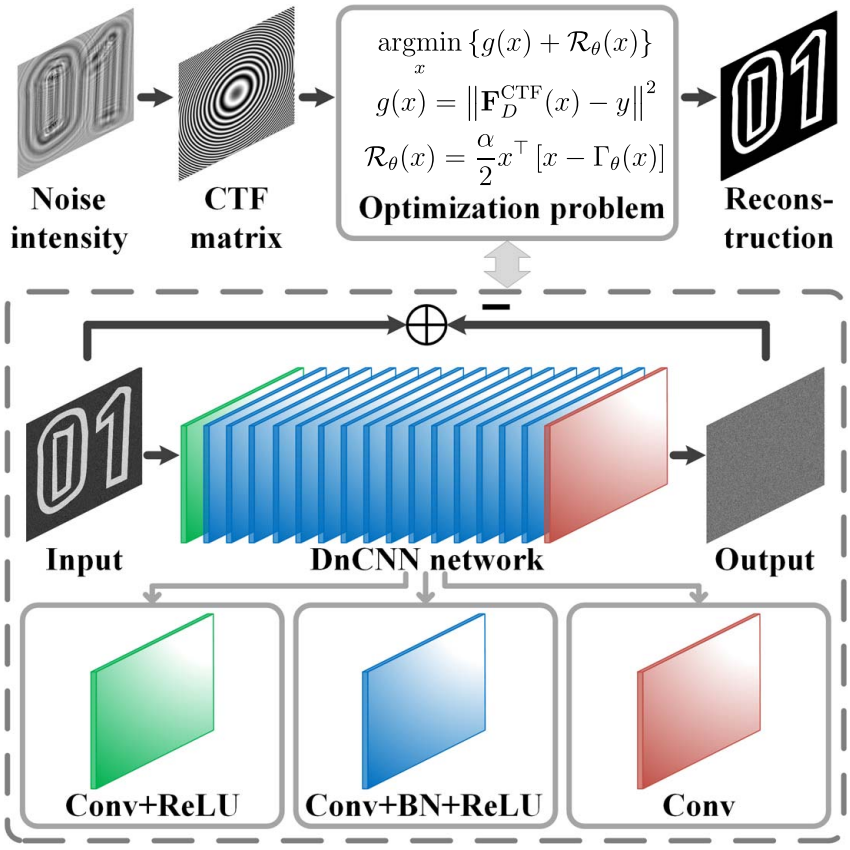


Figure 1.25.: Schematic approach of the CTF-Deep phase retrieval method, which consists of a CTF matrix estimation, a unified optimization problem, and an inner loop with a DnCNN denoiser (C. Bai et al. 2019).

algorithm. Another approach proposed to unroll an ADMM algorithm (UADMM) (Vial et al. 2022). They substituted the proximity operator found in this algorithm with learnable activation functions (Agostinelli et al. 2014). In experiments, UADMM surpasses the classical ADMM algorithm in performance while maintaining a lightweight and easily interpretable structure. They also compared the case where the weights are tied, i.e. the network is the same at each iteration, with the case where the weights are not tied. UADMM-untied achieves a lower loss value compared to UADMM-tied, its tied counterpart, which can be a consequence that this variant has a greater number of trainable parameters.

Finally, we cite the example of HIONet (Yuchi Yang et al. 2023), which, as its name suggests, was inspired by Fienup's HIO algorithm. But unlike the other approaches mentioned above, HIONet is an unrolled approach. Let us assume that support as the object domain constraint, from proposition V.1, we can write the HIO update as

$$x_{k+1} = \frac{1}{2}\left[R_S\left(R_M + (\beta-1)P_M\right) + \mathrm{Id} + (1-\beta)P_M\right](x_k) \tag{1.212}$$

This can be rewritten only in terms of projectors

$$x_{k+1} = \left[(1+\beta)P_S P_M + \mathrm{Id} - P_S - \beta P_M\right](x_k) \tag{1.213}$$

They suggested to keep the P_M operator which introduces the measurements constraint into the model, and to replace the P_S operator with a *deep prior projection operator* P_r.

Compared to conventional iterative approaches, which sometimes require a thousand iterations, the unrolling methods require just a few dozen of iterations, and often produce better quality images. These previous methods focus on the classic phase retrieval problem of recovering an image f from the magnitude of its Fourier transform $|\mathcal{F}(f)|^2$, which manifests in the far-field regime. A more in-depth study of unrolling methods for phase retrieval in a propagation-based set-up can be found in chapter 5.

2 | Mixed scale dense convolutional networks for X-ray phase contrast imaging

Deep learning has gained significant popularity and success in image reconstruction tasks due to its ability to learn complex and data-driven representations directly from large amounts of training data. Deep learning for direct reconstruction has demonstrated remarkable success in various image reconstruction tasks, including image denoising, super-resolution, inpainting, deblurring, and compressive sensing. In this context, the entire reconstruction process is learned by training a deep neural network to map the measurements directly to the desired output. The network architecture plays a crucial role in deep learning for direct reconstruction.

In this chapter, we propose supervised learning approaches using mixed scale dense convolutional neural networks to simultaneously retrieve the phase and the attenuation from X-ray phase contrast images. This network architecture uses dilated convolutions to capture features at different images scales and densely connects all feature maps. The long range information in images becomes quickly available and greater receptive field size can be obtained without losing resolution. This network architecture seems to account for the effect of the Fresnel operator very efficiently. We train the networks using simulated data of objects consisting of either homogeneous components, characterized by a fixed ratio of the induced refractive phase shifts and attenuation, or heterogeneous components, consisting of various materials. We also trained the networks in the image domain, by applying a simple initial reconstruction using the adjoint of the Fréchet derivative. We compare the results obtained with the MS-D network to reconstructions using the U-Net, another popular network architecture, as well as to reconstructions using the Contrast Transfer Function method, a direct phase and attenuation

retrieval method based on linearization of the direct problem. The networks are evaluated using simulated noisy data as well as images acquired at NanoMAX (MAX IV, Lund, Sweden). In all cases, large improvements of the reconstruction errors are obtained on simulated data compared to the linearized method. Moreover, on experimental data, the networks improve the reconstruction quantitatively, improving the low-frequency behavior and the resolution.

I Introduction

In order to obtain an approximate solution, direct inversion approaches based on linearization of the forward problem have been proposed. The Contrast Transfer Function (CTF) (Paganin 2006) is such a method which allows to retrieve both attenuation and phase. Others methods rather rely on assumptions on relationships between phase and attenuation (D. Paganin, S. C. Mayo, et al. 2002; Max Langer, Peter Cloetens, Hesse, Suhonen, Pacureanu, Raum, and Françoise Peyrin 2014). Iterative methods are not limited by these constraints. Among them, there are techniques which retrieve the object by alternating projections on constraints between the detector and object space (James R. Fienup 1982). These also include variational approaches based on the Fréchet derivative of the forward operator (Bruno Sixou et al. 2013) in conjunction with the Landweber algorithm. This type of algorithm permits a flexible inclusion of priors, based on non-negativity or Total Variation, for instance, but consider the attenuation and the phase as independent unknowns to retrieve. More recently, second-order type methods have been proposed such as iteratively regularized Gauss-Newton (IRGN) (Maretzke, Bartels, et al. 2016). Deep learning methods have been much developed in recent years for signal processing tasks (LeCun, Y. Bengio, and Hinton 2015). Recent approaches based on deep learning have yielded promising results for reducing the reconstruction error for several inverse problems (Arridge et al. 2019; Jonas Adler and Lunz 2018). Some approaches optimize a reconstruction network trained to map the measured data and the reconstructed image (Jin et al. 2017). Several iterative schemes have been proposed using deep learning methods to improve the results obtained with classical iterative approaches for inverse problems (Jonas Adler and O. Oktem 2017; Hauptmann et al. 2018). Some deep learning architectures applied to the phase problem have been proposed, for instance, in order to learn a regularization into a CTF-based optimization algorithm (C. Bai et al. 2019). Others like PhaseGAN (Y. Zhang et al. 2021) were able to recover both attenuation and phase from a single measured intensity by including explicitly the Fresnel propagator in the training. But few proposed to retrieve both attenuation and phase directly from the diffraction patterns. In recent developments, a novel network architecture known as the Mixed Scale Dense (MS-D) network was introduced (M.Pelt and Sethian 2018). This innovative architecture has demonstrated significant enhancements in reconstruction quality, surpassing traditional techniques and even outperforming other convolutional neural networks in the context of tomographic reconstruction problems (Pelt, Batenburg, and Sethian 2018).

Hence, the objective of this chapter is to create an end-to-end deep learning framework for extracting phase and attenuation information from X-ray phase contrast images through the utilization of MS-D neural networks. We assess the reconstruction outcomes against those obtained by the conventional model-based CTF linear approach. The neural network was trained using synthetic data representing objects composed of various homogeneous materials, both single and combined, at different signal-to-noise ratios and using a single or several propagation distances.

II Mixed Scale Dense Convolutional Neural Networks

II.1 Dilated convolutions

Dilated convolutions are a specialized type of convolutional operation that allow for an increased receptive field without significantly increasing the number of parameters or computational cost. In a traditional convolution, a filter slides over the input image or feature map with a fixed stride, resulting in a limited view of the input. However, dilated convolutions introduce gaps or "holes" within the filter, effectively enlarging its receptive field thanks to the introduction of a dilation ratio l. More formally, the discrete operation that defines the action of a l-dilated convolution filter K on an image I is given by:

$$(K *_l I)(i,j) = \sum_m \sum_n I(m,n)K(m+il, n+jl) \tag{2.1}$$

The familiar discrete convolution $*$ is simply the 1-dilation convolution.

By adjusting the dilation rate, which determines the spacing between the gaps, the receptive field can be expanded exponentially (F. Yu and Koltun 2016). This enables dilated convolutions to capture larger contextual information while preserving the spatial resolution of the input. This property has proven particularly useful in tasks such as semantic segmentation, where capturing fine-grained details and global context is crucial (L.-C. Chen et al. 2016). Additionally, dilated convolutions have been widely adopted in deep learning architectures to improve performance and efficiency (Oord et al. 2016; Z. Wu, Shen, and van den Hengel 2019; K. Sun et al. 2019).

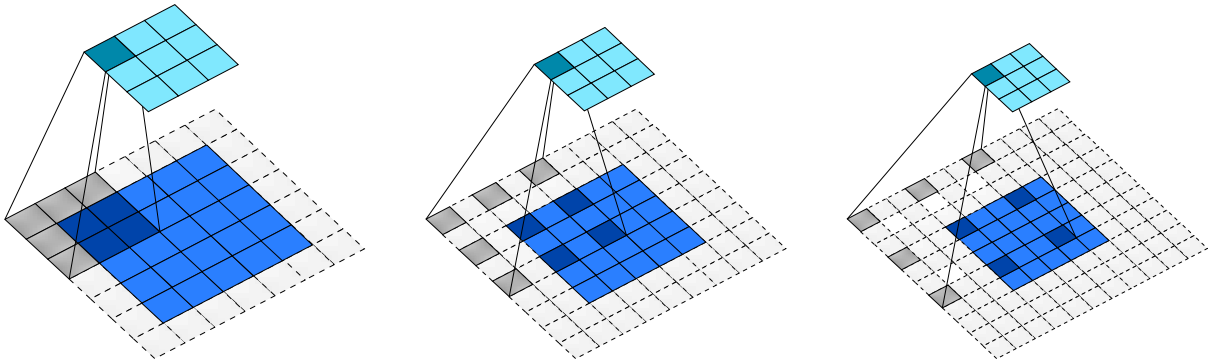


Figure 2.1.: Dilation convolutions using different dilation rates.

II.2 Dense connection

A densely connected convolutional network, also known as DenseNet (Huang, Z. Liu, et al. 2017), is a deep learning architecture that incorporates *dense connections* between layers. Unlike traditional convolutional neural networks where each layer is connected only to its subsequent layer, DenseNet introduces direct connections between every layer within the network. This connectivity pattern gives rise to a densely connected block structure, where each layer receives feature maps from all preceding layers and passes its own feature maps to all subsequent layers. Consequently, information flow within the network becomes more efficient, enabling better gradient propagation and feature reuse.

The key idea behind DenseNet is to encourage feature reuse by allowing each layer to have direct access to the feature maps of all preceding layers. This design has several advantages.

Firstly, it reduces the number of parameters by reusing features instead of learning them from scratch in each layer. Secondly, it enhances gradient flow and addresses the vanishing gradient problem, which can hinder the training of very deep networks. Additionally, the dense connections promote better feature propagation throughout the network, enabling better representation learning and enhancing the network's ability to capture complex patterns and dependencies.

Densely connected networks have demonstrated impressive performance in various computer vision tasks, such as image classification (Huang, S. Liu, et al. 2018) and segmentation (Dutta 2021). They have achieved state-of-the-art results while maintaining relatively compact model sizes. Dense connectivity and feature reuse properties make it an effective and efficient choice for deep learning tasks, especially in scenarios with limited training data or computational resources.

II.3 MS-D Net

The mixed scale dense (MS-D) neural networks has recently been proposed (M.Pelt and Sethian 2018). Compared to encoder-decoder networks, the MS-D network requires fewer trainable parameters and intermediate images while achieving precise reconstruction outcomes. It facilitates the handling of large images and reduces the requisite quantity of training data, making it particularly advantageous for enhancing tomographic reconstruction in scenarios with limited data availability. This network gives more accurate reconstruction results than traditional methods or others convolution neural networks (Pelt, Batenburg, and Sethian 2018). MS-D networks are densely connected : in order to compute an image of a certain layer, all previous layer images are used as input instead of only those of the previous layers. MS-D networks use dilated convolutions in order to retain image features at various scales. The scales are mixed by choosing adapted dilation factor distributions to avoid the gridding effect (Z. Wang and Ji 2018). Dilated convolutional filters allow the access to long-range information within images during the initial layers of the network. This early acquisition of a larger receptive field size enables the utilization of this information to improve the results in deeper layers, which seems in line with the action of the Fresnel operator.

Each feature map is the result of applying the same set of operations to all previous feature maps:

1. dilated convolutions with 3×3 filters with a dilation rate selected from the list $[d_{\min}, d_{\min}+1, \ldots, d_{\max}]$
2. summing resulting images
3. adding a constant bias
4. applying rectified linear unit (ReLU) activation function

Finally, the output of the network results of a linear combination involving all the generated feature maps and input channels, following the application of the ReLU activation function. The weights assigned to each feature map, including input channels, are learned according to the receptive field in the generated images that is in line with the desired output. This implies that feature maps or input channels whose receptive field closely aligns with the desired output will be given more weight in the final output formation with processing by point wise convolution.

III Experiments

III.1 Simulated datasets for training

In order to compare the performance of the MS-D network and of the CTF, we generated synthetic X-ray phase contrast images. The X-ray energy was set to 13 keV for a wavelength

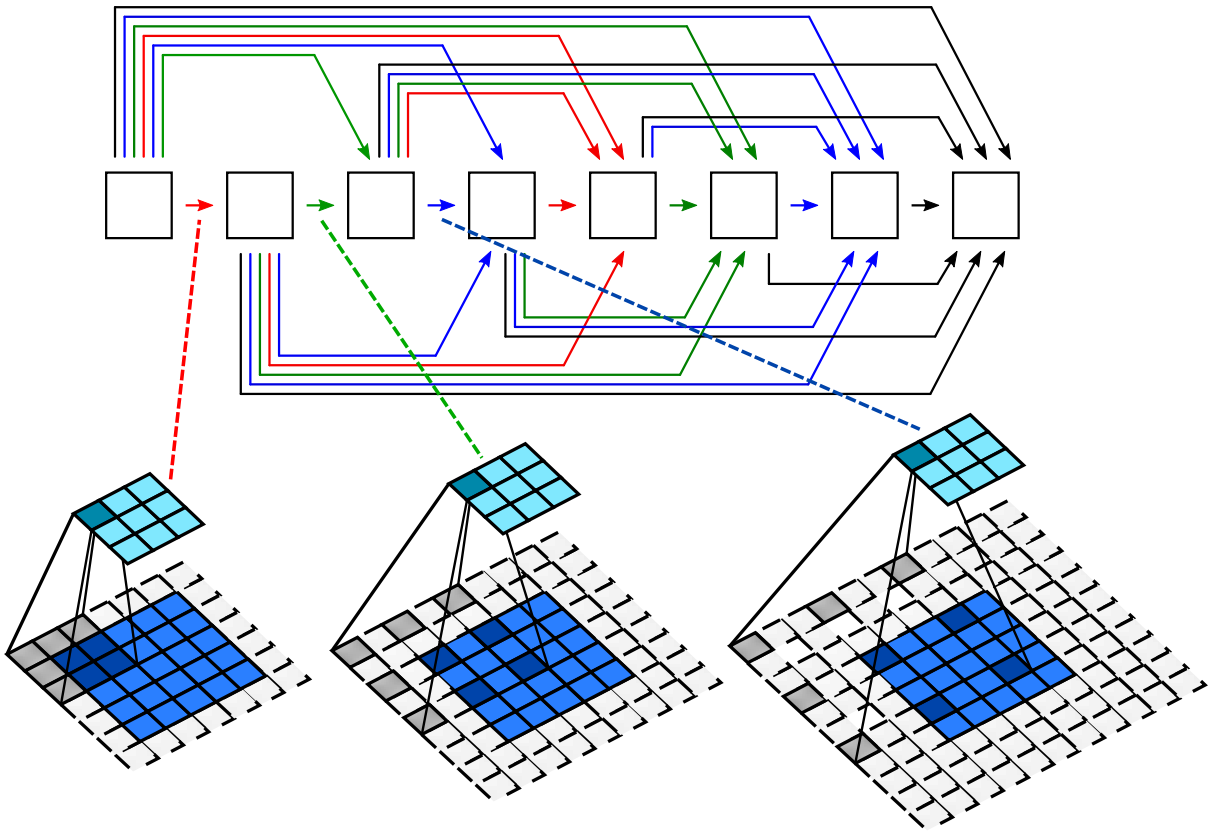


Figure 2.2.: The mixed scale dense network architecture using $L = 7$ layers and a dilation factor $l = 3$.

of $\lambda = 0.095$ nm, and the pixel size in the object space was set to 6 nm. We created projection datasets from 3D objects created from random combinations of 1 to 10 shapes, consisting with either one homogeneous material (to create homogenous objects) or three different materials (to create heterogeneous objects). The refractive indices (1.32) used for the materials are given in Table 2.1.

Tableau 2.1.: Complex refractive indices materials at 13 keV

Material	Symbol	$\mu(\mathrm{cm}^{-1})$	$\frac{2\pi}{\lambda}\delta_r(\mathrm{cm}^{-1})$	δ_r/β
Gold	Au	2 790	11 395	8.16
Palladium	Pd	615	8 251	26.83
Zinc	Zn	859	5 270	12.27

The shapes used were ellipsoids and paraboloids with random positions and orientations. 2D analytical tomographic projections of the real and imaginary part of the refractive index, corresponding to the phase (1.35) and the attenuation (1.34) respectively, were obtained from the 3D objects for an image size of 2048×2048 pixels. Objects and projections were generated using the software *TomoPhantom* (Kazantsev et al. 2018). Phase contrast images were generated from the projection images according to (1.42) at propagation distances $D = [10.1, 15.5, 17.8, 19, 20.3]$

mm and downsampled to 512×512 to avoid aliasing in the calculation of the diffraction patterns. The datasets were generated using different levels of white Gaussian noise to yield a certain peak to peak signal to noise ratio (PPSNR) in the longest distance. The noise level was kept the same in all distances, corresponding to usual experimental conditions. We generated two datasets. The first consisted of only homogeneous objects, the material used was gold, The second consisted of heterogeneous objects using the three materials given in Tab. 2.1. Each dataset consisted of 12 000 pairs of 5 input images (phase contrast images at different propagation distances) and 2 output images (attenuation and phase). From each dataset, 10 000 images were used for training, 1 000 for validation during training, and 1 000 for evaluation. An augmentation of the training data was performed, by random 90 degree rotations or flipping, to a factor of 2, yielding a total of 20 000 training images.

IV Results

For the simulations, we trained 6 MS-D networks corresponding to homogeneous and heterogeneous objects using a single distance ($D = 10.1$ mm) or the 5 distances. Overall, we used the same MS-D network architecture composed of $L = 100$ layers and 3×3 dilated convolutional kernel. The dilation rates was selected in the list $[1, 2, \ldots, 10, 1, 2, \ldots, 10, 1, 2 \ldots]$. The networks were trained using the ADAM optimizer with l_2 norm between labels and predictions as loss function. An independent set of image pairs was used as a validation set to monitor the network quality during training and provide a stopping criterion. The network parameters that yielded the lowest validation error were saved as output of the training procedure.

IV.1 Simulation results

In this section, we evaluate the different trained MS-D networks on synthetic data. We compare results of trained MS-D networks with the popular U-Net architecture (Jin et al. 2017) (see fig. 1.19) and to CTF using the normalized mean square error (NMSE) defined by :

$$\mathrm{NMSE}(x) = \frac{\|x - x_{\mathrm{true}}\|_2}{\|x_{\mathrm{true}}\|_2} \tag{2.2}$$

As quantitative measures of reconstruction quality, we computed the average NMSE on 1 000 images that were not used neither for training nor validation. In all cases, the regularization parameter for the CTF method was optimized upstream.

The results for homogeneous objects are summarized in Table 2.2. The MS-D network successfully reconstructs attenuation and phase as identical up to a constant factor, as shown by the corresponding reconstruction errors for both. Conversely, the U-Net model struggled to reconstruct both attenuation and phase; it exclusively retrieved the phase while setting the attenuation to zero. This is why our training strategy for U-Net involved outputting a single channel (phase) and treating the attenuation as proportional to the phase, which explain the identical reconstruction error.

The results obtained on heterogeneous objects are summarized in Tab.2.4. The performance of the MS-D network is slightly diminished in this context, because of the diversity of the dataset, nevertheless, the results remain very good. Notably, the network exhibits a somewhat superior proficiency in retrieving the phase compared to the attenuation. For a qualitative assessment, we present a selected example of reconstructed phase projections in Figure 2.3.

Tableau 2.2.: Normalized mean square error and standard deviation (in %) for 1 000 test images, homogeneous objects.

	#Distances	#Parameters	NMSE (in %)	
			Absorption	Phase
CTFHomo	5	-	13.5 (3.92)	13.5 (3.92)
CTFHomo	1	-	21.5 (14.0)	21.5 (14.0)
U-Net	5	31.10^6	4.29 (4.82)	4.29 (4.82)
MS-D Net	5	49.10^3	**3.95 (4.41)**	**3.95 (4.41)**
MS-D Net	1	45.10^3	4.37 (5.50)	4.37 (5.50)

Tableau 2.3.: Normalized mean square error and standard deviation (in %) for 1 000 test images, heterogeneous objects.

	#Distances	#Parameters	NMSE (in %)	
			Absorption	Phase
CTF	5	-	42.5 (19.8)	30.4 (8.99)
U-Net	5	31.10^6	11.1 (12.3)	7.65 (9.35)
MS-D Net	5	49.10^3	**7.67 (10.7)**	**5.33 (6.74)**
MS-D Net	1	45.10^3	11.9 (9.05)	7.76 (6.36)

IV.2 Experimental results

In this part, we applied the MS-D network to experimental data acquired at beamline NanoMAX at the MAX IV synchrotron (Lund, Sweden) (Max Langer, Y. Zhang, et al. 2021). The different diffraction patterns were magnified in order to have the same pixel size, which was measured to be 6 nm. The X-ray energy was set to 13 keV. The sample is placed at different positions relative to the focus and detector positions for different amounts of magnification and consequently different effective propagation distances corresponding to $D = [10.1, 15.5, 17.8, 19, 20.3]$ mm. Those phase contrast images were not directly used as input, they were magnified in order to have the same pixel size (Fig. 2.4).

The differents results in case of heterogenous assumption are displayed in Fig. 2.5. We see that CTF method retrieved well the shape of the object but left artifacts on the low frequency range. On the other hand, the U-Net seems to reduce those artifacts but roughly recover the edges. The MS-D Net reduces the artifacts while reconstructing well the object, even when a single distance is given as input.

Assuming homogeneous object composition, we compare the homogeneous version of CTF with the networks. We see in Fig. 2.6 that both CTF and MS-D Net outperformed the U-Net approach, whether we use one or several distances. The MS-D Net reconstructs with less artifacts in the low frequency range than the linearized method, and both were able to recover well the edges compared to U-Net.

For the experimental data, we used as quantitative evaluation the normalized error (NE) and

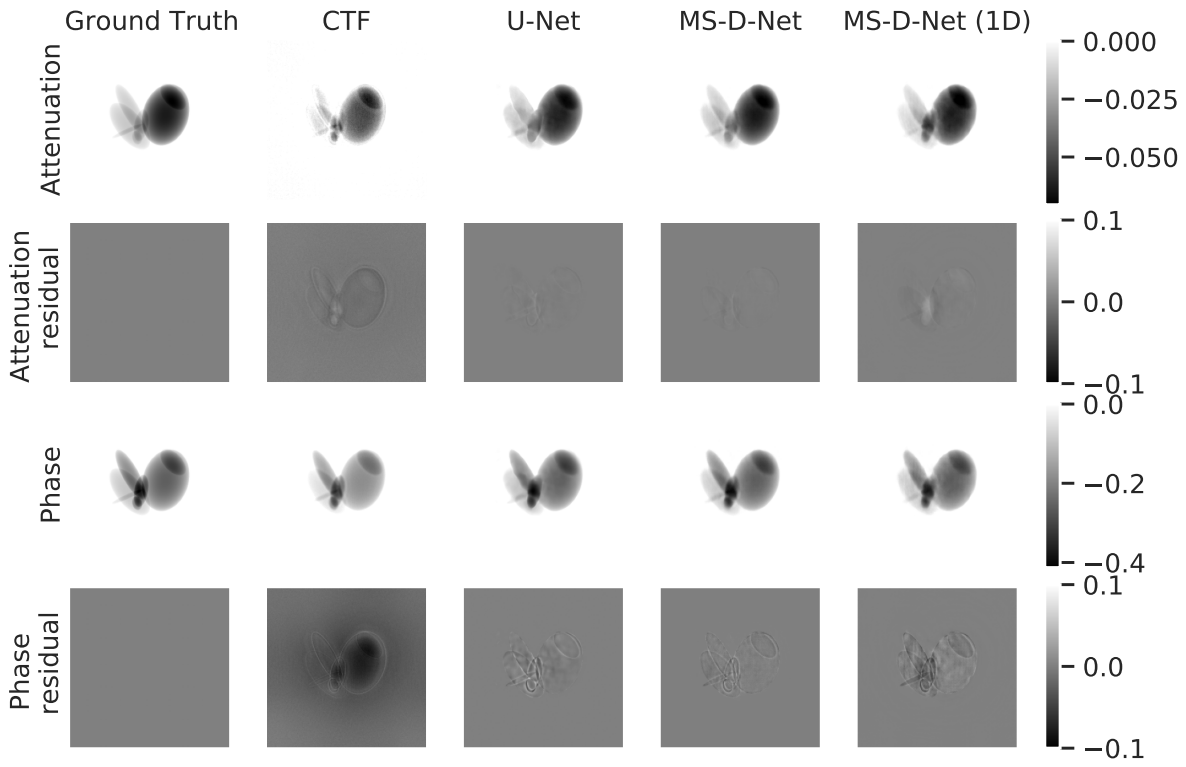


Figure 2.3.: Comparison of the different approaches on simulated heterogeneous objects.

relative standard deviation (RSD) calculated as

$$\mathrm{NE} = \frac{l_t - l_m}{l_t} \quad \text{and} \quad \mathrm{RSD} = \frac{s_m}{l_m} \tag{2.3}$$

where l_t is the expected value, l_m the measured mean value and s_m the standard deviation in the corresponding material. Calculation of l_m and s_m was done inside the object, to avoid the influence of the blur at the edges. We also measured the resolution by fitting an error function to a line profile across an object edge, and then calculating the corresponding Gaussian full width at half maximum (FWHM) based on the error function fitting parameters (Max Langer, Peter Cloetens, Hesse, Suhonen, Pacureanu, Raum, and Françoise Peyrin 2014). The result for experimental data are presented in table 2.4.

For heterogeneous data, the CTF method produces reconstructions with strong low-frequency noise. In contrast, all neural networks provide quantitatively more accurate reconstructions than CTF. It is worth noting that while the U-Net achieves the best quantitative reconstruction in terms of numerical values, the qualitative aspect is not as strong. It exhibits inaccuracies in reconstructing the star shapes, including disconnections, rounded corners, and waviness in contours. However, the U-Net excels in achieving higher resolution since the reconstructions tend to approach piecewise constant and thus works well for this particular sample. In the case of homogeneous data, the CTFHomo algorithm produces a quantitatively very good reconstruction with some remaining low frequency noise, and the edges are more blurred than in the reconstructions from the networks. It can be noted that this algorithm is precisely tailored to the imaged object, with the δ_r/β-ratio explicitly provided as input, while the networks

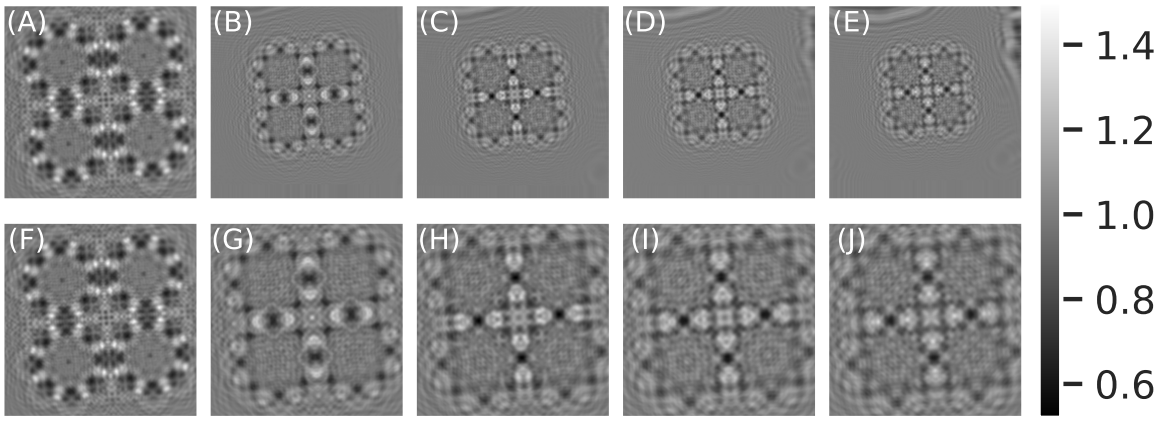


Figure 2.4.: (A)-(E) Phase-contrast images acquired at sample positions progressively further from the focus (and thus closer to the detector) showing the varying degree of magnification and phase contrast. (F)-(J) Phase-contrast images magnified to have the same pixel size (6 nm)

implicitly learn this parameter. Surprisingly, the U-Net performs somewhat less effectively than in the case of heterogeneous objects, exhibiting inaccuracies in certain parts of the star shapes, despite achieving a very clean background. Conversely, the MS-D networks consistently yield very good reconstructions, whether using multiple distances or just a single distance.

Overall, the MS-D network when trained using a single distance propagation achieves better resolution than using several distances. This may be due to uncertainty in the measurement of the physical propagation distances and difficulty to exactly align this kind of phase contrast images. The same remark applies if we compare the results obtained by CTFHomo with 1 or 5 distances, albeit with a bigger improvement in quantitative results when using several distances.

IV.3 MS-D Network as Post-Processing

Finally, we trained the networks on the reconstruction domain, by initializing the data based on the adjoint of Fréchet derivative (Bruno Sixou et al. 2013) of the forward operator, namely :

$$\left[\mathbf{F}'_D(B,\varphi)\right]^*(u) = \left(\left[\left(-u\mathrm{e}^{B-i\varphi} * P_D\right) * \overline{P_D}\right]\mathrm{e}^{-B+i\varphi}, \left[\left(u\mathrm{e}^{B-i\varphi} * P_D\right) * \overline{P_D}\right] i\mathrm{e}^{-B+i\varphi}\right) \tag{2.4}$$

More precisely, the inputs considered here consist of the average of (2.4) over all distances at the point $(B,\varphi) = (0,0)$:

$$(B_0,\varphi_0) = \sum_{i=1}^{N_D}\left[\mathbf{F}'_{D_i}(0,0)\right]^*(\mathbf{I}^{\mathrm{obs}}_{D_i}) \tag{2.5}$$

Peforming this operation before training alows us to take into account prior knowledge on the physics of the inverse problem. In Tab. 2.5, we see that this direct reconstruction given by the adjoint of the derivative does improves the reconstruction quality of the phase, but performs somewhat worse on the attenuation. Initialization given by (2.5), as well as the reconstruction for the networks are displayed in Fig. 2.7.

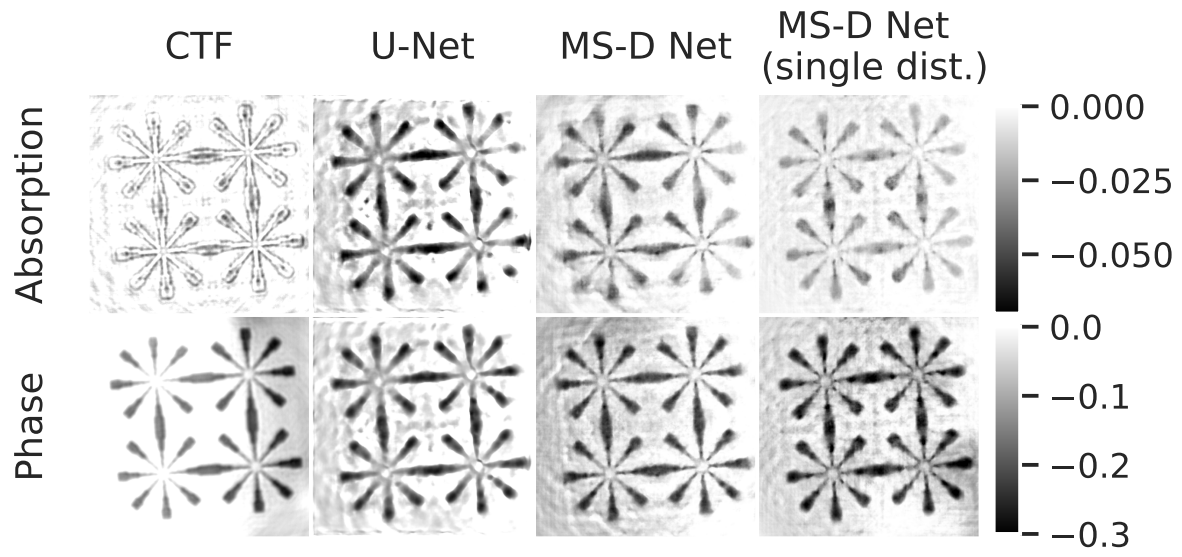


Figure 2.5.: Comparison of the different approaches on experimental data when trained on heterogeneous objects.

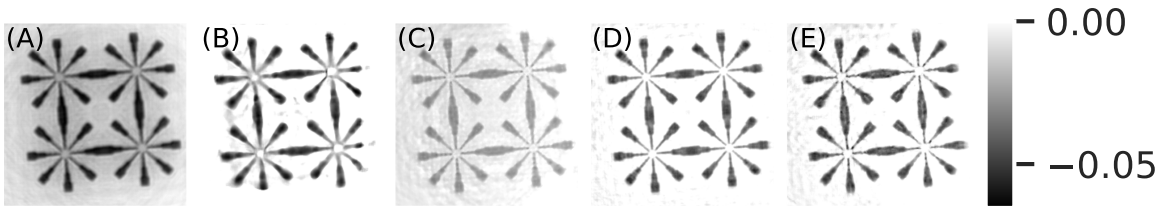


Figure 2.6.: Comparison of phase reconstructions using the different approaches on experimental data when trained on homogeneous objects. (A) CTFHomo (5 distances) (B) U-Net (C) MS-D Net (5 distances) (D) CTFHomo (1 distance) (E) MS-D Net (1 distance)

V Discussion

We used MS-D networks to perform attenuation and phase retrieval in X-ray in-line near-field phase contrast images. The network was trained and evaluated on simulated noisy data, utilizing relatively uncomplicated objects composed of combinations of ellipsoid and paraboloid shapes, involving one or several materials. None of the parameters of the physical model such as the energy of the X-ray, the propagation distances or the pixel size were given explicitly to the network. They were implicitly captured in the intensities data and the network's goal is to learn a proper reconstruction without such informations.

We have illustrated MS-D Net's potential on synthetic data using Tomophantom software and compared the results with the linearized CTF method and U-Net. The results are reported in tables 2.4 and 2.2 and in Fig. 2.3. We can conclude that both MS-D Net and U-Net outperformed the CTF method, quantitatively and qualitatively. MS-D Net was able to retrieve both attenuation and phase from a single diffraction pattern with similar quality as U-Net when using five measured intensities. On homogeneous objects, the MS-D network retrieves phase and attenuation as identical up to a constant factor, which means that it learned the constant

Tableau 2.4.: Reconstruction quality for the different algorithms for experimental data when trained on homogeneous/heterogeneous data.

		Homogeneous			
		Absorption		**Phase**	
	# Distances	NE (RSD) (in %)	Res. (in nm)	NE (RSD) (in %)	Res. (in nm)
CTFHomo	5	4.14 (11.5)	197	4.14 (11.5)	197
CTFHomo	1	25.3 (16.0)	128	25.3 (16.0)	128
U-Net	5	14.7 (27.5)	140	14.7 (27.5)	140
MS-D Net	5	**2.03 (11.6)**	165	**1.84 (11.2)**	165
MS-D Net	1	2.96 (12.5)	**93**	2.76 (12.3)	**93**
		Heterogeneous			
		Absorption		**Phase**	
	# Distances	NE (RSD) (in %)	Res. (in nm)	NE (RSD) (in %)	Res. (in nm)
CTF	5	81.3 (177)	102	21.6 (30.0)	213
U-Net	5	**6.83 (35.5)**	96	**2.30 (16.0)**	**159**
MS-D Net	5	33.7 (40.6)	98	3.22 (14.2)	202
MS-D Net	1	48.2 (32.0)	**92**	−11.5 (15.2)	208

ratio $\frac{\delta_r}{\beta}$ while U-Net did not, and both networks performed better than the homogeneous version of CTF whose ratio $\frac{\delta_r}{\beta}$ was given explicitly . On heterogeneous objects, both networks outperformed the linearized method and recover the phase better than the attenuation. When applied to homogeneous objects, the MS-D network successfully extracts phase and attenuation information, demonstrating an ability to implicitly learn the constant ratio $\frac{\delta_r}{\beta}$, whereas the U-Net did not acquire this knowledge. Both networks outperformed the homogeneous version of the CTF, where the $\frac{\delta_r}{\beta}$ ratio was explicitly provided. On heterogeneous objects, both networks surpassed the linearized method, with better phase information recovery than attenuation.

The different networks trained were also applied on experimental data. We saw that although U-Net showed less artifacts than CTF, it hardly recover the edge of the object. On the other hand, MS-D Net's reconstructions leaves some very strong ringing but showed less low frequency artifacts than the linearized method, and reconstructed well the sharp edges of the star, despite this kind of shape not being explicitly present in the training data.

Additionally, we have investigated the method that consists of using a CNN as post-processing of images from a direct reconstruction. The results reported in table 2.5 show that our approach was better than U-Net, that has been successfully used as denoising or artifact removal tool of direct reconstruction in tomography (Jin et al. 2017).

VI Conclusion

We presented the Mixed Scale Dense network, an architecture of convolutional neural networks. The MS-D Net allows to combine the short and long range information, which seems to be crucial to account for the action of the Fresnel propagator. We have compared the reconstruction results

Tableau 2.5.: Normalized mean square error (NMSE) and standard deviation for 1000 test images (in %), initialized with (2.5).

	Homogeneous		Heterogeneous	
	Absorption	Phase	Absorption	Phase
U-Net	9.60 (13.85)	4.74 (8.85)	16.22 (12.33)	5.77 (7.94)
MS-D Net	**1.96 (3.05)**	**1.96 (3.05)**	**8.19 (6.68)**	**4.83 (5.17)**

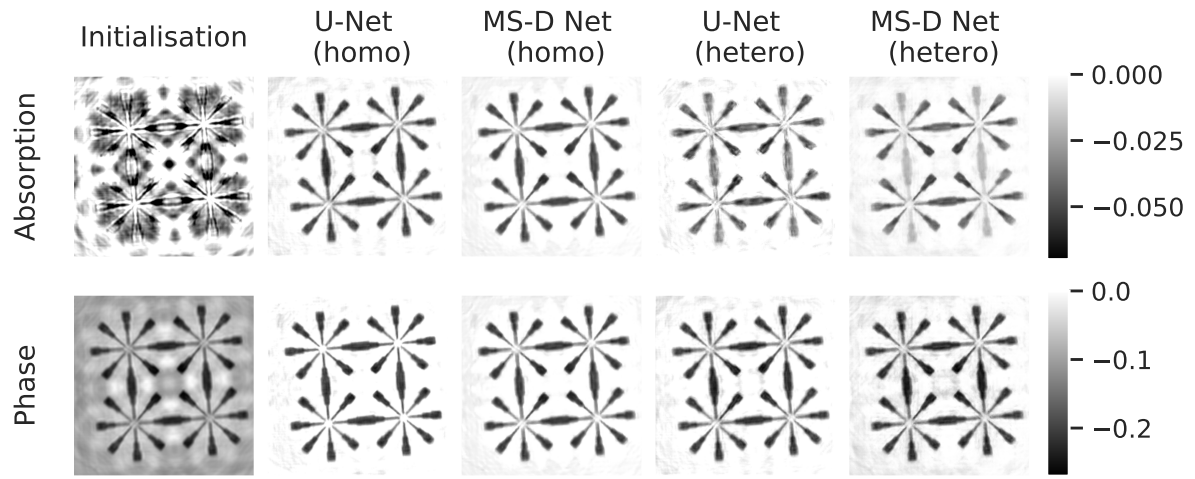


Figure 2.7.: Comparison of the different approaches on experimental data when trained on heterogeneous objects initialized with the adjoint of Fréchet derivative.

obtained with the deep learning method with the ones achieved with a classical regularization method. The network clearly outperforms the classical approach on the presented data.

On homogeneous objects, the network correctly retrieves phase and attenuation as being related by a constant factor. The network performs somewhat worse on heterogeneous objects, but its efficiency remains very good, with large decreases in reconstruction error compared to the linearized approach. Also, without relying on assumptions on relationships between the phase and attenuation, we were able to recover both from a single diffraction pattern.

Overall, the MS-D network showed excellent performance for phase and attenuation retrieval on both experimental and simulated data and improved the reconstruction quality compared to both the linearized method and the other network architecture. As with all learning approaches, however, the reconstruction efficiency is limited by the quality and precision of the training data. In particular, the network has a dependancy to the training data, which itself depend on different physical parameters such as the energy, the distances and the resolution. The implementation of the simulation might also introduce its own artifacts, for example from sampling and numerical precision.

In the chapter 4, we will consider learning an iterative scheme that incorporates some knowledge of the direct model into a data driven model. This will be done by unrolling a knowledge-driven iterative scheme and replacing the iterations with CNNs taking into account the forward operator.

3 | Nonlinear primal-dual algorithm

The primal-dual method provides a unified framework that incorporates both the primal problem, which focuses on fidelity to the observed data, and the dual problem, which encodes prior knowledge and regularizes the solution. By iteratively optimizing both the primal and dual variables, the primal-dual method balances data fidelity with regularization, leading to high-quality and visually good reconstructed images. This framework is particularly advantageous in cases where the reconstruction problem involves complex data fidelity terms, non-smooth regularizers, or additional constraints. It is known for its ability to handle large-scale problems efficiently and offers guarantees on convergence and solution optimality.

We report a non-linear primal-dual algorithm for the retrieval of phase shift and absorption from a single X-ray in-line phase contrast, or Fresnel diffraction, image. The algorithm permits to regularize phase and absorption separately. We demonstrate that taking into account the non-linearity in the reconstruction improves the reconstruction compared to linear methods. We also demonstrate that choosing different regularizers for absorption and phase can improve the reconstructions. The use of the Total Variation and its generalization in a primal-dual approach allows to exploit the sparsity of the investigated sample. On both simulated and real datasets, the proposed NL-PDHG method yields reconstructions with considerably less artifacts and improved the normalized mean squared error compared to its linearized version.

I Introduction

Several phase retrieval methods have been proposed to approximate the solution : direct inversion methods are based on the linearization of the forward model, they rely on Transport of Intensity Equation (TIE) (T. E. Gureyev and Nugent 1996; D. Paganin, S. C. Mayo, et al. 2002), Contrast Transfer Function (CTF) (Zabler et al. 2005) or on the Mixed approach (J. P. Guigay et al. 2007) between these two. All these approaches are only valid under some restrictive

assumptions on the propagation distance or on the object. Iterative methods are not limited by these constraints and some approaches have been proposed based on alternating projections on constraints between the detector and the object space (Gerchberg 1972; James R. Fienup 1982; Bauschke, Combettes, and D. Russell Luke 2002; Bauschke, Combettes, and D. Russel Luke 2003; Elser 2003; D Russel Luke 2005). These also include variational methods based on the Fréchet derivative of the forward operator in conjunction with the Landweber algorithm (V. Davidoiu et al. 2011). This kind of algorithm enables a flexible inclusion of priors, such as Tikhonov, Sobolev or sparsity regularization (Valentina Davidoiu et al. 2013; Maretzke, Bartels, et al. 2016). In order to include prior such as the Total Variation (TV), primal-dual schemes like Alternating Direction Method of Multipliers (ADMM) (Boyd, Parikh, et al. 2011) have been studied for the phase retrieval problem but rely on linearization of the forward model, either using TIE (Bostan et al. 2014) or CTF (Villanueva-Perez et al. 2017). The single-distance inverse problem is more severely ill-posed than the classical problems with several diffraction patterns (Beleggia et al. 2004; Maretzke and Hohage 2017) and few of the methods mentioned above propose to treat the case of a single-distance without any assumption on the object composition or the support. Data-driven methods based on neural networks are also attractive and have been widely studied for various problems in image processing. Several architectures have been proposed, for instance, the Mixed Scale Dense Network (Kannara Mom, Bruno Sixou, and Max Langer 2022) and PhaseGAN (Y. Zhang et al. 2021) were able to recover both attenuation and phase from a single measured intensity. Although deep learning methods can give impressive reconstructions, as with all learning approaches, the reconstruction quality is limited by the quality of the training data, moreover, the networks are quite dependent on the physical parameters such as the energy, propagation distances and pixel size used for the training data.

Here, we investigate a primal-dual approach based on the Primal-Dual Hybrid Gradient (PDHG) (Chambolle and Pock 2011) (see algorithm 7) which has so far not been considered to the simultaneous phase and absorption retrieval problem. We first propose an iterative method for the linearized CTF problem (PDHG-CTF) and then generalize it to the nonlinear case based on the NonLinear Primal-Dual Hybrid Gradient (NL-PDHG) (Valkonen 2014). We use different priors for absorption and phase to take into account the specificities of each quantity. The proposed method incorporates Total Variation (TV) regularization which allows for preservation of abrupt phase transitions, but also the Total Generalized Variation of second order (TGV^2) for getting rid of the staircasing effect on affine parts of the absorption retrieval. The suggested iterative algorithm is able to recover simultaneously the phase and absorption from a single diffraction pattern without homogeneity assumption or support constraint, moreover the algorithms does not need to be initialized with an approximated reconstruction. We demonstrate the accuracy of this approach on combinations of one or several different homogeneous materials at several signal to noise levels.

II Methods

We will assume that the functions B and φ belongs to the Hilbert space $L^2(\Omega, \mathbb{R})$ where Ω is a compact subset of $\mathbb{R}^2$. Recall that in this setting the forward operator (1.54) is a nonlinear mapping

$$\mathbf{F}_D : L^2(\Omega, \mathbb{R}) \times L^2(\Omega, \mathbb{R}) \to L^2(\mathbb{R}, \mathbb{R})$$

We consider reconstruction couples (B, φ) that are solutions of the following optimization problem:

$$(B^*, \varphi^*) = \underset{B, \varphi \in L^2(\Omega)}{\operatorname{argmin}} \left\{ d\left[\mathbf{I}_D^{\mathrm{obs}}, \mathbf{F}_D(B, \varphi)\right] + R(B, \varphi) \right\} \tag{3.1}$$

where $\mathbf{I}_D^{\text{obs}}$ is a given (noisy) measured intensity at a certain distance D, $d\left[I^{\text{obs}}, F(B,\varphi)\right]$ is the data fidelity term and $R(B,\varphi)$ a regularization term. The data fidelity term enforces the reconstructed couple to fit the acquired data and allows, through the choice of the loss function d, to add knowledge on statistical properties of the noise. The regularization term forces the solution to satisfy a priori information on the unknown object. When reconstructing a couple, one can use a joint regularization to penalize both channel with the same parameter or use a different regularization for each channel. The former has the advantage of having fewer parameters to manage, but unlike the latter, it does not take into account the specificities of each channel. In the following, we chose the latter approach, we penalize B and φ differently, using the Total Variation (TV) as well as the Total Variation of second order (TGV^2), which are introduced in the following paragraphs.

II.1 Total Variation

For a function $u \in L^1(\Omega)$, the Total Variation (TV) (Rudin, Osher, and Fatemi 1992) semi-norm is defined by:

$$\mathrm{TV}(u) = \sup\left\{\int_\Omega \varphi(x)\ \mathrm{div}(\psi(x))\mathrm{d}x \mid \psi \in C_c^1\left(\Omega, \mathbb{R}^2\right), |\psi(x)| \leq 1, \forall x \in \Omega\right\} \tag{3.2}$$

where $C_c^1\left(\Omega, \mathbb{R}^2\right)$ is the space of functions with compact support whose derivative is continuous. The quantity (3.2) is finite if and only if the derivative $\mathcal{D}u$ of u is a finite Radon measure on Ω (see (Chambolle 2004)), in this case, we say that the function u has bounded variation. We denotes by $\mathrm{BV}(\Omega, \mathbb{R})$ the set of functions that have bounded variation on Ω. For functions smooth enough $u \in W^{1,1}(\Omega)$, or equivalently $\nabla u \in L^1(\Omega)$, this quantity reduces to $\int_\Omega |\nabla u| \mathrm{d}x$.

The TV regularization has proven its effectiveness for image reconstruction and denoising tasks (Chambolle, Caselles, et al. 2010). It offers a valuable approach for preserving edges and fine details while suppressing noise and artifacts. It measures the amount of intensity variation across neighboring pixels. By promoting sparsity in the gradients of the reconstructed image, TV regularization effectively encourages piecewise constant regions while maintaining sharp transitions at edges.

We will give a discrete formulation of the Total Variation for an discrete image of size $N \times M$. To do this, we denote X the Euclidean space $\mathbb{R}^{N\times M}$ and $Y = X \times X$. We equip X with the usual inner product

$$\langle u|v\rangle_X = \sum_{i=1}^{N}\sum_{i=1}^{M} u_{ij}v_{ij},$$

and the associated norm $||\cdot||_X$.

We can now introduce a discrete version of the gradient operator. For $u \in X$, the gradient ∇u is a vector of Y given by

$$(\nabla u)_{i,j} = \left((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j}\right),$$

with

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if} \quad i < N \\ 0 & \text{if} \quad i = N \end{cases}, \qquad (\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if} \quad j < M \\ 0 & \text{if} \quad j = M \end{cases} \tag{3.3}$$

The discrete total variation is then given by the l^1 norm of the discrete gradient:

$$\mathrm{TV}(u) = \sum_{i=1}^{N}\sum_{j=1}^{M}\sqrt{|(\nabla u)^1_{i,j}|^2+|(\nabla u)^2_{i,j}|^2}, \tag{3.4}$$

We also introduce a discrete version of the divergence operator defini by analogy with the continuous framework by posing

$$\mathrm{div} = -\nabla^*$$

where ∇^* is the adjoint operator of ∇, i.e

$$\forall p = (p^1, p^2) \in Y, \forall u \in X, \quad \langle -\mathrm{div} p|u\rangle_X = \langle p|\nabla u\rangle_Y = \left\langle p^1|(\nabla u)^1\right\rangle_X + \left\langle p^2|(\nabla u)^2\right\rangle_X$$

Using this definition, the discrete divergent operator for a couple $p = (p^1, p^2)$ can be defined explicitly by:

$$(\mathrm{div} p)_{i,j} = \begin{cases} p^1_{i,j} & \text{if} \quad i = 1 \\ p^1_{i,j} - p^1_{i-1,j} & \text{if} \quad 1 < i < N \\ -p^1_{i-1,j} & \text{if} \quad i = N \end{cases} + \begin{cases} p^2_{i,j} & \text{if} \quad j = 1 \\ p^2_{i,j} - p^2_{i,j-1} & \text{if} \quad 1 < j < M \\ -p^2_{i,j-1} & \text{if} \quad j = M \end{cases} \tag{3.5}$$

II.2 Total Generalized Variation

Total generalized variation (TGV) has been introduced (Bredies, Kunisch, and Pock 2010) to address some limitations of traditional Total Variation regularization in image reconstruction tasks. While TV regularization effectively promotes piecewise constant regions and sharp edges, it can suffer from staircase artifacts and oversmoothing in regions with fine textures or complex structures. TGV extends the concept of TV by considering higher-order gradients of the image, allowing for better preservation of higher-frequency details and smoother reconstructions in regions with low-frequency variations.

For a function $u \in L^1(\Omega)$, the Total Generalized Variation (TGV) of order k and of parameters $\alpha = (\alpha_0, ..., \alpha_{k-1})$, is defined as

$$\begin{aligned}\mathrm{TGV}^k_\alpha(u) = \sup\Bigg\{ \int_\Omega u(x)\ \mathrm{div}^k \psi(x) \mathrm{d}x \mid \psi \in C^k_c\left(\Omega, \mathrm{Sym}^k(\mathbb{R}^2)\right), \\ ||\mathrm{div}^l \psi||_\infty \leq \alpha_l,\ l = 0, \ldots, k-1 \Bigg\}\end{aligned} \tag{3.6}$$

where $C^k_c\left(\Omega, \mathrm{Sym}^k(\mathbb{R}^2)\right)$ denotes the space of symmetric tensors of order k with arguments in $\mathbb{R}^2$ and $\alpha_l > 0$ are some fixed parameters. By choosing $k = 1$ and $\alpha_0 = 1$, we see that we recover the Total Variation definition (3.2). We'll focus here on the $k = 2$ case, i.e., the total generalized variation of second order (TGV^2_α), which has many properties and has recently been used for inverse problem (Bredies and Valkonen 2020).

For $k = 2$, the second-order total generalised variation for parameters $\alpha, \beta > 0$ can be written (Bredies and Valkonen 2020) as:

$$\mathrm{TGV}^2_{(\alpha,\beta)}(u) = \inf_v \left\{ \alpha \int_\Omega |\mathcal{D}v| + \beta \int_\Omega |\mathcal{D}u - v| \right\} \tag{3.7}$$

where $v = (v_1, v_2)$ and $v_1, v_2 \in \mathrm{BV}(\Omega)$. The idea is to force the vector field v to have a sparse gradient and to penalize the gradient $\mathcal{D}u$ to deviate only on a sparse set from v. We observe in

the definition of TGV^2 how it balances between first and second order features, controled with the ratio $\frac{\alpha}{\beta}$.

II.3 Gradient descent with smooth Total Variation

We consider the problem (3.1) with the loss function d equal to l^2 squared norm, and the regularization R equal to the Total Variation on B and φ. Assuming the solution to be smooth enough, we then seek to minimize the following functional:

$$\min_{B,\varphi}\left\{\frac{1}{2}\int_\Omega \left[\mathbf{F}_D(B,\varphi)(x)-\mathbf{I}_D^{\mathrm{obs}}(x)\right]^2 \mathrm{d}x+\eta\int_\Omega |\mathcal{D}B|+\mu\int_\Omega |\mathcal{D}\varphi|\right\} \tag{3.8}$$

where $\eta,\mu>0$ are regularization parameters. The point here is that the gradient of the TV semi-norm is given by $\mathrm{div}\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right)$ which is not defined at a pixel $x\in\Omega$ if $\nabla\varphi(x)=0$. We can avoid this problem by considering a smooth version of the TV regularization (Kalmoun 2018):

$$\mathrm{TV}^\epsilon(\varphi):=\int_\Omega\sqrt{\epsilon^2+|\nabla\varphi(x)|^2}\mathrm{d}x\approx\int_\Omega|\mathcal{D}\varphi| \quad\text{and}\quad \nabla\mathrm{TV}^\epsilon(\varphi)=\mathrm{div}\left(\frac{\nabla\varphi}{\sqrt{\epsilon^2+|\nabla\varphi|^2}}\right) \tag{3.9}$$

where $\epsilon>0$ is a smoothing parameter. Combining all this, this amounts to minimize the energy

$$J_{\eta,\mu,\epsilon}(B,\varphi)=\frac{1}{2}||\mathbf{F}_D(B,\varphi)-\mathbf{I}_D^{\mathrm{obs}}||_2^2+\eta\mathrm{TV}^\epsilon(B)+\mu\mathrm{TV}^\epsilon(\varphi) \tag{3.10}$$

This can be done with an algorithm of projected gradient descent, in order to constraint the sought solutions to be positive. We call this approach Gradient Descent with smooth Total Variation (GD-TV^ϵ). The GD-TV^ϵ method is summarized in algorithm 9, where P_+ is the projection onto positive values.

Algorithm 9 GD-TV^ϵ

Given :

- step size τ, smoothing factor ϵ and weighting parameters η,μ
- $(B_0,\varphi_0)\in\mathbb{R}^{n\times m}\times\mathbb{R}^{n\times m}$

for $k=0,\dots,N\mathrm{iter}$ **do** :

$$\begin{aligned}(B_{k+1},\varphi_{k+1}) &\leftarrow (B_k,\varphi_k)-\tau\nabla J_{\eta,\mu,\epsilon}(B_k,\varphi_k)\\ &= (B_k,\varphi_k)-\tau\left\{\left[\mathbf{F}_D'(B_k,\varphi_k)\right]^*\left(\mathbf{F}_D(B_k,\varphi_k)-\mathbf{I}_D^{\mathrm{obs}}\right)+\left[\eta\nabla\mathrm{TV}^\epsilon(B_k),\mu\nabla\mathrm{TV}^\epsilon(\varphi_k)\right]\right\}\\ (B_{k+1},\varphi_{k+1}) &\leftarrow P_+\left[(B_{k+1},\varphi_{k+1})\right]\end{aligned}$$

II.4 Primal Dual Hybrid Gradient method based on CTF linearization

The main drawback of the GD-TV^ϵ approach is that one cannot efficiently implement the TV semi-norm with a simple gradient descent without smoothing it, and this implies the choice of the ϵ parameter as well. To overcome this problem, we can use a primal-dual approach (see VI.4.2). Since the contribution of attenuation and phase to the phase contrast image is different, it may be more interesting to use different regularization for B and φ respectively.

For the absorption B, we choose to use second order total generalized variation regularization, and for the phase φ, Total Variation regularization. As PDHG algorithm is only valid for a

linear operator, we focus on CTF-linearized problem, using the CTF-forward operator $\mathbf{F}_D^{\mathrm{CTF}}$. Choosing regularization parameters to be $\alpha, \beta, \nu > 0$, and assuming our solution to be positive and sufficiently smooth, we then seek to solve the following minimization problem :

$$\min_{B,\varphi} \left\{ \frac{1}{2} \int_\Omega \left[\mathbf{F}_D^{\mathrm{CTF}}(B,\varphi)(x) - \mathbf{I}_D^{\mathrm{obs}}(x) \right]^2 \mathrm{d}x + \mathrm{TGV}^2_{(\alpha,\beta)}(B) + \nu \mathrm{TV}(\varphi) + \iota_+(B,\varphi) \right\} \tag{3.11}$$

where ι_+ denotes the indicator function which constraint the sought solutions to be positive :

$$\iota_+(B,\varphi) = \begin{cases} 0 & \text{if } B, \varphi > 0 \\ +\infty & \text{else} \end{cases} \tag{3.12}$$

Using the formulation of TGV^2 introduced in (3.7), the problem (3.11) is equivalent to:

$$\min_{B>0,\varphi>0} \inf_{v} \left\{ \frac{1}{2} \int_\Omega \left[\mathbf{F}_D^{\mathrm{CTF}}(B,\varphi)(x) - \mathbf{I}_D^{\mathrm{obs}}(x) \right]^2 \mathrm{d}x + \alpha \int_\Omega |\nabla v| + \beta \int_\Omega |\nabla B - v| + \nu \int_\Omega |\nabla \varphi| \right\} \tag{3.13}$$

By introducing the discrete scalar images $B, \varphi \in \mathbb{R}^{N\times M}$ and the vectorial image $\mathbf{v} = (v_1, v_2) \in \left(\mathbb{R}^{N\times M}\right)^2$, we can obtain the discrete version of (3.13), which is given by:

$$\min_{\substack{B,\varphi,\mathbf{v} \\ B>0,\varphi\geq 0}} \left\{ \left\| \mathbf{F}_D^{\mathrm{CTF}}(B,\varphi) - \mathbf{I}_D^{\mathrm{obs}} \right\|_2^2 + \alpha \left\| \mathbf{D}\mathbf{v} \right\|_1 + \beta \left\| \nabla B - \mathbf{v} \right\|_1 + \nu \left\| \nabla \varphi \right\|_1 \right\} \tag{3.14}$$

where

$$\mathbf{D} : \left(\mathbb{R}^{n\times m}\right)^2 \to \left(\mathbb{R}^{n\times m}\right)^4, \quad (v_1, v_2) \mapsto (\nabla v_1, \nabla v_2)$$

If we denote

$$\mathcal{X} = \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m} \times \left(\mathbb{R}^{n\times m}\right)^2$$
$$\mathcal{Y} = \mathbb{R}^{n\times m} \times \left(\mathbb{R}^{n\times m}\right)^4 \times \left(\mathbb{R}^{n\times m}\right)^2 \times \left(\mathbb{R}^{n\times m}\right)^2$$

to be the discretized primal and dual spaces, then the minimization problem (3.11) can be rewritten as:

$$\min_{B,\varphi,\mathbf{v}} \left\{ \mathcal{H}\left[\mathcal{K}_{\mathrm{CTF}}(B,\varphi,\mathrm{v}) \right] + \mathcal{G}(B,\varphi,\mathrm{v}) \right\} \tag{3.15}$$

with :

$$\begin{aligned} &\mathcal{K}_{\mathrm{CTF}}(B,\varphi,\mathbf{v}) = \left[\mathbf{F}_D^{\mathrm{CTF}}(B,\varphi), \mathbf{D}(\mathbf{v}), \nabla B - \mathbf{v}, \nabla \varphi \right] \\ &\mathcal{H}\left(h^1, h^2, h^3, h^4\right) = \left\| h^1 - \mathbf{I}_D^{\mathrm{obs}} \right\|_2^2 + \alpha \left\| h^2 \right\|_1 + \beta \left\| h^3 \right\|_1 + \nu \left\| h^4 \right\|_1 \\ &\mathcal{G}(B,\varphi,\mathbf{v}) = \iota_+(B,\varphi) \end{aligned} \tag{3.16}$$

Using this formulation, we define the PDHG-CTF algorithm (Algo. 10), which iterates over the triplet $x_k = (B_k, \varphi_k, \mathrm{v}_k)$, where B_k and φ_k represent the absorption and phase shift we are looking for, and $\mathrm{v_k}$ is the variable from the TGV formula (3.7), at the k-th iteration. Here, τ and σ are the step sizes of the primal and dual space, respectively, $\mathcal{K}^*$ denotes the adjoint operator of $\mathcal{K}$, $\mathrm{prox}_{\tau\mathcal{G}}$ the proximal operator of $\tau\mathcal{G}$ and $\mathcal{H}^*$ the conjugate of $\mathcal{H}$. The algorithm 10 is convergent (Chambolle and Pock 2016b) if the primal step τ and the dual step σ satisfiy the inequality:

$$\sigma\tau |||\mathcal{K}_{\mathrm{CTF}}|||^2 = \sigma\tau \sup_{x\in\mathcal{X}} \left\{ \frac{||\mathcal{K}_{\mathrm{CTF}}(x)||_{\mathcal{Y}}}{||x||_{\mathcal{X}}} \right\}^2 < 1 \tag{3.17}$$

Algorithm 10 PDHG-CTF

Given :

- step sizes σ, τ such that $\sigma\tau|||\mathcal{K}_{\text{CTF}}|||^2 < 1$ and relaxation parameter $\theta \in [0, 1]$
- regularization parameters α, β and ν
- an initial pair $x_0 = \{B_0, \varphi_0, v_0\} \in X$ (primal) and $h_0 = [h_0^1, h_0^2, h_0^3, h_0^4] \in Y$ (dual)

for $k = 0, \ldots,$ Niter **do** :

$h_{k+1} \leftarrow \text{prox}_{\sigma\mathcal{H}^*}\left(h_k + \sigma\mathcal{K}_{\text{CTF}}(\overline{x}_k)\right)$

$x_{k+1} \leftarrow \text{prox}_{\tau\mathcal{G}}\left(x_k - \tau\mathcal{K}_{\text{CTF}}{}^*(h_{k+1})\right)$

$\overline{x}_{k+1} \leftarrow x_{k+1} + \theta\left(x_{k+1} - x_k\right)$

II.5 Non Linear Primal Dual Hybrid Gradient

Originally designed for linear operator, the PDHG algorithm has been generalized to nonlinear cases (Valkonen 2014). By replacing the linear operator F_D^{CTF} by the nonlinear operator F_D (1.54) in the minimization problem (3.11), we obtain a new minimization problem that we can solve with the NL-PDHG method. The difference to the PDGH-CTF is in the data fidelity term, i.e., instead of (3.14), we now have to solve:

$$\min_{\substack{B,\varphi,\mathbf{v} \\ B>0,\varphi>0}} \left\{\left\|\mathbf{F}_D(B,\varphi) - \mathbf{I}_D^{\text{obs}}\right\|_2^2 + \alpha\left\|\mathbf{D}\mathbf{v}\right\|_1 + \beta\left\|\nabla B - \mathbf{v}\right\|_1 + \nu\|\nabla\varphi\|_1\right\} \tag{3.18}$$

As before, (3.18) can be rewritten in the form (3.15), this time using the non-linear operator $\mathcal{K}_{\text{NL}}$ defined by:

$$\mathcal{K}_{\text{NL}}\left(B,\varphi,\mathbf{v}\right) = \left[\mathbf{F}_D\left(B,\varphi\right), \mathbf{D}(\mathbf{v}), \nabla B - \mathbf{v}, \nabla\varphi\right]$$

This new problem (3.18) can be solved using the NonLinear Primal Dual Hybrid Gradient (NL-PDHG) algorithm, which is given in Algo. 11. We observe that unlike in Algo. 10, the

Algorithm 11 NL-PDHG

Given :

- step sizes $\sigma_0, \tau_0 > 0$ and relaxation parameter $\theta \in [0, 1]$
- regularization parameters α, β and ν
- an initial pair $x_0 = \{B_0, \varphi_0, v_0\} \in X$ (primal) and $h_0 = [h_0^1, h_0^2, h_0^3, h_0^4] \in Y$ (dual)

for $k = 0, \ldots, N_{\text{iter}}$ **do** :

$h_{k+1} \leftarrow \text{prox}_{\sigma\mathcal{H}^*}\left(h_k + \sigma_k\mathcal{K}_{\text{NL}}\overline{x}_k\right)$

$x_{k+1} \leftarrow \text{prox}_{\tau\mathcal{G}}\left(x_k - \tau_k\left[\mathcal{K}'_{\text{NL}}(x_k)\right]^* h_{k+1}\right)$

$\overline{x}_{k+1} \leftarrow x_{k+1} + \theta\left(x_{k+1} - x_k\right)$

$\sigma_{k+1}, \tau_{k+1} \leftarrow \sigma_k, \tau_k$ such that $\sigma_k\tau_k \sup_{n=0,1,\ldots,k}\left\{|||\mathcal{K}'_{\text{NL}}(x_n)|||^2\right\} < 1$

iterations are based on the operator $\mathcal{K}'_{\text{NL}}(x_k)$, the Fréchet derivative of the operator $\mathcal{K}_{\text{NL}}$ at the point x_k, for which we have an explicit formula. The main differences with the linear version is that to ensure the convergence of the algorithm, the gradient of $\mathcal{K}$ has to be Lipschitz in a neighborhood of a solution, the initial iterate has to be close enough to a solution and the step sizes must satisfy the local inequalities (Valkonen 2014)

$$\sigma_k\tau_k \sup_{n=0,1,\ldots,k}\left\{|||\mathcal{K}'_{\text{NL}}(x_n)|||^2\right\} < 1 \tag{3.19}$$

for all k.

III Experiments

III.1 Implementation details

In this part, we give details on the choice of the different regularization parameters and step sizes. We also discuss the stopping criterion and the choice of the initialization.

III.1.1 Step size parameters

The convergence conditions for the gradient descent have not been analyzed in detail, we have noticed that choosing a sufficiently small fixed step size $\tau = 0.01$ was enough to obtain convergence ot the iterates in practice. For the primal-dual methods, in order to satisfy the condition (3.19) (resp. (3.17)) on the primal and dual steps, one can estimate the linear operator norm $|||\mathcal{K}'_{\mathrm{NL}}(x)|||^2$ (resp. $|||\mathcal{K}_{\mathrm{CTF}}|||^2$) by constructing a sequence $\mathbf{y}_n = \left[\mathcal{K}'_{\mathrm{NL}}(x)\right]^* \mathcal{K}'_{\mathrm{NL}}(\mathbf{y}_{n-1})$ and computing the quotient $\rho_n = \frac{\|\mathbf{y}_n(x)\|}{\|\mathbf{y}_{n-1}\|}$. The sequence $(\rho_n)_{n\in\mathbb{N}}$ converges (Chaari et al. 2009) and we have $\lim_{n\to+\infty} \rho_n = |||\mathcal{K}'_{\mathrm{NL}}(x)|||^2$. One has simply to choose $\rho_N \approx |||\mathcal{K}'_{\mathrm{NL}}(x)|||^2$ for N sufficiently large and set

$$\sigma = \tau = \frac{0.99}{\sqrt{\rho_N}}$$

Since this process is quite costly, the variables σ_k and τ_k were updated only every 50 iterations for NL-PDHG.

III.1.2 Regularization parameters

The parameters for the different methods have been optimized by scanning a wide interval to obtain a large decrease of the data term. For GD-TV$^\epsilon$, the smoothing factor was set to 10^{-3}, with weighting parameters $\eta = 10^{-1}$ and $\mu = 10^{-3}$. For both PDHG-CTF and NL-PDHG, we used the same set of parameters: $\alpha = 10^{-2}, \beta = 5 \times 10^{-3}, \nu = 10^{-2}$ and the relaxation parameter $\theta = 1$.

III.1.3 Stopping criterion

A criterion often used for primal-dual methods is based on the computation of the *duality gap*, which, for the saddle-point problem

$$\max_{y\in\mathcal{Y}} \min_{x\in\mathcal{X}} \{\langle y|\mathcal{K}x\rangle - \mathcal{H}^*(y) + \mathcal{G}(x)\} \tag{3.20}$$

is defined by

$$\mathcal{H}(\mathcal{K}x) + \mathcal{G}(x) + \mathcal{G}^*(-\mathcal{K}^*y) + \mathcal{H}^*y \tag{3.21}$$

It can be used as a stopping criterion when it goes under a certain threshold. But the primal variable $\mathbf{v}$ is involved in TGV^2 (3.7) and the duality gap is in this case equal to infinity (Valkonen, Bredies, and Knoll 2013), so it is not useful as a stopping criterion in practice. This problem can be avoided by using a modified duality gap such as the pseudo-duality gap (Valkonen 2014). As the computation of this quantity represents an important computational cost (Knoll et al. 2017), and to have a proper comparison with the gradient descent method, we did not employ this kind of criteria, but rather used a fixed number of $N\mathrm{iter} = 1\,000$ iterations for the different experiments.

III.1.4 Initialization

Even if theoretically, one must make sure that the initialization is close enough to a solution to ensure convergence, in practice, we did not have a problem of convergence when initializing with $(B_0, \varphi_0) = (0, 0)$. In the following, we report values only from zero-initialization.

III.2 Evaluation metrics

In order to evaluate the reconstruction quality obtained by the different methods, we use different metrics:

- The normalized mean squared error (NMSE) defined by:

$$\mathrm{NMSE}(x, x_{\mathrm{true}}) = \frac{\|x - x_{\mathrm{true}}\|_2}{\|x_{\mathrm{true}}\|_2}$$

  The NMSE measures the average squared difference between the reconstructed image and the original (ground truth) image, normalized by the average squared intensity of the original image. A low NMSE value indicates that the reconstructed image is close to the ideal solution.
- The Structural Similarity Index Measure (SSIM) can be defined as:

$$\mathrm{SSIM}(x, x_{\mathrm{true}}) = \frac{(2m_x m_{x_{\mathrm{true}}} + C_1)(2\sigma_{x x_{\mathrm{true}}} + C_2)}{(m_x^2 m_{x_{\mathrm{true}}}^2 + C_1)(\sigma_x^2 \sigma_{x_{\mathrm{true}}}^2 + C_2)}$$

  where m_x and σ_x represent the mean and the standard deviation of x pixels, respectively, and $\sigma_{x x_{\mathrm{true}}}$ is the covariance between x and x_{true} and $C_1 = 10^{-4}$, $C_2 = 9{\times}10^{-4}$ are constants. SSIM measures the similarity between two images by considering their structural information, including luminance, contrast, and structural details. A higher SSIM value indicates a higher perceived similarity and better quality of the reconstructed image.
- The Peak Signal-to-Noise Ratio (PSNR) provides a measure of the ratio between the peak signal power and the power of the noise or distortion:

$$\mathrm{PSNR}(x, x_{\mathrm{true}}) = 10 \log_{10} \left(\frac{M^2}{\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (x_{i,j} - x_{\mathrm{true}_{i,j}})^2} \right)$$

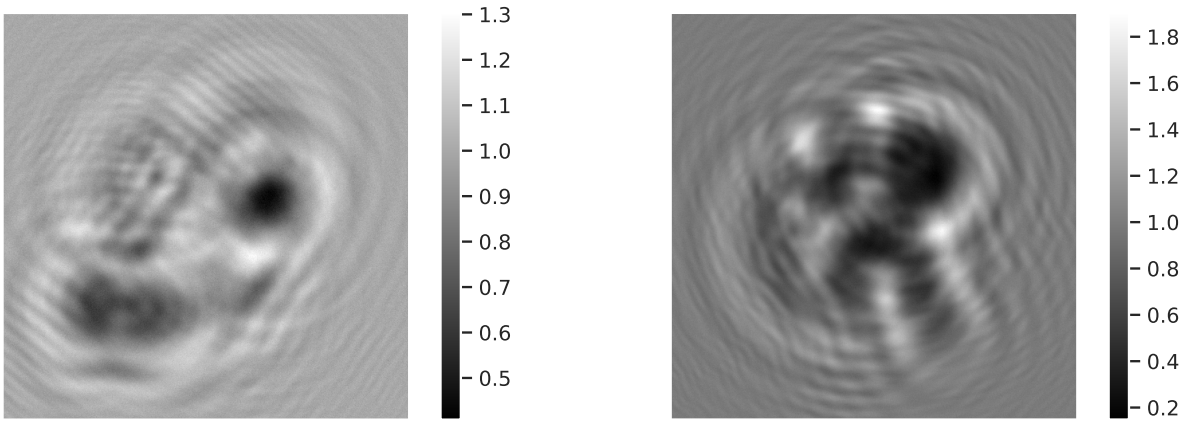
  where M is the maximum fluctuation in the input image data type (the maximum possible pixel value). PSNR is expressed in decibels (dB), a higher PSNR value indicates a smaller amount of noise or distortion and, thus, higher image quality.

III.3 Results and discussion

In order to evaluate the algorithms, a set of 3D object were generated by creating random combinations of one to ten ellipsoid or paraboloid shapes consisting of three different materials: gold (Au), palladium (Pd) and zinc (Zn). 2D projections of size 512×512 pixels were generated, yielding images of the phase and the absorption, which where subsequently used to generate phase contrast images by using (1.54). The X-ray energy was set to 13 keV for a wavelength of $\lambda = 0.095$ nm, the propagation distance $D = 20.3$ mm and the pixel size to 12 nm. As quantitative measures of reconstruction quality and to quantify the uncertainty of the reconstructions we generated a test dataset of 1 000 images to compare the methods. Examples of such diffraction are represented in Fig. 3.1, where the first image 3.1(a) displays the intensity measured when the object is rather *thin*, while the second 3.1(b) displays the intensity measured when the object is *thicker*.

III.3.1 Simulated data

For qualitative evaluation, examples of reconstructed phase and absorption projections from the diffraction patterns (Fig. 3.1) are displayed in Fig. 3.2 and 3.3 (with negative contrast). Figure 3.2 shows the reconstructions obtained with the various methods (and the associated



((a)) Intensity obtained with thin object.

((b)) Intensity obtained with thick object.

Figure 3.1.: Simulated intensities for two different objects of the dataset.

residuals $x - x\text{true}$) when the object considered is quite thin. In this case, the CTF assumptions are well satisfied, that is why PDHG-CTF gets the best reconstructions. The nonlinear approach has similar results, even if we see that some parts are missing inside the recovered projections compared to the linearized approach. The gradient descent performs somewhat worse, we observe that parts of the contour are missing in the reconstructions which seems to be caused by the projection on positive values and the approximation of the Total Variation. Looking at the absorption reconstruction more closely, we can see the staircasing effect that appears for GD-TV$^\epsilon$ because of the Total Variation contrary to the primal-dual approaches which use the Total Generalized Variation of order 2.

Figure 3.3 shows the reconstructions of the methods when the object considered is more thick than the previous one, as we can see on the values of the ground truths. We observe that for GD-TV$^\epsilon$ and PDHG-CTF, the reconstructions get worse, they lack a little contour for the absorption, and the information inside the phase is not well recovered. The same remarks apply for the NL-PDHG, although this time the reconstructions are better. Finally, we can explicitly see what the nonlinearity information brings in this case, by comparing the linearized approach by CTF and the one using the direct nonlinear model. In particular, for objects that deviate from the CTF assumptions, the NL-PDHG can still be expected to give satisfactory results.

For quantitative analysis, we computed the average NMSE, SSIM and PSNR on the test dataset, the results are summarized in Tab. 3.1. The evolution of the averages NMSE, as well as the standard deviation, are displayed in Fig. 3.4, we can see that the convergence for the phase is faster than the absorption, which suggests that different step sizes could probably be used for each of the channels to speed up the convergence. The primal-dual approaches perform better than the gradient descent, and overall the NL-PDHG method has the best reconstructions on average. We can see that applying the same regularization to B and φ gives worse reconstructions, and that the absorption is better recovered with a TGV^2 regularization while the phase has better results for the TV regularization. This can be explained by the fact that in our experiments, the retrieved absorption tends to have missing parts which, when reconstructed with TV leaves some staircasing effects. The TGV^2 regularization for B offers best performance, providing a good compromise between sharp boundaries and smoothness. While for φ, a better reconstruction cannot be expected by taking into account the second order. Moreover, if the considered objects has sufficiently weak attenuation and slowly varying phase,

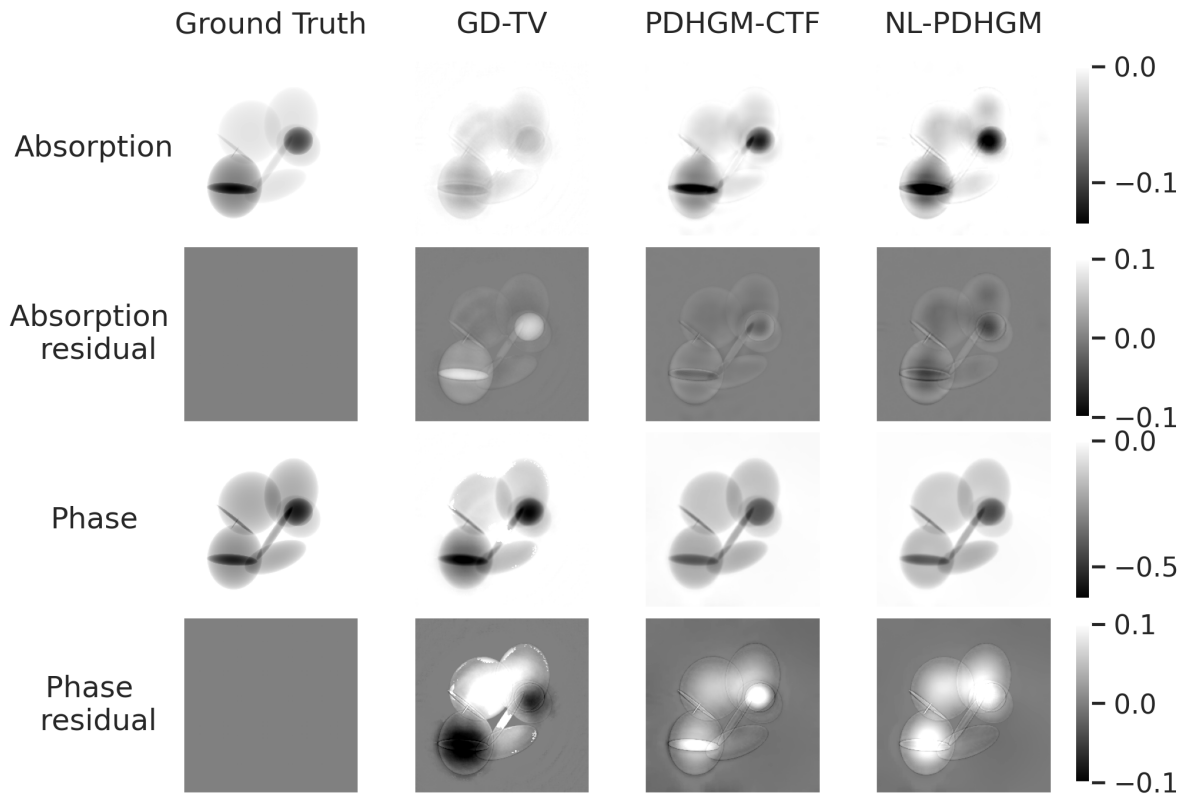


Figure 3.2.: Comparison of the different methods using the diffraction pattern 3.1(a)

then the PDHG-CTF method converges faster. It needs a hundred iterations to obtain the same quality of reconstruction on average as the nonlinear version, which needs a thousand iterations. In contrast, if we consider the most absorbent object of the dataset, which has a maximum absorption equal to 0.882, then the CTF approximation is less good and the linearized method gives a NMSE of 60.5 and 75.7, while NL-PDHG gives a NMSE of 47.9 and 61.4 for the absorption and phase, respectively.

III.3.2 Experimental data

To demonstrate the capability of this primal-dual framework, the methods were applied on experimental data acquired at beamline NanoMAX at the MAX IV synchrotron (Lund, Sweden) (Kalbfleisch et al. 2022). The diffraction pattern ($2\,048 \times 2\,048$ pixels) in Fig. 3.5 was obtained with an effective pixel size of 12 nm and a defocusing distance $D = 20.3$ mm. And since the object is piecewise constant, TV and TGV type regularizations are well suited. The X-ray energy was set to 13 keV for a wavelength of $\lambda = 0.095$ nm. The sample consists of a stack of Pd, Zn, Pd, Au metal layers with thicknesses of 21, 10, 11, 163 nm, respectively, deposited on a 1 mm-thick silicon nitride substrate, thus the expected values for absorption and phase are 0.0483 and 0.217. The execution time was 37 min for GD-TV$^\epsilon$ and 50 min for the primal-dual methods, after $1\,000$ iterations. As shown in Fig. 3.6 and Fig. 3.7, the GD-TV$^\epsilon$ algorithm yields the same problems of missing parts of the object as on simulated data in the recovered phase and absorption. Both primal-dual approaches seem to retrieve well the absorption, while reducing spurious fringes in the background, and the recovered phases are close to artifact-free. Note that here the object is not very absorbing since it is quite thin. Therefore, the linearized method also yields a very good

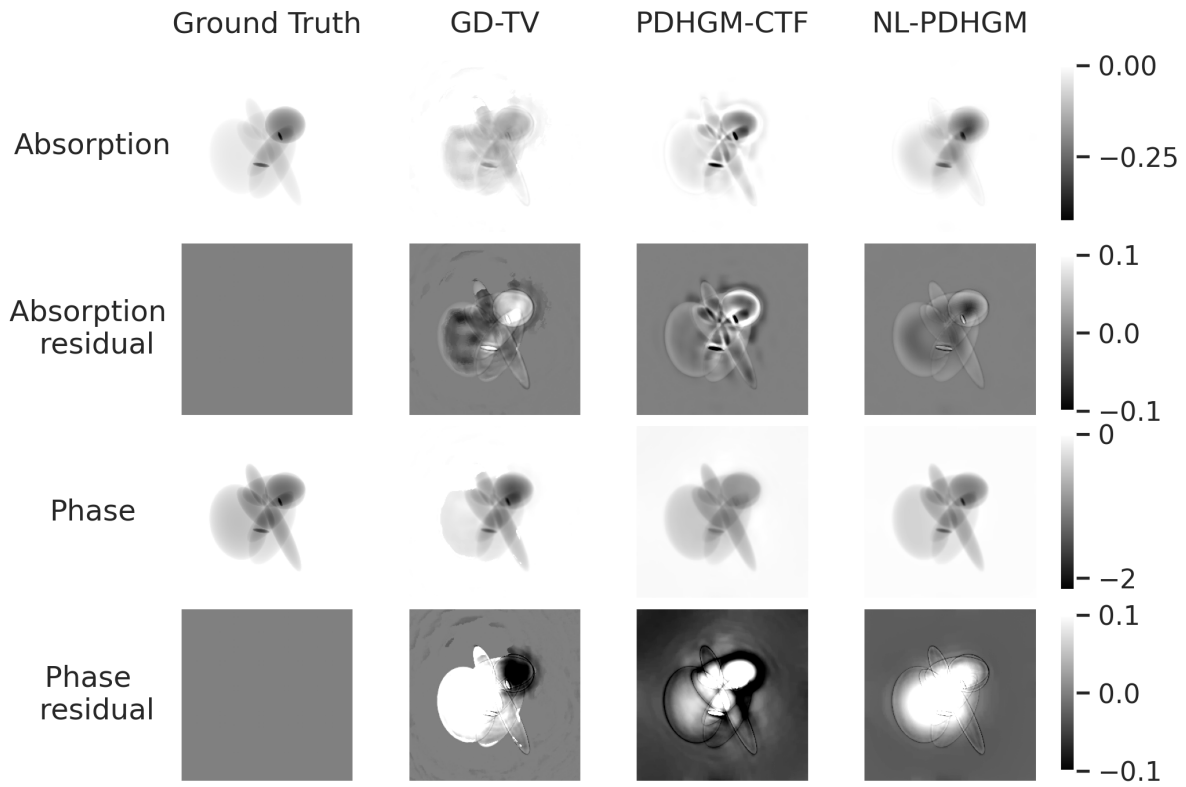


Figure 3.3.: Comparison of the different methods using the diffraction pattern 3.1(b)

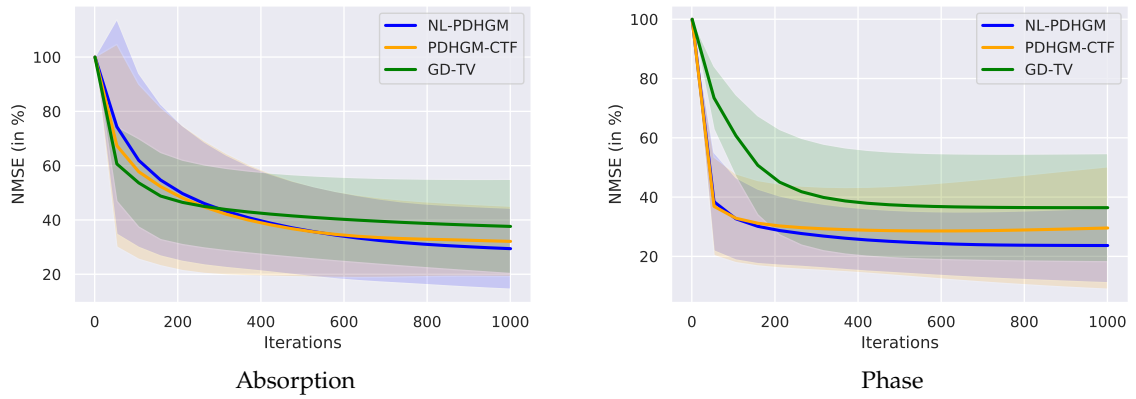


Figure 3.4.: Evolution of average NMSE (in %) for 1 000 test images. The transparent areas correspond to the standard deviation.

reconstruction. And since the object is piecewise constant, TV and TGV type regularizations are well suited.

III.3.3 Generalization to large ration of phase over absorption

To confirm that the NL-PDHG method can be generalized to cases with higher $\frac{\delta_r}{\beta}$ quotients, we performed simulations under the following conditions: the energy was set to 27 keV (corresponding to $\lambda = 0.046$ nm), the propagation distance to $D = 47.6$ mm and the pixel size

Tableau 3.1.: Average NMSE, SSIM, PSNR and standard deviation for 1 000 test images using different strategies for regularization.

	Regularization		NMSE (in %)		SSIM (in %)		PSNR	
	Absorption	Phase	Absorption	Phase	Absorption	Phase	Absorption	Phase
GD-TV$^\epsilon$	TV$^\epsilon$	TV$^\epsilon$	37.5 (17.4)	36.4 (18.2)	99.6 (0.440)	95.2 (6.95)	65.2 (9.43)	50.5 (11.1)
PDHG-CTF	TGV^2	TV	32.1 (12.9)	29.6 (20.9)	**99.8 (0.337)**	92.9 (7.87)	68.2 (9.10)	52.6 (8.19)
NL-PDHG	TGV^2	TV	**29.2 (14.8)**	**23.6 (12.6)**	**99.8 (0.237)**	**97.2 (3.12)**	**68.7. (8.63)**	**53.0 (6.40)**
NL-PDHG	TV	TV	41.3 (23.9)	25.6 (14.1)	99.7 (0.371)	94.8 (5.10)	65.0 (9.72)	52.8 (7.53)
NL-PDHG	TGV^2	TGV^2	32.4 (19.6)	29.3 (14.8)	**99.8 (0.249)**	92.3 (6.58)	66.8 (8.00)	51.4 (6.51)

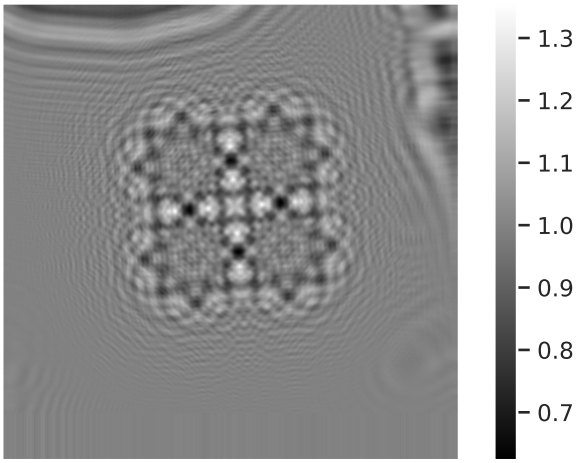


Figure 3.5.: Phase contrast image of experimental data.

is equal to 60 nm. The materials used are Al, Al_2O_3, poly(ethylene terephthalate) (PET) and polypropylene (PP) corresponding to

$$\frac{\delta_r}{\beta} \in [367, 570, 2\,200, 2\,930]\,,$$

as described in (Max Langer, Peter Cloetens, Hesse, Suhonen, Pacureanu, Raum, and Francoise Peyrin 2014). An example of an intensity and ground truth pair is shown in Fig. 3.8. For the reconstructed absorption B, we have NMSE = 42.5, SSIM = 99.6 and PSNR = 93.2, while for the reconstructed phase φ, we have NMSE = 15.7, SSIM = 96.8 and PSNR = 57.1. To obtain these results, we had to tune the different weighting parameters $\alpha = 5 \times 10^{-1}, \beta = 5 \times 10^{-2}, \nu = 10^{-4}$ (instead of $\alpha = 10^{-2}, \beta = 5 \times 10^{-3}, \nu = 10^{-2}$) and to increase the number of iterations up to 10 000 (Fig. 3.9). The first graph displays the evolution of the value of the functional

$$J(B, \varphi) = \left\| \mathrm{F}_D(B, \varphi) - \mathrm{I}_D^{\mathrm{obs}} \right\|_2^2 + \mathrm{TGV}^2_{(\alpha,\beta)}(B) + \nu \mathrm{TV}(\varphi)$$

and the second shows the evolution of the NMSE for both the absorption and phase. We observe that the NL-PDHG method manages to recover the desired outputs, but requires more iterations to converge. So this method can be generalized to cases with larger ratios of phase over absorption, in the settings of high resolution and without assumptions about the materials. For the absorption part, the results could probably be further improved and the convergence

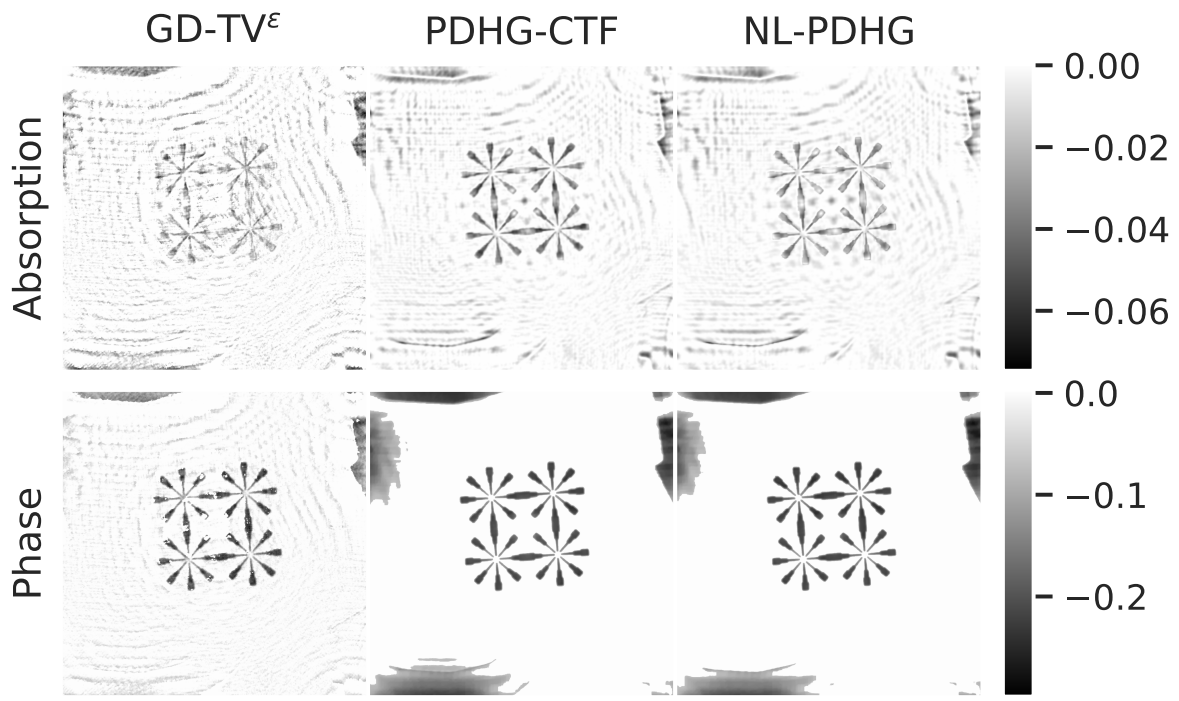


Figure 3.6.: Reconstructions from the experimental intensity.

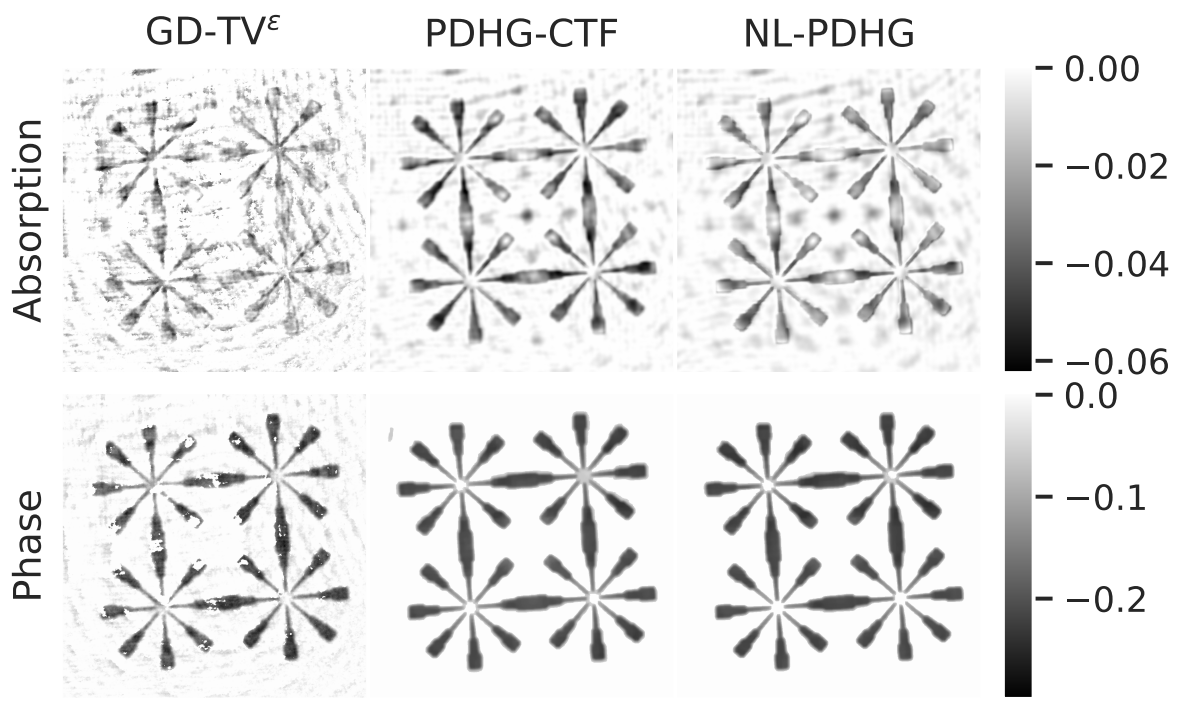
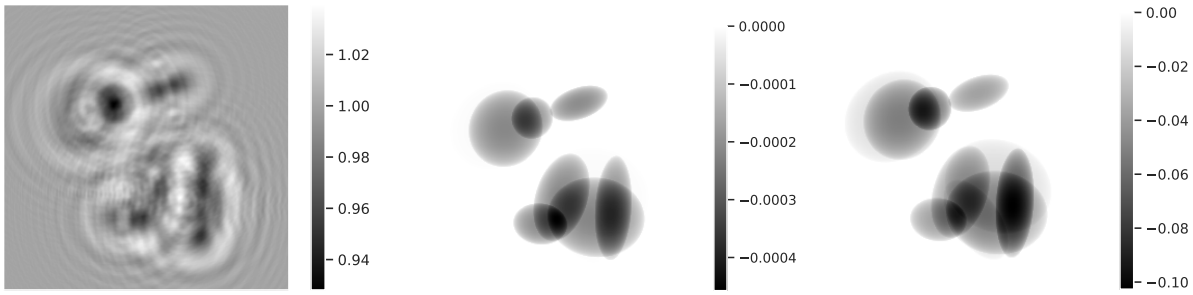


Figure 3.7.: Reconstructions from the experimental intensity (after cropping the image).



((a)) Simulated intensity. ((b)) Absorption ground truth. ((c)) Phase ground truth.

Figure 3.8.: Example of one simulated pair under the new experimental conditions.

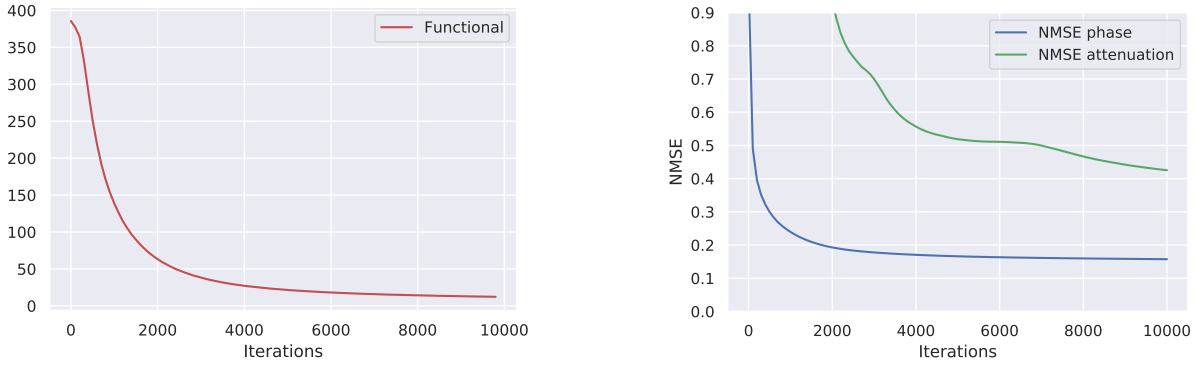


Figure 3.9.: Evolution of the functional and the normalized mean squared error.

could be accelerated by choosing better parameters but more in depth optimisation would have to be done.

IV Conclusion and perspectives

We presented two algorithms based on a primal-dual approach for the reconstruction of both the absorption and phase from a single diffraction pattern. The algorithms permit the use of different regularizations for absorption and phase, which was shown to improve reconstructions by selecting regularizations that take into account their regularity and the way the attenuation and phase contributes to the phase contrast image. In addition, we observed the significant contribution of the nonlinear information of the problem. This suggests that the NL-PDHG algorithm could be applied in a wide variety of cases where linear methods would fail. A drawback of the algorithms is that three regularization parameters must be chosen. The choice of parameters was shown to be robust by the application to a large set of images. However. we have seen that the method could be used for samples that have higher phase over absorption ratio (Alloo et al. 2022) or have more complex structures by adjusting the parameters and increasing the number of iterations. Further investigation should be carried out for non-sparse objects, such as those encountered in biological soft-tissue imaging. The case of X-rays from a laboratory environment can be studied by considering the Kullback-Leibler divergence instead of l_2 norm. A direct extension of this work would be to apply the proposed algorithms to phase contrast tomography (Kostenko et al. 2013), in particular when there is no assumption of multi-materials, considering the 3D version of TV or generalized TV. Finally, the algorithms could be extended

by using neural networks to learn the regularization parameters or the regularization itself.

4 | Deep Gauss-Newton for phase retrieval

In this chapter, we introduce the Deep Gauss-Newton (DGN) algorithm. The DGN allows to take into account the knowledge of the forward model in a deep neural network by unrolling a Gauss-Newton optimization method. No regularization or step size need to be chosen, they are learned through convolutional neural networks. The proposed algorithm does not require any good initial reconstruction and is able to retrieve simultaneously the phase and absorption from a single-distance diffraction pattern. The DGN method was applied to both simulated and experimental data and enable to achieve large improvements of the reconstruction error and of the resolution compared to a state of the art iterative method and to another neural network based reconstruction algorithm.

We then propose to extend this approach by unrolling the proximal Gauss-Newton scheme. We show that the choice of the optimization algorithm we unroll is important. By changing the architecture of DGN, we obtain the Deep Proximal Gauss-Newton (DPGN). We demonstrate that this new architecture improves the quality of reconstructions in terms of resolution and reconstruction error, although the number of parameters remains more or less the same.

I Introduction

The development of deep learning methods in recent years has led to many advances in image and signal processing (LeCun, Y. Bengio, and Hinton 2015). Specifically, deep neural networks (DNNs) have been used to solve a wide variety of inverse problems (Ongie et al. 2020). Despite this progress, the black-box nature of DNNs, i.e., their lack of interpretability, is one of the primary obstacles for their use. Several approaches to exploiting DNNs for the solution of inverse problems have been proposed. DNNs can be used to reconstruct the unknown image directly from an available measurements, or some parts of an algorithm can be replaced by DNNs. Algorithm unrolling is an emerging technique based on the incorporation of convolutional

neural networks (CNNs) into an iterative optimization scheme in order to give the DNN a specific role in the reconstruction enhancing the properties of the classical method (Arridge et al. 2019). Such approaches have found many applications (Monga, Y. Li, and Eldar 2021) and unrolling has been applied to several optimization approaches: gradient descent algorithms (Hauptmann et al. 2018), primal-dual schemes (Jonas Adler and Ozan Öktem 2018), and Alternating Direction Method of Multipliers (Yan Yang et al. 2020). For phase retrieval, several architectures have been proposed, including MS-D Net (Kannara Mom, Bruno Sixou, and Max Langer 2022) and PhaseGAN (Y. Zhang et al. 2021). This kind of network is trained to approximate the inverse operator and often require large training sets as well as long training time. Other methods have incorporated neural networks into iterative schemes, but they are either computationally demanding (Metzler et al. 2018; Işil, F. S. Oktem, and Koç 2019) or rely on a linearization of the forward model (C. Bai et al. 2019).

Here, we present a new learned iterative scheme, the Deep Gauss-Newton (DGN) algorithm, which is obtained by unrolling a Gauss-Newton iteration and which is able to retrieve both the absorption and the phase from a single diffraction pattern. The proposed method combines CNNs and knowledge of the imaging physics given by the forward operator and its Fréchet derivative. The rationale behind this choice is to take a well-known algorithm that is known to converge quickly and enhance it with machine learning by unrolling. We expect this scheme to inherit or improve the convergence properties of the Gauss-Newton method. Another advantage of this approach is that no regularization has to be chosen, instead it is adaptively learned from the data during training. We demonstrate the capability of the method to retrieve phase and attenuation from a single phase contrast image on simulated data as well as experimental data.

II Methods

In this section we present successively the deep Gauss-Newton and the deep proximal Gauss-Newton schemes. We detail the newtworks architectures and the layers that have been used to improve the variational iterative formula.

II.1 Deep Gauss-Newton

In the following, we will note $f = (B, \varphi)$ the couple we aim to retrieve. The inverse of F_D can be approximated using variational methods such as IRGN (Maretzke, Bartels, et al. 2016), corresponding to Tikhonov regularization of the Newton steps:

$$f_{k+1} = \underset{f}{\operatorname{argmin}} \left\{ \left|\left| \mathrm{F}_D(f_k) + \mathrm{F}'_D(f_k)\,(f - f_k) - \mathbf{I}_D^{\mathrm{obs}} \right|\right|_2^2 + \alpha_k \left|\left| f \right|\right|_2^2 \right\} \tag{4.1}$$

where $\mathrm{F}'_D(f_k)$ is the Fréchet dérivative (V. Davidoiu et al. 2011) of F_D at the point f_k, $\alpha_k > 0$ is a regularization parameter at iteration k, and $\mathbf{I}_D^{\mathrm{obs}}$ is a noisy measured intensity. The minimization problem (4.1) has a unique solution given by

$$f_{k+1} = f_k + \left[\mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k) + \alpha_k \mathrm{Id}\right]^{-1} \left\{ \mathrm{F}'_D(f_k)^* \left[\mathbf{I}_D^{\mathrm{obs}} - \mathrm{F}_D(f_k)\right] - \alpha_k f_k \right\} \tag{4.2}$$

where $\mathrm{F}'_D(f_k)^*$ is the adjoint of the linear map $\mathrm{F}'_D(f_k)$ and Id is the identity. Usually, a step size for the update of f_k is introduced. Here, we propose instead to learn a regularization by replacing the Tikhonov term $\alpha_k f_k$ with a CNN $\mathrm{G}_{\theta_k^g}$ with parameters θ_k^g, and to approximate the inverse of the operator $\left[\mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k) + \alpha_k \mathrm{Id}\right]$ with another CNN $\mathrm{H}_{\theta_k^h}$ with parameters θ_k^h, based on the current iterate f_k and on the approximate Hessian $\mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k)$. The network

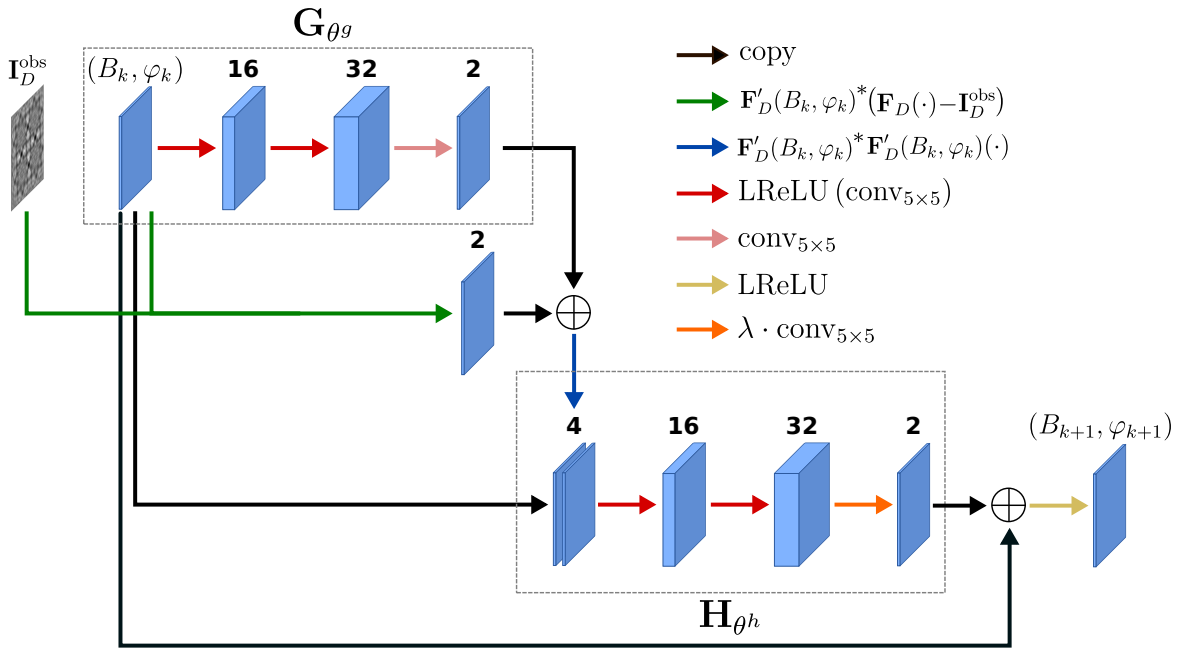


Figure 4.1.: Architecture of the network $\Gamma_\theta^{\text{DGN}}$, representing one iteration of the Deep Gauss-Newton method.

$\mathrm{H}_{\theta_k^h}$ then replaces the classical approximation of the inverse of the Hessian used in the traditional Gauss-Newton scheme by a potentially better and faster learned approximation. If the algorithm is stopped after N iterations, we get

$$f_N = \left(\Gamma_{\theta_N}^{\text{DGN}} \circ \cdots \circ \Gamma_{\theta_1}^{\text{DGN}}\right)(f_0, \mathbf{I}_D^{\text{obs}}) \tag{4.3}$$

where f_0 is the initial guess, $\theta_k = \left(\theta_k^g, \theta_k^h\right)$ and

$$\Gamma_{\theta_k}^{\text{DGN}}(f_k, \mathbf{I}_D^{\text{obs}}) = f_k + \mathrm{H}_{\theta_k^h}\left[f_k, \mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k)\left\{\mathrm{F}'_D(f_k)^*\left[\mathbf{I}_D^{\text{obs}} - \mathrm{F}_D(f_k)\right] + \mathrm{G}_{\theta_k^g}(f_k)\right\}\right] \tag{4.4}$$

Unrolling this scheme, we can consider $\Lambda_\Theta^{\text{DGN}} = \Gamma_{\theta_N}^{\text{DGN}} \circ \cdots \circ \Gamma_{\theta_1}^{\text{DGN}}$ as a deep neural network representing N iterations with $\Theta = (\theta_1, \ldots, \theta_N)$ its parameters. Recent work on unrolling schemes has shown that using the same transformation at each iteration so that $\theta_k = \theta$ for $k \in \{1, \ldots, N\}$, yields good results (Dabre and Fujita 2019). $\Lambda_\Theta^{\text{DGN}}$ can then be seen as a recurrent neural network. The architecture of the network $\Gamma_\theta^{\text{DGN}}$ used for each iteration is shown in Fig. 4.1, and the various operations are summarized in the algorithm. 12.

Algorithm 12 Deep Gauss-Newton

Given an initialization (B_0, φ_0).

for $k = 1, \ldots, N$ **do** :

$r_k \leftarrow \mathrm{G}_{\theta^g}(B_{k-1}, \varphi_{k-1})$

$d_k \leftarrow \mathrm{F}'_D(B_{k-1}, \varphi_{k-1})^* \mathrm{F}'_D(B_{k-1}, \varphi_{k-1})\left\{\mathrm{F}'_D(B_{k-1}, \varphi_{k-1})^*\left[\mathbf{I}_D^{\text{obs}} - \mathrm{F}_D(B_{k-1}, \varphi_{k-1})\right]\right\}$

$t_k \leftarrow \mathrm{H}_{\theta^h}\left[(B_{k-1}, \varphi_{k-1}), r_k + d_k\right]$

$(B_k, \varphi_k) \leftarrow (B_{k-1}, \varphi_{k-1}) + t_k$

The network G_{θ^g} takes the current iterate f_k as input, spreads it to 16 and then 32 channels by a convolutional layer with kernel size 5×5 using a leaky rectified linear unit (LReLU) as non-linearity, defined as $\mathrm{LReLU}_\alpha(x) = \max(x, \alpha x)$, $\alpha > 0$. The output of the network G_{θ^g} is added to $\mathrm{F}'_D(f_k)^* \left[\mathbf{I}_D^{\mathrm{obs}} - \mathrm{F}_D(f_k)\right]$, stacked with the current iterate and then fed to the network H_{θ^h} which consists of the same set of operations as G_{θ^g} (except it has four input channels instead of two). Finally, the output is added to the current iterate and projected onto positive numbers by a LReLU. The architectures of the networks are kept simple for several reasons. A shallow network added to the iterative update saves computational time and the memory required, while giving good reconstruction results.

II.2 Deep Proximal Gauss-Newton

More recently, an extension of the Gauss-Newton method, the Proximal Gauss-Newton method has been introduced (Salzo and Villa 2011) and it incorporates a proximal term to improve convergence and handle certain types of constraints or regularization requirements. The algorithm iteratively updates the solution by performing a Gauss-Newton step followed by a proximal regularization step. Instead of considering an iteration of the form (4.2), we can add a scaled proximity operator to this Newton step (Fu et al. 2017):

$$f_{k+1} = \mathrm{prox}_J^{H(f_k)} \left(f_k + H(f_k)^{-1} \left\{ \mathrm{F}'_D(f_k)^* \left[\mathbf{I}_D^{\mathrm{obs}} - \mathrm{F}_D(f_k) \right] - \alpha_k f_k \right\} \right) \tag{4.5}$$

where $H(f_k) = \mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k) + \alpha_k \mathrm{Id}$ and $\mathrm{prox}_J^{H(f_k)}$ is the proximity operator associated to a certain function J and scaled by $H(f_k)$:

$$\mathrm{prox}_J^H(z) = \inf_x \left\{ J(x) + \frac{1}{2} \|x - z\|_H^2 \right\} \tag{4.6}$$

The norm $\|\cdot\|_H$ is induced by the inner product $\langle x|z\rangle_H = \langle x|Hz\rangle$. It can be shown that the iteration (4.5) is equivalent to:

$$f_{k+1} = \underset{f}{\mathrm{argmin}} \left\{ \left\| \mathrm{F}_D(f_k) + \mathrm{F}'_D(f_k)\,(f - f_k) - \mathbf{I}_D^{\mathrm{obs}} \right\|_2^2 + \frac{\alpha_k}{2} \|f\|_2^2 + J(f) \right\} \tag{4.7}$$

The functional J acts like an additional regularization which is expected to improve the reconstruction. The idea behind the Deep Poximal Gauss-Newton (DPGN) is basically the same as DGN. The Tikhonov term is replaced by a CNN G_{θ^g}, the inverse of the operator $H(f_k)$ is approximate by a CNN H_{θ^h}, and the scaled proximity operator $\mathrm{prox}_J^{H(f_k)}$ is replaced by a CNN J_{θ^j}. Again, if the algorithm is stopped after a fixed number N of iterations, it can be written as in (4.3):

$$f_N = \left(\Gamma_{\theta_N}^{\mathrm{DPGN}} \circ \cdots \circ \Gamma_{\theta_1}^{\mathrm{DPGN}} \right) (f_0, \mathbf{I}_D^{\mathrm{obs}}) \tag{4.8}$$

Then, with $\theta = \left(\theta^g, \theta^h, \theta^j\right)$, one iteration of the DPGN can be described as:

$$\Gamma_\theta^{\mathrm{DPGN}}(f_k, \mathbf{I}_D^{\mathrm{obs}}) = \mathrm{J}_{\theta^j} \left(f_k + \mathrm{H}_{\theta^h} \left[f_k, \mathrm{F}'_D(f_k)^* \mathrm{F}'_D(f_k) \left\{ \mathrm{F}'_D(f_k)^* \left[\mathbf{I}_D^{\mathrm{obs}} - \mathrm{F}_D(f_k) \right] + \mathrm{G}_{\theta^g}(f_k) \right\} \right] \right) \tag{4.9}$$

For a fair comparison, DPGN's architecture differs slightly from DGN's in order to have approximatively the same number of parameters. The network G_{θ^g} is simplified and consists simply of 16 convolutional filters followed by 2 convolutional filters. The network H_{θ^h} is the

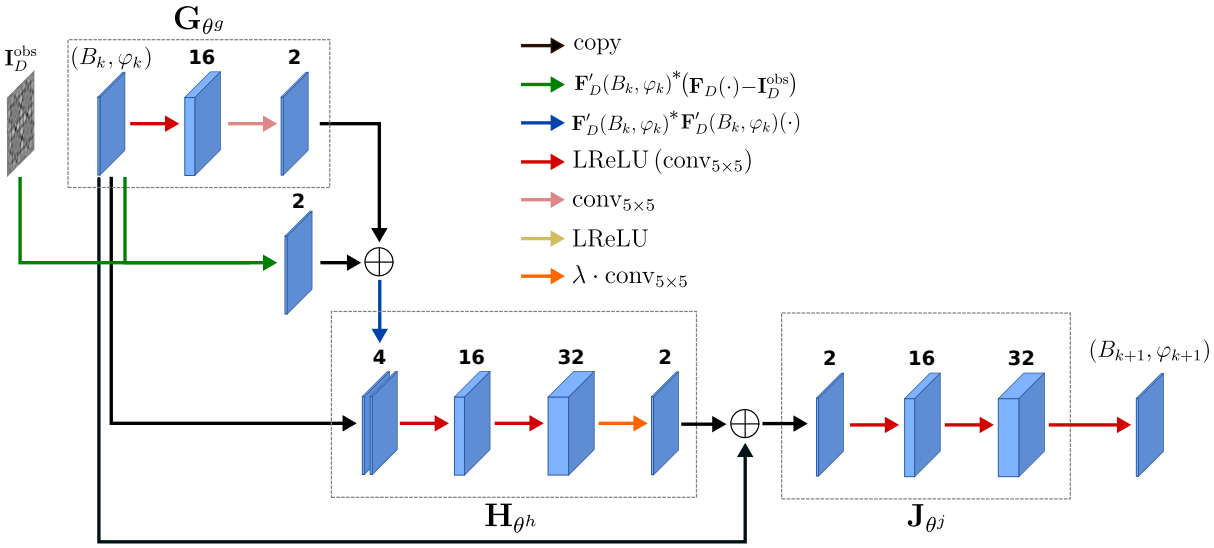


Figure 4.2.: Architecture of the network $\Gamma_\theta^{\text{DPGN}}$, representing one iteration of the Deep Proximal Gauss-Newton method.

Algorithm 13 Deep Proximal Gauss-Newton

Given :

- $(B_0, \varphi_0) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m}$

for $k = 1, \ldots, N$ **do** :

$r_k \leftarrow \mathrm{G}_{\theta^g}(B_{k-1}, \varphi_{k-1})$

$d_k \leftarrow \mathrm{F}'_D(B_{k-1}, \varphi_{k-1})^* \mathrm{F}'_D(B_{k-1}, \varphi_{k-1}) \left\{ \mathrm{F}'_D(B_{k-1}, \varphi_{k-1})^* \left[\mathbf{I}_D^{\text{obs}} - \mathrm{F}_D(B_{k-1}, \varphi_{k-1}) \right] \right\}$

$t_k \leftarrow \mathrm{H}_{\theta^h}\left[(B_{k-1}, \varphi_{k-1}), r_k + d_k\right]$

$(B_k, \varphi_k) \leftarrow \mathrm{J}_{\theta^j}\left[(B_{k-1}, \varphi_{k-1}) + t_k\right]$

same as in DGN. But the major difference here is that, once the output of H_{θ^h} is added to the current iterate, we apply the network J_{θ^j} which spreads it to 16 and then 32 channels by a convolutional layer, and the output is projected onto positive numbers by a LReLU. The architecture of the network $\Gamma_\theta^{\text{DPGN}}$ used for each iteration is shown in Fig. 4.2, and the DPGN algorithm is summarized in algorithm. 13.

III Results

III.1 Training data

Training and validation data were generated under the same conditions as described in III.3. The dataset consisted of 11 000 pairs, from which 10 000 images were used for training, 1 000 for validation during training. The 1 000 data used for evaluation in the previous chapter constitute our test datatset.

III.2 Training strategy

We compare the unrolling frameworks to the standard IRGN method (Maretzke, Bartels, et al. 2016) as well as to the Mixed-Scale Dense Network (MS-D Net), a direct reconstruction method that does not include prior knowledge on the imaging physics (Kannara Mom, Bruno Sixou, and Max Langer 2022). In the IRGN, the positive-definite linear operator in (4.2) can be inverted

Tableau 4.1.: Parameters for the Neural Networks

	MS-D Net	DGN	DPGN
Loss function	MSE	MSE	MSE
Training epochs	100	100	100
Learning rate	10^{-3}	5×10^{-4}	5×10^{-4}
Batch size	10	10	10
Activation function	ReLU	LReLU	LReLU
Training time	35h	21h	22h
Number of parameters	46×10^3	31×10^3	32×10^3

efficiently by a conjugate gradient (CG) method VII.2. We used 100 Newton steps and 10 iterations for the CG, corresponding to a total of 1 000 CG-iterations. For the DGN (and the DPGN) method, we used $N = 10$ iterations, which means that the derivative $\mathrm{F}'_D(f_k)$ and its adjoint $\mathrm{F}'_D(f_k)^*$ are evaluated 10 times. The number of iterations was chosen empirically so that the NMSE stagnates. As opposed to (Hauptmann et al. 2018), where several networks are trained sequentially, i.e. iteration by iteration, here, given a training set $\{y^i, f^i\}$ where y_i denotes the intensity $\mathrm{F}_D(f^i)$ corrupted by noise, we use one network $\mathbf{\Gamma}_\theta$ applied N times in a recurrent fashion to obtain the DNN $\mathbf{\Lambda}_\theta$, which is trained to perform end-to-end reconstruction. The unrolling methods were trained using 100 epochs with a batch size of 10, the ADAM optimizer, an initial learning rate of 5×10^{-4} and a cosine annealing learning rate schedule (Loshchilov and Hutter 2016). The LReLU activation function parameter was set to the default $\alpha = 0.3$. Warm-up initialization decreased training time but did not yield better final results. Therefore, for simplicity, zero initialization, $f_0 = (0, 0)$, was used throughout. For the MS-D Net, we used the same settings as in (Kannara Mom, Bruno Sixou, and Max Langer 2022). Using LReLU in the MS-D Net did not improve the reconstructions. The hyperparameters for the networks are summarized in Tab. 4.1 and were optimized using grid search. All three networks were trained on the same training set as described III.1.

III.3 Fourier Ring Correlation

The spatial resolution was evaluated using the Fourier Ring Correlation (FRC). FRC involves comparing two images by calculating the correlation between their Fourier transforms, specifically within rings in the frequency domain:

$$\mathrm{FRC}_{\{f, f_{\mathrm{true}}\}}(R_i) = \frac{\sum_{r \in C(R_i)} \widehat{f^*}(r) \widehat{f_{\mathrm{true}}}(r)}{\sqrt{\left(\sum_{r \in C(R_i)} \left|\widehat{f}(r)\right|^2\right)\left(\sum_{r \in C(R_i)} \left|\widehat{f_{\mathrm{true}}}(r)\right|^2\right)}} \tag{4.10}$$

where R_i is the radius of the ring $C(R_i)$ in the Fourier domain within which the correlation is computed, f^* is the conjugate of f and $\widehat{f}$ is the Fourier transform of f. The FRC analysis generates a plot called the FRC curve, which represents the correlation values as a function of spatial frequency (see Fig. 4.3). From this, we can build a metric to measure the information we are able to reconstruct at a certain frequency level. The resolution ρ of the reconstructed image can then be defined as

$$\rho(f, f_{\mathrm{true}}) = \left(R_{\{\mathrm{FRC}_{\{f, f_{\mathrm{true}}\}}(R) \leq \tau(R)\}}\right)^{-1} \tag{4.11}$$

where $R_{\{\mathrm{FRC}_{\{f,f_{\mathrm{true}}\}}(R)(R)\leq\tau(R)\}}$ is the radius for which the FRC is lower than a threshold τ. For the threshold, we used the 2σ-threshold, as defined in (Banterle et al. 2013), which depends on the radius, and is computed as

$$\tau(R) = \frac{2}{\sqrt{\frac{N_p(R)}{2}}} \tag{4.12}$$

with R the radius in the Fourier domain and $N_p(R)$ the number of pixels contained within the corresponding ring. From the FRC curve, we also compute the Fourier Ring Correlation Metric (FRCM), which is the mean square difference between the FRC and unity over all spatial frequencies:

$$\mathrm{FRCM}(f, f_{\mathrm{true}}) = \sum_i \left(1 - \mathrm{FRC}_{\{f,f_{\mathrm{true}}\}}(R_i)\right)^2 \tag{4.13}$$

A small FRCM implies a higher similarity in the Fourier domain.

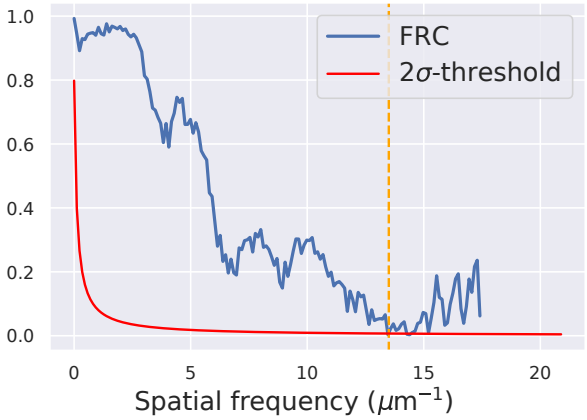
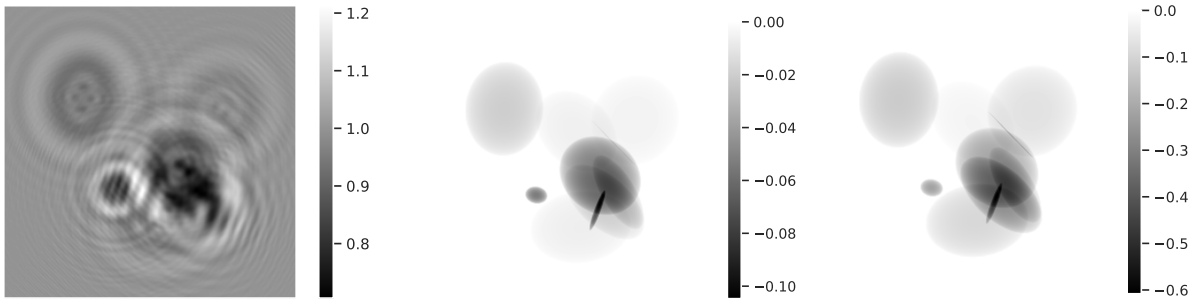


Figure 4.3.: Resolution evaluation using Fourier Ring Correlation. The resolution estimated by the 2σ-threshold criterion is 75 nm.

III.4 Simulated data



((a)) Simulated intensity. ((b)) Absorption ground truth. ((c)) Phase ground truth.

Figure 4.4.: Example of one simulated pair.

To quantify reconstruction quality on simulated data, we used the normalized mean square error (NMSE), the FRCM and the resolution criterion defined above. Phase and absorption

Tableau 4.2.: Results (mean and standard deviation) on 1 000 simulated images.

Method	NMSE (%)		FRCM (%)		Resolution (nm)		Time (s)
	Absorption	Phase	Absorption	Phase	Absorption	Phase	
IRGN	85.5 (40.7)	39.3 (15.0)	71.2 (9.95)	68.1 (5.45)	238 (136)	154 (43)	116
MS-D Net	13.6 (12.8)	10.6 (10.8)	48.8 (13.8)	47.8 (13.3)	102 (77.4)	98.5 (135)	**2.60**
DGN	12.1 (13.5)	4.61 (6.20)	35.7 (15.7)	23.0 (16.6)	72.2 (55.2)	62.3 (37.0)	5.88
DPRGN	**11.0 (12.3)**	**4.05 (5.25)**	**31.5 (13.9)**	**19.8 (15.8)**	**74.0 (63.4)**	**59.1 (28.3)**	6.31

reconstructions from one simulated image pair in Fig. 4.4 are shown in Fig. 4.5. Both deep learning methods performed better than the IRGN method, which tends to leave artifacts in the absorption and yields a blurred phase.

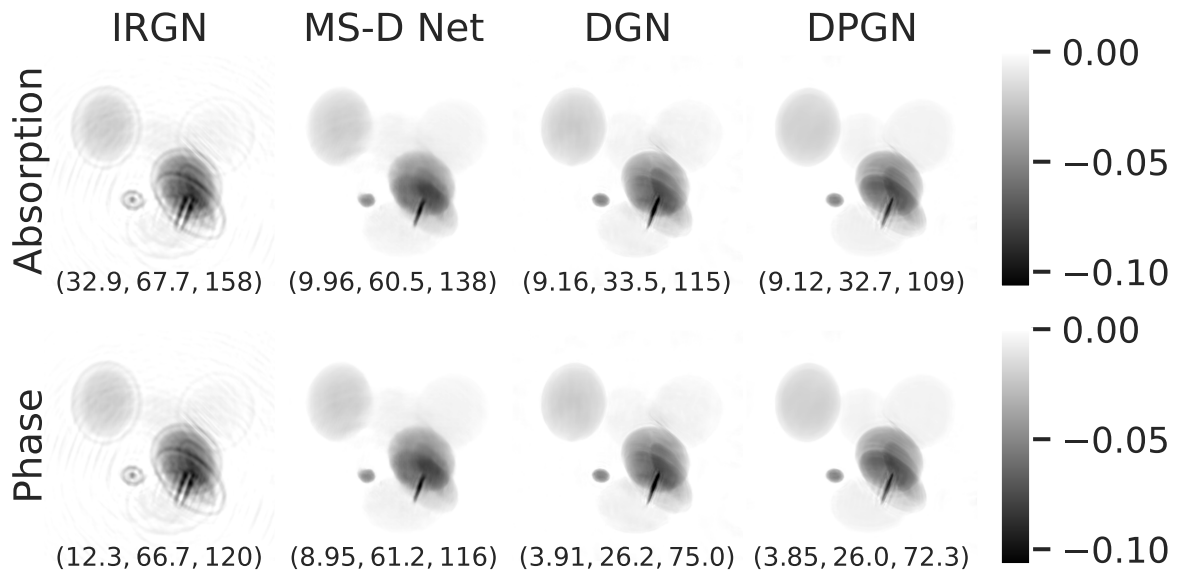
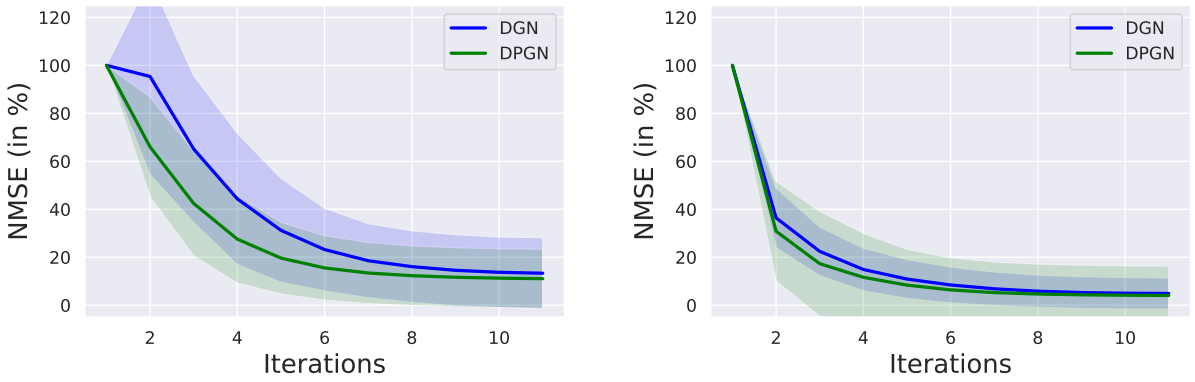


Figure 4.5.: Reconstructions from simulated data. Reconstruction quality is given as (NMSE (%), FRCM (%), resolution (nm)).

The mean and standard deviation of these metrics were calculated on the test dataset described in III.1. The average computation time for one reconstruction was measured to compare the execution time. The results are summarized in Tab. 4.2. On average, the deep learning approaches performed better than the IRGN, both in terms of reconstructed values and resolution. In terms of NMSE, the deep learning approaches gave similar results for the absorption, but the DGN performed better than the MS-D Net for phase recovery. Moreover, the DGN yielded better resolution as well as better correlation in the frequency domain. We can also see that the DPGN method has made it possible to further improve the reconstructions obtained by DGN, in terms of errors and resolution. As expected, the MS-D Net was fastest, since it only requires one application of a neural network. The unrolling approaches are efficient despite the need to compute the derivative of the forward operator as well as its adjoint several times and is 20 times faster than its standard iterative counterpart. The graphs 4.6 show the evolution of the average NMSE on the test dataset over the iterations for both DGN and DPGN. We observe that these approaches seem to have converged after $N = 10$ iterations, as they have been trained to do so. For the phase, a major improvement is achieved in the first step, whereas



((a)) Absorption ((b)) Phase

Figure 4.6.: Evolution of average NMSE (%) of the unrolling approaches for 1 000 test images. The transparent areas correspond to the standard deviation.

absorption requires a few more iterations. These two methods have almost converged after 5 iterations, but the curves continue to decrease.

III.5 Experimental data

The proposed approach was applied on data acquired at beamline NanoMAX at the MAX IV synchrotron (Lund, Sweden) (Kalbfleisch et al. 2022). The sample consists of a stack of palladium, zinc, palladium, gold layers with thicknesses of 21, 10, 11, 163 nm, respectively, deposited on a 1 mm-thick silicon nitride substrate, resulting in expected values for absorption and phase of 0.0483 and 0.217 respectively. To evaluate the reconstructions quantitatively, we used the normalized error (NE) and relative standard deviation (RSD), calculated as $\mathrm{NE} = \frac{l_t - l_m}{l_t}$ and $\mathrm{RSD} = \frac{s_m}{l_m}$, where l_t is the expected value, l_m the measured mean value and s_m the standard deviation in the corresponding material. Calculation of l_m and s_m was done in homogeneous parts of the object to avoid the influence of blur at the edges. The shape of the object was estimated from a phase reconstruction using an iterative method, and chosen so that the calculation of l_m and s_m was done in homogeneous parts to avoid the influence of blur at the edges. The resolution was measured by fitting an error function to a line profile across an object edge and calculating the corresponding Gaussian full width at half maximum based on the error function fitting parameters.

The reconstructions obtained in Fig. 4.7 show that both the DGN and the DPGN yield very high quality reconstructions of the object with almost no remaining visible artifacts. Although the MS-D Net performs well on simulations, it seems to not generalize as well as the DGN and the DPGN to the experimental data given the chosen training strategy. This could be due to the unrolling methods explicitly taking into account the physics model, learning the noise statistics from the data yielding an adapted regularization, while leveraging the convergence properties of the optimization method (Monga, Y. Li, and Eldar 2021). On the other hand, the MS-D Net was trained to reconstruct directly from the measurements, without knowledge of the physics model.

Figure 4.8 shows the reconstructions obtained for absorption and phase at each iteration of the unrolling methods. The same behavior can be seen as in the case of simulated data 4.6. After a single iteration, we were able to recover the phase shift, although some artifacts remain. But these artifacts are erased after a few more iterations, and just after 5 iterations, we've almost

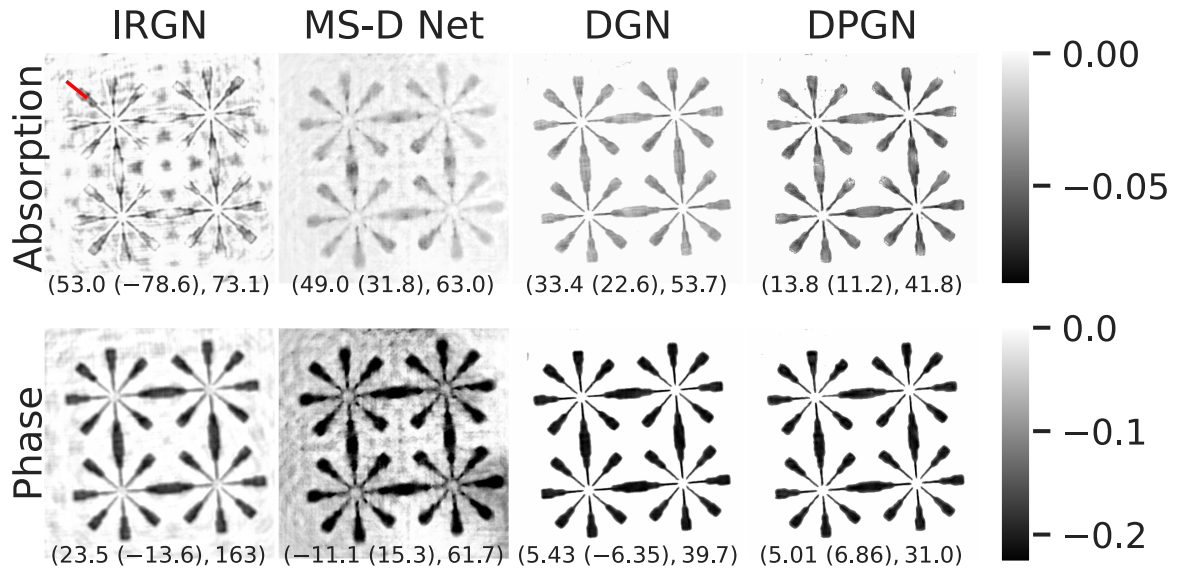


Figure 4.7.: Reconstructions for experimental data. The profiles along the red line were measured to estimate the resolution. Values correspond to (NE (%) (RSD (%)), resolution (nm)).

reached convergence, since the recovered phase has hardly changed at all. For absorption, it takes a few more iterations ($\approx 3-4$) to get the object with a few artifacts, but just like the phase, a few more iterations allow us to get the object with almost no artifacts.

IV Conclusion

A limitation of the proposed algorithms is that the forward model has to be fully known, e.g. the propagation distance has to be precisely measured. In future work we will study possibilites to correct errors in the forward model. We will also investigate out-of-distribution generalization error, e.g. with respect to different noise properties. The developed DGN algorithm allows to efficiently retrieve both the phase and absorption from a single phase contrast image. By exploiting recent developments in deep learning and integrating CNNs into a regularized Gauss-Newton type scheme, with the DGN and the DPGN, we overcome the limitations of classical iterative approaches while leveraging the power of neural networks. Since no regularization term needs to be chosen, the unrolling networks are trained to learn an optimal one for the absorption and the phase respectively, which improves the quality of the reconstructions. Taking into account the knowledge of the forward model in a simple network enhances the reconstructions and allows a better generalization on real data. Compared to the standard IRGN, the unrolling methods both substantially improved the reconstruction and reduced the calculation time. The choice of optimization scheme we unrolled seems to have an impact on reconstruction quality. We saw that by changing the architecture while keeping the same number of parameters, the DPGN approach improved on the already good results obtained with DGN.

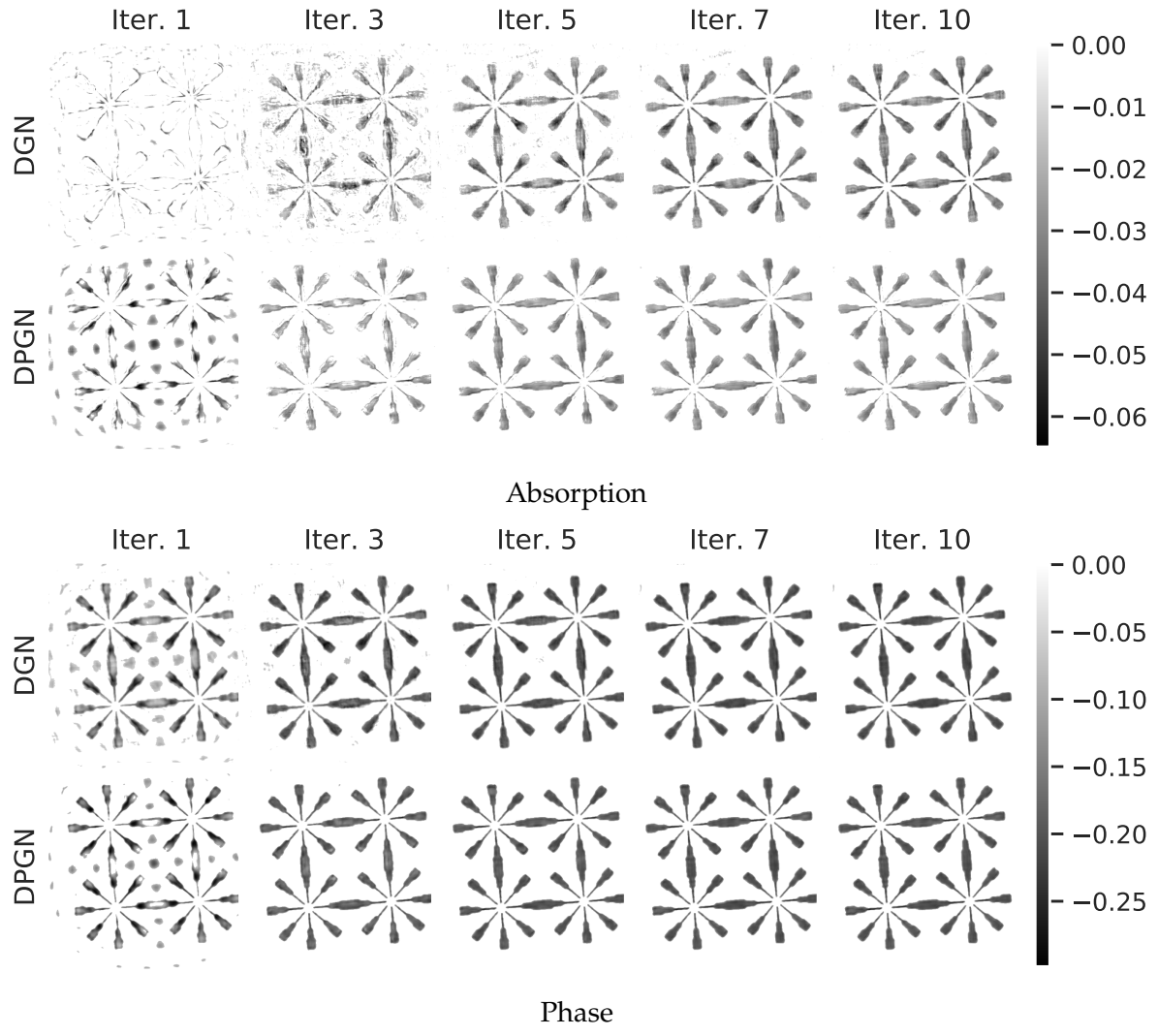


Figure 4.8.: Evolution of experimental data reconstructions with unrolling approaches over the course of iterations.

5 | Comparison of algorithms unrolling for phase retrieval

In the previous chapter, we introduced two algorithms based on the idea of unrolling an iterative algorithm. We have seen that the scheme on which we start influences the quality of the reconstruction obtained. Here, we take the analysis a step further, comparing unrolling approaches from different optimization schemes with their standard counterparts. In particular, we discuss the number of iterations used for unrolling, the effect produced at the training level and also what happened during the inference.

I Introduction

Unrolling approaches have been used in various fields of image and signal processing (Monga, Y. Li, and Eldar 2021), based on different optimization schemes: gradient descent (Hauptmann et al. 2018; Jonas Adler and O. Oktem 2017), primal-dual schemes (Jonas Adler and Ozan Öktem 2018), alternating direction method of multipliers (Yan Yang et al. 2020) and Gauss-Newton (Kannara Mom, Max Langer, and Bruno Sixou 2023). However, despite the growing interest and potential of unrolled schemes, a comprehensive comparative analysis of different variants and their performance characteristics is still lacking. This article aims to bridge that gap by providing an in-depth exploration and comparison of various unrolled schemes for the problem of phase retrieval.

Throughout this section, we will delve into the underlying principles of phase retrieval and its significance in practical applications. We will then introduce the concept of unrolled schemes in a general way, highlighting their potential benefits and the underlying mechanisms that enable them to address the challenges of phase retrieval. Subsequently, we will survey and compare different variations of unrolled schemes proposed in the literature. We'll see that the choice of optimization scheme has an influence on reconstruction quality and method stability.

II Unrolling iterative solutions

Reconstruction methods based on variational approaches lead to an optimization problem that consists in minimizing an energy parameterized by a measured data $y \in \mathcal{Y}$:

$$\begin{array}{rccc} \mathbf{J}: & \mathcal{Y} & \to & \mathcal{X} \\ & y & \mapsto & \underset{x \in \mathcal{X}}{\operatorname{argmin}} \{\mathcal{J}(x, y)\} \end{array} \tag{5.1}$$

where $\mathcal{J} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $\mathcal{X}$ represents the model parameter space and $\mathcal{Y}$ the data space. We assume that both $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces. In image reconstruction tasks, the functional $\mathcal{J}$ is usually written as

$$\mathcal{J}(x, y) = d\left[\mathbf{F}_D(x), y\right] + \mathcal{R}(x) \tag{5.2}$$

where $d : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is a function (not necessarily a distance) which measures the consistency with the data, and $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is a regularization term which allows us to give some prior knowledge on the desired reconstruction.

The general idea behind unrolling is to train a deep neural network $\Lambda_\theta : \mathcal{Y} \to \mathcal{X}$ parametrized by the parameters θ that is appropriate to approximate the operator $\mathbf{J}$ and is implicitly defined through an iterative scheme. In the following paragraphs, we will give examples of how to unroll different iterative schemes.

II.1 Deep Gradient Descent (DGD)

The first case we consider here is the gradient descent scheme. Formally, assuming that the applications d and $\mathcal{R}$ of (5.2) are differentiable, a gradient descent that we stop after N iterations can be written as:

$$\begin{cases} (B_0, \varphi_0) \text{ is given,} \\ (B_k, \varphi_k) = (B_{k-1}, \varphi_{k-1}) - \theta_k \nabla \mathcal{J}\left[(B_{k-1}, \varphi_{k-1}), \mathbf{I}_D^{\text{obs}}\right], \quad \text{for } k = 1, \dots, N \end{cases} \tag{5.3}$$

where θ_k is the gradient step at the k-th iteration. In the gradient descent scheme defined by (5.3), we can express the $N-$th iterate by:

$$(B_N, \varphi_N) = \left(\Gamma_{\theta_N} \circ \cdots \circ \Gamma_{\theta_1}\right)(B_0, \varphi_0) \tag{5.4}$$

where $\Gamma_{\theta_k} := \mathrm{Id} - \theta_k \nabla \mathcal{J}(\cdot, \mathbf{I}_D^{\text{obs}})$, for $k = 1, \dots, N$ and $\mathbf{I}_D^{\text{obs}} \in \mathcal{Y}$ represents the measured intensity. Unrolling this iteration, we can then consider that

$$\Lambda_\Theta = \Gamma_{\theta_N} \circ \cdots \circ \Gamma_{\theta_1} \tag{5.5}$$

is a convolutional neural network representing N iterations and that $\Theta = (\theta_1, \dots, \theta_N)$ represents the parameters of this network.

Let's consider the case where we have the same operation at each iteration, i.e. $\theta_1 = \cdots = \theta_N = \theta$, in which case Λ_Θ can be seen as a recurrent neural network . Instead of choosing a regularization and gradient steps, we propose to learn the iteration directly, which, for $k = 1, \dots, N$, corresponds to:

$$(B_k, \varphi_k) = \Gamma_\theta^{\text{DGD}}\left\{(B_{k-1}, \varphi_{k-1}), \nabla d\left[\mathbf{F}_D(B_{k-1}, \varphi_{k-1}), \mathbf{I}_D^{\text{obs}}\right]\right\} \tag{5.6}$$

In this case, the regularization term is implicitly learned during network training. This formula is very similar to the proximal forward-backward scheme (Pesquet et al. 2021) and to ISTA-Net (J. Zhang and Ghanem 2018), where the proximal operator is hard to compute and is either

replaced or approximated by a neural network. The $\Gamma_\theta^{\mathrm{DGD}}$ network architecture used here is inspired by (Hauptmann et al. 2018), and is shown in Fig. 5.1 The data discrepancy measure d

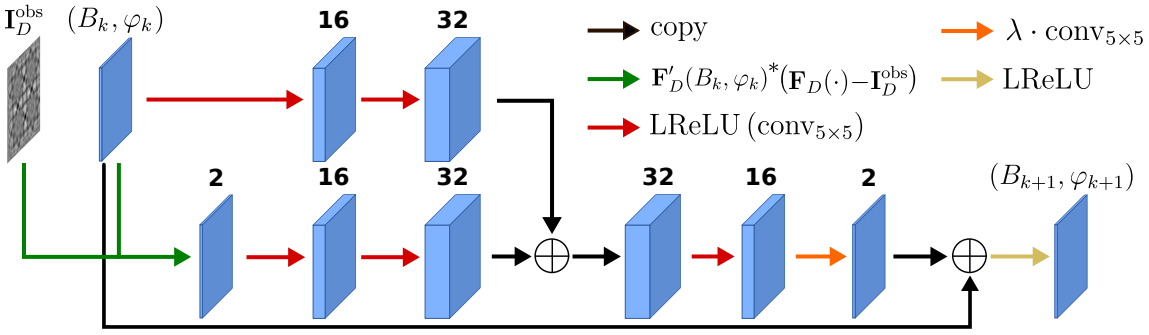


Figure 5.1.: Architecture of the network $\Gamma_\theta^{\mathrm{DGD}}$, representing one iteration of the Deep Gradient Descent.

we choose here is the L^2 norm: $d(x, y) = \|x - y\|_2^2$. In this case, the gradient is given by:

$$\nabla d\left[\mathbf{F}_D(B_k, \varphi_k), \mathbf{I}_D^{\mathrm{obs}}\right] = \mathbf{F}'_D(B_k, \varphi_k)^* \left(\mathbf{F}_D(B_k, \varphi_k) - \mathbf{I}_D^{\mathrm{obs}}\right) \tag{5.7}$$

The particular structure we've opted for $\Gamma_\theta^{\mathrm{DGD}}$ when implementing the equation (5.6) update is depicted in Figure 5.1. During each iteration, we feed the current iterate (B_k, φ_k) and $\nabla d\left[\mathbf{F}_D(B_k, \varphi_k), \mathbf{I}_D^{\mathrm{obs}}\right]$ into a comparable process, wherein both undergo expansion to 16 and subsequently 32 channels through a convolutional layer using a kernel size of $s = 5$, equipped with Leaky ReLU as nonlinearity. The outcomes from both pipelines are combined and initially reduced to 16 channels and are processed with a LReLU activation function. Following this, the channels are further reduced to 2 channels through a simple channel-wise scalar multiplication by $\lambda = (\lambda_B, \lambda_\varphi) \in \mathbb{R}^2$ without applying any nonlinearity. The resulting value is then added to the current iterate and transformed to positive values by passing it through another ReLU projection. The architecture is kept simple for several reasons: firstly, we aim to learn a combination of current iteration and gradient information in order to mimic a descent step, rather than post-processing with a large network; and secondly, to minimize the memory required for training. This network structure, similar to a gradient descent algorithm, is called Deep Gradient Descent (DGD).

II.2 Deep Primal-Dual (DPD)

In Chapter 3, we introduced a widely recognized primal-dual approach which is the primal-dual hybrid gradient (PDHG) algorithm (Esser, X. Zhang, and T. Chan 2010), alternatively referred to as the Chambolle-Pock algorithm (Chambolle and Pock 2016b). This algorithm has been extended to accommodate non-linear operators, which resulted in the nonlinear primal-dual hybrid gradient (NL-PDHG). This scheme, is specifically designed for addressing minimization problems exhibiting the following structure:

$$\min_{x \in \mathcal{X}} \{\mathcal{H}\left[\mathcal{K}(x)\right] + \mathcal{G}(x)\} \tag{5.8}$$

where $\mathcal{K} : \mathcal{X} \to \mathcal{Y}$ is a (possibly non-linear) operator and $\mathcal{H} : \mathcal{Y} \to \mathbb{R}$ and $\mathcal{G} : \mathcal{X} \to \mathbb{R}$ are convex, proper, lower semicontinuous functionals on the dual and primal spaces, respectively The different stages of the NL-PDHG algorithm are summarized in alg. 14. The operator $\mathcal{K}'(x_k)^*$

denotes the adjoint of the Fréchet derivative of $\mathcal{K}$ at the point x_k. Authors of (Jonas Adler and

Algorithm 14 Non-linear primal-dual hybrid gradient

Given :

- Initial step sizes σ_0, τ_0 and relaxation parameter $\gamma \in [0, 1]$
- Initial iterates $x_0 \in \mathcal{X}$ (primal) and $h_0 \in \mathcal{Y}$ (dual)

for $k = 1, 2, \ldots$ **do** :

$h_k \leftarrow \text{prox}_{\sigma \mathcal{H}^*}(h_{k-1} + \sigma \mathcal{K}(\overline{x}_{k-1}))$

$x_k \leftarrow \text{prox}_{\tau \mathcal{G}}(x_{k-1} - \tau \mathcal{K}'(x_{k-1})^*(h_k))$

$\overline{x}_k \leftarrow x_k + \gamma(x_k - x_{k-1})$

$\sigma_{i+1}, \tau_{i+1} \leftarrow \sigma_i, \tau_i$

such that $\sigma_i \tau_i \sup_{k=0,1,\ldots,i} \left\{ |||\mathcal{K}'(x_k)|||^2 \right\} < 1$

Ozan Öktem 2018) proposed to unroll this primal-dual scheme and learn the different steps of 14 through two networks, one updating the primal variable and the other the dual variable. The idea is then to learn the iterative process in both model parameter and data spaces. By doing so, we get the Deep Primal-Dual algorithm 15, where

$$\mathrm{D}_{\theta^d} : \mathcal{Y} \times \mathcal{Y} \to \mathcal{Y} \quad \text{and} \quad \mathrm{P}_{\theta^p} : \mathcal{X} \times \mathcal{X} \to \mathcal{X}$$

are the updating operators in the dual and primal spaces, respectively. For the dual variable h_k, instead of having an explicit update $(h_{k-1} + \sigma \mathcal{K}(\overline{x}_{k-1}))$, we feed the dual network D_{θ^d} with the couple $(h_{k-1}, \mathcal{K}(x_{k-1}))$ in order to allow the network to learn an optimal combination between the current iterate with the output of the operator $\mathcal{K}$. For the primal update x_k, in the same way, the primal network P_{θ^d} takes the couple $(x_{k-1}, \mathcal{K}'(x_{k-1})^*(h_k))$ as input and learn an optimal combination.

Note that the step corresponding to over-relaxation $\overline{x}_k \leftarrow x_k + \gamma(x_k - x_{k-1})$ is no longer necessary, instead, the network chooses at which point the $\mathcal{K}$ operator should be evaluated. And unlike (Jonas Adler and Ozan Öktem 2018), we used the same learned proximal operators

Algorithm 15 Deep Primal-Dual

Given :

- Initial iterates: $x_0 \in \mathcal{X}$ (primal) and $h_0 \in \mathcal{Y}$ (dual)

for $k = 1, \ldots, N$ **do**:

$h_k \leftarrow \mathrm{D}_{\theta^d}(h_{k-1}, \mathcal{K}(x_{k-1}))$

$x_k \leftarrow \mathrm{P}_{\theta^p}(x_{k-1}, \mathcal{K}'(x_{k-1})^*(h_k))$

in each iteration, so this reduces the number of parameters and still gives similar reconstruction quality than if they differ. In our case, the operator $\mathcal{K}$ is defined as $\mathcal{K}(x) = \mathbf{F}_D(x) - \mathbf{I}_D^{\text{obs}}$, with $x = (B, \varphi)$, so this corresponds to the residual of the image x.

The architectures of the primal and dual networks as well as the different operations are summarized in the scheme displayed in Fig. 5.2. Like DGD, the architecture of each network remains simple. At each iteration, we feed the dual network D_{θ^d} with the current iterate of the dual variable h_k and the residual $\mathbf{F}_D(B_k, \varphi_k) - \mathbf{I}_D^{\text{obs}}$, this is then extended to 16 and then 32 channels through a convolutional layer utilizing a kernel size of $s = 5$, followed by LReLU activation. The channels are further reduced to a single channel (which corresponds to the number of measurement) without nonlinearity, followed by a scalar multiplication by λ_d, which acts like a dual step size. The next iterate of the dual variable is then obtained by summing

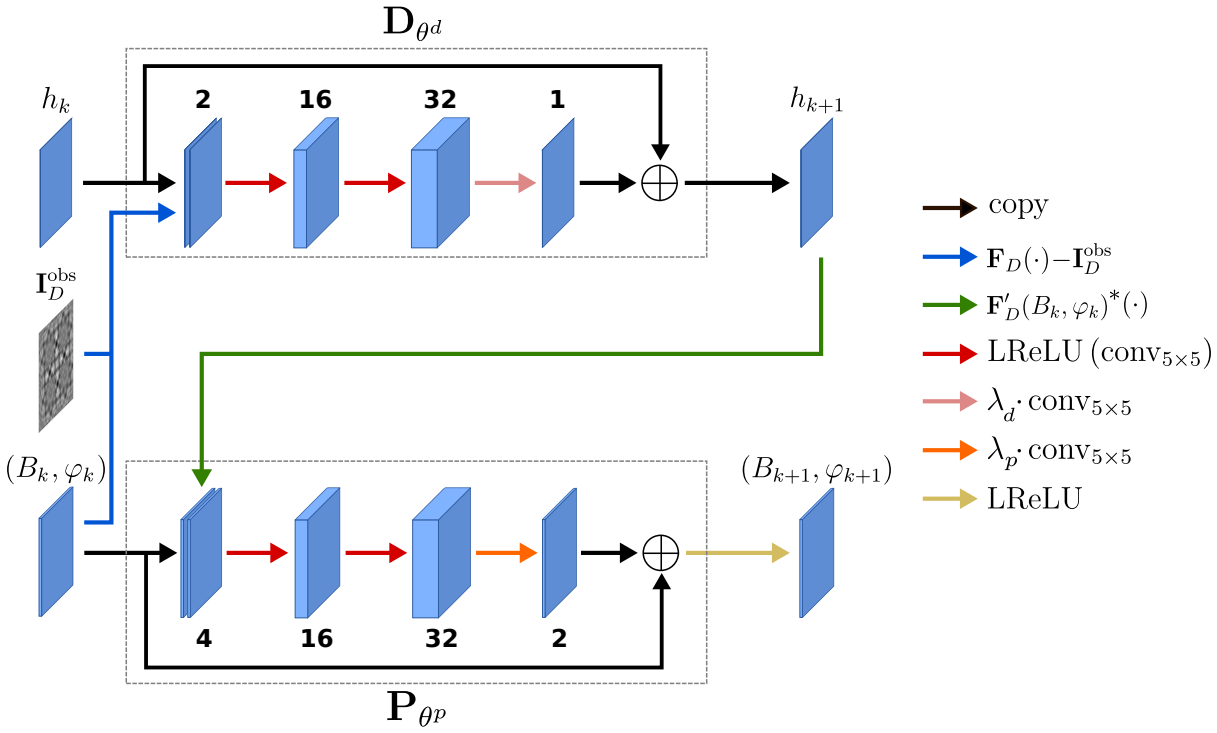


Figure 5.2.: Architecture of the network $\Gamma_\theta^{\text{DPD}}$, representing one iteration of the Deep Primal-Dual.

the current iterate with the output of this pipeline. After applying the adjoint of the Frechet derivative to the dual variable, this is concatenating with the current iterate of the primal variable (B_k, φ_k), and this is given as input to the primal network P_{θ^p} which consists of the same operations as the dual network, except that the ouptut is projected to the positive number by a LReLU activation.

This corresponds to defining each iteration as an operator

$$\begin{array}{rccc} \Gamma_\theta^{\text{DPD}} : & \mathcal{X} \times \mathcal{Y} & \to & \mathcal{X} \times \mathcal{Y} \\ & [h_k, (B_k, \varphi_k)] & \mapsto & [h_{k+1}, (B_{k+1}, \varphi_{k+1})] \end{array} \tag{5.9}$$

with $\theta = (\theta^d, \theta^p)$, which is given by

$$\Gamma_\theta^{\text{DPD}} \begin{bmatrix} h \\ (B, \varphi) \end{bmatrix} = \begin{bmatrix} \mathrm{D}_{\theta^d}\left(h, \mathbf{F}_D(B, \varphi) - \mathbf{I}_D^{\text{obs}}\right) \\ \mathrm{P}_{\theta^p}\left((B, \varphi), \mathbf{F}'_D(B, \varphi)^*\left[\mathrm{D}_{\theta^d}\left(h, \mathbf{F}_D(B, \varphi) - \mathbf{I}_D^{\text{obs}}\right)\right]\right) \end{bmatrix} \tag{5.10}$$

II.3 Deep Gauss-Newton (DGN)

The Deep Gauss-Newton method introduced in the section 4 is briefly recalled here. The IRGN update corresponding to Tikhonov regularization of the Newton steps can be written as:

$$(B_{k+1}, \varphi_{k+1}) = (B_k, \varphi_k) + H(B_k, \varphi_k)^{-1}\left\{\mathrm{F}'_D(B_k, \varphi_k)^*\left[\mathbf{I}_D^{\text{obs}} - \mathrm{F}_D(B_k, \varphi_k)\right] - \alpha_k(B_k, \varphi_k)\right\} \tag{5.11}$$

where α_k is the weighting parameter of the regularization and the operator H is defined as

$$H(B_k, \varphi_k) = \left[\mathrm{F}'_D(B_k, \varphi_k)^*\mathrm{F}'_D(B_k, \varphi_k) + \alpha_k \mathrm{Id}\right] \tag{5.12}$$

Tableau 5.1.: Parameters for the unrolling neural networks.

	DGD	DPD	DGN	DPGN
N (iterations)	10	10	10	10
Loss function	MSE	MSE	MSE	MSE
Training epochs	100	100	100	100
Learning rate	10^{-3}	5×10^{-4}	5×10^{-4}	5×10^{-4}
Batch size	10	10	10	10
Optimizer	ADAM	ADAM	ADAM	ADAM
Training time	19h	19h	21h	22h
Parameters	41×10^3	30×10^3	31×10^3	32×10^3

The main idea was to replace the part corresponding to the gradient of the Tikhonov regularization $\alpha_k(B_k, \varphi_k)$ by a network G_{θ^g} and to approximate $H(B, \varphi)^{-1}$ by another network H_{θ^h}. Following this, the IRGN update (5.11) can then be replaced by

$$(B_{k+1}, \varphi_{k+1}) = (B_k, \varphi_k) + \mathrm{H}_{\theta^h}\left[(B_k, \varphi_k), \tilde{H}(B_k, \varphi_k)\left\{\mathrm{F}'_D(B_k, \varphi_k)^*\left[\mathbf{I}_D^{\mathrm{obs}} - \mathrm{F}_D(B_k, \varphi_k)\right] + \mathrm{G}_{\theta^g}(B_k, \varphi_k)\right\}\right] \tag{5.13}$$

where $\tilde{H}(B_k, \varphi_k) = \mathrm{F}'_D(B_k, \varphi_k)^* \mathrm{F}'_D(B_k, \varphi_k)$ is an semi-definite positive operator. The update (5.13) defines the neural networks $\Gamma_\theta^{\mathrm{DGN}}$, where $\theta = (\theta^g, \theta^h)$ is the set of parameters. One iteration of the Deep Gauss-Newton method is displayed in Figure 5.3.

II.4 Deep Proximal Gauss-Newton (DPGN)

The iterations of the IRGN method (5.11) can be extended to the Proximal Gauss-Newton (PGN) method by adding a step, which corresponds to the application of a proximal operator scaled by the operator $H(B_k, \varphi_k)$ (5.12). If we unroll the PGN method, it means starting from the DGN architecture and adding a network J_{θ^j} which will act as a learned proximal operator (see Figure 5.4). Compared with DGN, the architecture of the network G_{θ^g} acting like a regularization term has been simplified, so that the number of parameters remains similar to DGN.

III Experiments

III.1 Implementation details

We compare the different unrolling structures with their counterparts conventional iterative methods: the gradient descent (II.3), the nonlinear primal-dual hybrid gradient (II.5) and the iteratively regularized Gauss-Newton. For GD-TV$^\epsilon$ and NL-PDHG we used 1 000 iterations, and for IRGN we used 100 Newton steps and 10 iterations for the CG. The regularization parameters for each methods have been optimized as described in III.1.2. For all unrolling approaches, we used $N = 10$ iterations, which, empirically, was enough for the NMSE to stagnate. As described in (III.2), for each unrolled method we use a single network $\mathbf{\Gamma}_\theta$ applied N times in a recurrent fashion to obtain the DNN $\mathbf{\Lambda}_\theta$, which is trained to perform end-to-end reconstruction. The unrolling methods were trained using the same hyperparameters 100 epochs with MSE loss, a batch size of 10, the ADAM optimizer, an initial learning rate of either 5×10^{-4} or 10^{-3} and a cosine annealing learning rate schedule (Loshchilov and Hutter 2016). The LReLU activation function parameter was set to the default $\alpha = 0.3$. Here, zero initialization, $(B_0, \varphi_0) = (0, 0)$, was used throughout. The details of the training are all summarized in Tab. 5.1. All unrolling methods were trained on the same dataset as described in .

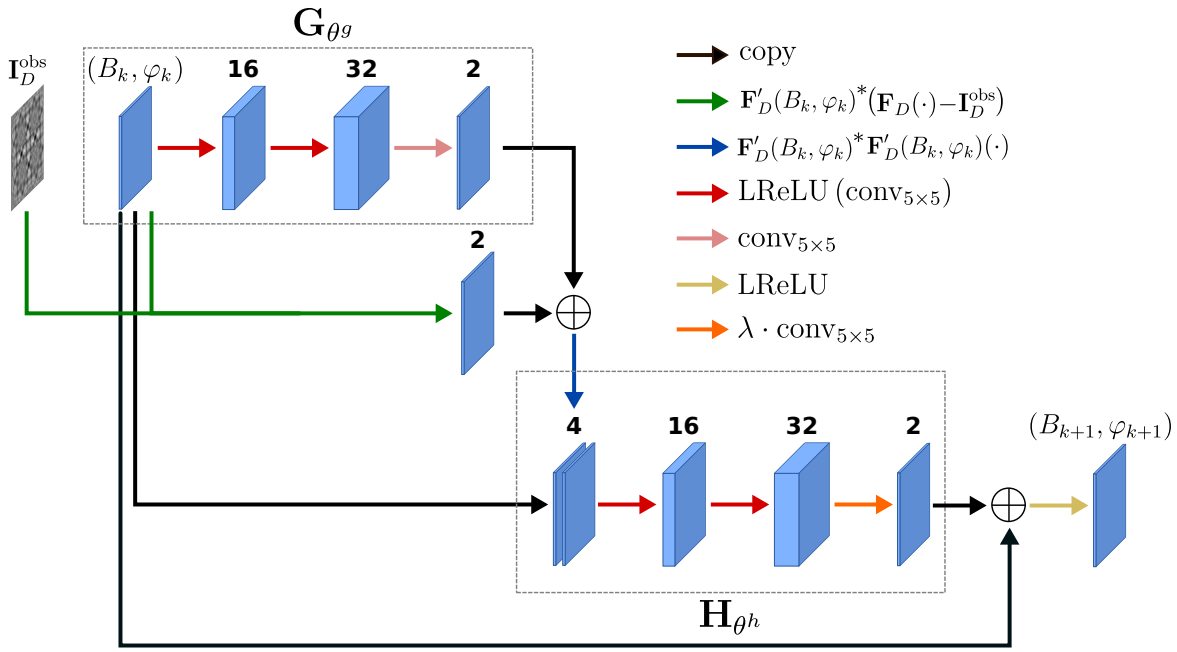


Figure 5.3.: Architecture of the network $\Gamma_\theta^{\text{DGN}}$, representing one iteration of the Deep Gauss-Newton.

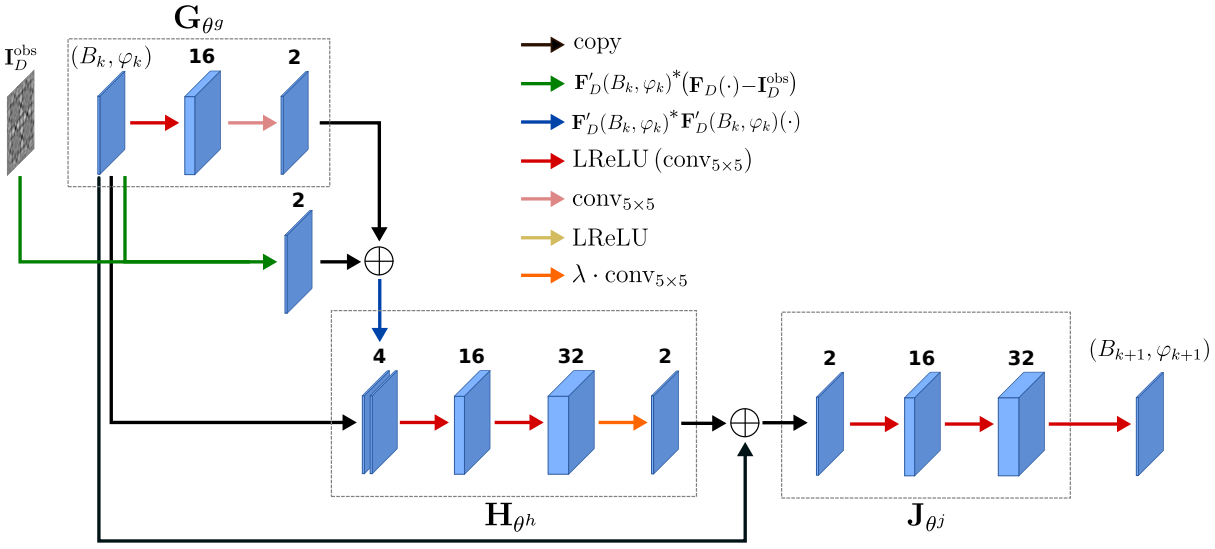


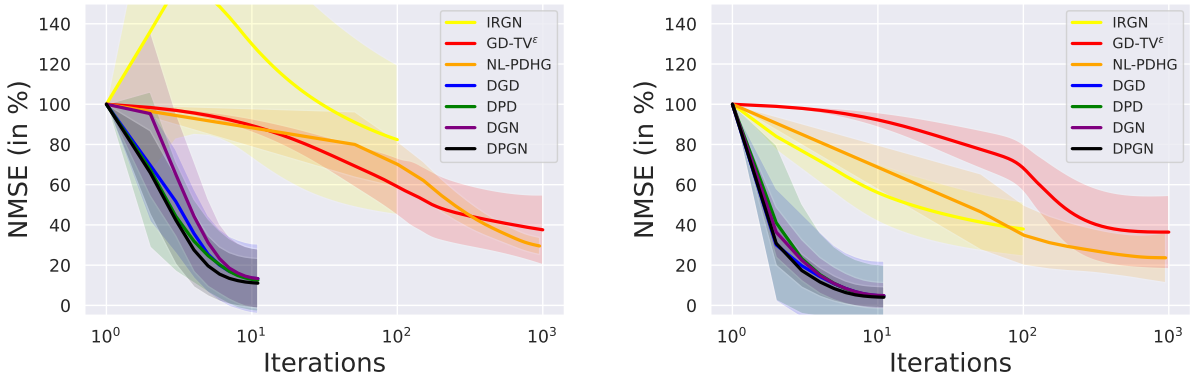
Figure 5.4.: Architecture of the network $\Gamma_\theta^{\text{DPGN}}$, representing one iteration of the Deep Proximal Gauss-Newton.

III.2 Simulated results

In order to quantify the quality of the reconstructions of each method on simulated data, we used the NMSE, the FRCM and the resolution criterion defined in III.3. The average and standard deviations of these metrics as well as the average computation time were computed on the test dataset containing 1 000 pairs, the results are summarized in Tab. 5.2.

	NMSE (in %)		FRCM (in %)		Resolution (in nm)		#Parameters	Time (in s)
Method	Absorption	Phase	Absorption	Phase	Absorption	Phase		
GD-TV$^\epsilon$	37.5 (17.4)	36.4 (18.2)	61.8 (12.2)	57.7 (13.2)	214 (101)	139 (78)	–	145
IRGN	85.5 (40.7)	39.3 (15.0)	71.2 (9.95)	68.1 (5.45)	238 (136)	154 (43)	–	116
NL-PDHG	29.2 (14.8)	23.6 (12.6)	58.4 (9.08)	50.7 (8.28)	146 (85.2)	99.7 (26.5)	–	147
DGD	13.2 (17.3)	4.74 (6.99)	37.6 (13.2)	23.8 (15.7)	82.2 (116)	64.3 (62.6)	41×10^3	**3.85**
DPD	12.5 (15.5)	4.48 (6.25)	39.2 (14.4)	24.3 (16.5)	107 (138)	75.5 (66.7)	31×10^3	4.63
DGN	12.1 (13.5)	4.61 (6.20)	35.7 (15.7)	23.0 (16.6)	76.8 (63.3)	63.0 (37.4)	31×10^3	5.88
DPGN	**11.0 (12.3)**	**4.05 (5.25)**	**31.5 (13.9)**	**19.8 (15.8)**	**74.0 (63.4)**	**59.1 (28.3)**	32×10^3	6.31

Tableau 5.2.: Comparison of different methods applied on the test dataset containing 1 000 images, according to different metrics.

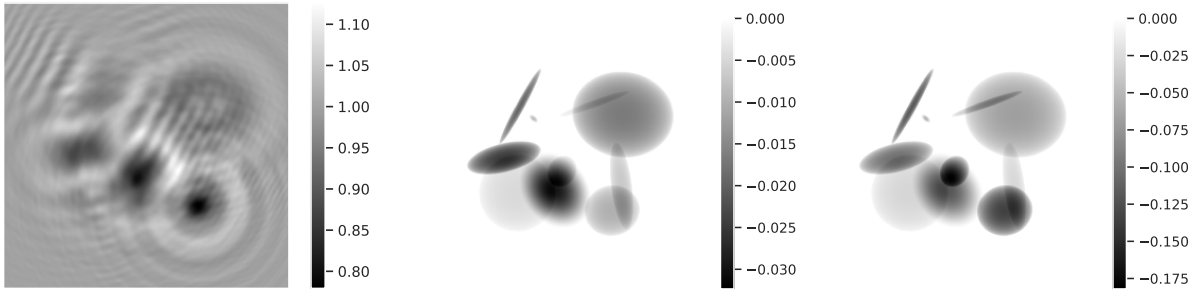


((a)) Absorption ((b)) Phase

Figure 5.5.: Evolution of average NMSE (%) of the unrolling approaches for 1 000 test images. The transparent areas correspond to the standard deviation.

The first result we can observe is that the deep learning-based methods outperform their classical iterative counterparts in terms of NMSE and resolution. Moreover, as the number of iterations for these approaches is fixed at $N = 10$, the reconstruction time is considerably faster. All unrolling approaches have similar reconstruction quality on average, but the Deep Gradient Descent has the worst NMSE among them, despite having the highest number of parameters. This is in line with our observation that we made with the MS-D Net, that more parameters don't necessarily mean better results. Here, we can see that the choice of the optimization scheme has an impact on the reconstruction quality. Unrolling a first order scheme such as the gradient descent leads to the smaller reconstruction time, but doesn't seem complex enough to achieve optimal results. Learning in both model parameter space and data space enables DPD to improve the normalized error, but seems to have poorer quality in the frequency domain. But overall, it's the unrolling methods based on Gauss-Newton type scheme that provide the best reconstructions, both in terms of error and frequency/resolution. They have a slightly longer

reconstruction time, as they require more operator evaluations within the network, but are still very fast. The graphs in Figure 5.5 show the evolution of the normalized error of absorption and phase. Unrolled approaches have similar behaviour, where the first iteration seems to do most of the work, and subsequent iterations refine it. We can see that after 5-6 iterations, deep learning approaches already have a smaller error than conventional approaches.



((a)) Simulated intensity. ((b)) Absorption ground truth. ((c)) Phase ground truth.

Figure 5.6.: Example of one simulated pair.

For a qualitative evaluation, we display in Figures 5.7 and 5.8 the reconstructions obtained using the simulated pair 5.6. When it comes to reconstructing the absorption, the conventional methods suffer the most: edges are hardly recovered cleanly, and there are still quite a few artifacts present. While the unrolled methods recover the sample well. We observe that DPGN is slightly cleaner than the others, and this can be seen in the metrics. For the phase map (Fig. 5.8), iterative methods are doing better, altough we still have some low frequency artefacts for the IRGN. And here, the unrolled methods have an almost perfect reconstruction.

III.3 Experimental results

Finally, the proposed algorithm was tested on real single-distance data acquired at beamline NanoMAX at the MAX IV synchrotron. In Fig. 5.9, we can see that, once again, classical iterative methods have trouble in recovering the absorption map, except for the NL-PDHG method, whose regularization seems better suited. The unrolling methods show very good reconstruction, with some artifacts present for DGD. The DPGN method gives the best visual reconstruction, and better approximates the mean value inside the object. In terms of phase recovery, Figure 5.10 shows that the unrolling methods give very similar results, and only the classical NL-PDHG method manages to do the same. However, it has poorer resolution, which may be due to the fact that the edges are a little smoothed out by TV regularization.

III.4 Effect of the number of iterations

For the above experiments, we set the number of *training* iterations of the unrolling approaches to $N = 10$. Here, we give some empirical results to justify this choice and the effect it can have on reconstruction quality on simulated data. In the following experiments, we focus on the DPGN method. To see the effect of this hyperparameter, we trained the DPGN network with different fixed numbers of iterations $N \in \{5, 10, 20\}$. For the three newtorks, we use the same parameters as described in Tab. 5.1. Since the $\Gamma_\theta^{\text{DPGN}}$ network is the same at each iteration, the total number of parameters learned is the same in all cases. Graphs of the evolution of the normalized error as a function of the number of iterations are shown in Fig. 5.11. The curves show the a similar behavior whatever the number of iterations over which the network has been

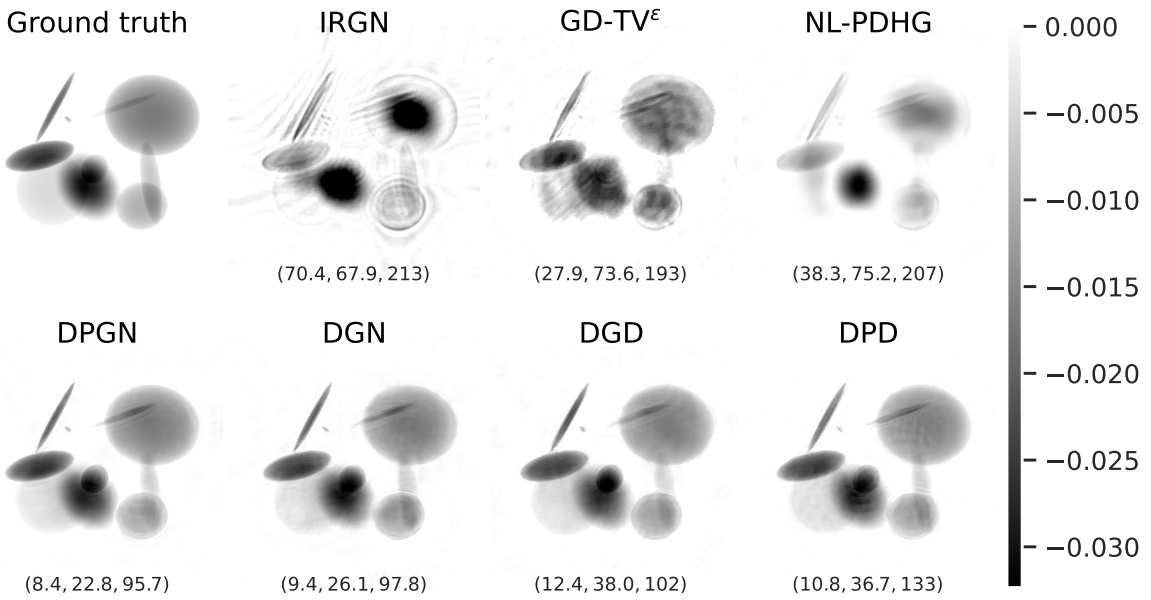


Figure 5.7.: Absorption reconstructions from simulated data. Reconstruction quality is given as (NMSE (%), FRCM (%), resolution (nm)).

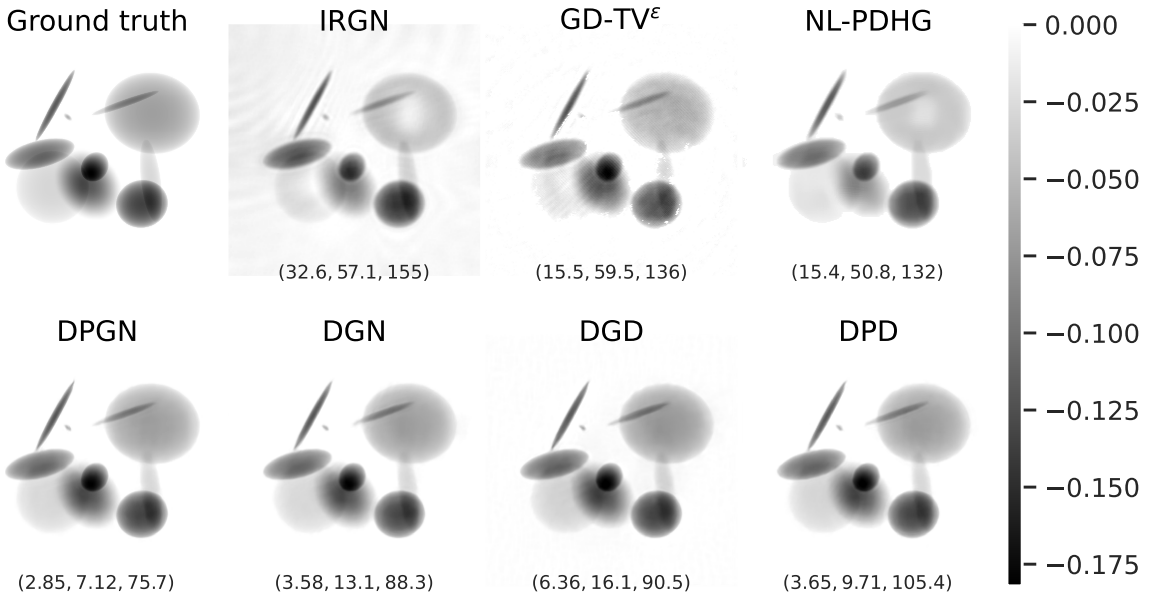


Figure 5.8.: Phase reconstructions from simulated data. Reconstruction quality is given as (NMSE (%), FRCM (%), resolution (nm)).

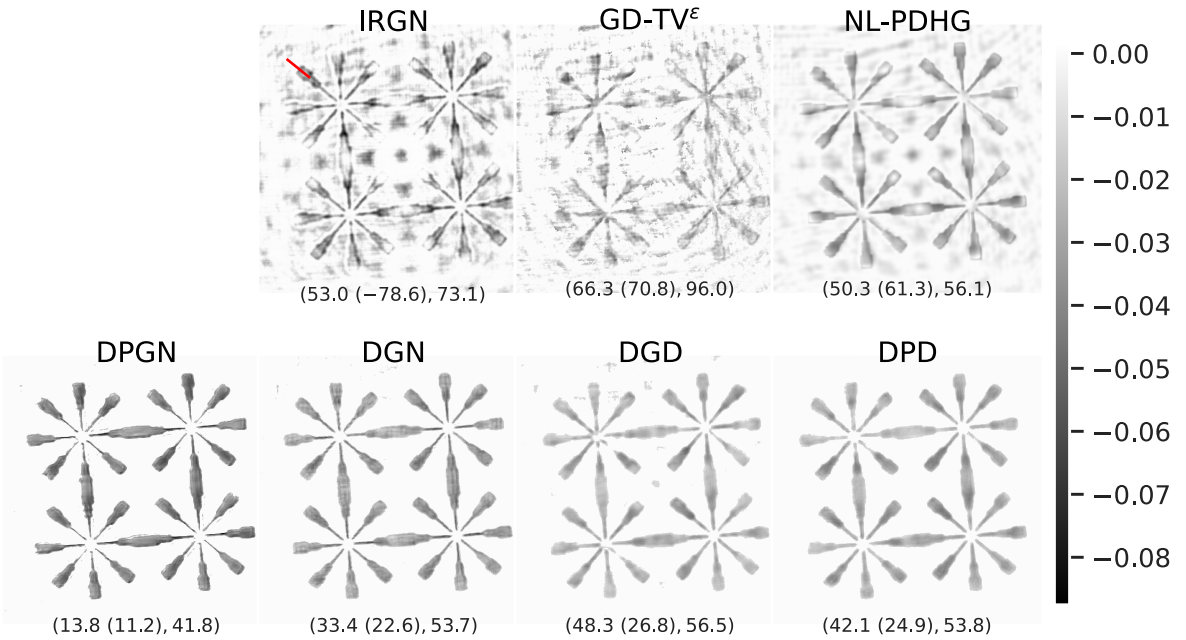


Figure 5.9.: Absorption reconstructions for experimental data. The profiles along the red line were measured to estimate the resolution. Values correspond to (NE (%) (RSD (%)), resolution (nm)).

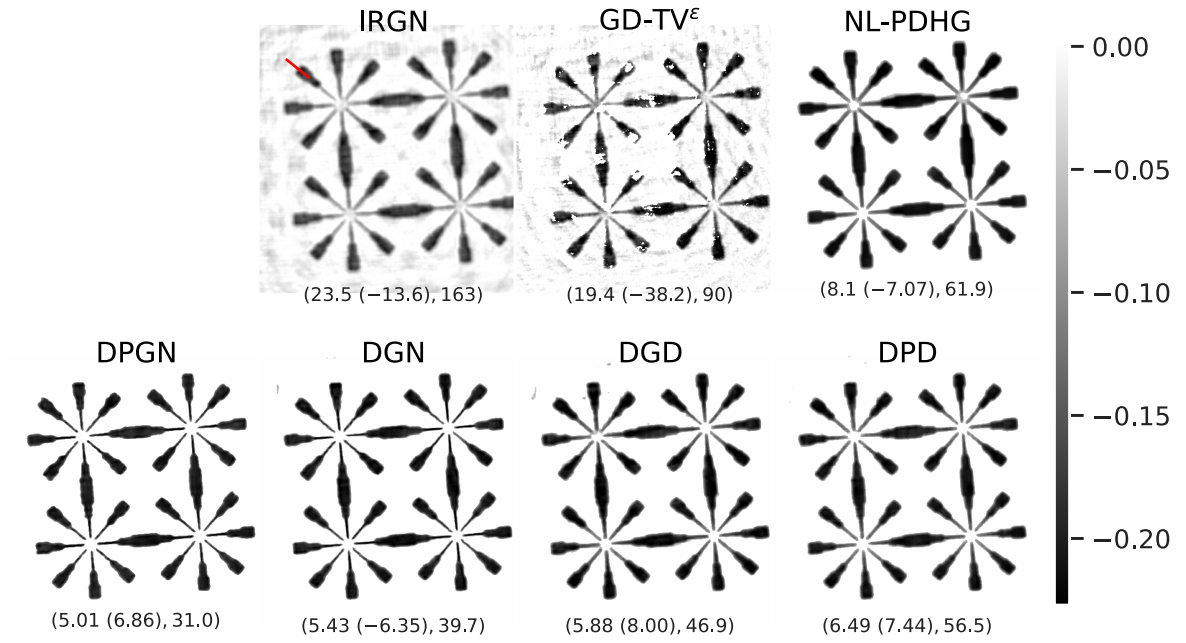
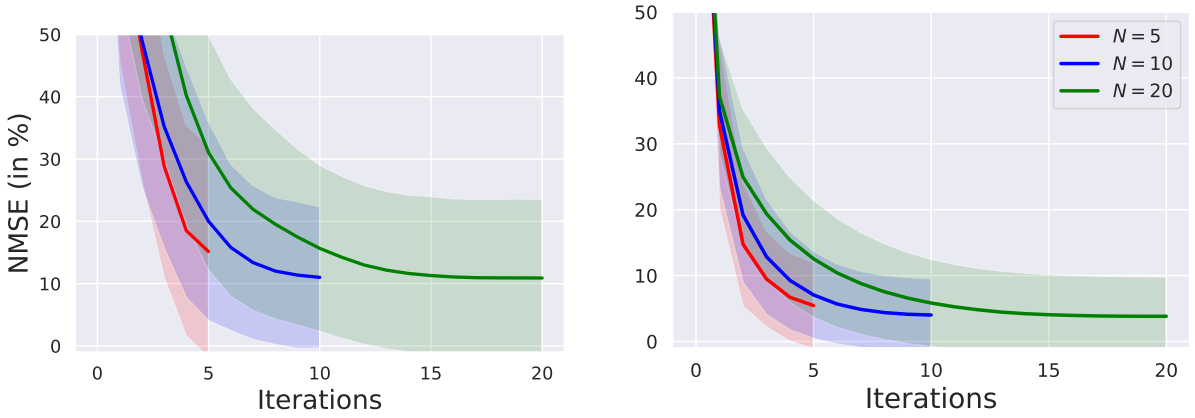


Figure 5.10.: Phase reconstructions for experimental data. The profiles along the red line were measured to estimate the resolution. Values correspond to (NE (%) (RSD (%)), resolution (nm)).

	NMSE (in %)		FRCM (in %)		Resolution (in nm)		Training time (h)
N	Absorption	Phase	Absorption	Phase	Absorption	Phase	
5	15.3 (15.6)	5.83 (7.51)	39.1 (13.7)	26.0 (15.9)	94.7 (75.3)	64.3 (59.7)	**12h**
10	11.0 (12.3)	4.05 (5.25)	31.5 (13.9)	**19.8 (15.8)**	**74.0 (63.4)**	**59.1 (28.3)**	22h
20	**10.7 (11.2)**	**4.01 (4.88)**	**31.2 (12.1)**	22.3 (14.9)	76.1 (61.6)	62.5 (52.3)	40h

Tableau 5.3.: Comparison of different methods applied on the test dataset containing 1 000 images, according to different metrics.

trained, with a large decrease of the NMSE. Although the networks used here have the same architecture, they seem to behave differently in each case. For $N = 5$, the network doesn't seem to reach convergence, whereas for $N = 10$ and $N = 20$ this seems to be the case. And for $N = 20$, the last 5 iterations have little impact on the reconstruction. For each network, we computed the averages of the metrics, which are summarized in Tab. 5.3. We can see that the training time



((a)) Absorption ((b)) Phase

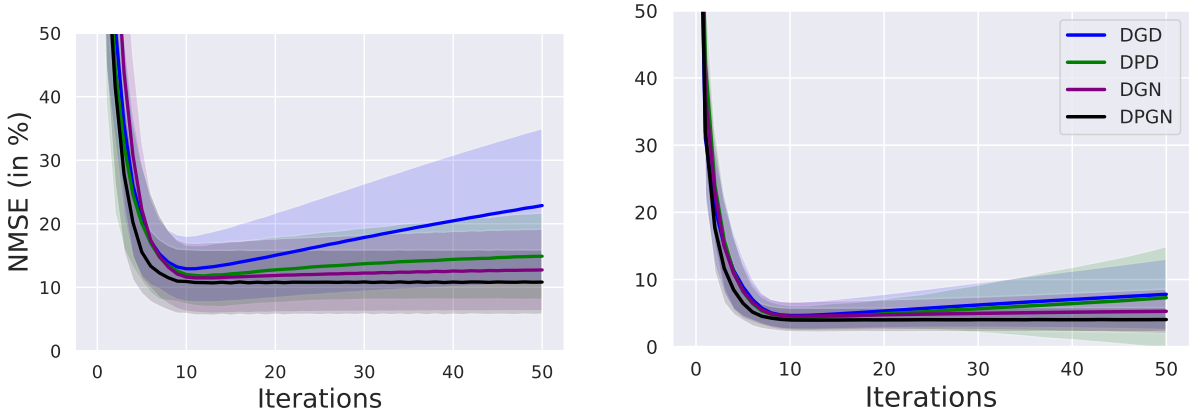
Figure 5.11.: Evolution of average NMSE (%) of the DPGN for 1 000 test images when trained using different number of iterations. The transparent areas correspond to the standard deviation.

required is proportional to the fixed number of iterations. This is because the iteration defined by $\Gamma_\theta^{\text{DPGN}}$ is copied N times to obtain the deep network, which is train in an end-to-end manner. When $N = 5$ we can see that the training time is relatively short, but that this does not seem sufficient to obtain optimum reconstruction quality. On the contrary, for $N = 20$ the training time is very long, and we don't necessarily improve on the results obtained with $N = 10$, so this seems to be a good compromise to reach convergence with a small computation time.

III.5 Running additional steps at inference

Conventional deep unrolled optimization networks necessitate the selection of a predetermined quantity of training iterations N during the training process. Straying from this fixed iteration count during inference generally results in a substantial decline in quality. The graphs shown in Fig. 5.12 represent the evolution of the NMSE of the unrolling networks trained in III.1 with $N = 10$ and tested during inference with $N\text{inference} = 50$ iterations. For both absorption and phase, the minimum error is reached for 10 iterations, which not surprising, since the networks

have been trained with this number of iteration. But if we continue to iterate afterwards, we see that the quality of reconstructions decreases for all methods except DPGN. Actually, the errors for DPGN do not increase, but do not decrease either, as if the trained network had reached convergence after the given number of iterations.



((a)) Absorption

((b)) Phase

Figure 5.12.: Evolution of average NMSE (%) of the unrolling approaches for 1 000 test images when running for 50 iterations. The transparent areas correspond to the standard deviation.

IV Discussion and perspectives

There are many works about unrolling for linear inverse problems and we showed in this section that unrolling methods can be useful for nonlinear inverse problems. Different optimization schemes has been described and we saw how they can be unrolled. We illustrated how the choice of the scheme we start with had an impact on the quality of the reconstructions. We have seen that, with approximatively the same number of parameters and the same training strategy, these approaches produce similar results, but some deliver better quality both visually and quantitatively. Network leveraging second order information, even with an approximate Hessian, like the DGN or DPGN give better reconstruction results and was the most stable. All unrolled approaches were more efficient than their classical counterparts both on simulated and experimental data. Coupling these approaches with tomography could give us more information on how they compare. In order to have a better understanding of the networks, we could do what is called in some papers an *ablation study* to suppress some parts of the network to understand its specific properties. However, an important choice we need to make is the fixed number of iterations used to unroll the algorithms. We've seen that too few iterations give poorer results, but too many don't improve quality and require a lot of training time. One way of optimizing this number N would be to use a regularized cost function with exponential weights to optimize the network parameters (Vishnevskiy, Walheim, and Kozerke 2020)

$$\min_{\theta} \mathbb{E}_{(B^*,\varphi^*)} \sum_{k=1}^{N} \mathrm{e}^{-\tau(K-k)} \|(B_k, \varphi_k) - (B^*, \varphi^*)\|_2^2 \tag{5.14}$$

where (B^*, φ^*) is the ground truth, $(B_k, \varphi_k) = (\Gamma_\theta \circ \cdots \circ \Gamma_\theta)\,(B_0, \varphi_0)$ the k-th iterate and τ is a parameter that controls layers penalization during iterations. Then we could choose the number N from which the error hardly evolves. This type of loss function where some weigts are added during the iterations may be useful to escape local minima for complex optimization problems (Kannara Mom, Lesaint, et al. 2023). Another method would be the one proposed by (Gilton, Ongie, and Willett 2021) where the authors proposed to solve linear inverse problems based on deep equilibrium models corresponding to an infinite number of iterations. In this case, the networks will be trained to converge towards a fixed point. An extension to the nonlinear case could be interesting for the DPGN approach, where the J network could be trained in such a way. The convergence properties of the iterative algorithms presented are not clear at present. We could try to show that the network J has properties similar to a proximal operator as in (Pesquet et al. 2021), or if not, we could try to enforce these properties. The difficulty here is that the phase retrieval problem is nonlinear and the step size depends on the current iterate. Moreover, the operator involved in (4.5) is not a classical proximal operator and depends on the operator $H(B, \varphi)$. Also, we used the same networks for the phase and absorption. It could be interesting to investigate more complex architecture where the different spatial regularity of the network can be exploited.

Material decomposition for spectral computed tomography using deep learning

The unrolling approach proposed for simultaneously retrieve the absorption and the phase can be generalized to other inverse problems, as long as the forward operator is differentiable. In this section, we investigate the problem of material decomposition in the projection space for spectral computed tomography. We propose to solve this nonlinear inverse problem by unrolling a gradient descent scheme (DGD). We show that the proposed approach improves the material decomposition compared to both the classical iterative variational method and a U-Net architecture which does not take into account the physics of the forward problem. The calculation time is also reduced fir the simultaneous retrieval of the asborption and phase.

Introduction

In tomography with conventional X-rays, the object of interest is modeled by its linear attenuation coefficient. This is reconstructed from a sequence of X-rays acquired from different viewing angles. In spectral tomography, the measurements are energy-resolved, enabling the linear attenuation coefficient to be decomposed on a material basis (R. E. Alvarez and Macovski 1976) (typically water/bone in medical imaging). The two stages of the problem – tomographic reconstruction and basis material decomposition – can be carried out jointly or consecutively. Joint methods, called *one-step* in the literature, directly invert the global measurement formation model (Foygel Barber et al. 2016; Mechlem et al. 2017; Cai et al. 2013; Long and Fessler 2014). Although theoretically optimal in terms of compromise between noise and spatial resolution of reconstructions, these methods are extremely costly in terms of computing time. Sequential methods (known as *two-step*) separate the material decomposition problem from the tomographic reconstruction problem. The decomposition step can be performed in image space (Ding et al. 2018) (i.e. after the tomographic reconstruction step) or in projection space (Ducros et al. 2017) (before the reconstruction). The basis material decomposition is a non-linear and ill-posed problem (see below). This chapter proposes a method for material decomposition in projection space. Many methods exist: early methods proposed to approximate the inverse operator by a simpler parametric model (e.g. polynomial (R. E. Alvarez and Macovski 1976; Robert E. Alvarez 2011)). Variational approaches are based on the estimation of a poissonian maximum likelihood (Schlomka et al. 2008) or a Gaussian one (Ducros et al. 2017). In recent years, many approaches based on deep learning (DL) techniques have been proposed to solve complex inverse problems. More recently, methods that combine physical models with DL techniques have been developed. These methods include model physics at different levels (Abascal, Ducros, and F. Peyrin 2018; Eguizabal, O. Öktem, and Persson 2022).

Here, we propose a method that relies on unrolling a gradient descent scheme to improve the decomposition results obtained with classical variational methods. The direct problem is presented, followed by a detailed description of the proposed deep learning inversion method. Simulation results are presented and discussed.

Material and methods

Measurement formation model

The linear attenuation coefficient μ of an object depends on the energy E of the incident radiation and can be written as a linear combination of the attenuation coefficients f_m of a small number M (typically $M = 2$) of materials (R. E. Alvarez 2010)

$$\mu(\mathbf{x}, E) = \sum_{m=1}^{M} f_m(E)\tau_m(\mathbf{x}). \tag{5.15}$$

When the object is subjected to polychromatic radiation, the measurement at a pixel $\mathbf{u}$ is modelled by the Beer-Lambert law

$$\bar{s}_b(\mathbf{u}) = \int_0^\infty S_b^{\text{eff}}(E) \exp\left(-\sum_{m=1}^{M} f_m(E)a_m(\mathbf{u})\right) \mathrm{d}E, \tag{5.16}$$

where $S_b^{\text{eff}}(E)$ is the *effective spectrum* (i.e. taking into account the detector response) in the b-th energy channel of the detector and $a_m(\mathbf{u})$ is the projected thickness of the material maps τ_m, i.e. the integral of the functions $\tau_m(\mathbf{x})$ along the X-ray (from the source to the pixel $\mathbf{u}$ of the detector).

$$a_m(\mathbf{u}) = \int_{\mathcal{L}(\mathbf{u})} \tau_m(\mathbf{x}) \, \mathrm{d}\mathbf{x}. \tag{5.17}$$

The quantity $\bar{s}_b(\mathbf{u})$ is actually the number of average photons. The actual number of photons s_b follows a Poisson distribution

$$s_b(\mathbf{u}) \sim \mathcal{P}\left(\bar{s}_b(\mathbf{u})\right). \tag{5.18}$$

By analogy with conventional tomography, we have chosen to work with attenuation

$$\hat{s}_b(\mathbf{u}) = -\log \frac{s_b(\mathbf{u})}{s_b^0(\mathbf{u})} \tag{5.19}$$

where $s_b^0(\mathbf{u}) = \int_0^\infty S_b^{\text{eff}}(E) \, \mathrm{d}E$ is the total number of photons measured in the absence of an object in the incident beam.

In practice, the digital detector provides a discretized image $\mathbf{u}_{x,y}$ of $P = P_x \times P_y$ pixels. A measurement is therefore a set of B images $\hat{\mathbf{s}} = (\hat{\mathbf{s}}_1, \ldots, \hat{\mathbf{s}}_B) \in \mathbb{R}^{P \times B}$ where $\hat{\mathbf{s}}_b = \left\{\hat{\mathbf{s}}_b(\mathbf{u}_{x,y})\right\}_{1 \le x \le P_x, 1 \le y \le P_y}$ is an attenuation image. The quantities $\mathbf{a}_m$ and $\hat{\mathbf{s}}_b$ are both two-dimensional images. To distinguish them, the term *projection* will be reserved for $\mathbf{a}_m$ images, which denotes the projected lengths of material maps τ_m (see (5.17)). The decomposition in the projection domain consists in finding the M projections $\mathbf{a} = (\mathbf{a}_1, \ldots, \mathbf{a}_M) \in \mathbb{R}^{P \times M}$ from B attenuation images $\hat{\mathbf{s}}$. In dual-energy CT, $B = 2$: attenuation is measured at two different energy levels. Denoting $\mathcal{F}$ the direct, non-linear, discretized model of the data, we have

$$\begin{array}{rccc} \mathcal{F}: & \mathbb{R}^{P \times M} & \to & \mathbb{R}^{P \times B} \\ & \mathbf{a} & \mapsto & \hat{\mathbf{s}} \end{array} \tag{5.20}$$

The basis material decomposition consists in inverting the operator $\mathcal{F}$. Estimating the inverse of the operator $\mathcal{F}$ can be done using a variational approach, which minimizes an energy defined as the sum of a data fidelity term and a regularization $\mathcal{R}$. This amounts to minimizing, with

respect to the variable $\mathbf{a}$, the following functional

$$F(\mathbf{a}, \hat{\mathbf{s}}) := \frac{1}{2}\|\mathcal{F}(\mathbf{a}) - \hat{\mathbf{s}}\|_2^2 + \mathcal{R}(\mathbf{a}) \tag{5.21}$$

Although photon noise is poissonian (Equation 5.18), typical flux levels in spectral tomography make it legitimate to use the L^2 norm of the residual as a data fidelity criterion. Iterative methods such as the regularized weighted least squares Gauss–Newton algorithm have been proposed in (Ducros et al. 2017), but the choice of regularization $\mathcal{R}$ and hyperparameters (specific to a noise level) is tricky. We propose here a learned iterative scheme, the Deep Gradient Descent (DGD) algorithm, obtained by unrolling a gradient descent. Using training data, DL-based methods learn an optimal decomposition

$$\Lambda_\theta : \mathbb{R}^{P\times B} \to \mathbb{R}^{P\times M}$$

associated with a neural network parameterized by a set of learnable parameters θ to be optimized. Among supervised learning methods, *unrolling* methods rely on an iterative scheme to minimize the data fidelity term in Equation 5.21. The iterative scheme is truncated after a fixed number of iterations K and some updates are modified by adding neural networks. The network corresponding to Λ_θ thus combines neural networks with the direct and adjoint operators associated with the physics of the problem (Arridge et al. 2019; J. Adler et al. 2017; Monga, Y. Li, and Eldar 2021). Recently, we have shown that this type of approach can be used for a nonlinear inverse problem using the derivative of the operator and its adjoint (K. Mom, M. Langer, and B. Sixou 2022).

We consider here the unrolling method obtained by adapting a gradient descent. If $\mathcal{R}$ is sufficiently smooth, the gradient descent scheme for solving the problem (5.21) can be written, at iteration k

$$\mathbf{a}^{(k+1)} = \mathbf{a}^{(k)} - \lambda_k \nabla F(\mathbf{a}^{(k)}, \hat{\mathbf{s}}). \tag{5.22}$$

If we stop the descent algorithm after K iterations, we obtain

$$\mathbf{a}^{(K)} = \left(G_{\lambda_K} \circ \cdots \circ G_{\lambda_1}\right)(\mathbf{a}^{(0)}, \hat{\mathbf{s}}), \tag{5.23}$$

where $G_{\lambda_k}(\cdot, \hat{\mathbf{s}}) := \mathrm{Id} - \lambda_k \nabla F(\cdot, \hat{\mathbf{s}})$ for $1 \le k \le K$. In the unrolling approach, the idea is to replace the functions G_{λ_k}, representing an iteration of the schema (5.22), by a convolutional neural network $\Gamma^\theta_{\lambda_k}$. The networks $\left\{\Gamma^\theta_{\lambda_k}\right\}_{k=1}^K$ share a set of parameters θ, which will be optimized by training. The descent step λ_k is also optimized, but is specific to each iteration. A network $\Gamma^\theta_{\lambda_k}$ takes as input the current iterate $\mathbf{a}^{(k)}$ and the attenuations measured $\hat{\mathbf{s}}$. From these, the gradient of the data fidelity term is calculated

$$\nabla\left(\frac{1}{2}\|\mathcal{F}(\mathbf{a}^{(k)}) - \hat{\mathbf{s}}\|_2^2\right) = \mathbf{J}^{(k)\top}\left[\mathcal{F}(\mathbf{a}^{(k)}) - \hat{\mathbf{s}}\right] \tag{5.24}$$

where $\mathbf{J}^{(k)}$ represents the Jacobian of the operator $\mathcal{F}$ at the point $\mathbf{a}^{(k)}$. The architecture of the network $\Gamma^\theta_{\lambda_k}$ is shown in Figure 5.13. The architecture is inspired by (Hauptmann et al. 2018), and has been adapted to the decomposition problem. Unlike (Hauptmann et al. 2018), the filters are of size 3×3 and the activation functions used are Leaky ReLU, defined by

$$\mathrm{LReLU}_\alpha(x) = \max(x, 0) + \alpha \min(x, 0).$$

In addition, the steps $\lambda_k \in \mathbb{R}^M$ are different at each iteration, with a descent step for each channel.

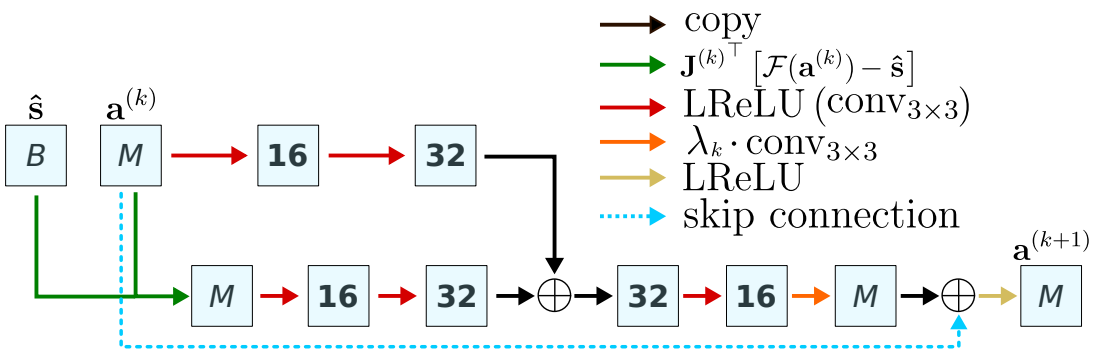


Figure 5.13.: Architecture of the network $\Gamma^{\theta}_{\lambda_k}$, which represents an iteration of the Deep Gradient Descent.

This network structure, similar to a gradient descent algorithm, is called Deep Gradient Descent (DGD). We thus define the convolutional neural network

$$\Lambda^{\theta}_{\lambda} = \Gamma^{\theta}_{\lambda_K} \circ \cdots \circ \Gamma^{\theta}_{\lambda_1}$$

representing K iterations and $(\theta, \lambda = \{\lambda_k\}_{k=1}^{K})$ represent the parameters of this network.

Experimental protocol

Data generation

The dataset used is based on 1,600 three-dimensional digital phantoms. Each phantom comprises between 4 and 7 ellipsoids, themselves made of water or bone (the two basic materials used for decomposition). Each phantom is thus made up of a pair (τ_1, τ_2) of 3D material maps (see §IV). The size, orientation and position of the ellipsoids are chosen randomly within limits that ensure the phantom is entirely contained within a 200×200 mm^3 bounding box, and that variability is sufficient. Each phantom (τ_1, τ_2) is then projected from 10 different viewing angles to give reference projections $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2)$. Then the full spectral model described in §IV is applied with two different effective spectrum S_1^{eff} and S_2^{eff} corresponding to X-ray tube voltages of 80 kVp and 120 kVp respectively. The resulting images $\hat{\mathbf{s}} = (\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$ constitute the input image pairs for the decomposition algorithms. In total, the dataset is thus made up of 16,000 (1600 phantoms taken from 10 different angles of view) image pairs $\hat{\mathbf{s}}^n$ $(1 \leq n \leq 16000)$. Each pair is associated with the corresponding reference projection pair $\mathbf{a}^n$. Figure 5.14 shows an example pair from the dataset. The dataset is then split into 14,000 pairs for training the newtwork, 1,000 pairs for validation and 1,000 pairs for testing (all from phantoms not used for training).

Four datasets $\mathcal{T}(\varepsilon) = \left\{\left(\mathbf{a}^n, \hat{\mathbf{s}}^n\right), n \leq 16000\right\}$ were generated, corresponding to distinct noise levels. By convention, $\mathcal{T}(0)$ is a noise-free dataset: the equation (5.18) is not applied. The values $\varepsilon = 10^4$, 10^5 and 10^6 correspond to incident photon numbers (the total number of photons emitted by the emission spectrum S_1^{eff} and S_2^{eff}), which determine the noise level in the images $\hat{\mathbf{s}} = (\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2)$. The higher the number of incident photons ε, the less noise there is in the data.

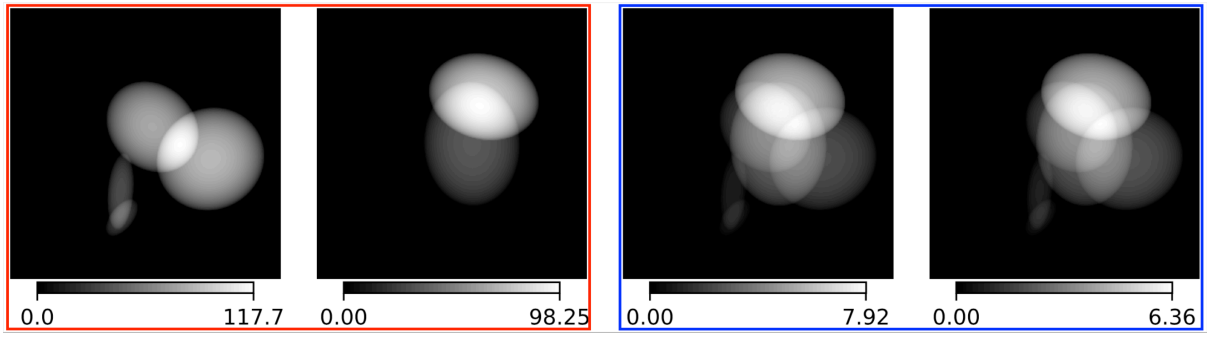


Figure 5.14.: An example of pair $(\mathbf{a}, \hat{\mathbf{s}})$ of the dataset. *Red box:* Water (*left*) and bone (*right*) reference projection. *Blue box:* Low-energy (*left*) and high-energy (*right*) measurements.

Training and implementation

Our training strategy consists in using a regularized cost function with exponential weights to optimize the network parameters (Vishnevskiy, Walheim, and Kozerke 2020)

$$\min_{\theta} \mathbb{E}_{(\hat{\mathbf{s}},\mathbf{a}^*)} \sum_{k=1}^{K} \mathrm{e}^{-\tau(K-k)} \|\mathbf{a}^{(k)} - \mathbf{a}^*\|_2^2 \tag{5.25}$$

where K denotes the number of iterations of the unrolled network, $\mathbf{a}^*$ is the ground truth, $\mathbf{a}^{(k)} = \left(\Gamma^{\theta}_{\lambda_k} \circ \cdots \circ \Gamma^{\theta}_{\lambda_1}\right)(\mathbf{a}^{(0)}, \hat{\mathbf{s}})$, and τ is a parameter that controls layer penalization along iterations. Penalizing the cost function therefore forces the outputs of intermediate iterations to be close to the ground truth, which has the effect of changing the network function. When $\tau \to +\infty$, only the network output $\mathbf{a}^{(K)}$ is optimized, which improves the fit on the training data.

The number of unrolled iterations is set to $K = 10$: for each batch, the operator $\mathcal{F}$ and the Jacobian $\mathbf{J}$ are evaluated ten times. The choice of parameter τ was set at 10^{-1}, which corresponds to the best validation error for $\tau \in \{10^{-2}, 10^{-1}, 1, 10\}$. The total number of parameters learned by the network $\Lambda^{\theta}_{\lambda}$ is about 28×10^3. The network is trained over 100 epochs, with a batch size fixed at 10, an Adam optimizer, a learning step initialized at 5×10^{-4} with a cosine annealing learning rate schedule (Loshchilov and Hutter 2016). The value of the parameter α of the LReLU activation function was left at 0.3 by default.

We compared the DGD algorithm with a regularized Gauss-Newton method (Ducros et al. 2017), with first- and second-order Tikhonov regularization for bone and water, respectively, and a positivity constraint. The regularization parameters were chosen following the procedure described in (Ducros et al. 2017), which gave $\alpha = 10^{-1.5}$ and $\beta_{\text{water}} = \beta_{\text{bone}} = 1$. Another network, the U-Net, was trained to decompose directly from input attenuation measurements. This method does not include model physics. The U-Net architecture used corresponds to the one describes in Figure 1.19. It has around 30×10^6 parameters, and has been trained with 100 epochs with a batch size of 10. The cost function is the squared error. Training lasted around 20 h and 15 h for DGD and U-Net, respectively. The results presented here for DGD are obtained considering zero initializations $\mathbf{a}^{(0)} = 0_{\mathbb{R}^{P \times M}}$. The networks were trained on each of the databases, making a total of eight trained networks.

In order to evaluate the decomposition quality obtained by the different methods, we use two

Tableau 5.4.: Average (standard deviation) of NMSE and SSIM metrics for test databases at different noise levels.

	$\mathcal{T}(0)$		$\mathcal{T}(10^6)$		$\mathcal{T}(10^5)$		$\mathcal{T}(10^4)$	
	Water	Bone	Water	Bone	Water	Bone	Water	Bone
				NMSE				
U-Net	0.032 (0.073)	0.016 (0.018)	0.037 (0.062)	0.017 (0.016)	0.056 (0.111)	0.019 (0.019)	0.064 (0.102)	0.029 (0.020)
DGD	**0.027 (0.076)**	**0.014 (0.021)**	**0.036 (0.127)**	**0.015 (0.021)**	**0.034 (0.067)**	**0.015 (0.018)**	**0.053 (0.091)**	**0.019 (0.021)**
				SSIM				
U-Net	0.367 (0.067)	0.592 (0.064)	0.331 (0.071)	0.552 (0.044)	0.292 (0.067)	0.424 (0.059)	0.239 (0.071)	0.319 (0.062)
DGD	**0.988 (0.008)**	**0.982 (0.011)**	**0.986 (0.009)**	**0.974 (0.013)**	**0.985 (0.010)**	**0.973 (0.017)**	**0.965 (0.017)**	**0.953 (0.028)**

metrics: The normalized mean squared error (NMSE) defined by:

$$\mathrm{NMSE}(\mathbf{a}, \mathbf{a}^*) = \frac{\|\mathbf{a} - \mathbf{a}^*\|_2}{\|\mathbf{a}^*\|_2}$$

The lower the NMSE, the better the decomposition. In addition, we used the Structural Similarity Index Measure (SSIM), which can be defined as:

$$\mathrm{SSIM}(\mathbf{a}, \mathbf{a}^*) = \frac{(2m_{\mathbf{a}}m_{\mathbf{a}^*} + C_1)(2\sigma_{\mathbf{a}\mathbf{a}^*} + C_2)}{(m_{\mathbf{a}}^2 m_{\mathbf{a}^*}^2 + C_1)(\sigma_{\mathbf{a}}^2\sigma_{\mathbf{a}^*}^2 + C_2)}$$

where $m_{\mathbf{a}}$ and σ_a represent the mean and the standard deviation of $\mathbf{a}$ pixels, respectively, and $\sigma_{\mathbf{a}\mathbf{a}^*}$ is the covariance between $\mathbf{a}$ and $\mathbf{a}^*$ and $C_1 = 10^{-4}, C_2 = 9 \times 10^{-4}$ are constants. A higher SSIM value indicates a higher similarity between the images, thus a better reconstruction.

Results

The results obtained are summarized in Table 5.4. When noise is high, decomposition quality deteriorates. In all cases, the performance of the DGD unrolled method exceeds that of U-Net. Moreover, although the normalized errors of the two methods are roughly similar, we observe that in terms of similarity, the DGD approach outperforms U-Net, having an almost perfect index. For the Gauss-Newton method and the noiseless dataset, the average NMSE is 0.035 (0.12) and 0.022 (0.021) and the mean SSIM index 0.351 (0.075) and 0.437 (0.072), for water and bone respectively. For noisy databases, the Gauss-Newton method also gives poorer results on average, with some cases diverging for fixed regularization parameters.

The evolution of metric averages and standard deviations for DGD on the $\mathcal{T}(10^4)$ test dataset are shown in Fig. 5.15. In terms of the logarithm of the NMSE, the very first iterations show the greatest improvement. It does not seem to improve much after the fifth iteration. However, the SSIM index continues to improve with each iteration. This behavior in the evolution of the metrics is due to the regularization imposed on the cost function (5.25), which forces the outputs of intermediate iterations to be close to the desired output.

An example of water/bone decomposition, based on the test data Figure 5.14, is given in Fig 5.16. Although the regularization parameters for Gauss-Newton have been optimized, the noise is too high to obtain a satisfactory decomposition. As for those obtained by neural networks, the decomposition and NMSE are similar, but DGD has a higher similarity index than U-Net for both materials. The average time taken to decompose a case was 30 s, 0.6 s and 1.7 s for Gauss-Newton, U-Net and DGD, respectively. By shifting the computational load to the

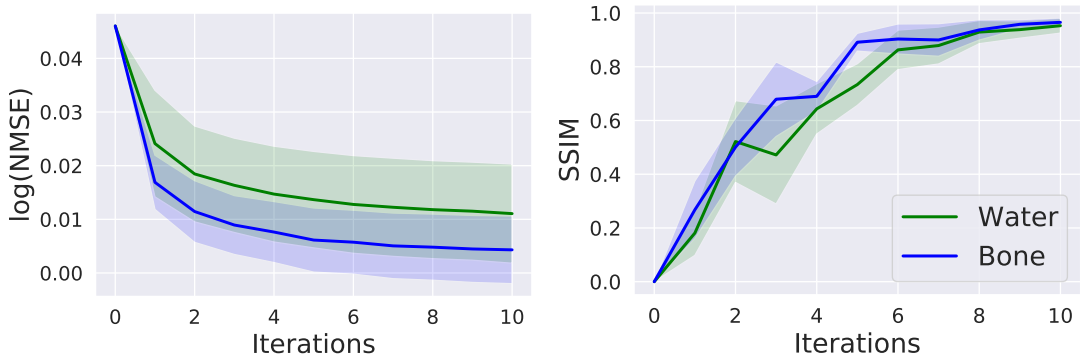


Figure 5.15.: Evolution of metric averages log(NMSE) and SSIM over iterations for the DGD method on the $\mathcal{T}(10^4)$ test dataset. Transparent areas represent standard deviation.

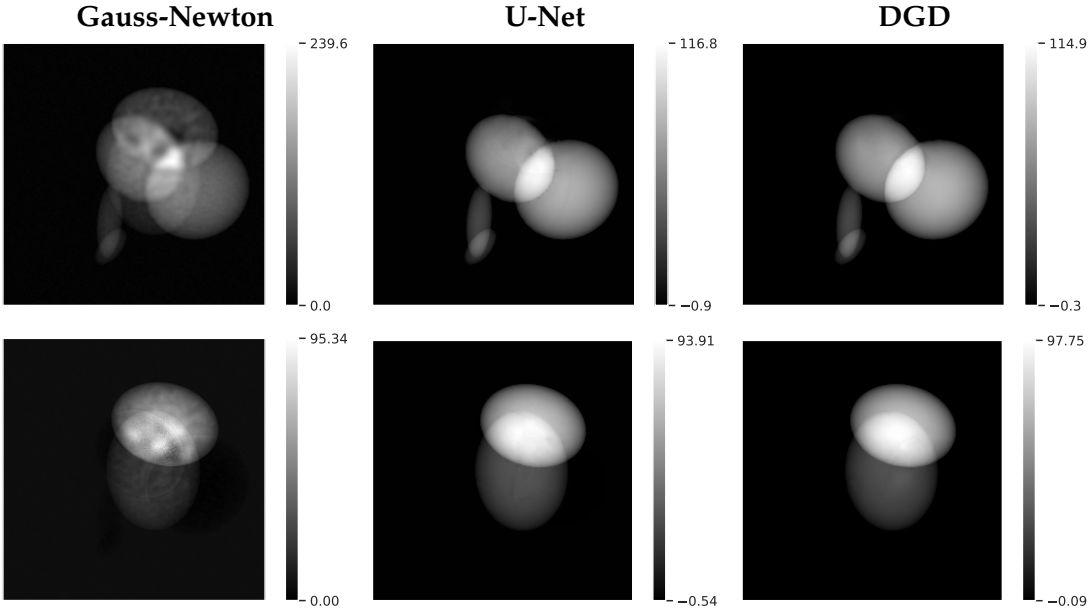


Figure 5.16.: Decomposition obtained for the test pair shown in Figure 5.14 with $\epsilon = 10^4$. The first line corresponds to water and the second to bone.

learning phase, networks enable decomposition 20 to 50 times faster than the GN variational method, which also requires around ten iterations to converge.

Conclusion and discussion

In this work, we have presented an unrolling approach applied to a gradient descent for the nonlinear inverse basis material decomposition problem. This method consists in unrolling a simple gradient descent by modifying it with a convolutional neural network. The method improves the quality of decompositions and shortens the calculation time. It preserves the convergence properties of the gradient descent and enables an adapted regularization to be learned. An extension of the proposed work would be to apply the same type of scheme to more complex iterative algorithms such as the Gauss-Newton method or a nonlinear primal-dual method.

Conclusion and perspectives

In this thesis, we have explored classical and deep learning algorithms for the phase retrieval problem. The main contribution of this work is the development of new methods allowing the retrieval of phase shift and absorption from a single X-ray in-line phase contrast. For this purpose, we introduced (Chapter 2) the mixed scale dense convolutional neural networks (MS-D Net), which uses dilated convolutions to extract features at different scales and densely connect all these feature maps. This network architecture efficiently incorporates long-range information in images, allowing for larger receptive field sizes without sacrificing resolution, effectively addressing the impact of the Fresnel operator. The MS-D Net was trained to reconstruct directly from the intensity measurements, without knowledge of the forward model. On homogeneous objects, the MS-D network successfully captures phase and attenuation as proportional, differing only by a constant factor, indicating that it learned the $\frac{\delta_r}{\beta}$ ratio. This network was able to retrieve both attenuation and phase of a heterogeneous object, from a single diffraction pattern, without any assumption on the object composition or on the support of the object.

Evaluation of such a network to experimental data was not entirely conclusive. This led us to the study of iterative methods. Among these methods, there are the variational approaches, based on the Fréchet derivative of the forward operator in conjunction with the Landweber algorithm. These approaches offer flexibility for incorporating various prior information about the solution through regularization. Recently, the total variation (TV) regularization has been used for the phase retrieval problem through an alternating direction method of multipliers (ADMM) scheme and provided good reconstruction. However, it has only been proposed for the linearized case, either by the contrast transfer function (CTF) or the transport of intensity equation (TIE). We therefore proposed the nonlinear primal-dual hybrid gradient (NL-PDHG) method, which takes account of nonlinear information, but also allows absorption and phase to be regularized separately. This method uses the TV regularization as well as its second-order generalization (TGV^2). A quantitative comparison showed that the best results were obtained with TV regularization for phase and TGV^2 regularization for absorption. In particular, this has resulted in high quality reconstruction on real data. However, this still hides some difficulties such as the choice of the priors and the parameters.

To unite the best of both worlds - neural networks and iterative methods, we have then studied the unrolled approaches. We introduce a new learned iterative scheme, the Deep Gauss-Newton (DGN). This algorithm is obtained by unrolling the iteratively regularized Gauss-Newton (IRGN) method. The proposed DGN combines convolutional neural networks and knowledge of the physical model given by the forward operator and its Fréchet derivative. The IRGN is known to converge quickly with satisfactory reconstruction and the DGN algorithm is likely to inherit, or even enhance it with deep learning. An additional benefit of the DGN is the absence of the need for manual regularization selection, as it is automatically learned from the data during the training process. This association results in fast reconstructions with almost no artifacts on both simulated and experimental data. Through a modification of the DGN architecture, we have developed the Deep Proximal Gauss-Newton (DPGN) model. Experiments show that this innovative architecture enhances the quality of reconstructions by improving both resolution and reconstruction error, all while maintaining a similar parameter count.

In view of the promising results of the unrolling method, we took the study a step further (Chapter 5) The quality of the reconstructions was influenced by the initial choice of the scheme we unrolled. When the number of parameters is roughly equivalent and the learning strategy

is the same, these methods give comparable results, with some offering superior visual and quantitative quality. Models that leverage second-order information, such as DGN or DPGN, consistently deliver improved reconstruction results and exhibit greater stability. Furthermore, all unrolled approaches prove to be more efficient than their traditional counterparts and demonstrate strong generalization capabilities in experimental conditions.

There are several perspectives to the work presented in this thesis. A first approach is to couple the proposed algorithms with tomographic reconstruction, this last step could also be replaced by a neural network. In our model, ideal conditions were assumed, the case of X-rays from a laboratory environment could be studied. For the NL-PDHG and the unrolling methods, this amounts to consider the Kullback–Leibler divergence instead of the L^2 norm. A limitation of model-based learning method is that the forward model has to be fully known. Bayesian approaches are an interesting way of quantifying these uncertainties, the possibilities to correct errors in the forward model might be interesting as well. An extent of deep equilibrium approaches for non linear inverse problems could open doors to study the convergent properties. A further development is the study of dynamic imaging with various spatial and temporal regularizations using specific iterative methods that could be unrolled. Finally, a challenge is the management of free physical parameters in the problem, such as X-ray energy, propagation distance, and imaging resolution, which could be included in the learning process.

List of publications

Articles in international peer-reviewed journals

Max Langer, Yuhe Zhang, Diogo Figueirinhas, Jean-Baptiste Forien, Kannara Mom, Claire Mouton, Rajmund Mokso, and Pablo Villanueva-Perez (July 2021). "*PyPhase* – a Python package for X-ray phase imaging". In: *Journal of Synchrotron Radiation* 28.4, pp. 1261–1266. DOI: 10.1107/S1600577521004951. URL: https://doi.org/10.1107/S1600577521004951

Kannara Mom, Bruno Sixou, and Max Langer (Apr. 2022). "Mixed scale dense convolutional networks for x-ray phase contrast imaging". In: *Appl. Opt.* 61.10, pp. 2497–2505. DOI: 10.1364/AO.443330. URL: http://opg.optica.org/ao/abstract.cfm?URI=ao-61-10-2497

Kannara Mom, Max Langer, and Bruno Sixou (Oct. 2022b). "Nonlinear primal–dual algorithm for the phase and absorption retrieval from a single phase contrast image". In: *Opt. Lett.* 47.20, pp. 5389–5392. DOI: 10.1364/OL.469174. URL: https://opg.optica.org/ol/abstract.cfm?URI=ol-47-20-5389

Kannara Mom, Max Langer, and Bruno Sixou (Mar. 2023). "Deep Gauss-Newton for phase retrieval". In: *Opt. Lett.* 48.5, pp. 1136–1139. DOI: 10.1364/OL.484862. URL: https://opg.optica.org/ol/abstract.cfm?URI=ol-48-5-1136

Communications in international conferences

Kannara Mom, Max Langer, and Bruno Sixou (Apr. 2022c). "Nonlinear primal-dual method for X-ray in-line phase contrast imaging". In: *SPIE Photonics Europe 2022*. Vol. 12136. Proceedings of SPIE. Strasbourg, France. DOI: 10.1117/12.2619732. URL: https://hal.science/hal-03706059

Communications in national conferences

Kannara Mom, Max Langer, and Bruno Sixou (Sept. 2022a). "Apprentissage de méthodes itératives pour l'imagerie de contraste de phase des rayons X". in: *XXVIIIème Colloque Francophone de Traitement du Signal et des Images (Gretsi 2022)*. Nancy, France. URL: https://hal.science/hal-03706103

Kannara Mom, Jerome Lesaint, Nicolas Ducros, Bruno Sixou, and Max Langer (Aug. 2023). "Décomposition en matériaux de base par apprentissage profond pour la tomographie spectrale". In: *XXIXème Colloque Francophone de Traitement du Signal et des Images (Gretsi 2023)*. Grenoble, France

Bibliography

Röntgen, W. (1896). "On a new kind or rays". In: *Journal of the Franklin Institute* 141.3, pp. 183–191. ISSN: 0016-0032. URL: https://eurekamag.com/research/086/142/086142574.php (cit. on p. 1).

McCulloch, Warren S and Walter Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5, pp. 115–133 (cit. on pp. 82, 85).

Rose, Albert (1946). "A Unified Approach to the Performance of Photographic Film, Television Pickup Tubes, and the Human Eye * –>". In: *Journal of the Society of Motion Picture Engineers* 47, pp. 273–294. URL: https://api.semanticscholar.org/CorpusID:124928474 (cit. on p. 49).

Landweber, Louis (1951). "An iteration formula for Fredholm integral equations of the first kind". In: *American Journal of Mathematics* 73, pp. 615–624 (cit. on p. 75).

Hestenes, Magnus R. and Eduard Stiefel (1952). "Methods of conjugate gradients for solving linear systems". In: *Journal of research of the National Bureau of Standards* 49, pp. 409–435 (cit. on p. 80).

Arrow, K.J. (1958). *Studies in Linear and Non-linear Programming*. Stanford mathematical studies in the social sciences. Stanford University Press. ISBN: 9780804705622. URL: https://books.google.fr/books?id=jWi4AAAAIAAJ (cit. on p. 77).

Rosenblatt, F. (1958). "The perceptron: A probabilistic model for information storage and organization in the brain." In: *Psychological Review* 65.6, pp. 386–408. ISSN: 0033-295X. DOI: 10.1037/h0042519. URL: http://dx.doi.org/10.1037/h0042519 (cit. on p. 83).

Minty, George J. (1962). "Monotone (nonlinear) operators in Hilbert space". In: *Duke Mathematical Journal* 29, pp. 341–346. URL: https://api.semanticscholar.org/CorpusID:121956938 (cit. on p. 73).

Fletcher, R. and C. M. Reeves (1964). "Function minimization by conjugate gradients". In: *Comput. J.* 7, pp. 149–154 (cit. on p. 80).

Bonse, U. and M. Hart (Nov. 1965). "AN X-RAY INTERFEROMETER". In: *Applied Physics Letters* 6.8, pp. 155–156. ISSN: 0003-6951. DOI: 10.1063/1.1754212. eprint: https://pubs.aip.org/aip/apl/article-pdf/6/8/155/7785327/155_1_online.pdf. URL: https://doi.org/10.1063/1.1754212 (cit. on pp. 1, 7).

Moreau, Jean Jacques (1965). "Proximité et dualité dans un espace hilbertien". In: *Bulletin de la Société Mathématique de France* 93, pp. 273–299 (cit. on p. 72).

Polak, E. and G. Ribiere (1969). "Note sur la convergence de méthodes de directions conjuguées". fr. In: *Revue française d'informatique et de recherche opérationnelle. Série rouge* 3.R1, pp. 35–43. URL: http://www.numdam.org/item/M2AN_1969__3_1_35_0/ (cit. on p. 80).

Polyak, Boris (Dec. 1969). "The conjugate gradient method in extreme problem". In: *USSR Computational Mathematics and Mathematical Physics* 9, pp. 94–112. DOI: 10.1016/0041-5553(69)90035-4 (cit. on p. 80).

Martinet, B. (1970). "Brève communication. Régularisation d'inéquations variationnelles par approximations successives". fr. In: *Revue française d'informatique et de recherche opérationnelle. Série rouge* 4.R3, pp. 154–158. URL: http://www.numdam.org/item/M2AN_1970__4_3_154_0/ (cit. on p. 76).

Gerchberg, R. W. (1972). "A practical algorithm for the determination of phase from image and diffraction plane pictures". In: *Optik* 35, pp. 237–246 (cit. on pp. 3, 9, 15, 63, 116).

Rockafellar, R. Tyrrell (1974). *Conjugate Duality and Optimization*. Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9781611970524. eprint: https://epubs.siam.org/

doi/pdf/10.1137/1.9781611970524. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611970524 (cit. on p. 72).
Alvarez, R. E. and A. Macovski (1976). "Energy-selective reconstructions in X-ray computerized tomography." In: *Phys. Med. and Biol.* ISSN: 0031-9155. URL: http://www.ncbi.nlm.nih.gov/pubmed/967922 (cit. on p. 157).
Keller, Joseph B. (1976). "Inverse Problems". In: *The American Mathematical Monthly* 83.2, pp. 107–118. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2976988 (visited on 08/07/2023) (cit. on p. 38).
Guigay, Jean-Pierre (Oct. 1977). "FOURIER TRANSFORM ANALYSIS OF FRESNEL DIFFRACTION PATTERNS AND IN-LINE HOLOGRAMS." In: *Optik* 49, pp. 121–125 (cit. on pp. 3, 8, 49).
Tikhonov, Andrey N. and Vasiliy Y. Arsenin (1977). *Solutions of ill-posed problems*. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics. Washington, D.C.: John Wiley & Sons, New York: V. H. Winston & Sons (cit. on pp. 41, 68).
Fienup, J. R. (July 1978). "Reconstruction of an object from the modulus of its Fourier transform". In: *Opt. Lett.* 3.1, pp. 27–29. DOI: 10.1364/OL.3.000027. URL: https://opg.optica.org/ol/abstract.cfm?URI=ol-3-1-27 (cit. on pp. 3, 9, 63).
Fienup, James R. (1982). "Phase retrieval algorithms: a comparison". In: *Applied Optics* 21 (15), pp. 2758–2769 (cit. on pp. 3, 9, 63–65, 104, 116).
Teague, Michael Reed (Sept. 1982). "Irradiance moments: their propagation and use for unique retrieval of phase". In: *J. Opt. Soc. Am.* 72.9, pp. 1199–1209. DOI: 10.1364/JOSA.72.001199. URL: https://opg.optica.org/abstract.cfm?URI=josa-72-9-1199 (cit. on p. 60).
Nesterov, Yurii (1983). "A method for solving the convex programming problem with convergence rate $O(1/k^2)$". In: *Proceedings of the USSR Academy of Sciences* 269, pp. 543–547. URL: https://api.semanticscholar.org/CorpusID:145918791 (cit. on pp. 74, 87).
Teague, Michael Reed (Nov. 1983). "Deterministic phase retrieval: a Green's function solution". In: *J. Opt. Soc. Am.* 73.11, pp. 1434–1441. DOI: 10.1364/JOSA.73.001434. URL: https://opg.optica.org/abstract.cfm?URI=josa-73-11-1434 (cit. on pp. 3, 8, 14, 60).
Fromovitz, Stan (1984). *Methods for Solving Incorrectly Posed Problems*. Springer New York, NY (cit. on p. 42).
Levi, Aharon and Henry Stark (Sept. 1984). "Image restoration by the method of generalized projections with application to restoration from magnitude". In: *J. Opt. Soc. Am. A* 1.9, pp. 932–943. DOI: 10.1364/JOSAA.1.000932. URL: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-1-9-932 (cit. on p. 65).
Ciarlet, Philippe G. (1989). *Introduction to Numerical Linear Algebra and Optimisation*. Cambridge Texts in Applied Mathematics. Cambridge University Press. DOI: 10.1017/9781139171984 (cit. on p. 40).
Rudin, Leonid I., Stanley Osher, and Emad Fatemi (1992). "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1, pp. 259–268. ISSN: 0167-2789. DOI: https://doi.org/10.1016/0167-2789(92)90242-F. URL: https://www.sciencedirect.com/science/article/pii/016727899290242F (cit. on pp. 68, 117).
Guinier, A. (1994). *X-ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies*. Dover Books on Physics Series. Dover Publications. ISBN: 9780486680118. URL: https://books.google.fr/books?id=vjyZFo5nGUoC (cit. on p. 47).
Davis, Timothy J. et al. (1995). "Phase-contrast imaging of weakly absorbing materials using hard X-rays". In: *Nature* 373, pp. 595–598. URL: https://api.semanticscholar.org/CorpusID:4287341 (cit. on pp. 1, 7).

Hanke, Martin, Andreas Neubauer, and Otmar Scherzer (1995). "A convergence analysis of the Landweber iteration for nonlinear ill-posed problems". In: *Numerische Mathematik* 72, pp. 21–37 (cit. on p. 79).
Momose, Atsushi and J Fukuda (1995). "Phase-contrast radiographs of nonstained rat cerebellar specimen." In: *Medical physics* 22 4, pp. 375–9. URL: https://api.semanticscholar.org/CorpusID:23108769 (cit. on pp. 1, 7).
Snigirev, A. et al. (Dec. 1995). "On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation". In: *Review of Scientific Instruments* 66.12, pp. 5486–5492. DOI: 10.1063/1.1146073 (cit. on pp. 1, 7, 8).
Takeda, Tohoru et al. (1995). "Phase-contrast imaging with synchrotron x-rays for detecting cancer lesions". In: *Academic Radiology* 2.9, pp. 799–803. ISSN: 1076-6332. DOI: https://doi.org/10.1016/S1076-6332(05)80490-8. URL: https://www.sciencedirect.com/science/article/pii/S1076633205804908 (cit. on pp. 1, 7).
Cloetens, Peter et al. (Jan. 1996). "Phase objects in synchrotron radiation hard X-ray imaging". In: *Journal of Physics D: Applied Physics* 29.1, pp. 133–146. DOI: 10.1088/0022-3727/29/1/023. URL: https://doi.org/10.1088/0022-3727/29/1/023 (cit. on pp. 1, 3, 7, 8).
Goodman, J.W. (1996). *Introduction to Fourier Optics*. Electrical Engineering Series. McGraw-Hill. ISBN: 9780070242548. URL: https://books.google.fr/books?id=QllRAAAAMAAJ (cit. on pp. 2, 8, 13, 48).
Gureyev, T. E. and K. A. Nugent (Aug. 1996). "Phase retrieval with the transport-of-intensity equation. II. Orthogonal series solution for nonuniform illumination". In: *J. Opt. Soc. Am. A* 13.8, pp. 1670–1682. DOI: 10.1364/JOSAA.13.001670. URL: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-13-8-1670 (cit. on pp. 3, 8, 60, 115).
Momose, Atsushi, Tohoru Takeda, et al. (1996). "Phase–contrast X–ray computed tomography for observing biological soft tissues". In: *Nature Medicine* 2, pp. 473–475 (cit. on pp. 1, 7).
Chapman, D et al. (Nov. 1997). "Diffraction enhanced x-ray imaging". In: *Physics in Medicine Biology* 42.11, p. 2015. DOI: 10.1088/0031-9155/42/11/001. URL: https://dx.doi.org/10.1088/0031-9155/42/11/001 (cit. on pp. 1, 2, 7).
Cloetens, P., M. Pateyron-Salomé, et al. (May 1997). "Observation of microstructure and damage in materials by phase sensitive radiography and tomography". In: *Journal of Applied Physics* 81.9, pp. 5878–5886. ISSN: 0021-8979. DOI: 10.1063/1.364374. eprint: https://pubs.aip.org/aip/jap/article-pdf/81/9/5878/10586750/5878_1_online.pdf. URL: https://doi.org/10.1063/1.364374 (cit. on pp. 3, 8).
Lecun, Yann et al. (Dec. 1998). "Gradient-Based Learning Applied to Document Recognition". In: *Proceedings of the IEEE* 86, pp. 2278–2324. DOI: 10.1109/5.726791 (cit. on p. 88).
Paganin, D. and K. A. Nugent (Mar. 1998). "Noninterferometric Phase Imaging with Partially Coherent Light". In: *Phys. Rev. Lett.* 80.12, pp. 2586–2589. DOI: 10.1103/PhysRevLett.80.2586. URL: http://link.aps.org/doi/10.1103/PhysRevLett.80.2586 (cit. on pp. 3, 8, 60).
Scherzer, Otmar (1998). "A Modified Landweber Iteration for Solving Parameter Estimation Problems". In: *Applied Mathematics and Optimization* 38, pp. 45–68. DOI: https://doi.org/10.1007/s002459900081 (cit. on p. 79).
Bonse, U (1999). *Developments in X-ray tomography II : 22-23 July, 1999, Denver, Colorado*. eng. SPIE proceedings series Developments in X-ray tomography II. Place of publication not identified: SPIE. ISBN: 0-8194-3258-X (cit. on pp. 1, 7).
Cloetens, P., W. Ludwig, et al. (1999). "Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays". In: *Applied Physics Letters* 75.19, pp. 2912–2914. DOI: 10.1063/1.125225. eprint: https://doi.org/10.1063/1.125225. URL: https://doi.org/10.1063/1.125225 (cit. on pp. 2, 8, 53).

Ekeland, Ivar and Roger Témam (1999). *Convex Analysis and Variational Problems*. Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9781611971088. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611971088. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611971088 (cit. on p. 77).

Engl, H.W., M. Hanke, and A. Neubauer (2000). *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands. ISBN: 9780792361404. URL: https://books.google.fr/books?id=VuEV-Gj1GZcC (cit. on p. 80).

Hansen, Per Christian (2000). "The L-curve and its use in the numerical treatment of inverse problems". English. In: *InviteComputational Inverse Problems in Electrocardiology*. InviteComputational Inverse Problems in Electrocardiology ; Conference date: 01-01-2000. WIT Press (cit. on p. 42).

Allen, L.J. and M.P. Oxley (2001). "Phase retrieval from series of images obtained by defocus variation". In: *Optics Communications* 199.1, pp. 65–75. ISSN: 0030-4018. DOI: https://doi.org/10.1016/S0030-4018(01)01556-5. URL: https://www.sciencedirect.com/science/article/pii/S0030401801015565 (cit. on p. 60).

Bauschke, Heinz H., Patrick L. Combettes, and D. Russell Luke (July 2002). "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization". In: *J. Opt. Soc. Am. A* 19.7, pp. 1334–1345. DOI: 10.1364/JOSAA.19.001334. URL: http://www.osapublishing.org/josaa/abstract.cfm?URI=josaa-19-7-1334 (cit. on pp. 64–66, 68, 116).

David, C. et al. (Oct. 2002). "Differential x-ray phase contrast imaging using a shearing interferometer". In: *Applied Physics Letters* 81.17, pp. 3287–3289. ISSN: 0003-6951. DOI: 10.1063/1.1516611. eprint: https://pubs.aip.org/aip/apl/article-pdf/81/17/3287/10195235/3287_1_online.pdf. URL: https://doi.org/10.1063/1.1516611 (cit. on pp. 1, 7).

Paganin, D., S. C. Mayo, et al. (2002). "Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object". In: *Journal of Microscopy* 206, pp. 33–40 (cit. on pp. 3, 8, 14, 56, 60, 104, 115).

Bauschke, Heinz H., Patrick L. Combettes, and D. Russel Luke (2003). "A Hybrid Projection Reflection Method for Phase Retrieval". In: *Journal of the Optical Society of America A* 20 (6), pp. 1025–1034 (cit. on pp. 15, 65, 66, 68, 116).

Elser, Veit (2003). "Phase retrieval by iterated projections". In: *Journal of the Optical Society of America A* 20 (1), pp. 40–55. URL: https://doi.org/10.1364/JOSAA.20.000040 (cit. on pp. 15, 67, 116).

Mayo, S.C. et al. (Sept. 2003). "X-ray phase-contrast microscopy and microtomography". In: *Opt. Express* 11.19, pp. 2289–2302. DOI: 10.1364/OE.11.002289. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-11-19-2289 (cit. on pp. 1, 7).

Beleggia, M. et al. (2004). "On the transport of intensity technique for phase retrieval". In: *Ultramicroscopy* 102.1, pp. 37–49. ISSN: 0304-3991. DOI: https://doi.org/10.1016/j.ultramic.2004.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0304399104001664 (cit. on p. 116).

Boyd, Stephen and Lieven Vandenberghe (2004). *Convex optimization*. Cambridge university press (cit. on p. 70).

Chambolle, Antonin (2004). "An Algorithm for Total Variation Minimization and Applications". In: *Journal of Mathematical Imaging and Vision* 20.1, pp. 89–97. ISSN: 1573-7683. DOI: 10.1023/B:JMIV.0000011325.36760.1e. URL: https://doi.org/10.1023/B:JMIV.0000011325.36760.1e (cit. on pp. 74, 117).

Gureyev, T.E et al. (2004). "Linear algorithms for phase retrieval in the Fresnel region". In: *Optics Communications* 231.1, pp. 53–70. ISSN: 0030-4018. DOI: https://doi.org/10.1016/

j.optcom.2003.12.020. URL: https://www.sciencedirect.com/science/article/pii/S0030401803023320 (cit. on p. 61).
Turner, L.D. et al. (2004). "X-ray phase imaging: Demonstration of extended conditions with homogeneous objects". In: *Optics Express* 12 (cit. on p. 59).
Luke, D Russel (2005). "Relaxed averaged alternating reflections for diffraction imaging". In: *Inverse Problems* 21, pp. 37–50. URL: http://dx.doi.org/10.1088/0266-5611/21/1/004 (cit. on pp. 15, 66–68, 116).
Pagot, Elodie et al. (Feb. 2005). "Quantitative comparison between two phase contrast techniques: diffraction enhanced imaging and phase propagation imaging". In: *Physics in Medicine Biology* 50.4, p. 709. DOI: 10.1088/0031-9155/50/4/010. URL: https://dx.doi.org/10.1088/0031-9155/50/4/010 (cit. on pp. 1, 7).
Weitkamp, Timm et al. (Aug. 2005). "X-ray phase imaging with a grating interferometer". In: *Opt. Express* 13.16, pp. 6296–6304. DOI: 10.1364/OPEX.13.006296. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-13-16-6296 (cit. on pp. 1, 2, 7).
Zabler, S. et al. (2005). "Optimization of phase contrast imaging using hard X-rays". In: *Review of Scientific Instruments* 76.7, p. 073705. DOI: 10.1063/1.1960797. eprint: https://doi.org/10.1063/1.1960797. URL: https://doi.org/10.1063/1.1960797 (cit. on pp. 3, 8, 59, 115).
Paganin (2006). *Coherent X-ray optics*. Oxford Series on Synchrotron Radiation (cit. on pp. 3, 8, 53, 56, 60, 104).
Ramlau, Ronny and Gerd Teschke (July 2006). "A Tikhonov-Based Projection Iteration for Nonlinear Ill-Posed Problems with Sparsity Constraints". In: *Numer. Math.* 104.2, pp. 177–203. ISSN: 0029-599X. DOI: 10.1007/s00211-006-0016-3. URL: https://doi.org/10.1007/s00211-006-0016-3 (cit. on p. 68).
Cancedda, R. et al. (2007). "Bulk and interface investigations of scaffolds and tissue-engineered bones by X-ray microtomography and X-ray microdiffraction". In: *Biomaterials* 28.15. Imaging Techniques for Biomaterials Characterization, pp. 2505–2524. ISSN: 0142-9612. DOI: https://doi.org/10.1016/j.biomaterials.2007.01.022. URL: https://www.sciencedirect.com/science/article/pii/S0142961207000531 (cit. on pp. 3, 8).
Dabov, Kostadin et al. (2007). "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering". In: *IEEE Transactions on Image Processing* 16.8, pp. 2080–2095. DOI: 10.1109/TIP.2007.901238 (cit. on p. 94).
Guigay, Jean Pierre et al. (June 2007). "Mixed transfer function and transport of intensity approach for phase retrieval in the Fresnel region". In: *Optics Letters* 32.12, pp. 1617–1619. DOI: 10.1364/OL.32.001617. URL: http://ol.osa.org/abstract.cfm?URI=ol-32-12-1617 (cit. on pp. 14, 61, 115).
Meng, Fanbo, Hong Liu, and Xizeng Wu (June 2007). "An iterative phase retrieval algorithm for in-line x-ray phase imaging". In: *Opt. Express* 15.13, pp. 8383–8390. DOI: 10.1364/OE.15.008383. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-15-13-8383 (cit. on p. 61).
Wang, Yi (2007). "Intuitive dimensional analyses of the energy and atomic number dependences of the cross sections for radiation interaction with matter". In: *Journal of X-ray Science and Technology* 15, pp. 169–175. URL: https://api.semanticscholar.org/CorpusID:117800758 (cit. on p. 44).
Kaltenbacher, Barbara, Andreas Neubauer, and Otmar Scherzer (2008). *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (cit. on pp. 16, 70, 81).
Langer, Max (Nov. 2008). "Phase Retrieval in the Fresnel Region for Hard X-ray Tomography". PhD thesis. DOI: 10.13140/RG.2.2.28752.02569 (cit. on pp. 3, 9).
Langer, Max, Peter Cloetens, Jean-Pierre Guigay, et al. (2008). "Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography". In: *Medical Physics* 35.10, pp. 4556–

4566. DOI: https://doi.org/10.1118/1.2975224. eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1118/1.2975224. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2975224 (cit. on p. 62).
Schlomka, J. P. et al. (2008). "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography". In: *Phys. Med. Biol.* ISSN: 00319155 (cit. on p. 157).
Zhu, Mingqiang and Tony F. Chan (2008). "An Efficient Primal-Dual Hybrid Gradient Algorithm For Total Variation Image Restoration". In: (cit. on p. 78).
Chaari, Lotfi et al. (Aug. 2009). "Solving inverse problems with overcomplete transforms and convex optimization techniques". In: *SPIE*. San Diego, California, United States. URL: https://hal.archives-ouvertes.fr/hal-00826119 (cit. on p. 122).
Deng, Jia et al. (2009). "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848 (cit. on p. 91).
Langer, Max, Peter Cloetens, and Françoise Peyrin (Aug. 2009). "Fourier-wavelet regularization of phase retrieval in x-ray in-line phase tomography". In: *J. Opt. Soc. Am. A* 26.8, pp. 1876–1881. DOI: 10.1364/JOSAA.26.001876. URL: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-26-8-1876 (cit. on p. 62).
Alvarez, R. E. (2010). "Near optimal energy selective x-ray imaging system performance with simple detectors". In: *Med. Phys.* ISSN: 00942405. DOI: 10.1118/1.3284538 (cit. on p. 158).
Beltran, M.A. et al. (Mar. 2010). "2D and 3D X-ray phase retrieval of multi-material objects using a single defocus distance". In: *Opt. Express* 18.7, pp. 6423–6436. DOI: 10.1364/OE.18.006423. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-18-7-6423 (cit. on p. 61).
Bredies, Kristian, Karl Kunisch, and Thomas Pock (2010). "Total Generalized Variation". In: *SIAM Journal on Imaging Sciences* 3.3, pp. 492–526. DOI: 10.1137/090769521. eprint: https://doi.org/10.1137/090769521. URL: https://doi.org/10.1137/090769521 (cit. on p. 118).
Chambolle, Antonin, Vicent Caselles, et al. (2010). "An Introduction to Total Variation for Image Analysis:" in: *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Ed. by Massimo Fornasier. De Gruyter, pp. 263–340. DOI: doi:10.1515/9783110226157.263. URL: https://doi.org/10.1515/9783110226157.263 (cit. on p. 117).
Esser, Ernie, Xiaoqun Zhang, and Tony Chan (Jan. 2010). "A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science". In: *SIAM J. Imaging Sciences* 3, pp. 1015–1046. DOI: 10.1137/09076934X (cit. on pp. 78, 145).
Langer, M., P. Cloetens, and F. Peyrin (2010). "Regularization of Phase Retrieval with Phase-Attenuation Duality Prior for 3D Holotomography". In: *IEEE Trans Image Proces* 19, pp. 2428–2436. DOI: 10.1109/TIP.2010.2048608 (cit. on p. 62).
Als-Nielsen, J. and D. McMorrow (2011). *Elements of Modern X-ray Physics*. Wiley. ISBN: 9781119970156. URL: https://books.google.fr/books?id=rlqlboWlTRMC (cit. on pp. 46, 50).
Alvarez, Robert E. (2011). "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis". In: *Medical Physics*. ISSN: 00942405. DOI: 10.1118/1.3570658 (cit. on p. 157).
Bauschke, Heinz H. and Patrick L. Combettes (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 1st. Springer Publishing Company, Incorporated. ISBN: 1441994661 (cit. on pp. 70, 74).
Boyd, Stephen, Neal Parikh, et al. (Jan. 2011). "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". In: *Foundations and Trends in Machine Learning* 3, pp. 1–122. DOI: 10.1561/2200000016 (cit. on pp. 78, 116).

Buades, Antoni, Bartomeu Coll, and Jean-Michel Morel (2011). "Non-Local Means Denoising". In: *Image Processing On Line* 1. https://doi.org/10.5201/ipol.2011.bcm_nlm, pp. 208–212 (cit. on p. 94).

Chambolle, Antonin and Thomas Pock (May 2011). "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging". In: *Journal of Mathematical Imaging and Vision* 40, pp. 120–145. DOI: https://doi.org/10.1007/s10851-010-0251-1 (cit. on pp. 78, 116).

Davidoiu, V. et al. (Nov. 2011). "Non-linear iterative phase retrieval based on Frechet derivative". In: *Opt. Express* 19.23, pp. 22809–22819. DOI: 10.1364/OE.19.022809. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-19-23-22809 (cit. on pp. 4, 9, 116, 132).

Salzo, Saverio and Silvia Villa (Mar. 2011). "Convergence analysis of a proximal Gauss-Newton method". In: *Computational Optimization and Applications* 53. DOI: 10.1007/s10589-012-9476-9 (cit. on p. 134).

Weitkamp, T. et al. (July 2011). "*ANKAphase*: software for single-distance phase retrieval from inline X-ray phase-contrast radiographs". In: *Journal of Synchrotron Radiation* 18.4, pp. 617–629. DOI: 10.1107/S0909049511002895. URL: https://doi.org/10.1107/S0909049511002895 (cit. on p. 60).

Langer, Max, Alexandra Pacureanu, et al. (Aug. 2012). "X-Ray Phase Nanotomography Resolves the 3D Human Bone Ultrastructure". In: *PLOS ONE* 7.8, pp. 1–7. DOI: 10.1371/journal.pone.0035691. URL: https://doi.org/10.1371/journal.pone.0035691 (cit. on pp. 3, 8, 9, 62).

Mayo, Sheridan C., Andrew W. Stevenson, and Stephen W. Wilkins (2012). "In-Line Phase-Contrast X-ray Imaging and Tomography for Materials Science". In: *Materials* 5, pp. 937–965 (cit. on pp. 3, 8).

Banterle, Niccolò et al. (2013). "Fourier ring correlation as a resolution criterion for super-resolution microscopy". In: *Journal of Structural Biology* 183.3, pp. 363–367. ISSN: 1047-8477. DOI: https://doi.org/10.1016/j.jsb.2013.05.004. URL: https://www.sciencedirect.com/science/article/pii/S1047847713001184 (cit. on p. 137).

Cai, C. et al. (2013). "A full-spectral Bayesian reconstruction approach based on the material decomposition model applied in dual-energy computed tomography". In: *Medical Physics*. ISSN: 00942405. DOI: 10.1118/1.4820478 (cit. on p. 157).

Davidoiu, Valentina et al. (June 2013). "Nonlinear approaches for the single-distance phase retrieval problem involving regularizations with sparsity constraints". In: *Appl. Opt.* 52.17, pp. 3977–3986. DOI: 10.1364/AO.52.003977. URL: http://ao.osa.org/abstract.cfm?URI=ao-52-17-3977 (cit. on p. 116).

Kostenko, Alexander et al. (May 2013). "Total variation minimization approach in in-line x-ray phase-contrast tomography". In: *Opt. Express* 21.10, pp. 12185–12196. DOI: 10.1364/OE.21.012185. URL: http://opg.optica.org/oe/abstract.cfm?URI=oe-21-10-12185 (cit. on p. 129).

Sixou, Bruno et al. (2013). "Absorption and phase retrieval with Tikhonov and joint sparsity regularizations". In: *Inverse Problems and Imaging* 7.1, pp. 267–282. ISSN: 1930-8337. DOI: 10.3934/ipi.2013.7.267. URL: https://www.aimsciences.org/article/id/fbbedec0-adfa-43a8-829a-967293fbb5f2 (cit. on pp. 4, 9, 57, 104, 111).

Valkonen, Tuomo, Kristian Bredies, and Florian Knoll (2013). "Total Generalized Variation in Diffusion Tensor Imaging". English. In: *SIAM journal on imaging sciences* 6.1, pp. 487–525. ISSN: 1936-4954. DOI: 10.1137/120867172 (cit. on p. 122).

Venkatakrishnan, Singanallur V., Charles A. Bouman, and Brendt Wohlberg (2013). "Plug-and-Play priors for model based reconstruction". In: *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948. DOI: 10.1109/GlobalSIP.2013.6737048 (cit. on pp. 93, 94).

Agostinelli, Forest et al. (2014). "Learning Activation Functions to Improve Deep Neural Networks". In: *arXiv: Neural and Evolutionary Computing*. URL: https://api.semanticscholar.org/CorpusID:7179166 (cit. on p. 101).
Bostan, Emrah et al. (2014). "Phase retrieval by using transport-of-intensity equation and differential interference contrast microscopy". In: *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 3939–3943. DOI: 10.1109/ICIP.2014.7025800 (cit. on p. 116).
Goldstein, Tom, Christoph Studer, and Richard Baraniuk (2014). "A Field Guide to Forward-Backward Splitting with a FASTA Implementation". In: *ArXiv* abs/1411.3406. URL: https://api.semanticscholar.org/CorpusID:9037325 (cit. on p. 100).
Langer, Max, Peter Cloetens, Bernhard Hesse, Heikki Suhonen, Alexandra Pacureanu, Kay Raum, and Françoise Peyrin (2014). "Priors for X-ray in-line phase tomography of heterogeneous objects". In: *Philolosophical Transactions of the Royal Society A* 372: 20130129. DOI: 10.1098/rsta.2013.0129. URL: https://doi.org/10.1098/rsta.2013.0129 (cit. on pp. 104, 110).
Langer, Max, Peter Cloetens, Bernhard Hesse, Heikki Suhonen, Alexandra Pacureanu, Kay Raum, and Francoise Peyrin (Jan. 2014). "Priors for X-ray in-line phase tomography of heterogeneous objects". In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 372, p. 20130129. DOI: 10.1098/rsta.2013.0129 (cit. on p. 127).
Long, Yong and Jeffrey A. Fessler (2014). "Multi-material decomposition using statistical image reconstruction for spectral CT". In: *IEEE Transactions on Medical Imaging*. ISSN: 1558254X. DOI: 10.1109/TMI.2014.2320284 (cit. on p. 157).
Nesterov, Yurii (2014). *Introductory Lectures on Convex Optimization: A Basic Course*. 1st ed. Springer Publishing Company, Incorporated. ISBN: 1461346916 (cit. on p. 70).
Valkonen, Tuomo (May 2014). "A primal–dual hybrid gradient method for nonlinear operators with applications to MRI". In: *Inverse Problems* 30.5, p. 055012. DOI: 10.1088/0266-5611/30/5/055012. URL: https://doi.org/10.1088/0266-5611/30/5/055012 (cit. on pp. 24, 116, 121, 122).
LeCun, Yann, Y. Bengio, and Geoffrey Hinton (May 2015). "Deep Learning". In: *Nature* 521, pp. 436–44. DOI: 10.1038/nature14539 (cit. on pp. 4, 10, 16, 81, 88, 104, 131).
Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. Cham: Springer International Publishing, pp. 234–241. ISBN: 978-3-319-24574-4 (cit. on p. 90).
Chambolle, Antonin and Thomas Pock (2016a). "An introduction to continuous optimization for imaging". In: *Acta Numerica* 25, pp. 161–319. DOI: 10.1017/S096249291600009X (cit. on p. 16).
– (2016b). "An introduction to continuous optimization for imaging". In: *Acta Numerica* 25, pp. 161–319. DOI: 10.1017/S096249291600009X (cit. on pp. 120, 145).
Chen, Liang-Chieh et al. (June 2016). "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP. DOI: 10.1109/TPAMI.2017.2699184 (cit. on p. 105).
Clevert, Djork-Arné, Thomas Unterthiner, and Sepp Hochreiter (2016). *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. arXiv: 1511.07289 [cs.LG] (cit. on p. 85).
Eggl, Elena et al. (Sept. 2016). "The Munich Compact Light Source: Initial performance measures". In: *Journal of Synchrotron Radiation* 23. DOI: 10.1107/S160057751600967X (cit. on p. 1).
Foygel Barber, Rina et al. (2016). "An algorithm for constrained one-step inversion of spectral CT data". In: *Physics in Medicine and Biology*. ISSN: 13616560. DOI: 10.1088/0031-9155/61/10/3784 (cit. on p. 157).

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press (cit. on p. 81).
Loshchilov, Ilya and Frank Hutter (Aug. 2016). "SGDR: Stochastic Gradient Descent with Warm Restarts". In: *ICLR 2017* (cit. on pp. 136, 148, 161).
Maretzke, Simon, Matthias Bartels, et al. (Mar. 2016). "Regularized Newton methods for X-ray phase contrast and general imaging problems". In: *Opt. Express* 24.6, pp. 6490–6506. DOI: 10.1364/OE.24.006490. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-24-6-6490 (cit. on pp. 4, 9, 104, 116, 132, 135).
Oord, Aäron van den et al. (2016). "WaveNet: A Generative Model for Raw Audio". In: *Arxiv*. URL: https://arxiv.org/abs/1609.03499 (cit. on p. 105).
Yu, Fisher and Vladlen Koltun (2016). "Multi-scale context aggregation by dilated convolutions". In: *International Conference on Learning Representations* (cit. on p. 105).
Adler, J. et al. (Oct. 2017). "Learning to solve inverse problems using Wasserstein loss". In: *arXiv:1710.10898*. DOI: 10.48550/arxiv.1710.10898. arXiv: 1710.10898. URL: https://arxiv.org/abs/1710.10898v1 (cit. on p. 159).
Adler, Jonas and Ozan Oktem (2017). "Solving ill-posed inverse problems using iterative deep neural networks". In: *Inverse Problems* 33 (12) (cit. on pp. 104, 143).
Chen, Hu et al. (2017). "Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN)". In: *ArXiv* abs/1702.00288. URL: https://api.semanticscholar.org/CorpusID:231697457 (cit. on p. 92).
Chollet, Francois (2017). *Deep Learning with Python*. 1st. USA: Manning Publications Co. ISBN: 1617294438 (cit. on p. 81).
Ducros, N. et al. (Sept. 2017). "Regularization of nonlinear decomposition of spectral x-ray projection images". In: *Medical Physics*. ISSN: 00942405. DOI: 10.1002/mp.12283. URL: http://doi.wiley.com/10.1002/mp.12283 (cit. on pp. 157, 159, 161).
Fu, Hongsun et al. (2017). In: *Journal of Inverse and Ill-posed Problems* 25.3, pp. 341–356. DOI: doi:10.1515/jiip-2015-0092. URL: https://doi.org/10.1515/jiip-2015-0092 (cit. on p. 134).
Huang, Gao, Zhuang Liu, et al. (2017). "Densely Connected Convolutional Networks". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. DOI: https://doi.org/10.1109/CVPR.2017.243 (cit. on p. 105).
Jin, K. H. et al. (2017). "Deep convolutional neural network for inverse problems in imaging." In: *IEEE Trans. Image Process.* 26, pp. 4509–4522 (cit. on pp. 90–92, 104, 108, 113).
Kamilov, Ulugbek S., Hassan Mansour, and Brendt Wohlberg (2017). "A Plug-and-Play Priors Approach for Solving Nonlinear Imaging Inverse Problems". In: *IEEE Signal Processing Letters* 24, pp. 1872–1876. URL: https://api.semanticscholar.org/CorpusID:5606563 (cit. on p. 95).
Kingma, Diederik P. and Jimmy Ba (2017). *Adam: A Method for Stochastic Optimization*. arXiv: 1412.6980 [cs.LG] (cit. on p. 87).
Klambauer, Günter et al. (2017). *Self-Normalizing Neural Networks*. arXiv: 1706.02515 [cs.LG] (cit. on p. 85).
Knoll, Florian et al. (2017). "Joint MR-PET Reconstruction Using a Multi-Channel Image Regularizer". In: *IEEE Transactions on Medical Imaging* 36, pp. 1–16 (cit. on p. 122).
Maretzke, Simon and Thorsten Hohage (2017). "Stability Estimates for Linearized Near-Field Phase Retrieval in X-ray Phase Contrast Imaging". In: *SIAM Journal on Applied Mathematics* 77.2, pp. 384–408. DOI: 10.1137/16M1086170. eprint: https://doi.org/10.1137/16M1086170. URL: https://doi.org/10.1137/16M1086170 (cit. on p. 116).

Mechlem, Korbinian et al. (2017). "A post-processing algorithm for spectral CT material selective images using learned dictionaries". In: *Biomedical Physics & Engineering Express*. DOI: 10.1088/2057-1976/aa6045 (cit. on p. 157).
Meinhardt, Tim et al. (Oct. 2017). "Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE. DOI: 10.1109/iccv.2017.198. URL: https://doi.org/10.1137/16M1102884 (cit. on p. 94).
Rivenson, Yair et al. (Oct. 2017). "Phase recovery and holographic image reconstruction using deep learning in neural networks". In: *Light: Science &amp Applications* 7.2, pp. 17141–17141. DOI: 10.1038/lsa.2017.141. URL: https://doi.org/10.1038/lsa.2017.141 (cit. on pp. 97, 98).
Romano, Yaniv, Michael Elad, and Peyman Milanfar (2017). "The Little Engine That Could: Regularization by Denoising (RED)". In: *SIAM Journal on Imaging Sciences* 10.4, pp. 1804–1844. DOI: 10.1137/16M1102884. eprint: https://doi.org/10.1137/16M1102884. URL: https://doi.org/10.1137/16M1102884 (cit. on p. 100).
Sandino, Christopher M. et al. (2017). "Deep convolutional neural networks for accelerated dynamic magnetic resonance imaging". In: URL: https://api.semanticscholar.org/CorpusID:26883002 (cit. on p. 92).
Villanueva-Perez, Pablo et al. (Mar. 2017). "Contrast-transfer-function phase retrieval based on compressed sensing". In: *Opt. Lett.* 42.6, pp. 1133–1136. DOI: 10.1364/OL.42.001133. URL: http://ol.osa.org/abstract.cfm?URI=ol-42-6-1133 (cit. on p. 116).
Zhang, Kai et al. (2017). "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising". In: *IEEE Transactions on Image Processing* 26.7, pp. 3142–3155. DOI: 10.1109/TIP.2017.2662206 (cit. on p. 100).
Zhu, Bo et al. (2017). "Image reconstruction by domain-transform manifold learning". In: *Nature* 555, pp. 487–492. URL: https://api.semanticscholar.org/CorpusID:4173387 (cit. on p. 91).
Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A Efros (2017a). "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". In: *Computer Vision (ICCV), 2017 IEEE International Conference on* (cit. on p. 99).
– (2017b). "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks". In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251. DOI: 10.1109/ICCV.2017.244 (cit. on p. 99).
Abascal, J. F. P. J., N. Ducros, and F. Peyrin (Dec. 2018). "Nonlinear material decomposition using a regularized iterative scheme based on the Bregman distance". In: *Inverse Probl.* ISSN: 0266-5611. DOI: 10.1088/1361-6420/aae1e7. URL: http://stacks.iop.org/0266-5611/34/i=12/a=124003?key=crossref.08a054bc65014d6b99846d5683714ee3 (cit. on p. 157).
Adler, Jonas and Sebastian Lunz (2018). "Banach Wasserstein GAN". In: *Advances in Neural Information Processing Systems (NIPS 2018)*, pp. 6754–6763 (cit. on p. 104).
Adler, Jonas and Ozan Öktem (2018). "Learned Primal-Dual Reconstruction". In: *IEEE Trans. on Medical Imaging* 37.6, pp. 1322–1332. DOI: 10.1109/TMI.2018.2799231 (cit. on pp. 96, 132, 143, 146).
Antholzer, Stephan, Markus Haltmeier, and Johannes Schwab (2018). *Deep Learning for Photoacoustic Tomography from Sparse Data*. arXiv: 1704.04587 [cs.CV] (cit. on p. 92).
Buzzard, Gregery T. et al. (2018). "Plug-and-Play Unplugged: Optimization-Free Reconstruction Using Consensus Equilibrium". In: *SIAM Journal on Imaging Sciences* 11.3, pp. 2001–2020. DOI: 10.1137/17M1122451. eprint: https://doi.org/10.1137/17M1122451. URL: https://doi.org/10.1137/17M1122451 (cit. on p. 94).

Ding, Q. et al. (2018). "Image-domain multimaterial decomposition for dual-energy CT based on prior information of material images". In: *Med. Phys.* ISSN: 2473-4209. DOI: 10.1002/MP.13001. URL: https://pubmed.ncbi.nlm.nih.gov/29807395/ (cit. on p. 157).
Hauptmann, Andreas et al. (2018). "Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography". In: *IEEE Trans. on Medical Imaging* 37.6, pp. 1382–1393. DOI: 10.1109/TMI.2018.2820382 (cit. on pp. 32, 96, 104, 132, 136, 143, 145, 159).
Huang, Gao, Shichen Liu, et al. (2018). *CondenseNet: An Efficient DenseNet using Learned Group Convolutions*. arXiv: 1711.09224 [cs.CV] (cit. on p. 106).
Kalmoun, El Mostafa (2018). "An Investigation of Smooth TV-Like Regularization in the Context of the Optical Flow Problem". In: *Journal of Imaging* 4.2. ISSN: 2313-433X. DOI: 10.3390/jimaging4020031. URL: https://www.mdpi.com/2313-433X/4/2/31 (cit. on p. 119).
Kazantsev, Daniil et al. (2018). "TomoPhantom, a software package to generate 2D–4D analytical phantoms for CT image reconstruction algorithm benchmarks". In: *SoftwareX* 7, pp. 150–155. DOI: https://doi.org/10.1016/j.softx.2018.05.003 (cit. on p. 107).
M.Pelt, D. and J. A. Sethian (2018). "A mixed-scale dense convolutional neural network for image analysis". In: *Proceedings of the National Academy of Sciences USA* 115, pp. 254–259 (cit. on pp. 104, 106).
Metzler, Christopher A. et al. (2018). "prDeep: Robust Phase Retrieval with Flexible Deep Neural Networks". In: *CoRR* abs/1803.00212. arXiv: 1803.00212. URL: http://arxiv.org/abs/1803.00212 (cit. on pp. 100, 132).
Miyato, Takeru et al. (2018). "Spectral Normalization for Generative Adversarial Networks". In: *ArXiv* abs/1802.05957. URL: https://api.semanticscholar.org/CorpusID:3366315 (cit. on p. 94).
Pelt, D. M., K. J. Batenburg, and J. A. Sethian (2018). "Improving tomographic reconstrucion from limited data using mixed-scale dense convolutional neural networks". In: *Journal of Imaging* 4, p. 128 (cit. on pp. 104, 106).
Wang, Zhengyang and Shuiwang Ji (2018). "Smoothed Dilated Convolutions for Improved Dense Prediction". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, pp. 2486–2495 (cit. on p. 106).
Wu, Sitong et al. (2018). "Direct Reconstruction of Ultrasound Elastography Using an End-to-End Deep Neural Network". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by Alejandro F. Frangi et al. Cham: Springer International Publishing, pp. 374–382. ISBN: 978-3-030-00928-1 (cit. on p. 92).
Yu, Boliang et al. (2018). "Evaluation of phase retrieval approaches in magnified X-ray phase nano computerized tomography applied to bone tissue". In: *Optics Express, Optical Society of America* 26 (9), pp. 11110–11124 (cit. on pp. 59, 60).
Zhang, Jian and Bernard Ghanem (2018). "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing". In: *CVPR*, pp. 1828–1837 (cit. on p. 144).
Arridge, S. et al. (May 2019). "Solving inverse problems using data-driven models". en. In: *Acta Numer.* ISSN: 0962-4929, 1474-0508. DOI: 10.1017/S0962492919000059. URL: https://www.cambridge.org/core/journals/acta-numerica/article/solving-inverse-problems-using-datadriven-models/CE5B3725869AEAF46E04874115B0AB15# (visited on 01/27/2023) (cit. on pp. 4, 10, 16, 81, 104, 132, 159).
Bai, Chen et al. (2019). "Robust contrast-transfer-function phase retrieval via flexible deep learning networks". In: *Optics Letters* 44 (21), pp. 5141–5144 (cit. on pp. 100, 101, 104, 132).
Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun (2019). "Deep Equilibrium Models". In: *ArXiv* abs/1909.01377. URL: https://api.semanticscholar.org/CorpusID:202539738 (cit. on p. 96).

Dabre, Raj and Atsushi Fujita (July 2019). "Recurrent Stacking of Layers for Compact Neural Machine Translation Models". In: *AAAI Conference on Artificial Intelligence* 33, pp. 6292–6299. DOI: 10.1609/aaai.v33i01.33016292 (cit. on p. 133).
Geron, Aurelien (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd. O'Reilly Media, Inc. ISBN: 1492032646 (cit. on p. 81).
Işil, Ca**u**gatay, Figen S. Oktem, and Aykut Koç (July 2019). "Deep iterative reconstruction for phase retrieval". In: *Appl. Opt.* 58.20, pp. 5422–5431. DOI: 10.1364/AO.58.005422. URL: http://opg.optica.org/ao/abstract.cfm?URI=ao-58-20-5422 (cit. on pp. 98, 132).
Ryu, Ernest K. et al. (2019). "Plug-and-Play Methods Provably Converge with Properly Trained Denoisers". In: *International Conference on Machine Learning* (cit. on pp. 94, 97).
Sun, Ke et al. (2019). "Deep High-Resolution Representation Learning for Human Pose Estimation". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5686–5696. DOI: 10.1109/CVPR.2019.00584 (cit. on p. 105).
Sun, Yu, Brendt Wohlberg, and Ulugbek S. Kamilov (2019). "An Online Plug-and-Play Algorithm for Regularized Image Reconstruction". In: *IEEE Transactions on Computational Imaging* 5.3, pp. 395–408. DOI: 10.1109/TCI.2019.2893568 (cit. on p. 94).
Sun, Yu, Shiqi Xu, et al. (2019). "Regularized Fourier Ptychography Using an Online Plug-and-play Algorithm". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7665–7669. DOI: 10.1109/ICASSP.2019.8683057 (cit. on p. 95).
Wu, Zifeng, Chunhua Shen, and Anton van den Hengel (2019). "Wider or Deeper: Revisiting the ResNet Model for Visual Recognition". In: *Pattern Recognition* 90, pp. 119–133. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2019.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0031320319300135 (cit. on p. 105).
Aragón Artacho, Francisco Javier, Rubén Campoy, and Matthew Tam (Apr. 2020). "The Douglas–Rachford algorithm for convex and nonconvex feasibility problems". In: *Mathematical Methods of Operations Research* 91. DOI: 10.1007/s00186-019-00691-9 (cit. on p. 66).
Bredies, Kristian and Tuomo Valkonen (2020). *Inverse problems with second-order Total Generalized Variation constraints*. arXiv: 2005.09725 [math.NA] (cit. on p. 118).
Günther, Benedikt et al. (Sept. 2020). "The versatile X-ray beamline of the Munich Compact Light Source: design, instrumentation and applications". In: *Journal of Synchrotron Radiation* 27.5, pp. 1395–1414. DOI: 10.1107/S1600577520008309. URL: https://doi.org/10.1107/S1600577520008309 (cit. on p. 1).
He, Ji, Yongbo Wang, and Jianhua Ma (2020). "Radon Inversion via Deep Learning". In: *IEEE Transactions on Medical Imaging* 39.6, pp. 2076–2087. DOI: 10.1109/TMI.2020.2964266 (cit. on p. 91).
Liu, Chang et al. (2020). "Robustness Investigation on Deep Learning CT Reconstruction for Real-Time Dose Optimization". In: *ArXiv* abs/2012.03579. URL: https://api.semanticscholar.org/CorpusID:227340372 (cit. on p. 91).
Manekar, Raunak et al. (2020). "End to end learning for phase retrieval". In: *ICML workshop on ML Interpretability for Scientific Discovery* (cit. on p. 99).
Naimipour, Naveed, Shahin Khobahi, and Mojtaba Soltanalian (2020). "Unfolded algorithms for deep phase retrieval". In: *arXiv preprint arXiv:2012.11102* (cit. on p. 100).
Ongie, Gregory et al. (2020). "Deep Learning Techniques for Inverse Problems in Imaging". In: *IEEE Journal on Selected Areas in Information Theory* 1.1, pp. 39–56. DOI: 10.1109/JSAIT.2020.2991563 (cit. on pp. 16, 81, 94, 131).
Vásárhelyi, L. et al. (2020). "Microcomputed tomography–based characterization of advanced materials: a review". In: *Materials Today Advances* 8, p. 100084. ISSN: 2590-0498. DOI: https:

//doi.org/10.1016/j.mtadv.2020.100084. URL: https://www.sciencedirect.com/science/article/pii/S259004982030031X (cit. on pp. 1, 7).
Vishnevskiy, Valery, Jonas Walheim, and Sebastian Kozerke (Apr. 2020). "Deep variational network for rapid 4D flow MRI reconstruction". In: *Nature Machine Intelligence* 2. DOI: 10.1038/s42256-020-0165-6 (cit. on pp. 155, 161).
Yang, Yan et al. (2020). "ADMM-CSNet: A Deep Learning Approach for Image Compressive Sensing". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.3, pp. 521–538. DOI: 10.1109/TPAMI.2018.2883941 (cit. on pp. 96, 132, 143).
Zhang, Hai-Miao and Bin Dong (Jan. 2020). "A Review on Deep Learning in Medical Image Reconstruction". In: *Journal of the Operations Research Society of China* 8.2, pp. 311–340. DOI: 10.1007/s40305-019-00287-4. URL: https://doi.org/10.1007%2Fs40305-019-00287-4 (cit. on p. 81).
Celestre, Rafael (Feb. 2021). "Investigations of the effect of optical imperfections on partially coherent X-ray beam by combining optical simulations with wavefront sensing experiments". PhD thesis. DOI: 10.13140/RG.2.2.28752.02569 (cit. on pp. 12, 47).
Combettes, Patrick L. and Jean-Christophe Pesquet (2021). "Fixed Point Strategies in Data Science". In: *IEEE Transactions on Signal Processing* 69, pp. 3878–3905. DOI: 10.1109/TSP.2021.3069677 (cit. on p. 74).
Dutta, Kaushik (2021). "Densely Connected Recurrent Residual (Dense R2UNet) Convolutional Neural Network for Segmentation of Lung CT Images". In: *ArXiv* abs/2102.00663 (cit. on p. 106).
Gilton, Davis, Gregory Ongie, and Rebecca Willett (2021). "Deep Equilibrium Architectures for Inverse Problems in Imaging". In: *IEEE Transactions on Computational Imaging* 7, pp. 1123–1133. DOI: 10.1109/TCI.2021.3118944 (cit. on pp. 36, 96, 97, 156).
Heaton, Howard et al. (2021). *Feasibility-based Fixed Point Networks*. arXiv: 2104.14090 (cit. on p. 97).
Langer, Max, Yuhe Zhang, et al. (July 2021). "*PyPhase* – a Python package for X-ray phase imaging". In: *Journal of Synchrotron Radiation* 28.4, pp. 1261–1266. DOI: 10.1107/S1600577521004951. URL: https://doi.org/10.1107/S1600577521004951 (cit. on pp. 5, 11, 109, 167).
Monga, Vishal, Yuelong Li, and Yonina C. Eldar (2021). "Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing". In: *IEEE Signal Proc. Magazine* 38.2, pp. 18–44. DOI: 10.1109/MSP.2020.3016905 (cit. on pp. 96, 132, 139, 143, 159).
Pesquet, Jean-Christophe et al. (2021). "Learning Maximally Monotone Operators for Image Recovery". In: *SIAM Journal on Imaging Sciences* 14.3, pp. 1206–1237. DOI: 10.1137/20M1387961. eprint: https://doi.org/10.1137/20M1387961. URL: https://doi.org/10.1137/20M1387961 (cit. on pp. 36, 86, 94, 144, 156).
Zhang, Yuhe et al. (June 2021). "PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets". In: *Opt. Express* 29.13, pp. 19593–19604. DOI: 10.1364/OE.423222. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-29-13-19593 (cit. on pp. 99, 104, 116, 132).
Alloo, S. J. et al. (Apr. 2022). "Tomographic phase and attenuation extraction for a sample composed of unknown materials using x-ray propagation-based phase-contrast imaging". In: *Opt. Lett.* 47.8, pp. 1945–1948. DOI: 10.1364/OL.445802. URL: http://opg.optica.org/ol/abstract.cfm?URI=ol-47-8-1945 (cit. on p. 129).
Deshpande, Rucha et al. (2022). "Investigating the robustness of a deep learning-based method for quantitative phase retrieval from propagation-based x-ray phase contrast measurements under laboratory conditions". In: *Physics in Medicine & Biology* 68. URL: https://api.semanticscholar.org/CorpusID:253265293 (cit. on pp. 16, 81).

Eguizabal, A., O. Öktem, and M. U. Persson (Aug. 2022). "Deep Learning for Material Decomposition in Photon-Counting CT". In: *arXiv:2208.03360* (cit. on p. 157).
Gugel, Leon and Shai Dekel (2022). "PR-DAD: Phase Retrieval Using Deep Auto-Decoders". In: *2022 7th International Conference on Frontiers of Signal Processing (ICFSP)*, pp. 165–172. DOI: 10.1109/ICFSP55781.2022.9924739 (cit. on p. 98).
Hoop, Maarten V de, Matti Lassas, and Christopher A Wong (2022). "Deep learning architectures for nonlinear operator functions and nonlinear inverse problems". In: *Mathematical Statistics and Learning* 4.1, pp. 1–86 (cit. on pp. 16, 81).
Kalbfleisch, Sebastian et al. (Jan. 2022). "X-ray in-line holography and holotomography at the NanoMAX beamline". In: *Journal of Synchrotron Radiation* 29. DOI: 10.1107/S1600577521012200 (cit. on pp. 125, 139).
Li, Samuel Z. et al. (2022). "Shallow U-Net deep learning approach for phase retrieval in propagation-based phase-contrast Imaging". In: *Developments in X-Ray Tomography XIV*. Ed. by Bert Müller and Ge Wang. Vol. 12242. International Society for Optics and Photonics. SPIE, 122421Q. DOI: 10.1117/12.2644579. URL: https://doi.org/10.1117/12.2644579 (cit. on p. 98).
Mom, K., M. Langer, and B. Sixou (2022). "Apprentissage de méthodes itératives pour l'imagerie de contraste de phase des rayons X". In: *GRETSI 2022* (cit. on p. 159).
Mom, Kannara, Max Langer, and Bruno Sixou (Sept. 2022a). "Apprentissage de méthodes itératives pour l'imagerie de contraste de phase des rayons X". In: *XXVIIIème Colloque Francophone de Traitement du Signal et des Images (Gretsi 2022)*. Nancy, France. URL: https://hal.science/hal-03706103 (cit. on pp. 5, 11, 167).
– (Oct. 2022b). "Nonlinear primal–dual algorithm for the phase and absorption retrieval from a single phase contrast image". In: *Opt. Lett.* 47.20, pp. 5389–5392. DOI: 10.1364/OL.469174. URL: https://opg.optica.org/ol/abstract.cfm?URI=ol-47-20-5389 (cit. on pp. 5, 11, 167).
– (Apr. 2022c). "Nonlinear primal-dual method for X-ray in-line phase contrast imaging". In: *SPIE Photonics Europe 2022*. Vol. 12136. Proceedings of SPIE. Strasbourg, France. DOI: 10.1117/12.2619732. URL: https://hal.science/hal-03706059 (cit. on p. 167).
Mom, Kannara, Bruno Sixou, and Max Langer (Apr. 2022). "Mixed scale dense convolutional networks for x-ray phase contrast imaging". In: *Appl. Opt.* 61.10, pp. 2497–2505. DOI: 10.1364/AO.443330. URL: http://opg.optica.org/ao/abstract.cfm?URI=ao-61-10-2497 (cit. on pp. 5, 11, 116, 132, 135, 136, 167).
Vial, Pierre-Hugo et al. (2022). "Learning the proximity operator in unfolded admm for phase retrieval". In: *IEEE Signal Processing Letters* 29, pp. 1619–1623 (cit. on p. 101).
Wu, Yue et al. (2022). "Enhanced phase retrieval via deep concatenation networks for in-line X-ray phase contrast imaging". In: *Physica Medica* 95, pp. 41–49. ISSN: 1120-1797. DOI: https://doi.org/10.1016/j.ejmp.2021.12.017. URL: https://www.sciencedirect.com/science/article/pii/S1120179721003719 (cit. on p. 98).
Xiang, Mingjun et al. (2022). *Amplitude/Phase Retrieval for Terahertz Holography with Supervised and Unsupervised Physics-Informed Deep Learning*. arXiv: 2212.06725 [physics.optics] (cit. on p. 99).
Ye, Qiuliang, Li-Wen Wang, and Daniel P. K. Lun (Aug. 2022). "SiSPRNet: end-to-end learning for single-shot phase retrieval". In: *Opt. Express* 30.18, pp. 31937–31958. DOI: 10.1364/OE.464086. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-18-31937 (cit. on p. 98).
Ye, Qiuliang, Li-Wen Wang, and Daniel Pak-Kong Lun (2022). *Towards Practical Single-shot Phase Retrieval with Physics-Driven Deep Neural Network*. arXiv: 2208.08604 [eess.IV] (cit. on p. 99).

Mom, Kannara, Max Langer, and Bruno Sixou (Mar. 2023). "Deep Gauss-Newton for phase retrieval". In: *Opt. Lett.* 48.5, pp. 1136–1139. DOI: 10.1364/OL.484862. URL: https://opg.optica.org/ol/abstract.cfm?URI=ol-48-5-1136 (cit. on pp. 5, 11, 143, 167).

Mom, Kannara, Jerome Lesaint, et al. (Aug. 2023). "Décomposition en matériaux de base par apprentissage profond pour la tomographie spectrale". In: *XXIXème Colloque Francophone de Traitement du Signal et des Images (Gretsi 2023)*. Grenoble, France (cit. on pp. 5, 12, 156, 167).

Yang, Yuchi et al. (2023). "HIONet: Deep priors based deep unfolded network for phase retrieval". In: *Digital Signal Processing* 132, p. 103797. ISSN: 1051-2004. DOI: https://doi.org/10.1016/j.dsp.2022.103797. URL: https://www.sciencedirect.com/science/article/pii/S1051200422004146 (cit. on p. 101).

Maretzke, Simon (n.d.). "Inverse Problems in Propagation-based X-ray Phase Contrast Imaging and Tomography: Stability Analysis and Reconstruction Methods". PhD thesis. der Georg-August University School of Science (GAUSS) (cit. on pp. 55, 56, 58).

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : MOM **DATE de SOUTENANCE** : 20/11/2023

Prénoms : Kannara

TITRE : Deep learning based phase retrieval for X-ray phase contrast imaging

NATURE : Doctorat **Numéro d'ordre** : 2023ISAL0087

Ecole doctorale : Electronique, Electrotechnique, Automatique (EEA)

Spécialité : Traitement du Signal et de l'Image

RESUME : The development of highly coherent X-ray sources, such as third-generation synchrotron radiation facilities, has significantly contributed to the advancement of phase contrast imaging. The high degree of coherence of these sources enables efficient implementation of phase contrast techniques, and can increase sensitivity by several orders of magnitude.
This novel imaging technique has found applications in a wide range of fields, including material science, paleontology, bone research, medicine, and biology. It enables the imaging of samples with low absorption constituents, where traditional absorption-based methods may fail to provide sufficient contrast. Several phase-sensitive imaging techniques have been developed, among them, propagation-based imaging requires no equipment other than the source, object and detector. Although the intensity can be measured at one or several propagation distances, the phase information is lost and must be estimated from those diffraction patterns, a process called phase retrieval. Phase retrieval in this context is a nonlinear ill-posed inverse problem. Various classical methods have been proposed to retrieve the phase, either by linearizing the problem to obtain an analytical solution, or by iterative algorithms. The main purpose of this thesis was to study what new deep learning approaches could bring to this phase retrieval problem. Various deep learning algorithms have been proposed and evaluated to address this problem. In particular, the case of a single distance, while taking into account the non-linear information of the direct model, was considered.

In the first part of this work, we show how neural networks can be used to reconstruct directly from measurements data, without model information. The architecture of the Mixed Scale Dense Network (MS-D Net) is introduced, combining dilated convolution and dense connection.
In the second part of this thesis, we propose a nonlinear primal–dual algorithm for the retrieval of phase shift and absorption from a single X-ray in-line phase contrast. We showed that choosing different regularizers for absorption and phase can improve the reconstructions.
In the third part, we propose to integrate neural networks into an existing optimization scheme using so-called unrolling approaches, in order to give the convolutional neural networks a specific role in the reconstruction.

The performance of theses algorithms are evaluated using simulated noisy data as well as images acquired at NanoMAX (MAX IV, Lund, Sweden). Most of these algorithms are available through the *Pyphase* package.

MOTS-CLÉS : Deep learning, iterative methods, Fresnel diffraction, inverse problem, nonlinear problem, X-ray imaging, nonlinear optimization, unrolling methods

Laboratoire (s) de recherche : Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé (CREATIS)

Directeur de thèse:

Sixou, Bruno	Maître de conférence, INSA Lyon	Directeur de thèse

Composition du jury :

Soussen, Charles	Professeur, CentralSupélec	Rapporteur
Bousse, Alexandre	Chargé de recherche, LaTIM, INSERM	Rapporteur
Denis, Loïc	Professeur, Université de Saint-étienne	Examinateur
Desbat, Laurent	Professeur, Université Grenoble Alpes	Examinateur
Pustelnik, Nelly,	Chargée de recherche, CNRS CRCN, ENS de Lyon	Examinatrice
Rolland Du Roscoat, Sabine	Professeur, Université Grenoble Alpes	Examinatrice
Villanueva Perez, Pablo	Associate senior lecturer, Lund university	Examinateur
Langer, Max	Chargé de recherche, CNRS, TIMC	Invité