Nonlinear primal-dual method for X-ray in-line phase contrast imaging

Kannara Mom, Max Langer, Bruno Sixou

▶ **To cite this version:**

Kannara Mom, Max Langer, Bruno Sixou. Nonlinear primal-dual method for X-ray in-line phase contrast imaging. SPIE Photonics Europe 2022, Apr 2022, Strasbourg, France. 10.1117/12.2619732 . hal-03706059

HAL Id: hal-03706059
https://hal.science/hal-03706059v1

Submitted on 4 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonlinear primal-dual method for X-ray in-line phase contrast imaging.

Kannara Mom[a], Max Langer[b], and Bruno Sixou[a]

[a]Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69621 Villeurbanne, France
[b]Université Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, F-38000 Grenoble, France

ABSTRACT

The in-line X-ray phase contrast imaging technique relies on the measurement of Fresnel diffraction intensity patterns due to the phase shift and the absorption induced by the object. The recovery of both phase and absorption is an ill-posed non-linear inverse problem. In this work, we address this problem with an iterative algorithm based on a primal-dual method, which allows us to introduce the non-linearity of the forward operator. We used a variational approach with different regularizations for the phase and absorption, in order to take into account the specificities of each quantity. Assuming the solution to be piecewise constant, the functional used involves the Total Generalized Variation (TGV) as well as the classical Total Variation (TV), which enables a compromise between sharp discontinuities and smoothness in the solution. This optimization problem is solved efficiently using primal-dual approach such as Primal-Dual Hybrid Gradient Method (PDHGM). From this approach, we propose an algorithm called PDGHGM-CTF, which is based on the linearized Contrast Transfer Function model, that we generalize for the nonlinear problem to get the Non-Linear Primal-Dual Hybrid Gradient Method (NL-PDHGM). The proposed iterative algorithms are able to recover simultaneously the phase and absorption from a single diffraction pattern without homogeneity assumption or support constraint, and the nonlinear version is valid without restriction on the object. Moreover, we show that the approach is robust with respect to the initialization. While giving a good approximation as starting point reduced the convergence time, it did not improve the reconstruction results. We demonstrate the potential of the proposed algorithms on simulated datasets. We show that it produces reconstructions with fewer artifacts and improved normalized mean squared error compared to a gradient descent scheme. We evaluate the robustness of the proposed algorithm by evaluating the reconstruction on simulated images of 1 000 random objects, given the same hyperparameters.

Keywords: Inverse problems, Iterative methods, X-ray diffraction, Holography, Non-linear operator.

1. INTRODUCTION

Phase-sensitive X-ray imaging techniques have developed significantly recently due to the high sensitivity offered by phase contrast imaging,[1] which has several applications in material science[2] and biomedical imaging.[3–7] When using sufficiently coherent X-ray beams, phase contrast can be achieved by letting the wave propagate into free space after interaction with the sample.[8] The relationship between the absorption and phase shift induced by a sample and the diffraction pattern relies on the Fresnel diffraction theory. The phase information is not explicitly recorded in the measured intensity and must be estimated from the diffraction patterns, a processus called phase retrieval. This phase retrieval problem sets an inverse nonlinear ill-posed inverse problem. 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),[9,10] Contrast Transfer Function (CTF)[11] or on the Mixed approach[12] 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

Further author information: (Send correspondence to Kannara Mom)
Kannara Mom: E-mail: kannara.mom@creatis.insa-lyon.fr

the detector and the object space.[13–18] These also include variational methods based on the Fréchet derivative of the forward operator in conjunction with the Landweber algorithm.[19] This kind of algorithm enables a flexible inclusion of priors, such as Tikhonov, Sobolev or sparsity regularization.[20,21] In order to include prior such as the Total Variation (TV) semi-norm, primal-dual schemes like Alternating Direction Method of Multipliers (ADMM)[22] have been studied for the phase retrieval problem but rely on linearization of the forward model, either using TIE[23] or CTF.[24] The single-distance inverse problem is more severely ill-posed than the classical problems with several diffraction patterns[25,26] 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[27] and PhaseGAN[28] 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.

In this work, we investigate a primal-dual approach based on the Primal-Dual Hybrid Gradient Method (PDHGM)[29] 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 (PDHGM-CTF) and then generalize it to the nonlinear case based on the NonLinear Primal-Dual Hybrid Gradient Method (NL-PDHGM).[30] 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. The paper is organized as follows, we first detail the image formation and the CTF method, then, we present the different methods, including primal-dual hybrid gradient method and the Total Generalized Variation. The last section details the simulation methodology, and the results obtained on those simulated data.

2. IMAGE FORMATION AND INVERSE PROBLEM

2.1 Direct problem definition

The interaction of coherent and parallel X-ray beam with an object is related to its complex refractive index:

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

where δ_r is the refractive index decrement and β is the absorption index for the spatial coordinate (x,y,z). Both δ_r and β depend on the material as well as the X-ray wavelength λ. The phase shift and the absorption are projections of the refraction and absorption index respectively defined with the following line integrals:

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

$$\varphi(\mathbf{x}) = \frac{2\pi}{\lambda}\int (1-\delta_r(\mathbf{x},z))\mathrm{d}z \tag{3}$$

where $\mathbf{x} = (x,y)$ represents the spatial coordinates in the perpendicular plane to the propagation direction z. For thin objects and straight-line propagation of the beam, the interaction of coherent and parallel X-ray beam with matter can be described by a transmittance function T :

$$T(\mathbf{x}) = \exp[-B(\mathbf{x}) + i\varphi(\mathbf{x})] \tag{4}$$

where $B(\mathbf{x})$ is the absorption and $\varphi(\mathbf{x})$ the phase shift induced by the object.

In the framework of the Fresnel diffraction theory, letting the beam propagates in free space over a relatively short distance D after interaction with the object can be described as a 2D convolution of the transmittance and of the Fresnel propagator for a distance D :

$$u_D(\mathbf{x}) = T(\mathbf{x}) * P_D(\mathbf{x}) \tag{5}$$

where

$$P_D(\mathbf{x}) = \frac{1}{i\lambda D} \exp(i\frac{\pi}{\lambda D}|\mathbf{x}|^2) \tag{6}$$

The intensity measured at a distance D downstream of the object is thus given by

$$I_D(\mathbf{x}) = |u_D(\mathbf{x})|^2 \tag{7}$$

The forward model describing the nonlinear relationship between the absorption and phase shift induced by a sample and the diffraction pattern can be written as

$$F_D\left(B, \varphi\right) = |\mathrm{e}^{-B+i\varphi} * P_D|^2 \tag{8}$$

Estimating the phase shift (resp. absorption) from these intensities, or diffraction patterns, is called phase (resp. absorption) retrieval. Our aim is to estimate both the phase and the absorption from a single intensity measurement.

2.2 Contrast Transfer function

The contrast transfer function method is based on the assumption of weak absorption and slowly varying phase shift :

$$B(\mathbf{x}) \ll 1, \quad |\varphi(\mathbf{x}) - \varphi(\mathbf{x} + \lambda D\mathbf{f})| \ll 1 \tag{9}$$

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

$$T(\mathbf{x}) \approx 1 - B(\mathbf{x}) + i\varphi(\mathbf{x}) \tag{10}$$

Substituting this approximation into (7) and again keeping only first order terms gives

$$\tilde{I}_D(\mathbf{f}) = \delta(\mathbf{f}) - 2\cos(\pi\lambda D\,|\mathbf{f}|^2)\tilde{B}(\mathbf{f}) + 2\sin(\pi\lambda D\,|\mathbf{f}|^2)\tilde{\varphi}(\mathbf{f}) \tag{11}$$

where $\mathbf{f}$ is the variable in the Fourier domain, $\delta(\mathbf{f})$ is the unit impulse function, $\tilde{B}(\mathbf{f})$ is the Fourier transform of the absorption and $\tilde{\varphi}(\mathbf{f})$ is the Fourier transform of the phase. The CTF-linearized forward model can thus be written as

$$F_D^{\mathrm{CTF}}\left(B, \varphi\right) = \left\{\mathcal{F}^{-1}\left(-2\cos(\pi\lambda D\,|\mathbf{f}|^2)\ ;\ 2\sin(\pi\lambda D\,|\mathbf{f}|^2)\right)\mathcal{F}\right\}(B, \varphi) \tag{12}$$

Since the phase contrast factor before $\tilde{\varphi}(\mathbf{f})$ in (11) has zero crossings, several distances have to be used in order to cover as much of the Fourier domain as possible.

3. METHODS

Different approaches have been proposed to simultaneously recover the phase and absorption from an image at a single distance but none of them proposed to use Total Variation (TV) without homogeneity assumption. Approaches based on the ADMM algorithm have been studied to incorporate TV regularization, but they rely on the linearization of the problem, either by TIE[23] or by CTF.[24] We propose to overcome these constraints by using the Primal-Dual Hybrid Gradient Method (PDHGM).[29] We show that this method, in the framework of the CTF linearized problem, works well even in the case of a single distance, and allows us to penalize the absorption and phase with the Total Variation and its higher order generalization. Finally, we use a generalization of this approach for nonlinear operators, known as NonLinear Primal-Dual Hybrid Gradient Method (NL-PDHGM), which enables us to take into account the nonlinearity information of the model.

3.1 The twofold problem of absorption and phase retrieval

In this work, we consider reconstruction couples that are solutions of the following optimization problem:

$$(B^*, \varphi^*) = \underset{B,\varphi}{\operatorname{argmin}} \left\{ d\left[I_D^{\text{obs}}, F_D(B,\varphi)\right] + R(B,\varphi) \right\} \tag{13}$$

where I_D^{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.

3.2 Gradient descent with smooth Total Variation (GD-TV$^\epsilon$)

The general idea is to consider the problem (13) with the loss function d to be the l^2 squared norm, and the regularization R to be the Total Variation, we then seek to minimize the following functional :

$$\min_{B,\varphi} \left\{ \frac{1}{2} \int_\Omega \left[F_D(B,\varphi)(x) - I_D^{\text{obs}}(x)\right]^2 \mathrm{d}x + \eta \int_\Omega |\mathcal{D}B| + \mu \int_\Omega |\mathcal{D}\varphi| \right\} \tag{14}$$

where $\Omega \subset \mathbb{R}^2$ is the image domain and $\eta, \mu > 0$ are regularization parameters. The term $\int_\Omega |\mathcal{D}\varphi|$ is the Total Variation of the image φ. For a function $\varphi \in L^1(\Omega)$, the Total Variation semi-norm can be defined as :

$$\int_\Omega |\mathcal{D}\varphi| := \sup \left\{ -\int_\Omega \varphi(x)\ \mathrm{div}\psi(x)\mathrm{d}x\ ;\ \psi \in \mathcal{C}_c^\infty\left(\Omega, \mathbf{R}^2\right), |\psi(x)| \leq 1, \forall x \in \Omega \right\} \tag{15}$$

For functions smooth enough $\varphi \in W^{1,1}(\Omega)$, or equivalently $\nabla\varphi \in L^1(\Omega)$, this quantity reduces to $\int_\Omega |\nabla\varphi| \mathrm{d}x$. It is well known that using Total Variation as regularization allows to reconstruct images while preserving sharp edges. The point here is that the gradient of the Total Variation semi-norm is given by $\mathrm{div}\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right)$ which is not defined at a pixel x if $\nabla\varphi(x) = 0$. We avoid this problem by considering a smooth version of the TV:

$$\text{TV}^\epsilon(\varphi) := \int_\Omega \sqrt{\epsilon^2 + |\nabla\varphi(x)|^2}\mathrm{d}x \approx \int_\Omega |\mathcal{D}\varphi| \quad \text{and} \quad \nabla\text{TV}^\epsilon(\varphi) = \mathrm{div}\left(\frac{\nabla\varphi}{\sqrt{\epsilon^2 + |\nabla\varphi|^2}}\right) \tag{16}$$

Combining all this, we have to minimize the following energy

$$J_{\eta,\mu,\epsilon}(B,\varphi) = \frac{1}{2}\left|\left|F_D(B,\varphi) - I_D^{\text{obs}}\right|\right|_2^2 + \eta\text{TV}^\epsilon(B) + \mu\text{TV}^\epsilon(\varphi) \tag{17}$$

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$^\epsilon$).

The term $[F_D'(B,\varphi)]^*$ in algorithm 1 represents the adjoint of the Fréchet derivative

$$F_D'(B,\varphi) = \begin{bmatrix} \partial_B F_D(B,\varphi) \\ \partial_\varphi F_D(B,\varphi) \end{bmatrix} \tag{18}$$

of the forward operator at the point (B,φ), of which we have an analytical expression.[19]

Algorithm 1 GD-TV$^\epsilon$

Given :

- step size τ, smooth factor ϵ and regularization parameters η, μ
- $(B_0, \varphi_0) \in \mathbf{R}^{n\times m} \times \mathbf{R}^{n\times m}$

for $i = 0, \dots, N_{\text{iter}}$ **do** :

$$\begin{aligned}(B_{i+1}, \varphi_{i+1}) &\leftarrow (B_i, \varphi_i) - \tau \nabla J_{\eta,\mu,\epsilon}(B_i, \varphi_i) \\ &= (B_i, \varphi_i) - \tau \left\{ [F_D'(B_i, \varphi_i)]^* \left(F_D(B_i, \varphi_i) - I_D^{\text{obs}} \right) + [\eta \nabla \text{TV}^\epsilon(B_i), \mu \nabla \text{TV}^\epsilon(\varphi_i)] \right\} \\ (B_{i+1}, \varphi_{i+1}) &\leftarrow [(B_{i+1}, \varphi_{i+1})]_+\end{aligned}$$

3.3 Primal Dual Hybrid Gradient Method based on CTF linearization (PDHGM-CTF)

The main drawback of the GD-TV$^\epsilon$ approach is that one cannot efficiently implement the TV semi-norm with a simple gradient descent without having to smooth it. To overcome this problem, we can use a primal-dual approach. A well-known algorithm for linear inverse problems is the Chambolle-Pock algorithm,[29] aka Primal Dual Hybrid Gradient Method (PDHGM)[31] which aims to solve a saddle-point problem of the form

$$\min_{x\in X} \max_{y\in Y} \{\langle \mathcal{K}x, y\rangle + \mathcal{G}(x) - \mathcal{F}^*(y)\} \tag{19}$$

where X is the primal space, Y the dual space, $\mathcal{K} : X \to Y$ a continuous linear operator, $\mathcal{F}$ and $\mathcal{G}$ are convex and possibly non-smooth functions, and $\mathcal{F}^*$ represents the Fenchel conjugate of $\mathcal{F}$. This saddle-point problem (19) is a primal-dual formulation of the primal problem

$$\min_{x\in X} \{\mathcal{F}(\mathcal{K}x) + \mathcal{G}(x)\} \tag{20}$$

The PDHGM algorithm iteratively solves (20) by alternating proximal gradient ascent in the dual space and proximal gradient descent in the primal space (see algorithm 2).

In order to avoid the staircase effects that can appear on affine parts of the reconstructed absorption B while still preserving the possibility of having sharp edges, we decided to use a generalization of the Total Variation, the so-called Total Generalized Variation (TGV).[32] More precisely, we use second-order Total Generalized Variation (TGV^2)[33] :

$$\text{TGV}^2_{(\alpha,\beta)}(B) = \inf_{v\in X} \left\{ \alpha \int_\Omega |\mathcal{D}v| + \beta \int_\Omega |\mathcal{D}B - v| \right\} \tag{21}$$

with $v = (v_1, v_2)$ where v_1 and v_2 are functions with finite Total Variation on Ω. The idea is to force the vector field v to have a sparse gradient and to penalize the gradient $\mathcal{D}B$ 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 $\frac{\alpha}{\beta}$.

Let's now formulate our problem as in (20), for the absorption B, we choose to use second order Total Generalized Variation regularization, and for the phase φ, Total Variation regularization. As PDHGM algorithm is only valid for linear operator, we focus on CTF-linearized problem, using the CTF-forward operator F_D^{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 :

$$\min_{B,\varphi} \inf_v \left\{ \frac{1}{2} \int_\Omega \left[F_D^{\text{CTF}}(B, \varphi)(x) - I_D^{\text{obs}}(x) \right]^2 \mathrm{d}x + \alpha \int_\Omega |\nabla v| + \beta \int_\Omega |\nabla B - v| + \nu \int_\Omega |\nabla \varphi| + \iota_+(B, \varphi) \right\} \tag{22}$$

where ι_+ denotes the indicator function which constraint the sought solutions to be positive :

$$\iota_+(B, \varphi) = \begin{cases} 0 & \text{if } B, \varphi > 0 \\ +\infty & \text{else} \end{cases} \tag{23}$$

By introducing the discrete scalar images $B, \varphi \in \mathbf{R}^{n\times m}$ and the vectorial image $\mathbf{v} = (v_1, v_2) = \left(\mathbf{R}^{n\times m}\right)^2$, we can obtain the discrete version of (22), which is given by :

$$\min_{\substack{B,\varphi,\mathbf{v}\\ B\geq 0,\varphi\geq 0}} \left\{ \left\| F_D^{\mathrm{CTF}}(B,\varphi) - I_D^{\mathrm{obs}} \right\|_2^2 + \alpha \left\| \mathbf{Dv} \right\|_1 + \beta \left\| \nabla B - \mathbf{v} \right\|_1 + \nu \left|\left| \nabla \varphi \right|\right|_1 \right\} \tag{24}$$

where $\mathbf{D} : \left(\mathbf{R}^{n\times m}\right)^2 \to \left(\mathbf{R}^{n\times m}\right)^4, \mathbf{v} \mapsto (\nabla v_1, \nabla v_2)$. If we denote $X = \mathbf{R}^{n\times m} \times \mathbf{R}^{n\times m} \times \left(\mathbf{R}^{n\times m}\right)^2$ and $Y = \mathbf{R}^{n\times m} \times \left(\mathbf{R}^{n\times m}\right)^4 \times \left(\mathbf{R}^{n\times m}\right)^2 \times \left(\mathbf{R}^{n\times m}\right)^2$ to be the discretized primal and dual spaces. Then the minimization problem (24) can be written in the form of :

$$\min_{(B,\varphi,\mathbf{v})\in X} \left\{ \mathcal{F}\left[\mathcal{K}_{\mathrm{CTF}}\left(B,\varphi,\mathbf{v}\right)\right] + \mathcal{G}\left(B,\varphi,\mathbf{v}\right) \right\} \tag{25}$$

with :

- $\mathcal{K}_{\mathrm{CTF}}\left(B,\varphi,\mathbf{v}\right) = \left[F_D^{\mathrm{CTF}}\left(B,\varphi\right), \mathbf{D}(\mathbf{v}), \nabla B - \mathbf{v}, \nabla\varphi\right]$
- $\mathcal{F}\left(h^1,h^2,h^3,h^4\right) = \left\|h^1 - I_D^{\mathrm{obs}}\right\|_2^2 + \alpha\left\|h^2\right\|_1 + \beta\left\|h^3\right\|_1 + \nu\left\|h^4\right\|_1$
- $\mathcal{G}\left(B,\varphi,\mathbf{v}\right) = \iota_+(B,\varphi)$

Using this formulation, we derive the PDHGM-CTF algorithm given in 2. Here, ${\mathcal{K}_{\mathrm{CTF}}}^*$ is the adjoint operator

Algorithm 2 PDHGM-CTF

Given :

- step sizes σ, τ such that $\sigma\tau|||\mathcal{K}_{\mathrm{CTF}}|||^2 < 1$ and relaxation parameter $\gamma \in [0,1]$
- regularization parameters α, β and ν
- $x_0 = \{(B_0, \varphi_0), v_0\} \in X$ (primal)
- $h_0 = [h_0^1, h_0^2, h_0^3, h_0^4] \in Y$ (dual)

for $i = 0, \ldots, N_{\mathrm{iter}}$ **do** :
$h_{i+1} \leftarrow \mathrm{prox}_{\sigma\mathcal{F}^*}\left(h_i + \sigma\mathcal{K}_{\mathrm{CTF}}\overline{x}_i\right)$
$x_{i+1} \leftarrow \mathrm{prox}_{\tau\mathcal{G}}\left(x_i - \tau{\mathcal{K}_{\mathrm{CTF}}}^* h_{i+1}\right)$
$\overline{x}_{i+1} \leftarrow x_{i+1} + \gamma\left(x_{i+1} - x_i\right)$

of $\mathcal{K}_{\mathrm{CTF}}$, and $\mathrm{prox}_{\sigma f}$ is the proximal operator of f with parameter σ defined as :

$$\mathrm{prox}_{\sigma f}(y) = \underset{x}{\mathrm{argmin}} \left\{ f(x) + \frac{1}{2\sigma}\left|\left|x-y\right|\right|^2 \right\} \tag{26}$$

The algorithm 2 is convergent[34] if the primal step τ and the dual step σ satisfiy the inequality

$$\sigma\tau|||\mathcal{K}_{\mathrm{CTF}}|||^2 = \sigma\tau \sup_{x\in X}\left\{ \frac{||\mathcal{K}_{\mathrm{CTF}}(x)||_Y}{||x||_X} \right\}^2 < 1 \tag{27}$$

3.4 Non Linear Primal Dual Hybrid Gradient Method (NL-PDHGM)

More recently, the PDHGM algorithm has been generalized to nonlinear operators, where it was shown that convergence of the algorithm is still guaranteed even for non-linear operators, under assumptions on the initialization and on the nonlinear operator.[30] In this case, the only difference to the PDGHM-CTF is in the data fidelity term, i.e., instead of (24), we seek to solve the following minimization problem :

$$\min_{\substack{B,\varphi,\mathbf{v}\\ B\geq 0,\varphi\geq 0}} \left\{ \left\| F_D(B,\varphi) - I_D^{\mathrm{obs}} \right\|_2^2 + \alpha \left\| \mathbf{Dv} \right\|_1 + \beta \left\| \nabla B - \mathbf{v} \right\|_1 + \nu \left|\left| \nabla \varphi \right|\right|_1 \right\} \tag{28}$$

where F_D is defined as the forward operator (8). As before, (28) can be rewritten in the form (25), this time using the non-linear operator defined by:

$$\mathcal{K}_{\mathrm{NL}}\left(B,\varphi,\mathbf{v}\right)=\left[F_D\left(B,\varphi\right),\mathbf{D}(v),\nabla B-v,\nabla\varphi\right]$$

This new problem (28) can be solved using the NonLinear Primal Dual Hybrid Gradient Method (NL-PDHGM), where the algorithm is given in Algo. 3. We observe that unlike Algo. 2, we have the appearance of $\mathcal{K}'_{\mathrm{NL}}(x_i)$,

Algorithm 3 NL-PDHGM

Given :

- $x_0=\{(B_0,\varphi_0),v_0\}\in X$ (primal)
- $h_0=[h_0^1,h_0^2,h_0^3,h_0^4]\in Y$ (dual)
- step sizes $\sigma_0,\tau_0>0$ and relaxation parameter $\gamma\in[0,1]$
- regularization parameters α,β and ν

for $i=0,\ldots,N_{\mathrm{iter}}$ **do** :
$h_{i+1}\leftarrow\mathrm{prox}_{\sigma\mathcal{F}^*}\left(h_i+\sigma_i\mathcal{K}_{\mathrm{NL}}\overline{x}_i\right)$
$x_{i+1}\leftarrow\mathrm{prox}_{\tau\mathcal{G}}\left(x_i-\tau_i\left[\mathcal{K}'_{\mathrm{NL}}(x_i)\right]^*h_{i+1}\right)$
$\overline{x}_{i+1}\leftarrow x_{i+1}+\gamma\left(x_{i+1}-x_i\right)$
$\sigma_{i+1},\tau_{i+1}\leftarrow\sigma_i,\tau_i$ such that $\sigma_i\tau_i\sup_{k=0,1,\ldots,i}\left\{|||\mathcal{K}'_{\mathrm{NL}}(x_k)|||^2\right\}<1$

the Fréchet derivative of the operator $\mathcal{K}_{\mathrm{NL}}$ at the point x_i, for which we have an explicit formula. The main differences with the linear version is that to ensure the convergence of the algorithm, the operator $\mathcal{K}_{\mathrm{NL}}$ has to be locally Lipschitz differentiable, i.e. it has to be Fréchet differentiable and its gradient has to be Lipschitz in a neighborhood of a solution. Also, the initial iterate has to be close enough to a solution and we need to make sure that at every step i, the step sizes satisfy

$$\sigma_i\tau_i\sup_{k=0,1,\ldots,i}\left\{|||\mathcal{K}'_{\mathrm{NL}}(x_k)|||^2\right\}=\sigma_i\tau_i\sup_{\substack{x\in X\\k=0,1,\ldots,i}}\left\{\frac{||\mathcal{K}'_{\mathrm{NL}}(x_k)(x)||_Y}{||x||_X}\right\}^2<1 \tag{29}$$

4. EXPERIMENTS

4.1 Implementation details

In this part, we give details on the choice of the different regularization parameters and steps size. We also discuss the stopping criterion and the choice of the initialization.

4.1.1 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 primal-dual methods, in order to satisfy the condition (29) (resp. (27)) on the primal and dual steps, one can estimate $|||\mathcal{K}'_{\mathrm{NL}}(x)|||^2$ (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 compute the quotient $\rho_n=\frac{||\mathbf{y}_n(x)||}{||\mathbf{y}_{n-1}||}$. The sequence $(\rho_n)_{n\in\mathbb{N}}$ converges[35] 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

$$\sigma=\tau<\frac{1}{\sqrt{\rho_N}}$$

Since this process is quite costly, we only update the variables σ and τ every 50 iterations for NL-PDHGM.

4.1.2 Regularization parameters

The regularization parameters have been chosen empirically, which is suboptimal. We have chosen $\eta = 10^{-1}, \mu = 10^{-3}$ and the smoothing factor ϵ was set to 10^{-3} for GD-TV$^{\epsilon}$. For both PDHGM-CTF and NL-PDHGM, we choose the same set of parameters : $\alpha = 10^{-2}, \beta = 5 \times 10^{-3}, \nu = 10^{-2}$ and the relaxation parameter to be $\gamma = 1$.

4.1.3 Stopping criterion

A criterion often used for primal-dual methods is based on the computation of the duality gap, which, for the problem (19) is defined by

$$\mathcal{F}(\mathcal{K}x) + \mathcal{G}(x) + \mathcal{G}^*(-\mathcal{K}^* y) + \mathcal{F}^* y \tag{30}$$

and use it as a stopping criterion when it goes under a certain threshold. But the primal variable v involves in TGV^2 (see eq. (21)) and the duality gap is in this case equal to infinity,[36] 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.[30] As the computation of this quantity represents an important computational cost,[37] 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 $N_{\text{iter}} = 1\,000$ iterations for the different experiments.

4.1.4 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.

4.2 Simulated dataset

In order to compare the performance of the different algorithms, we generated synthetic X-ray phase contrast images. The X-ray energy was set to 13 keV for a wavelength of $\lambda = 0.095$ nm, and the pixel size in the object space was set to 48 nm. We created projection datasets from 3D objects created from random combinations of one to 10 shapes, consisting with three different materials to create heterogeneous objects. The refractive indices (1) used for the materials are given in Table 1.

Table 1: Complex refractive indices materials at 13 keV

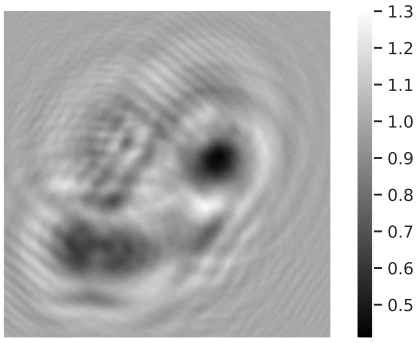
Material	μ (cm^{-1})	$\frac{2\pi}{\lambda}\delta_r$ (cm^{-1})	δ_r/β
Gold	2 790	11 395	8.16
Palladium	615	8 251	26.83
Zinc	859	5 270	12.27

The shapes used were ellipsoids and paraboloids with random positions and orientations. Then, 2D analytical tomographic projections of the real and imaginary part of the refractive index, corresponding to the phase (3) and the absorption (2) respectively, were obtained from the 3D objects for an image size of 2048×2048 pixels. Objects and projections were generated using the software *TomoPhantom*.[38] Phase contrast images were generated from the projection images according to (7) at propagation distances $D = 20.3$ mm and downsampled to 512×512 to avoid aliasing in the calculation of the diffraction pattern. The dataset was generated using different levels of white Gaussian noise to yield a certain peak to peak signal to noise ratio (PPSNR). Examples of such diffraction are represented in Fig. 1, where the first image 1a displays the intensity measured when the object is rather *thin*, while the second 1b displays the intensity measured when the object is *thicker*.

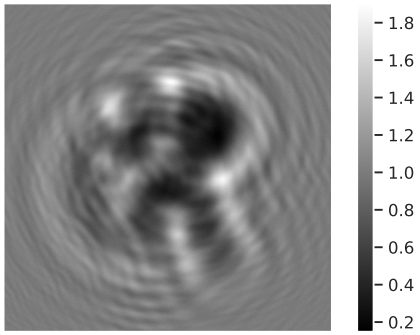
4.3 Results

We compare the different methods on synthetic dataset containing 1 000 images, generated as described in sec. 4.2. We compare the results using the normalized mean square error (NMSE) defined by :

$$\text{NMSE}(x, x_{\text{true}}) = \frac{||x - x_{\text{true}}||_2}{||x_{\text{true}}||_2} \tag{31}$$



(a) Intensity obtained with thin object.



(b) Intensity obtained with thick object.

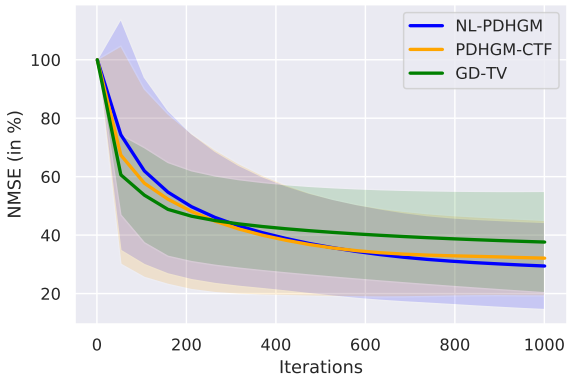
Figure 1: Simulated intensities at distance $D = 20.3$ mm for two different objects of the dataset.

Table 2: Average NMSE and standard deviation (in %) for 1 000 test images

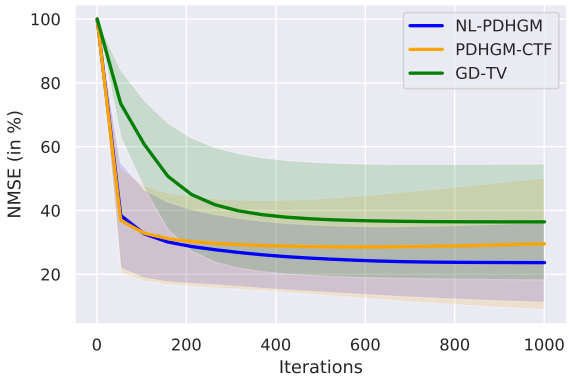
	Absorption	Phase
GD-TV	37.5 (17.4)	36.4 (18.2)
PDHGM-CTF	32.1 (12.9)	29.6 (20.9)
NL-PDHGM	**29.2 (14.8)**	**23.6 (12.6)**

In order to evaluate the robustness of each method, the regularization parameters were fixed for all cases (see 4.1), the averages NMSE obtained are summarized in Tab. 2

Overall, the NL-PDHGM method performed better than PDHGM-CTF and GD-TV$^\epsilon$, achieving better absorption and phase reconstruction in average. We find that all methods retrieved the phase better than the absorption. The evolution of the averages NMSE, as well as the standard deviation, are displayed in Fig. 2, 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.



Absorption



Phase

Figure 2: Evolution of average NMSE (in %) for 1 000 test images. The transparent areas correspond to the standard deviation.

For qualitative evaluation, examples of reconstructed phase and absorption projections from the diffraction patterns (Fig. 1) are displayed in Fig. 3 & 4 (with negative contrast). Figure 3 shows the reconstructions of the methods (and the associated residuals $x - x_{\text{true}}$) when the object considered is quite thin. In this case, the CTF assumptions are well satisfied, that is why PDHGM-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.

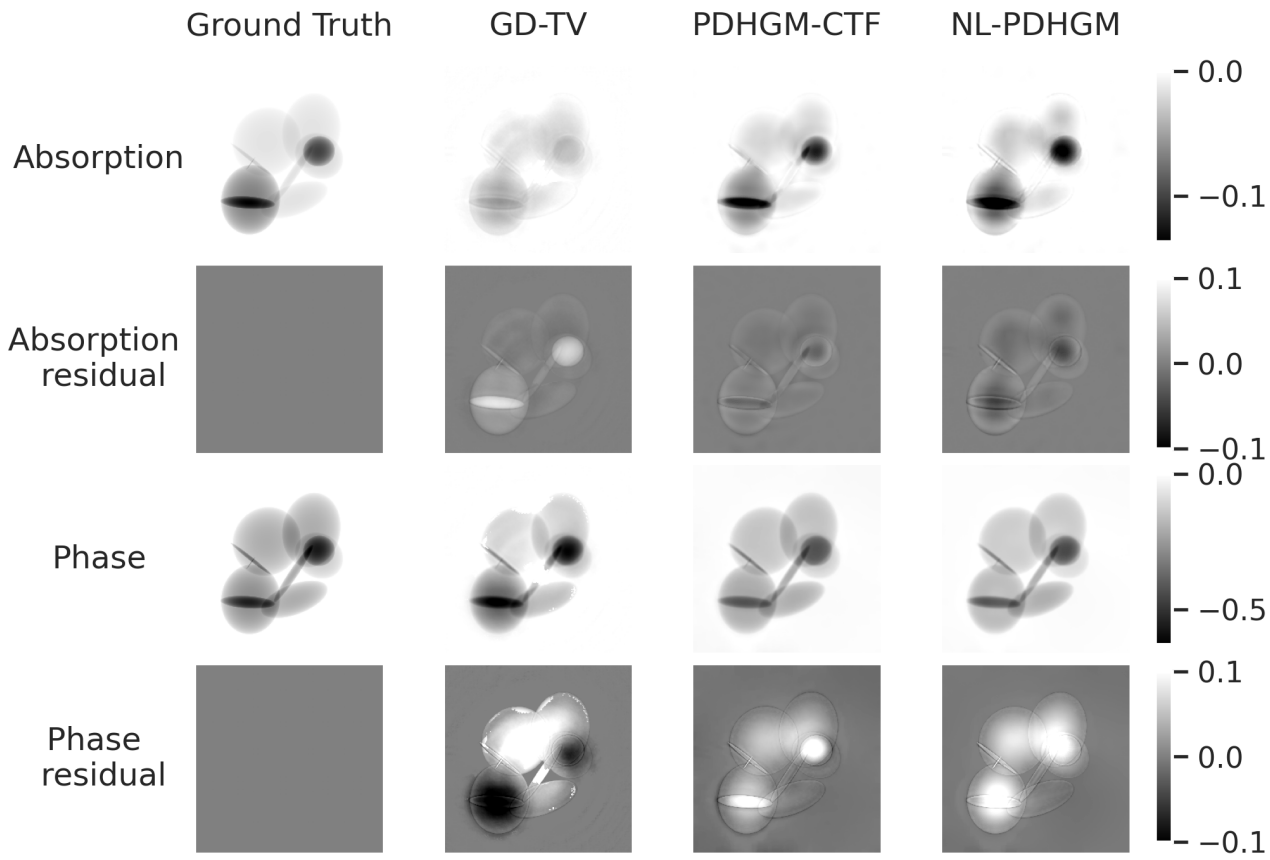


Figure 3: Comparison of the different methods using the diffraction pattern 1a

Figure 4 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 PDHGM-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-PDHGM, 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-PDHGM can still be expected to give satisfactory results.

5. CONCLUSION

We presented new methods based on a primal-dual approach that allowed us to reconstruct both the absorption and phase from a single diffraction pattern. We have proposed an iterative algorithm based on CTF-linearization (PDHGM-CTF) but also a more general one (NL-PDHGM) which allows to take into account the nonlinearity of the forward model. We have seen that primal-dual approaches allow to efficiently implement regularizations such as the Total Variation or its higher order generalization. With simpler approaches such as gradient descent, an approximation must be used for the gradient of Total Variation to be well defined, however, information is lost in making such a choice. The use of different regularizations for absorption and phase has proven to be more effective in taking into account the intrinsic features of each. In particular, the use of Total Generalized

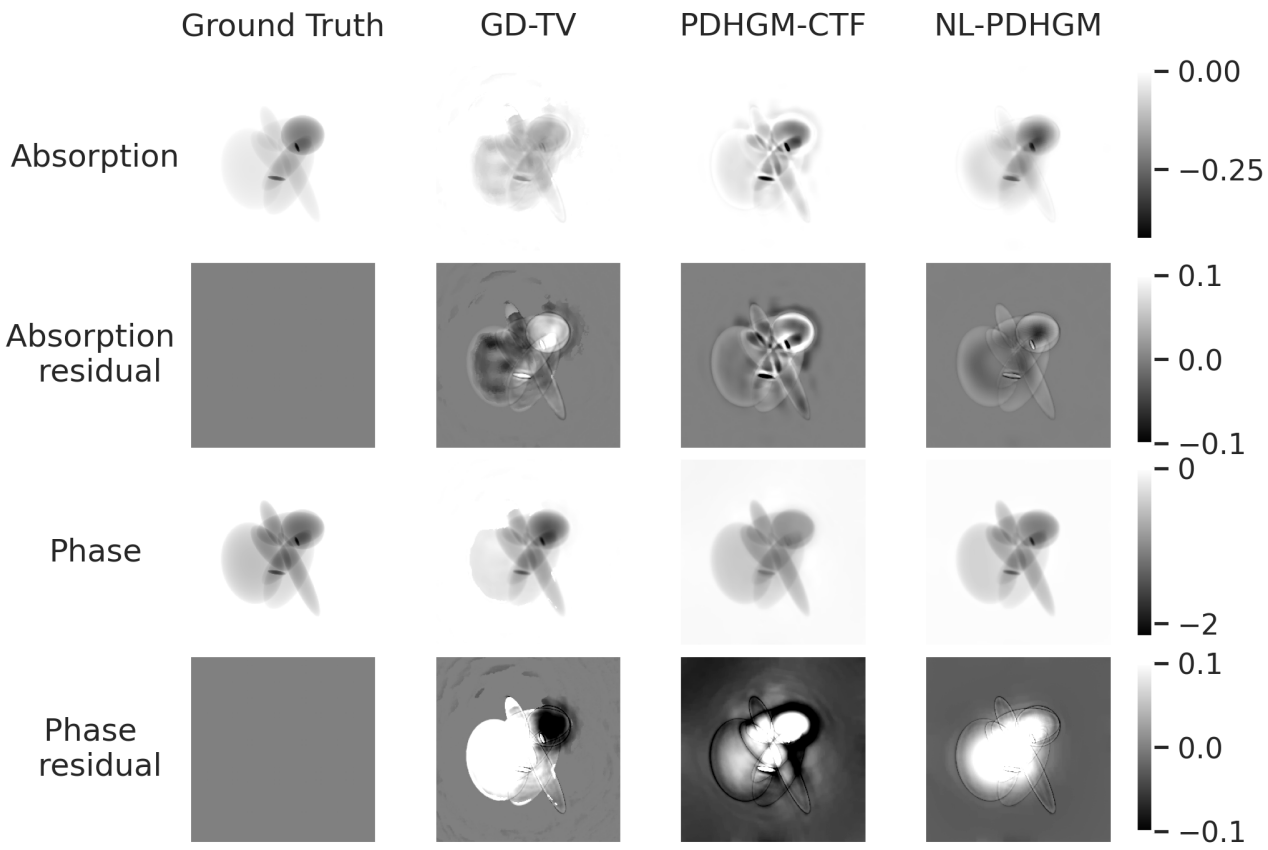


Figure 4: Comparison of the different methods using the diffraction pattern 1b

Variation on absorption allowed us to avoid the staircasing effect that can occurs when using Total Variation. For the phase, this problem does not occur, so the use of the Total Variation was enough to have a good recovery. In addition, we have noticed several things, firstly, when the CTF conditions are satisfied, the linearized version allows for more accurate reconstructions, and in practice, may often requires fewer iterations. But in the opposite case, we could observe the significant contribution of the nonlinear information of the problem, this suggests that the NL-PDHGM algorithm could be applied in a wide variety of cases where linear methods would fail. And although different regularization parameters must be chosen, we have seen that the choice of the latter is robust when applied over a thousand of cases. A direct extension of the proposed work would be to apply this approach to phase contrast tomography. Another approach would be to use neural networks to learn the regularization parameters or the regularization itself.

REFERENCES

[1] Momose, A., Takeda, T., Itai, Y., and Hirano, K., "Phase–contrast x–ray computed tomography for observing biological soft tissues," *Nature Medicine* **2**, 473–475 (1996).

[2] Mayo, S. C., Stevenson, A. W., and Wilkins, S. W., "In-line phase-contrast x-ray imaging and tomography for materials science," *Materials* **5**, 937 – 965 (2012).

[3] Langer, M., Pacureanu, A., Suhonen, H., Grimal, Q., Cloetens, P., and Peyrin, F., "X-ray phase nanotomography resolves the 3d human bone ultrastructure," *PLOS ONE* **7**, 1–7 (08 2012).

[4] Varga, P., Pacureanu, A., Langer, M., Suhonen, H., Hesse, B., Grimal, Q., Cloetens, P., Raum, K., and Peyrin, F., "Investigation of the three-dimensional orientation of mineralized collagen fibrils in human lamellar bone using synchrotron x-ray phase nano-tomography," *Acta Biomaterialia* **9**(9), 8118–8127 (2013).

[5] Langer, M. and Peyrin, F., "3d x-ray ultra-microscopy of bone tissue," *Osteoporosis International* **27** (09 2015).

[6] Giuliani, A., Mazzoni, S., Mele, L., Liccardo, D., Tromba, G., and Langer, M., "Synchrotron phase tomography: An emerging imaging method for microvessel detection in engineered bone of craniofacial districts," *Frontiers in Physiology* **8** (2017).

[7] Kalbfleisch, S., Kahnt, M., Buakor, K., Langer, M., Dreier, T., Dierks, H., Stjärneblad, P., Larsson, E., Gordeyeva, K., Chayanun, L., Söderberg, D., Wallentin, J., Bech, M., and Villanueva-Perez, P., "X-ray in-line holography and holotomography at the nanomax beamline," *Journal of Synchrotron Radiation* **29** (01 2022).

[8] Cloetens, P., Barrett, R., Baruchel, J., Guigay, J.-P., and Schlenker, M., "Phase objects in synchrotron radiation hard x-ray imaging," *Journal of Physics D: Applied Physics* **29**, 133–146 (jan 1996).

[9] Gureyev, T. E. and Nugent, K. A., "Phase retrieval with the transport-of-intensity equation. ii. orthogonal series solution for nonuniform illumination," *J. Opt. Soc. Am. A* **13**, 1670–1682 (Aug 1996).

[10] Paganin, D., Mayo, S. C., Gureyev, T. E., Miller, P. R., and Wilkins, S. W., "Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object," *Journal of Microscopy* **206**(1), 33–40 (2002).

[11] Zabler, S., Cloetens, P., Guigay, J.-P., Baruchel, J., and Schlenker, M., "Optimization of phase contrast imaging using hard x-rays," *Review of Scientific Instruments* **76**(7), 073705 (2005).

[12] Guigay, J. P., Langer, M., Boistel, R., and Cloetens, P., "Mixed transfer function and transport of intensity approach for phase retrieval in the fresnel region," *Optics Letters* **32**, 1617–1619 (Jun 2007).

[13] Gerchberg, R. W., "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* **35**, 237–246 (1972).

[14] Fienup, J. R., "Phase retrieval algorithms: a comparison," *Applied Optics* **21 (15)**, 2758–2769 (1982).

[15] Bauschke, H. H., Combettes, P. L., and Luke, D. R., "Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization," *J. Opt. Soc. Am. A* **19**, 1334–1345 (Jul 2002).

[16] Bauschke, H. H., Combettes, P. L., and Luke, D. R., "A hybrid projection reflection method for phase retrieval," *Journal of the Optical Society of America A* **20 (6)**, 1025–1034 (2003).

[17] Elser, V., "Phase retrieval by iterated projections," *Journal of the Optical Society of America A* **20 (1)**, 40–55 (2003).

[18] Luke, D. R., "Relaxed averaged alternating reflections for diffraction imaging," *Inverse Problems* **21**, 37–50 (2005).

[19] Davidoiu, V., Sixou, B., Langer, M., and Peyrin, F., "Non-linear iterative phase retrieval based on frechet derivative," *Opt. Express* **19**, 22809–22819 (Nov 2011).

[20] Davidoiu, V., Sixou, B., Langer, M., and Peyrin, F., "Nonlinear approaches for the single-distance phase retrieval problem involving regularizations with sparsity constraints," *Appl. Opt.* **52**, 3977–3986 (Jun 2013).

[21] Maretzke, S., Bartels, M., Krenkel, M., Salditt, T., and Hohage, T., "Regularized newton methods for x-ray phase contrast and general imaging problems," *Opt. Express* **24**, 6490–6506 (Mar 2016).

[22] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning* **3**, 1–122 (01 2011).

[23] Bostan, E., Froustey, E., Rappaz, B., Shaffer, E., Sage, D., and Unser, M., "Phase retrieval by using transport-of-intensity equation and differential interference contrast microscopy," in [*2014 IEEE International Conference on Image Processing (ICIP)*], 3939–3943 (2014).

[24] Villanueva-Perez, P., Arcadu, F., Cloetens, P., and Stampanoni, M., "Contrast-transfer-function phase retrieval based on compressed sensing," *Opt. Lett.* **42**, 1133–1136 (Mar 2017).

[25] Beleggia, M., Schofield, M., Volkov, V., and Zhu, Y., "On the transport of intensity technique for phase retrieval," *Ultramicroscopy* **102**(1), 37–49 (2004).

[26] Maretzke, S. and Hohage, T., "Stability estimates for linearized near-field phase retrieval in x-ray phase contrast imaging," *SIAM Journal on Applied Mathematics* **77**(2), 384–408 (2017).

[27] Mom, K., Sixou, B., and Langer, M., "Mixed scale dense convolutional networks for X-ray phase-contrast imaging," *Applied Optics* (In press).

[28] Zhang, Y., Noack, M. A., Vagovic, P., Fezzaa, K., Garcia-Moreno, F., Ritschel, T., and Villanueva-Perez, P., "PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets," *Opt. Express* **29**, 19593–19604 (Jun 2021).

[29] Chambolle, A. and Pock, T., "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision* **40**, 120–145 (May 2011).

[30] Valkonen, T., "A primal–dual hybrid gradient method for nonlinear operators with applications to MRI," *Inverse Problems* **30**, 055012 (may 2014).

[31] Esser, E., Zhang, X., and Chan, T., "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sciences* **3**, 1015–1046 (01 2010).

[32] Bredies, K., Kunisch, K., and Pock, T., "Total generalized variation," *SIAM Journal on Imaging Sciences* **3**(3), 492–526 (2010).

[33] Bredies, K. and Valkonen, T., "Inverse problems with second-order total generalized variation constraints," (2020).

[34] Chambolle, A. and Pock, T., "An introduction to continuous optimization for imaging," *Acta Numerica* **25**, 161–319 (2016).

[35] Chaari, L., Pustelnik, N., Chaux, C., and Pesquet, J.-C., "Solving inverse problems with overcomplete transforms and convex optimization techniques," in [*SPIE*], (Aug. 2009).

[36] Valkonen, T., Bredies, K., and Knoll, F., "Total generalized variation in diffusion tensor imaging," *SIAM journal on imaging sciences* **6**(1), 487–525 (2013).

[37] Knoll, F., Holler, M., Kösters, T., Otazo, R., Bredies, K., and Sodickson, D. K., "Joint mr-pet reconstruction using a multi-channel image regularizer," *IEEE Transactions on Medical Imaging* **36**, 1–16 (2017).

[38] Kazantsev, D., Pickalov, V., Nagella, S., Edoardo, P., and Withers, P. J., "TomoPhantom, a software package to generate 2D–4D analytical phantoms for CT image reconstruction algorithm benchmarks," *SoftwareX* **7**, 150–155 (2018).