Inverse problems with nonnegative and sparse solutions : algorithms and application to the phase retrieval problem

Muoi, Pham Quy; Hào, Dinh Nho; Sahoo, Sujit Kumar; Tang, Dongliang; Cong, Nguyen Huu; Dang, Cuong

2018

Muoi, P. Q., Hào, D. N., Sahoo, S. K., Tang, D., Cong, N. H., & Dang, C. (2018). Inverse problems with nonnegative and sparse solutions: algorithms and application to the phase retrieval problem. *Inverse Problems*, 34(5), 055007-. doi:10.1088/1361-6420/aab6c9

<https://hdl.handle.net/10356/106501>

<https://doi.org/10.1088/1361-6420/aab6c9>

© 2018 IOP Publishing Ltd. All rights reserved. This is an author-created, un-copyedited version of an article accepted for publication in *Inverse Problems*. IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any version derived from it. The definitive publisher authenticated version is available online at <https://doi.org/10.1088/1361-6420/aab6c9>.

Downloaded on 26 Apr 2023 18:44:23 SGT

Inverse Problems with Nonnegative and Sparse Solutions: Algorithms and Application to the Phase Retrieval Problem

Pham Quy Muoi¹, Dinh Nho Hào², Sujit Kumar Sahoo^{3,4},
Dongliang Tang³, Nguyen Huu Cong⁵, and Cuong Dang³

March 2, 2019

¹ Department of Mathematics, The University of Danang - University of Education, 459 Ton Duc Thang, Danang, Vietnam.

Email: pqmuoi@ud.edu.vn

² Hanoi Institute of Mathematics, VAST, 18 Hoang Quoc Viet Road, Hanoi, Vietnam

E-mail: hao@math.ac.vn

³ Centre for OptoElectronics and Biophotonics (COEB), School of Electrical and Electronic Engineering, The Photonics Institute (TPI), Nanyang Technological University Singapore, 50 Nanyang Avenue, 639798, Singapore.

Email: HCDang@ntu.edu.sg, s.k.sahoo@nus.edu.sg, dltang@ntu.edu.sg

⁴ Department of Statistics and Applied Probability, National University of Singapore, Singapore

⁵ Thai Nguyen University, Tan Thinh Ward, Thai Nguyen, Vietnam *Email:* congln@tnu.edu.vn

Abstract

In this paper, we study a gradient-type method and a semismooth Newton method for minimization problems in regularizing inverse problems with nonnegative and sparse solutions. We propose a special penalty functional forcing the minimizers of regularized minimization problems to be nonnegative and sparse, and then we apply the proposed algorithms in a practical the problem. The strong convergence of the gradient-type method and the local superlinear convergence of the semismooth Newton method are proven. Then, we use these algorithms for the phase retrieval problem and illustrate their efficiency in numerical examples, particularly in the practical problem of optical imaging through scattering media where all the noises from experiment are presented.

Keywords: *Gradient-type algorithm, Inverse problems, Nonnegative and sparse solutions, Nonnegative sparse regularization, Phase retrieve problem.*

1 Introduction

Many mathematical models lead to find a *nonnegative and sparse solution* u of the operator equation

$$K(u) = y, \quad (1)$$

where $K : X \rightarrow Y$ is a smooth operator (linear or nonlinear) between two (real Hilbert) spaces X and Y , and y is the exact data. Note that if $u = \sum_{i \in \Lambda} u_i \varphi_i$ with $\{\varphi_i\}$ being a basis or frame of the real Hilbert space X , then the *nonnegativity* means that $u_i \geq 0$ for all i and the *sparsity* means that there are finite nonzero elements u_i .

Due to this definition, the nonnegativity of u does not mean that the function u is nonnegative. It depends on the chosen basis. For examples, in $\mathbb{R}^n$ with canonical basis, the nonnegativity of $u \in \mathbb{R}^n$ means that u is a nonnegative function. In $L^2(0, 1)$, if the chosen basis consists of "hat functions" as in theory of the finite element method (piecewise linear functions), the nonnegativity of $u \in L^2(0, 1)$ also means that u is a nonnegative function. However, if we choose the Fourier basis for $L^2(0, 1)$, the nonnegativity of u_i does not imply that u is a nonnegative function.

The problem of finding a nonnegative solution u in usual meaning of equations (i.e. $u \geq 0$) falls into class of Problem (1) and appears in many fields such as image processing [1, 2, 36], and nonnegative matrix factorization [17, 25, 34], whereas the our definition of nonnegativity is a generalization.

We further assume that Problem (1) is ill-posed and the exact data is unknown. In practice, we have only noisy data y^δ of y such that

$$\|y - y^\delta\|_Y \leq \delta. \quad (2)$$

Therefore, the core problem is to find an approximation of the nonnegative and sparse solution given the operator K and a noisy data y^δ . The ill-posedness of Problem (1) enforces us to use a regularization method. Tikhonov-type regularization methods such as Tikhonov regularization [7], sparse regularization [4, 6, 15, 26, 27, 28], and total variational regularization [1, 36] are often used, lead to the minimization problems of the form

$$\min_{u \in X, u \geq 0} F(u; y^\delta) + \alpha \Phi(u), \quad (3)$$

where $F(u; y^\delta)$ is a functional in u measuring the difference between $K(u)$ and noise data y^δ , e.g., $F(u; y^\delta) = \frac{1}{2} \|K(u) - y^\delta\|^2$, and $\Phi : X \rightarrow \mathbb{R}$ is a penalty functional that is often chosen based on a prior information on exact solutions.

It is well-known that the efficiency of Tikhonov-type regularization depends on choices of functionals F and Φ , for example, if F and Φ are chosen such that

- The minimizers of Problem (3) fast converge to the exact solution u as $\delta \rightarrow 0$ (with responding to some choices of $\alpha = \alpha(\delta)$).
- Problem (3) is easy to solve, e.g., a convex problem, and there are simple and efficient algorithms to solve.

Many different choices of F and Φ have been studied. F is often the square norm of X [7], but other different norms are also considered [15]. If F is a convex function, then Problem (3) is convex and we have more tools and efficient algorithms to work with. This situation often happens if K is a linear operator. For nonlinear operators, such a choice for F is much more challenging and only available for some specific operators. For example, recently used functionals F in [19, 20, 21] are energy functionals, which are convex for nonlinear operators K . Generally, a good choice for F depends on Problem (1), but we do not aim at here. Our aim is to give a simple and good choice for Φ under a prior information on the sparsity and negativeness of the exact solution to (1).

Note that the sparsity of the exact solution encourages us to use sparsity regularization, which has been studied recently, e.g., [4, 6, 15, 14, 16, 26, 27, 31, 28]. This method applied for Problem (1) leads to considering Problem (3) with Φ being given by

$$\Phi(u) = \Phi_p(u) := \sum_{i \in \Lambda} \omega_i |\langle u, \varphi_i \rangle|^p, \quad (4)$$

where $\omega_i \geq \omega_0 > 0$ for all i and $\{\varphi_i\}_{i \in \Lambda}$ is an orthonormal basis (or frame) of the real Hilbert space X . We like to emphasize that the efficiency of sparse regularization analyzed in [4, 6, 15, 26, 27, 28] is not clear to the problem under consideration, because we are now working with the constrained minimization problem. To best of our knowledge, there are no simple and efficient algorithms available for solving them.

Therefore, our approach does not follow the above outline, i.e., we do not consider Problem (3) with Φ of the form (4). Instead, regularized approximate solutions of (1)–(2) are expressed as minimizers of the unconstrained minimization problem

$$\min_{u \in X} \Theta(u) := F(u; y^\delta) + \alpha \Phi(u), \quad (5)$$

where $F(\cdot; y^\delta) : X \rightarrow \mathbb{R}$ is a smooth functional measuring the difference between $K(u)$ and noise data y^δ , and $\Phi : X \rightarrow \mathbb{R}$ is defined by

$$\Phi(u) := \begin{cases} \sum_{i \in \Lambda} u_i, & \text{if } u_i \geq 0 \text{ for all } i, \\ +\infty, & \text{otherwise.} \end{cases} \quad (6)$$

Here $u = \sum_{i \in \Lambda} u_i \varphi_i$ and $\{\varphi_i\}_{i \in \Lambda}$ is an orthonormal basis of X . The proposed regularization method is called *nonnegative sparse regularization* (NSR) in distinction with sparse regularization.

We want to emphasize the advantage of NSR is not only suitable for the prior information of nonnegativity and sparsity of the solution, but also there exist efficient algorithms to solve it even for nonconvex $F(\cdot; y^\delta)$. We will be present such algorithms in the next sections. Note that it is easy to modify the penalty functional to obtain the non-positive and sparse solutions for inverse problems. All algorithms in this paper also work for *nonpositive sparse regularization*. Our main contributions in the paper are:

- We have proven that the regularized solutions, i.e. minimizers of Problem (5), are nonnegative and sparse.
- We have proven the strong convergence of the Gradient-type algorithm for Problem (5). This result improves our previous result in [29], where we could prove only the weak convergence.

- We have proven the local superlinear convergence of the semismooth Newton method for solving the first order equation for Problem (5).
- We have applied the algorithms to the phase retrieval problem, and we have tested on numerical simulation as well as on real experimental data (i.e. imaging through strongly scattering media).

The paper is organized as follows: in Section 2, we indicate that the minimizers of Problem (5) are nonnegative and sparse. Section 3 is devoted to present numerical algorithms for Problem (5) and prove the strong convergence of the gradient-type method as well as the semismooth Newton method. Finally, we illustrate the nonnegative sparse regularization and algorithms via the phase retrieval problem to do imaging through strongly scattering media.

2 Nonnegativity and sparsity of minimizers

From now we assume that X and Y are real Hilbert spaces, $\{\varphi_k\}_{k \in \Lambda}$ an orthonormal basis of X . The following lemma highlights the prosperities of Φ defined by (6).

Lemma 2.1 *The functional Φ defined by (6) has the following properties*

- 1) Φ is nonnegative, convex and weakly lower semi-continuous.
- 2) For any $u \in X$,

$$\Phi(u) \geq \|u\|.$$

This implies that Φ is weakly coercive, i.e. $\Phi(u) \rightarrow \infty$ as $\|u\| \rightarrow \infty$.

- 3) If $\{u^n\}_{n \in \mathbb{N}} \subset X$ weakly converges to $u \in X$ and $\Phi(u^n)$ converges to $\Phi(u) < \infty$, then $\Phi(u^n - u)$ converges to zero. This also implies that u^n strongly converges to u in X .

Proof. 1) Φ can be rewrite as

$$\Phi(u) = \sum_{i \in \Lambda} \psi(u_i), \tag{7}$$

where $\psi(\cdot)$ is defined by

$$\psi(x) = \begin{cases} x, & \text{if } x \geq 0, \\ +\infty, & \text{if } x < 0. \end{cases}$$

Since ψ is a nonnegative, convex and weakly semicontinuous functional, Φ is nonnegative, convex and weakly lower semi-continuous.

2) The inequality is automatically true if u is zero or have any negative component u_i . Thus, we prove it when u is positive, i.e. $u = \sum_{i \in \Lambda} u_i \varphi_i$ with $u_i \geq 0$ for all $i \in \Lambda$ and $\Phi(u) > 0$. We have

$$\sum_{i \in \Lambda} \left(\frac{u_i}{\Phi(u)} \right)^2 \leq \sum_{i \in \Lambda} \left(\frac{u_i}{\Phi(u)} \right) = 1.$$

Thus,

$$\Phi(u) \geq \left(\sum_{i \in \Lambda} u_i^2 \right)^{1/2} = \|u\|.$$

3) Since $\{u^n\}_{n \in \mathbb{N}} \subset X$ weakly converges to $u \in X$ and $\Phi(u^n)$ converges to $\Phi(u) < \infty$, by definition of Φ we deduce that u^n and u are nonnegative for n large enough. Without lost of generality, we can assume that u^n is nonnegative for all n . Since $\Phi(u^n)$ converges to $\Phi(u)$, we have

$$\begin{aligned} \limsup_n \Phi(u^n - u) &= \limsup_n (2(\Phi(u^n) + \Phi(u)) - 2(\Phi(u^n) + \Phi(u)) + \Phi(u^n - u)) \\ &= 4\Phi(u) - \liminf_n \sum_{i \in \Lambda} (2u_i^n + 2u_i - \psi(u_i^n - u_i)). \end{aligned}$$

Since $\{u^n\}_{n \in \mathbb{N}}$ weakly converges to u (i.e. $u_i^n \rightarrow u_i$ for every $i \in \Lambda$) and the weak semicontinuity of ψ , we have

$$-\liminf_n \sum_{i \in \Lambda} (2u_i^n + 2u_i - \psi(u_i^n - u_i)) \leq -\sum_{i \in \Lambda} 4u_i = -4\Phi(u).$$

Combining the above inequalities, we have

$$\limsup_n \Phi(u^n - u) = 0 \Rightarrow \lim_n \Phi(u^n - u) = 0.$$

Finally, from 2) we have $\Phi(u^n - u) \geq \|u^n - u\|$, which implies that u^n strongly converges to u in X . ■

In the following, we will show that Problem (5) has at least a minimizer and every minimizer is nonnegative and sparse. We also give the first order optimality condition. At this point, some conditions on the functional F are required, and given in the following assumption.

Assumption 1 (1) *Dom $(F(\cdot; y^\delta)) = X$ for any $y^\delta \in Y$, and for each $y^\delta \in Y$ $F(\cdot; y^\delta)$ is bounded from below and weakly lower semicontinuous. Without loss of generality, we assume $F(u; y^\delta) \geq 0, \forall u \in X$ and for all $y^\delta \in Y$.*

(2) *$F(\cdot; y^\delta)$ is Fréchet differentiable.*

Theorem 2.1 *Under Assumption 1,*

(a) *Problem (5) has at least a nonnegative solution, i.e., the functional Θ in (5) has at least a global nonnegative minimizer.*

(b) *Any local minimizer u of Θ in (5) satisfies the following condition*

$$u = \mathbb{P}_{s\alpha}(u - sF'(u; y^\delta)), \quad (8)$$

for any $s > 0$.

(c) *If u is a local minimizer of Θ in (5), then u is nonnegative and sparse.*

(d) *if F is convex and u is a local minimizer of Θ in (5), then u is a solution of (5). In addition, the solution of (5) is unique if F is strictly convex.*

Here, the proximal operator $\mathbb{P}_\lambda$ is defined by

$$\mathbb{P}_\lambda(u) = \sum_{i \in \Lambda} \max(u_i - \lambda, 0) \varphi_i, \quad (9)$$

with $u_i = \langle u, \varphi_i \rangle$.

Proof. (a) We have

$$0 \leq \inf_{u \in X} \Theta(u) \leq \Theta(0) = F(0; y^\delta) < \infty.$$

Denote $\{u^n\}$ a minimizing sequence. Without loss of generality, we can assume that u^n is nonnegative for all n . Since F is bounded from below and Φ is weakly coercive, the functional Θ is also weakly coercive. This implies that the sequence $\{u^n\}$ is bounded, and thus it has a weakly convergent subsequence (also denoted by $\{u^n\}$), weakly converges to $u^* \geq 0$. Furthermore, the objective functional Θ is weakly lower semicontinuous, because F and Φ are weakly lower semicontinuous. Therefore,

$$\Theta(u^*) \leq \lim_{n \rightarrow \infty} \inf \Theta(u^n) = \inf_u \Theta(u).$$

This means that u^* is a global nonnegative minimizer of Θ in (5).

(b) Now if u is a local minimizer of Θ in (5), then u is nonnegative and $-F'(u) \in \alpha \partial \Phi(u)$ or $u - sF'(u) \in (I + s\alpha \partial \Phi)(u)$ for any $s > 0$. Since Φ is proper, convex and lower semicontinuous, the proximal operator $\mathbb{P}_{s\alpha} := (I + s\alpha \partial \Phi)^{-1}$ is well defined [29]. Therefore, we have

$$u = \mathbb{P}_{s\alpha}(u - sF'(u)) = (I + s\alpha \partial \Phi)^{-1}(u - sF'(u)).$$

Note that if $u = (I + \lambda \partial \Phi)^{-1}(v)$, then u is the minimizer of problem

$$\min_g \frac{1}{2} \|g - v\|^2 + \lambda \Phi(g) = \min_g \sum_{i \in \Lambda} \left(\frac{1}{2} (g_i - v_i)^2 + \lambda \psi(g_i) \right).$$

Solving this minimization problem, we have $u = \mathbb{P}_\lambda(v) = \sum_{i \in \Lambda} \max(v_i - \lambda, 0) \varphi_i$.

(c) From part (b), we have u is nonnegative. Now we show the sparsity of u . Note that $u = \mathbb{P}_{s\alpha}(v)$ with $v = u - sF'(u) \in X$. Since $\|v\| = \sum_{i \in \Lambda} v_i^2 < \infty$, the sequence $\{v_i\}$ converges to zero. Thus, for i large enough, we have

$$v_i \leq |v_i| \leq s\alpha \Rightarrow \max(v_i - s\alpha, 0) = 0. \quad (10)$$

This implies that u has a finite number of nonzero components.

(d) This is evident since Problem (5) is (strictly) convex if F is (strictly) convex. $\blacksquare$

From part (a) and (c) of Theorem 3.1, it implies that any solution of Problem (5) is nonnegative and sparse. This proves that the new proposed approach in this paper forces regularized solutions to be non-negative and sparse. We have stated this advantage in the introduction. From (10) we can notice that the sparsity of minimizers of (5) depends on αs . Since (8) is true for all $s > 0$, the sparsity depends only on the regularized parameter α . We need a suitable choice of α such that the regularized solution is a good approximation of the exact solution of (1). Therefore, we must use a priori information of sparsity of the exact solution in choosing the value of α , which is evident and well-known in the theory of inverse and ill-posed problems.

3 Numerical Algorithms

This section is devoted to the gradient-type algorithm and the semismooth semi-Newton method for Problem (5), the algorithms for the unconstrained minimization problem

$$\min_{u \in X} \Theta(u) := F(u) + \alpha \Phi(u). \quad (11)$$

Here, $F(u) := F(u, y^\delta)$.

3.1 Gradient-type method

The starting point for the gradient-type method is a quadratic approximation method. The main idea is to replace Problem (11) by a sequence of problems $\{\min_{v \in X} \Theta_{s^n}(v, u^n)\}$ with $\Theta_{s^n}(\cdot, u^n)$ being strictly convex, which are easy to solve. Furthermore, Assumption (1) will allow us to prove that the sequence of minimizers $\{u^{n+1} = \operatorname{argmin}_{v \in X} \Theta_{s^n}(v, u^n)\}$ converges to a minimizer of Problem (11).

Based on this idea, for a fixed $s > 0$, we define the following quadratic approximation of $\Theta(v)$ at a given point u :

$$\Theta_s(v, u) := F(u) + \langle F'(u), v - u \rangle + \frac{s}{2} \|v - u\|^2 + \alpha \Phi(v). \quad (12)$$

From Theorem 2.1, the sequence of minimizers $\{u^{n+1} = \operatorname{argmin}_{v \in X} \Theta_{s^n}(v, u^n)\}$ is given by the alternative descriptions

$$u^{n+1} = \operatorname{argmin}_{v \in X} \Theta_{s^n}(v, u^n) = \mathbb{P}_{\frac{\alpha}{s^n}} \left(u^n - \frac{1}{s^n} F'(u^n) \right). \quad (13)$$

Now we prove the convergence of the sequence defined by (13). In doing so, we need the following results.

Lemma 3.1 *Let Assumption 1 hold. Assume that the sequence $\{u^n\}$ is defined by (13) where the sequence $\{s^n\}$ satisfies $s^n \in [\underline{s}, \bar{s}]$ ($0 < \underline{s} \leq \bar{s}$), and*

$$\Theta(u^{n+1}) \leq \Theta_{s^n}(u^{n+1}, u^n).$$

Then, the sequence $\{\Theta(u^n)\}$ is monotonically decreasing, $\lim_{n \rightarrow \infty} \|u^{n+1} - u^n\| = 0$ and the sequence $\{u^n\}$ is bounded.

Proof. We note that the functional Φ given by (6) satisfies the properties (1)-(3) of [29, Assumption 2.1] and thus the result is implied from [29, Lemma 2.3]. $\blacksquare$

Theorem 3.1 (Strong convergence) *Let Assumption 1 hold. In addition, we assume that F is the Frechét differential with Lipschitz constant L . Let the sequence $\{u^n\}$ be defined by (13) where the sequence $\{s^n\}$ satisfies $s^n \in [\underline{s}, \bar{s}]$ ($0 < \underline{s} \leq L \leq \bar{s}$) and*

$$\Theta(u^{n+1}) \leq \Theta_{s^n}(u^{n+1}, u^n).$$

Then, there exists a subsequence $\{u^{n_j}\}$ of $\{u^n\}$ strongly converging to some $u^ \in X$ as $j \rightarrow \infty$. Furthermore, u^* is a stationary point of Θ .*

Proof. By Lemma 3.1 the sequence $\{u^n\}_n$ is bounded. Since F' is Lipschitz continuous the sequence $\{F'(u^n)\}_n$ is bounded as well. Therefore, there exists a subsequence $\{u^{n_j}\}$ of $\{u^n\}$, which weakly converges to some $u^* \in X$, $\{F'(u^{n_j})\}$ weakly converges to some $v' \in X$, and $\{s^{n_j}\}$ converges to some $s \in [\underline{s}, \bar{s}]$ as $j \rightarrow \infty$.

By (13), we have

$$u_i^{n_j+1} = \max \left(u_i^{n_j} - \frac{1}{s^{n_j}} (F'(u^{n_j}))_i - \frac{\alpha}{s^{n_j}}, 0 \right), \forall i \in \Lambda.$$

Letting $j \rightarrow \infty$ and noting that $\|u^{n_j+1} - u^{n_j}\| \rightarrow 0$ by Lemma 3.1, we deduce that

$$u_i^* = \max \left(u_i^* - \frac{1}{s} v'_i - \frac{\alpha}{s}, 0 \right), \forall i \in \Lambda$$

or

$$u^* = \mathbb{P}_{\frac{\alpha}{s}} \left(u^* - \frac{1}{s} v' \right).$$

On the other hand, since $d^{n_j} = u^{n_j} - \frac{1}{s^{n_j}} (F'(u^{n_j}))$ is uniformly bounded there exists $N_0 \in \mathbb{N}$ (do not depend on n_j) such that

$$\sum_{i \geq N_0} |d_i^{n_j}|^2 < \left(\frac{\alpha}{\underline{s}} \right)^2 \Rightarrow u_i^{n_j+1} = \max \left(d_i^{n_j} - \frac{\alpha}{s^{n_j}}, 0 \right) = 0, \forall i \geq N_0.$$

As a consequence, the sequence $\{u^{n_j+1}\}_j$ has uniformly bounded support for all n_j . Thus up to a subsequence it strongly converges to u^* (the weak convergence and strong convergence in finite dimensional spaces are equivalent). Again, we also have that u^{n_j} strongly converges to u^* as $j \rightarrow \infty$.

Finally, since F' is continuous $F'(u^{n_j})$ strongly converges to $v' = F'(u^*)$, i.e.,

$$u^* = \mathbb{P}_{\frac{\alpha}{s}} \left(u^* - \frac{1}{s} F'(u^*) \right).$$

Thus, u^* is a stationary point of Θ . ■

Remark 3.1 We would like to emphasize that one of our main contributions is Theorem 3.1, which proves the strong convergence of the gradient-type method without the following assumption:

Assumption 2 "If $\{u^n\}$ converges weakly to u such that $\{\Theta(u^n)\}$ is monotonically decreasing, then there exists a subsequence $\{u^{n_j}\}$ such that $\{F'(u^{n_j})\} \rightarrow F'(u)$."

In [29], Assumption 2 is used to obtain the weak convergence of the gradient-type method for general minimization problem of the form (11). However, the strong convergence of the algorithm for general minimization problem in [29] has not been proved. In [27] Assumption 2 is also used to prove the strong convergence of the gradient-type method for sparse regularization, i.e., for Problem (3) with Φ given by (4).

By this theorem the iteration (13) converges to a stationary point of Problem (11) if and only if the stepsizes s^n satisfy the conditions: for all n ,

- $\{s^n\}$ is bounded, e.g. $s^n \in [\underline{s}, \bar{s}]$ ($0 < \underline{s} \leq L \leq \bar{s}$),
- and

$$\Theta(u^{n+1}) \leq \Theta_{s^n}(u^{n+1}, u^n).$$

Adding these two requirements to iteration (13), the gradient-type method is presented in Algorithm 3.1.

Remark 3.2 The first guess for s^{n+1} in Step 8 of Algorithm 3.1 is very important. Some strategies for doing this suggested in [3, 27, 29, 30], these are constant stepsize rule, backtracking rule, modified backtracking rule, Barzilai-Borwein rule. These strategies of choosing stepsizes have been analyzed in [28]. Each of them has advantages and disadvantages. For examples, the first rule makes the "repeat" iteration do only one time since the condition $\Theta(u^{n+1}) \leq \Theta_{s^n}(u^{n+1}, u^n)$ is always satisfied. However, if L is quite large, then stepsize $\frac{1}{s^n}$ is too small. As a result, Algorithm [28] is slow. Furthermore, in many cases, the Lipschitz constant L is not available. For the backtracking rule, it generates the increasing sequence $\{s^n\}$, which converges to a number greater than or equal to L , thus Algorithm [28] become slow using this rule for larger value of L . The last two rules are more flexible that makes the Algorithm faster as shown in [28].

Remark 3.3 If the functional F in Problem (11) is convex, then we can apply two accelerated versions (one is call Beck's accelerated algorithm and the other is called Nesterov's accelerated algorithm) in [29] for Problem (11). Two accelerated algorithms have the convergence rates of order $O(\frac{1}{n^2})$, which is much faster than the gradient-type algorithm. Detailed these algorithms for Problem (11) are straight forward from [3, 29, 30].

Algorithm 3.1 Gradient-type method

Input: Initial guess $u^0 : \Phi(u^0) < \infty, \mu \in (1, \infty)$ and $s^0 \in [\underline{s}, \bar{s}]$ ($0 < \underline{s} \leq L/\mu \leq \bar{s}$)

```
1: for  $n = 0, 1, 2, \dots$  do
2:   repeat
3:      $u^{n+1} \leftarrow \mathbb{P}_{\frac{\alpha}{s^n}}(u^n - \frac{1}{s^n} F'(u^n))$ 
4:     if  $\Theta(u^{n+1}) > \Theta_{s^n}(u^{n+1}, u^n)$  then
5:        $s^n \leftarrow s^n \cdot \mu$ 
6:     end if
7:   until  $\Theta(u^{n+1}) \leq \Theta_{s^n}(u^{n+1}, u^n)$  or  $s^n \notin [\underline{s}, \bar{s}]$ 
8:   Compute an initial guess for  $s^{n+1}$ 
9: end for
```

Output: $u = \lim u^n$

3.2 Semismooth Newton method

The first order optimality condition of problem (11) is given by

$$D(u) := u - \mathbb{P}_{\beta\alpha}(u - \beta F'(u)) = 0, \quad (14)$$

for any fixed $\beta > 0$. Note that $D(\cdot)$ is not Gâteaux differentiable. However, as shown below, $D(\cdot)$ is Newton differentiable, thus the semismooth Newton method can be applied.

In this section, we prove the Newton differentiability of the operator $\mathbb{P}_\lambda(\cdot)$ and present the semismooth Newton method to solve Equation (14) numerically. To this end, we first recall the definition of Newton differentiability and obtain the formula for Newton derivative of $\mathbb{P}_\lambda(\cdot)$. Here, we follow the outline of [29]. For more details on Newton differentiability, we refer to [5, 22, 35] and references therein.

Definition 3.1 Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Banach spaces and $U \subset \mathcal{H}_1$ be an open subset. A mapping $\psi : U \rightarrow \mathcal{H}_2$ is called Newton differentiable at $u \in U$ if there exists a family of mappings $\chi : U \rightarrow L(\mathcal{H}_1, \mathcal{H}_2)$ such that

$$\lim_{h \rightarrow 0} \frac{\|\psi(u+h) - \psi(u) - \chi(u+h)h\|_{\mathcal{H}_2}}{\|h\|_{\mathcal{H}_1}} = 0. \quad (15)$$

The function χ is called a Newton derivative of ψ at u .

Furthermore, χ is called a Newton derivative of ψ in U if it is a Newton derivative of ψ for every $u \in U$.

The Newton derivative is a generalization of Fréchet derivative. It is shown in [5, 22, 35] that if a function has Fréchet derivative then that derivative is also a Newton derivative. Furthermore, the Newton derivative has similar basic properties as Fréchet derivative such as the sum of two Newton differentiable functions is Newton differentiable and the composition of a Newton differential function with a Fréchet differential function is Newton differentiable [29].

Lemma 3.2 Let $\mathbb{P}_\lambda(\cdot)$ be defined by (9) with $\lambda > 0$. Let us define a family of operators $G(u) : X \rightarrow X$ by

$$(G(u)v)_k = \begin{cases} v_k, & \text{for } u_k > \lambda \\ 0, & \text{for } u_k \leq \lambda \end{cases}, \quad (16)$$

where $u_k = \langle u, \varphi_k \rangle$ and $v_k = \langle v, \varphi_k \rangle$.

(a) $G(u)(\cdot)$ is a continuous linear operator with $\|G(u)\| \leq 1$ for all $u \in X$. Then

(b) $\mathbb{P}_\lambda$ is Newton differentiable and $G(u)$ is a Newton derivative of $\mathbb{P}_\lambda$ at u .

Remark 3.4 $G(u)$ is the projection operator onto a set of non-zero or active coefficients of $\mathbb{P}_\lambda(u)$. In matrix notation, we can express the derivative $G(u)$ as

$$G(u) = \begin{pmatrix} I_{\mathbb{A}} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\mathbb{A} = \{k \in \Lambda : u_k > \lambda\}$.

With this notation, we are now in a position to calculate a Newton derivative of D .

Theorem 3.2 *Let F be twice Fréchet differentiable in an open set U and assume F'' is Lipschitz continuous in a neighborhood of $u \in U$. Then, D defined by (14) is Newton differentiable at u and a Newton derivative of D at u is given by*

$$D'(u) = I - G(u - \beta F'(u))(I - \beta F''(u)).$$

Furthermore, if the active and inactive sets at u respectively are

$$\begin{aligned}\mathbb{A}(u) &= \{k \in \Lambda : [u - \beta F'(u)]_k > \beta\alpha\}, \\ \mathbb{I}(u) &= \{k \in \Lambda : [u - \beta F'(u)]_k \leq \beta\alpha\}.\end{aligned}$$

and represent the operator $F''(u)$ as

$$F''(u) = \begin{pmatrix} \mathcal{M}_{\mathbb{A}\mathbb{A}} & \mathcal{M}_{\mathbb{A}\mathbb{I}} \\ \mathcal{M}_{\mathbb{I}\mathbb{A}} & \mathcal{M}_{\mathbb{I}\mathbb{I}} \end{pmatrix}.$$

Then, a Newton derivative of D at u is given by

$$D'(u) = \begin{pmatrix} 0 & 0 \\ 0 & I_{\mathbb{I}} \end{pmatrix} + \begin{pmatrix} I_{\mathbb{A}} & 0 \\ 0 & 0 \end{pmatrix} (\beta F''(u)) = \begin{pmatrix} \beta \mathcal{M}_{\mathbb{A}\mathbb{A}} & \beta \mathcal{M}_{\mathbb{A}\mathbb{I}} \\ 0 & I_{\mathbb{I}} \end{pmatrix}. \quad (17)$$

Proof. Using [29, Lemma 2.7], we claim that $D(u)$ is Newton differentiable and its Newton derivative is given by

$$D'(u) = I - G(u - \beta F'(u))(I - \beta F''(u)).$$

Then, using the matrix representation of G and F'' , after simple computation, we obtain the formula (17). $\blacksquare$

We now analyze the semismooth Newton method for solving equation (14) and propose sufficient conditions for local convergence. This method has already been used in the context of inverse problems by a number of authors [5, 22, 35]. Recently, the approach has been applied to sparse regularization problems [18, 29]. As discussed in the previous section, instead of directly approximating minimizers of problem (11) we rather aim at solving equation (14). The semismooth Newton method for this equation has the form

$$u^{n+1} = u^n - D'(u^n)^{-1} D(u^n). \quad (18)$$

For getting a more convenient algorithm, we introduce active and inactive sets as

$$\mathbb{A}^n = \{k \in \Lambda : [u^n - \beta F'(u^n)]_k > \beta\alpha\}, \quad \mathbb{I}^n = \{k \in \Lambda : [u^n - \beta F'(u^n)]_k \leq \beta\alpha\}.$$

Then by Theorem 3.2, we have

$$\begin{aligned}u^{n+1} &= u^n - \begin{pmatrix} \frac{1}{\beta} \mathcal{M}_{\mathbb{A}^n \mathbb{A}^n}^{-1} & -\mathcal{M}_{\mathbb{A}^n \mathbb{A}^n}^{-1} \mathcal{M}_{\mathbb{A}^n \mathbb{I}^n} \\ 0 & I_{\mathbb{I}^n} \end{pmatrix} (u^n - \mathbb{P}_{\beta\alpha}(u^n - \beta F'(u^n))) \\ &= u^n - \begin{pmatrix} \frac{1}{\beta} \mathcal{M}_{\mathbb{A}^n \mathbb{A}^n}^{-1} & -\mathcal{M}_{\mathbb{A}^n \mathbb{A}^n}^{-1} \mathcal{M}_{\mathbb{A}^n \mathbb{I}^n} \\ 0 & I_{\mathbb{I}^n} \end{pmatrix} \begin{pmatrix} \beta [F'(u^n)]|_{\mathbb{A}^n} \\ u_{\mathbb{I}^n}^n \end{pmatrix} \\ &= \begin{pmatrix} u_{\mathbb{A}^n}^n - \mathcal{M}_{\mathbb{A}^n \mathbb{A}^n}^{-1} ([F'(u^n)]|_{\mathbb{A}^n} - \mathcal{M}_{\mathbb{A}^n \mathbb{I}^n} u_{\mathbb{I}^n}^n) \\ 0 \end{pmatrix}. \quad (19)\end{aligned}$$

Here, we have implicitly assumed that $\mathcal{M}_{\mathbb{A}^n \mathbb{A}^n}^{-1}$ exists. Hence, instead of calculating the Newton update, we can set $u_{\mathbb{I}^n}^{n+1} = 0$, then solve the equation $\mathcal{M}_{\mathbb{A}^n \mathbb{A}^n} \delta u_{\mathbb{A}^n} = [F'(u^n)]|_{\mathbb{A}^n} - \mathcal{M}_{\mathbb{A}^n \mathbb{I}^n} u_{\mathbb{I}^n}^n$ and compute $u_{\mathbb{A}^n}^{n+1} = u_{\mathbb{A}^n}^n - \delta u_{\mathbb{A}^n}$.

We require some conditions for F to prove local convergence of semismooth Newton methods given in iteration (18) or (19).

Assumption 3 *Assume that*

- 1) Equation (14) has a solution $u^* \in U$, where U is an open set.
- 2) F is twice Fréchet differentiable, F' and F'' are Lipschitz continuous in U .
- 3) For an arbitrary but finite index set $\mathbb{A} \subset \Lambda$, the operator $F''(u)$ restricted to $\mathbb{A}$ has a bounded inverse, i.e. if F'' is represented by

$$F''(u) = \begin{pmatrix} \mathcal{M}_{\mathbb{A}\mathbb{A}} & \mathcal{M}_{\mathbb{A}\mathbb{I}} \\ \mathcal{M}_{\mathbb{I}\mathbb{A}} & \mathcal{M}_{\mathbb{I}\mathbb{I}} \end{pmatrix},$$

then there exists a $\rho > 0$ such that $\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}$ exists, and is uniformly bounded on the ball $B_\rho(u^*) \subset U$.

Under Assumption 3, following [18] we will show that the semismooth Newton method locally superlinearly converges. In doing so we need some auxiliary results.

Lemma 3.3 *Assume that F' is Lipschitz continuous with Lipschitz constant L in a neighborhood of u^* . Then there exist $k_0 \in \Lambda$ and $\rho > 0$ such that the condition $\|u - u^*\| < \rho$ implies the active set $A(u)$ satisfies the following relation:*

$$\mathbb{A}(u) \subset [1, k^0].$$

Moreover, k^0 and ρ depend on β, u^*, L and α .

Proof. The triangle inequality implies

$$|u_k - \beta F'(u)_k| \leq |u_k^* - \beta F'(u^*)_k| + |u_k - u_k^* - \beta(F'(u)_k - F'(u^*)_k)|. \quad (20)$$

The first term converges to zero as k tends to infinity, because u^* and $F'(u^*)$ are in X . In particular, there exists a k_0 , depending only on u^* and β , such that

$$|u_k^* - \beta F'(u^*)_k| < \beta\alpha/2 \text{ for all } k \geq k_0. \quad (21)$$

The second term can be estimated as follows,

$$\begin{aligned} |u_k - u_k^* - \beta(F'(u)_k - F'(u^*)_k)| &\leq |u_k - u_k^*| + \beta|F'(u)_k - F'(u^*)_k| \\ &\leq (1 + \beta L)\|u - u^*\|, \end{aligned}$$

where L is the Lipschitz constant of F' . Thus there exists $\rho > 0$ depending only on β, u^*, L and α such that

$$|u_k - u_k^* - \beta(F'(u)_k - F'(u^*)_k)| \leq \beta\alpha/2 \text{ for all } k \in \Lambda. \quad (22)$$

The proof of the lemma follows from (20)–(22). ■

Lemma 3.4 *If $F''(u) = \begin{pmatrix} \mathcal{M}_{\mathbb{A}\mathbb{A}} & \mathcal{M}_{\mathbb{A}\mathbb{I}} \\ \mathcal{M}_{\mathbb{I}\mathbb{A}} & \mathcal{M}_{\mathbb{I}\mathbb{I}} \end{pmatrix}$ and $\mathcal{M}_{\mathbb{A}\mathbb{A}}$ is injective, then $D'(u) : X \rightarrow X$ is invertible and*

$$\|D'(u)^{-1}\| \leq \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| \left(\frac{1}{\beta} + \|\mathcal{M}_{\mathbb{A}\mathbb{I}}\| \right) + 1,$$

where $\mathbb{A}$ and $\mathbb{I}$ respectively are the active and inactive sets at u .

Proof. Let $r \in X$ and consider the equation $D'(u)\delta u = r$, i.e. the equation

$$\begin{pmatrix} \beta\mathcal{M}_{\mathbb{A}\mathbb{A}} & \beta\mathcal{M}_{\mathbb{A}\mathbb{I}} \\ 0 & I_{\mathbb{I}} \end{pmatrix} \begin{pmatrix} \delta u_{\mathbb{A}} \\ \delta u_{\mathbb{I}} \end{pmatrix} = \begin{pmatrix} r_{\mathbb{A}} \\ r_{\mathbb{I}} \end{pmatrix}.$$

This equation is equivalent to $\delta u_{\mathbb{I}} = r_{\mathbb{I}}$ and

$$\beta\mathcal{M}_{\mathbb{A}\mathbb{A}}\delta u_{\mathbb{A}} = r_{\mathbb{A}} - \beta\mathcal{M}_{\mathbb{A}\mathbb{I}}r_{\mathbb{I}}. \quad (23)$$

Since the active set $\mathbb{A}$ is finite, $\mathcal{M}_{\mathbb{A}\mathbb{A}}$ is an injective operator (by hypothesis) on a finite dimensional space. Therefore, it is also surjective. We conclude that (23) has a unique solution and thus $D'(u)^{-1}$ exists. Furthermore, we have the estimate

$$\begin{aligned} \|D'(u)^{-1}r\| &= \left\| \begin{pmatrix} \frac{1}{\beta}\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1} & \mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\mathcal{M}_{\mathbb{A}\mathbb{I}} \\ 0 & I_{\mathbb{I}} \end{pmatrix} \begin{pmatrix} r_{\mathbb{A}} \\ r_{\mathbb{I}} \end{pmatrix} \right\| \\ &\leq \frac{1}{\beta}\|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| \|r_{\mathbb{A}}\| + \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| \|\mathcal{M}_{\mathbb{A}\mathbb{I}}\| \|r_{\mathbb{I}}\| + \|r_{\mathbb{I}}\| \\ &\leq \left(\frac{1}{\beta}\|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| + \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| \|\mathcal{M}_{\mathbb{A}\mathbb{I}}\| + 1 \right) \|r\|. \end{aligned}$$

■

Theorem 3.3 *If Assumption 3 is satisfied, then there exists a $r > 0$ such that the condition $\|u^0 - u^*\| < r$ implies the sequence u^n defined by (18) satisfy $\|u^n - u^*\| < r$ for all n , and $u^n \rightarrow u^*$ superlinearly.*

Proof. In order to apply classical results on superlinear convergence of generalized Newton methods ([5, Remark 2.7] or [22, Theorem 1.1]) we need to prove that $D'(u)^{-1}$ is uniformly bounded in a ball around u^* .

Let $\rho > 0$ be a number such that Assumption 3 and Lemma 3.3 are satisfied in the ball $B_\rho(u^*)$ with the center u^* and the radius ρ . By Lemma 3.3 there exists k_0 such that the active set satisfies $\mathbb{A}(u) \subset [0, k_0]$ for all $u \in B_\rho(u^*)$. We define

$$c(k_0) = \max_{\emptyset \neq \mathbb{A} \subset [0, k_0]} \sup_{u \in B_\rho(u^*)} \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| > 0.$$

Note that for every $\mathbb{A} \subset [0, k_0]$, $\mathbb{A} \neq \emptyset$, $\sup_{u \in B_\rho(u^*)} \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\|$ is finite by Assumption 3. Since $c(k_0)$ is the maximum of finitely many positive numbers, it exists, and it is a finite positive number.

On the other hand, again using Assumption 3 (the Lipschitz continuity implies uniformly boundedness), we have $\|\mathcal{M}_{\mathbb{A}\mathbb{I}}\| \leq \|F''(u)\| \leq \tau < \infty$ for all choices of $\mathbb{A}$ and $\mathbb{I}$.

Lemma 3.4 implies that

$$\|D'(u)^{-1}\| \leq c(k_0)\left(\frac{1}{\beta} + \tau\right) + 1.$$

Therefore, the inverse operator of the Newton derivative $D'(u)^{-1}$ is uniformly bounded in $B_\rho(u^*)$. The result now follows standard results for generalized Newton methods, see [5, Remark 2.7], [22, Theorem 1.1]. ■

4 Phase retrieval problem

In electron microscopy, optical imaging, astronomy, crystallography, and in other fields it often requires to recover an object from intensity measurements. In more detail, one is usually interested in determining an nonnegative and sparse object, $0 \leq f(x) \in L^2(\Omega)$, from $|F(u)|$, where $F(u)$ is the Fourier transform of $f(x)$, i.e.,

$$\mathcal{F}(u) = \int_{-\infty}^{\infty} f(x) \exp(-i2\pi u \cdot x) dx.$$

Here, x is an M -dimensional spatial coordinate, u is an M -dimensional spatial frequency coordinate and $x \cdot u$ is the inner product of two vectors x and u .

In practice, we are faced with the case $M = 2$ and deal with sampled data in computer, which are assumed to be square arrays, $u = (u_1, u_2)$ and $x = (x_1, x_2)$ where u_1, u_2, x_1 and $x_2 = 0, 1, 2, \dots, N-1$. For this data, the discrete Fourier transform (we use notation A for short) has the form

$$Af(u) = \sum_{x=0}^{N-1} f(x) \exp(-i2\pi u \cdot x/N) \quad (24)$$

with inverse

$$f(x) = \frac{1}{N^2} \sum_{u=0}^{N-1} Af(u) \exp(i2\pi u \cdot x/N). \quad (25)$$

A discretized mathematical model for the problem of phase retrieval is:

$$\begin{cases} \text{Find a real, nonnegative and sparse vector } f \in \mathbb{R}^{N^2} \text{ such that} \\ |Af| = b, \end{cases} \quad (26)$$

where $A \in \mathbb{R}^{N^2 \times N^2}$, $b \in \mathbb{R}^{N^2}$ and $|Af|$ is componentwise magnitude of vector Af .

The applications of this problem can be founded in [23] and the references therein. One special application, imaging through strongly scattering media, will be investigated in the next section. We emphasize that the constraints of nonnegativity and sparsity of solutions have not been investigated. In [23], without nonnegativity and sparsity of solutions, the problem is experimentally indicated to be ill-posed. The authors have indicated that two imaging systems of one object result in two different images, but their autocorrelations are the same. This shows that recovering image of an object from its autocorrelation or equivalently its FFT amplitude is not unique (the ill-posed of problem). In [24], the authors have investigated the uniqueness of the similar one as Problem (26) under mild conditions. There are several algorithms for solving this problem without considering nonnegativity and sparsity of solutions, but the most popular one is given by G.A. Fienup in [9]. Recently, the problem with a general linear operator F is investigated in [10] and some regularization algorithms based on the Moore-Penrose pseudoinverse of A are presented.

In the theory of inverse problems, a prior information of solutions is known, approximation solutions can be found with high fidelity. Such an information for Problem (26) is that it has a sparse and nonnegative solution. Sparsity and nonnegativity suggest to use the nonnegative sparse regularization:

$$\min_{f \in \mathbb{R}^{N^2}, y \in \mathbb{C}^{N^2}, |y|=b} |Af - y|^2 + \alpha \Phi(f), \quad (27)$$

where Φ is given by (6) and $\alpha > 0$ is a regularization parameter.

Now, we are going to approach to the solution using the two algorithms described in the previous sections, which utilize all of the three prior constraints (i.e. real, sparsity, and nonnegativity). We start with a f^0 and perform the iterations:

$$y^n = \operatorname{argmin}_{y \in \mathbb{C}^{N^2}, |y|=b} |Af^n - y|^2, \quad (28)$$

$$f^{n+1} = \operatorname{argmin}_{f \in \mathbb{R}^{N^2}} |Af - y^n|^2 + \alpha \Phi(f). \quad (29)$$

The solution of (28) is the orthogonal projection of Af^n on the set $\{y \in \mathbb{C}^{N^2} \mid |y| = b\}$ and is simply computed by

$$y^n(u) = b \exp(i\psi_n(u)) \text{ if } Af^n = |Af^n| \exp(i\psi_n(u)).$$

The solution of (29) can be approximated by one of two algorithms in Section 3.1 and Section 3.2. Note that the problem (29) falls into class of Problem (5) with

$$F(f) = |Af - y^n|^2,$$

and thus the first and second Fréchet derivatives of F respectively are

$$F'(f) = \operatorname{Re}(2A^*(Af - y^n)) \text{ and } F''(f) = A^*A = I,$$

because $A^{-1} = A^*$ for the Fourier transform. Therefore, all conditions for convergence of two algorithms in previous sections are satisfied. Applying these algorithms for the problem (26), we present in Algorithms 4.1 and 4.2, respectively.

Algorithm 4.1 Algorithm for (26) with (29) solved by a Gradient-type method

Input: Initial guess $f^0 \equiv 0$, outer iteration number N and inner iteration number m

```

1: for k=0:N do
2:    $g^k = Af^k = |Af^k| \exp(i\psi_k)$ 
3:    $y^k = b \exp(i\psi_k)$ 
4:    $f^{k,0} = f^k$ 
5:   for n = 0 : m do
6:      $f^{k,n+1} \leftarrow \mathbb{P}_\alpha(f^{k,n} - \operatorname{Re}(2A^{-1}(Af^{k,n} - y^k)))$ 
7:   end for
8:    $f^{k+1} = f^{k,m+1}$ 
9: end for
```

Output: $f = f^{N+1}$

Remark 4.1 (a) For Problem (26) since $F'' = I$ (the identity matrix), the Lipschitz constant of F' is one ($L = 1$). Thus, in Algorithm 4.1 we have used the gradient-type algorithm with constant stepsizes $\frac{1}{s^n} = 1$ for all n .

(b) Problem (26) has at least a solution (see Theorem 3.1). Note that since Problem (26) is nonconvex with the variable $u = (f, y)$, the objective functional may have some local minimizers. Therefore, Algorithm 4.1 and Algorithm 4.2 converge to the solution of Problem (26) only if a starting point is close enough to the solution. In our case, it happens if the exact solution of (1) is sparse enough. As a result, our approach is suitable if the exact solution of (1) is nonnegative and sparse.

5 Imaging through scattering media

In this section, we study the problem of imaging through the scattering media, such as frosted glass, optical diffuser or biological skin. The light passing through scattering media generates seemingly random speckle

Algorithm 4.2 Algorithm for (26) with (29) solved by SSN

Input: Initial guess $f^0 \equiv 0$, outer iteration number N , inner iteration number m and $\beta > 0$

```
1: for k=0:N do
2:    $g^k = Af^k = |Af^k| \exp(i\psi_k)$ 
3:    $y^k = b \exp(i\psi_k)$ 
4:    $f^{k,0} = f^k$ 
5:   for  $n = 0 : m$  do
6:      $grad^{k,n} = \text{Re}(2A^{-1}(Af^{k,n} - y^k))$ 
7:      $\mathbb{A}^n = \{i : [f^{k,n} - \beta grad^{k,n}]_i > \beta\alpha\}$ 
8:      $\mathbb{I}^n = \{i : [f^{k,n} - \beta grad^{k,n}]_i \leq \beta\alpha\}$ 
9:      $[f^{k,n+1}]_{\mathbb{A}^n} \leftarrow [f^{k,n} - grad^{k,n}]_{\mathbb{A}^n}$ 
10:     $[f^{k,n+1}]_{\mathbb{I}^n} \leftarrow 0$ 
11:   end for
12:    $f^{k+1} = f^{k,m+1}$ 
13: end for
```

Output: $f = f^{N+1}$

on the detector (camera). However, a special property of scattering media is an optical memory effect which allows us to recover the hidden object behind the scattering media. The concept of optical memory effect in speckle imaging states that any point on object plane would generate a random speckle pattern and other points on the object plane would generate identical but shifted speckle patterns with respect to the point positions [32, 33]. This refers to shift-invariant point spreading function, PSF, then the final intensity image on a camera should be equal to the superposition of these random speckle patterns [8, 11, 12, 23, 32, 33]. Therefore, the imaging intensity I on a camera could be expressed as the convolution of the object intensity pattern O with the shift-invariant random PSF :

$$I(x, y) = O(x, y) * PSF(x, y),$$

where $*$ represents the convolution operator, x or y corresponds to the transverse coordinate along x or y direction at the object or camera plane. The auto-correlation operation of the imaging field could be expressed as the following formula with the convolution theorem:

$$I \star I = (O * PSF) \star (O * PSF) = (O \star O) * (PSF \star PSF) \approx O \star O,$$

where $\star$ represents the auto-correlation operation. The auto-correlation of the random PSF can be approximated as a delta function ($PSF \star PSF \approx \delta$)[13]. Therefore, the autocorrelation of the object can be approximated from the autocorrelation of the speckle image:

$$I \star I = (O \star O) * \delta \approx O \star O.$$

It is known that the Fourier transform of the autocorrelation is the energy-spectrum. Therefore, the object Fourier transform amplitude can be estimated as the following:

$$\sqrt{A(I \star I)} \approx \sqrt{A(O \star O)} = |A(O)|,$$

With an appropriate mean removal, one can extract the Fourier amplitude of the object from the auto-correlation of the speckle image as shown in Figure 1(b). Therefore phase retrieval has become an essential requirement for imaging through scattering medium. And this problem is one specific case of Problem (26). Applying our approach leads to Problem (27).

Note that without considering the sparsity and non-negativity of the exact solution, there are some algorithms for the problem (Speckle Imaging) [9, 10]. Thus, a good comparison between these algorithms and Algorithms 4.1 and 4.2 should be considered from the points of convergence rate, quality of recovered solutions, sparsity and non-negativity of recovered solutions, and so on. However, such comparisons between these algorithms are beyond the scope of this paper. A detailed review of the existing algorithms for speckle imaging and their analysis is kept as a future work.

Here we demonstrate the use of our algorithms with this phase retrieval application. We are going to showcase three scenarios. First, the ideal discrete Fourier amplitude information is given to the algorithm. Second, an imaging system is simulated with an ideal speckle PSF , and the estimated Fourier amplitude of the object is given to the algorithm. Third, we have taken the actual speckle image captured through a

scattering medium in presence of all possible hardware noises, and the estimated Fourier amplitude is given to the algorithm. Except the first scenario, the input Fourier amplitude of the object is fairly deformed as describe in equation (2).

For the first part, we have taken an image of letter 4 as shown in Figure 1(a). The Fourier amplitude is extracted from the ideal image's auto-correlation, as shown in Figure 1(b). The Fourier phase is recovered from both Algorithm 4.1 and Algorithm 4.2, and the reconstructed images are shown in Figure 1(c) and (d). For the second part, we simulated a speckle *PSF* using Rayleigh distribution, and a speckle image is formed by convolving the *PSF* with the digit 4 image as shown in Figure 1(e). The Fourier amplitude is estimated as the aforementioned speckle auto-correlation principle, and it is shown in Figure 1(f). The Fourier phases are recovered by the proposed algorithms and the reconstructions are presented in Figure 1(g) and (h). Next we are going to describe our experimental for a real speckle imaging scenario.

Figure 1: Reconstruction results from the Fourier amplitude of both the ideal image and a simulated speckle image. (a) Ideal input image. (b) Autocorrelation of the ideal image. (c) Reconstructed ideal image by Algorithm 4.1 ($\alpha = 8e^{-3}, m = 7, N = 5000$). (d) Reconstructed ideal image by Algorithm 4.2 ($\alpha = 9e^{-4}, \beta = 1e^{-3}, m = 5, N = 1000$). (e) Simulated speckle image. (f) Autocorrelation of the speckle image. (g) Reconstructed image by Algorithm 4.1. (h) Reconstructed image by Algorithm 4.2. Autocorrelation area equals to 120×120 (pixels) in both algorithms.

The schematic of the experimental setup to measure the speckle image and a reconstruction demonstration are depicted in the Figure 2. The setup in Figure 2(a) contains a coherent laser (*DL660-050*), a home-built $10\times$ beam expander, a rapidly rotating diffuser (*ED1-C20-MD*), a target composed of different letters, a fixed diffuser (*DG10-1500-MD*), iris with 4mm diameter closed to a fixed diffuser, and a camera (Andor Neo 5.5, 25602160, pixel size $6.5\mu m$). The target with 'NTU' letters is shown in Figure 2(b), is positioned at a distance $u \approx 640mm$ from the diffuser and iris. The beam expander (consisted of 2 lenses) is to generate large area illumination to cover the entire sample. With nearly uniform incoherent illumination, the scattering diffraction pattern shown in Figure 2(c) is captured on the camera, which is placed at a distance $v \approx 160mm$ from the other side of the diffuser. The illumination used in the experimental setup is pseudo incoherent source where the optical phase is randomized by rotation diffuser. Therefore, the measure speckle intensity image is superposition of PSF intensities [12, 8, 11, 23].

The Fourier phase would be recovered from both Algorithm 4.1 and Algorithm 4.2, and the reconstructed images are shown in Figure 2(e,f). In addition, two different targets with number '0' and '5' are demonstrated with other experimental parameters: $u \approx 800mm, v \approx 160mm$ (i.e. it means the magnification is ~ 0.2). Figure 3 shows the original targets on the object plane, experimental speckle images on the camera, their corresponding auto-correlations and the recovered reconstructions. The simulation and experiment results suggest that the proposed algorithms take advantage of all the available priors, and they can be utilized for the purpose of an effective phase retrieval in real applications.

Figure 2: Experimental setup and reconstructed objects through the scattering medium. (a) Experimental setup: BE, beam expander. RD, rotating diffuser. (b) Microscope image of 'NTU' letters, which are the opening area on a chromium film. (c) Central pattern of the raw speckle image on the camera. (d) Auto-correlation of (c). (e) Reconstructed image from (d) by Algorithm 4.1 ($\alpha = 8e^{-4}$, $m = 7$, $N = 6000$). (f) Reconstructed from (d) by Algorithm 4.2 ($\alpha = 1e^{-3}$, $\beta = 1e^{-3}$, $m = 5$, $N = 5000$). Autocorrelation area equals to 180×180 (pixels) in both algorithms.

Figure 3: Experimental reconstruction from different targets when $u \approx 800mm$, $v \approx 160mm$. (a, f) Microscope images of number 0 and 5. (b, g) Measured speckle distributions on CCD. (c, h) Corresponding auto-correlations in b and g. (d, i) Reconstructed images by Algorithm 4.1 ($\alpha = 8e^{-4}$, $m = 7$, $N = 6000$) from their auto-correlations. (e, j) Reconstructed images by Algorithm 4.2 ($\alpha = 1e^{-3}$, $\beta = 1e^{-3}$, $m = 5$, $N = 1000$) from their auto-correlations. Autocorrelation Area = 150 in both algorithms.

6 Conclusion

In summary, we have proposed a functional for nonnegative sparse regularization (NSR) to find the solution of the inverse problems when such prior knowledge is available. The minimizers with the proposed functional provides regularized approximate solutions for an ill-posed problem. The advantage of the proposed NSR is that the numerical minimizers can be found by some efficient algorithms. Here, we have proposed such two methods (algorithms): gradient-type method and Semismooth Newton method. We also have been proven the convergence of these algorithms. In addition, we have applied them to the phase retrieval problem, and we have tested on numerical simulation as well as on real experimental data. The proposed algorithms have recovered the solutions successfully, and thus confirms their suitability for real-life applications.

Acknowledgement

This research was supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 101.02-2017.318 and by the Projects: KHCN-TB.12C/13-18 and B2016-DNA-44-TT. We also gratefully acknowledges the support by Singapore MOE - AcRF - Tier1 grant (RG70/15) and Singapore Ministry of Health's National Medical Research Council under new investigator grant (NMRC/BNIG/2039/2016).

References

- [1] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Springer, 2010.
- [2] J.F. Aujol. Some first-order algorithms for total variation based image restoration. *J. Math. Imaging Vision*, 34(3):307–327, 2009.
- [3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
- [4] T. Bonesky, K. Bredies, D.A. Lorenz, and P. Maass. A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. *Inverse Problems*, 23(5):2041–2058, 2007.
- [5] X. Chen. Superlinear convergence of smoothing quasi-Newton methods for nonsmooth equations. *J. Comput. Appl. Math.*, 80(1):105–126, 1997.
- [6] I. Daubechies, M. Defrise, and Ch. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.
- [7] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375. Springer Science & Business Media, 1996.
- [8] S. Feng, C. Kane, P.A. Lee, and A Douglas Stone. Correlations and fluctuations of coherent wave transmission through disordered media. *Phys. Rev. Lett.*, 61(7):834, 1988.
- [9] J.R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):2758–2769, 1982.
- [10] F. Fogel, I. Waldspurger, and A. d’Aspremont. Phase retrieval for imaging problems. *Math. Program. Comput.*, 8(3):311–335, 2016.
- [11] I. Freund. Looking through walls and around corners. *Phys. A*, 168(1):49–65, 1990.
- [12] I. Freund, M. Rosenbluh, and S. Feng. Memory effects in propagation of optical waves through disordered media. *Phys. Rev. Lett.*, 61(20):23–28, 1988.
- [13] J.W. Goodman. *Speckle Phenomena in Optics: Theory and Applications*. Roberts and Company Publishers, 2007.
- [14] M. Grasmair. Non-convex sparse regularisation. *J. Math. Anal. Appl.*, 365(1):19–28, 2010.
- [15] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with ℓ^q penalty term. *Inverse Problems*, 24(5):055020, 2008.
- [16] M. Grasmair, O. Scherzer, and M. Haltmeier. Necessary and sufficient conditions for linear convergence of ℓ^1 -regularization. *Comm. Pure Appl. Math.*, 64(2):161–182, 2011.

- [17] D. Greene, G. Cagney, N. Krogan, and P. Cunningham. Ensemble non-negative matrix factorization methods for clustering protein–protein interactions. *Bioinformatics*, 24(15):1722–1728, 2008.
- [18] R. Griesse and D.A. Lorenz. A semismooth Newton method for Tikhonov functionals with sparsity constraints. *Inverse Problems*, 24(3):035007, 19, 2008.
- [19] D.N. Hào and T.N.T. Quyen. Convergence rates for total variation regularization of coefficient identification problems in elliptic equations I. *Inverse Problems*, 27(7):075008, 2011.
- [20] D.N. Hào and T.N.T. Quyen. Convergence rates for tikhonov regularization of a two-coefficient identification problem in an elliptic boundary value problem. *Numerische Mathematik*, 120(1):45–77, 2012.
- [21] D.N. Hào and T.N.T. Quyen. Convergence rates for total variation regularization of coefficient identification problems in elliptic equations II. *J. Math. Anal. Appl.*, 388(1):593–616, 2012.
- [22] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13(3):865–888 (2003), 2002.
- [23] O. Katz, P. Heidmann, M. Fink, and S. Gigan. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature Photonics*, 8(10):784–790, 2014.
- [24] M.V. Klibanov. On the recovery of a 2-D function from the modulus of its Fourier transform. *J. Math. Anal. Appl.*, 323(2):818–843, 2006.
- [25] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [26] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16(5):463–478, 2008.
- [27] D.A. Lorenz, P. Maass, and P.Q. Muoi. Gradient descent for Tikhonov functionals with sparsity constraints: theory and numerical comparison of step size rules. *Electron. Trans. Numer. Anal.*, 39:437–463, 2012.
- [28] P.Q. Muoi, D.N. Hào, P. Maass, and M. Pidcock. Semismooth newton and quasi-newton methods in weighted 1-regularization. *J. Inverse and Ill-Posed Probl.*, 21(5):665–693, 2013.
- [29] P.Q. Muoi, D.N. Hào, P. Maass, and M. Pidcock. Descent gradient methods for nonsmooth minimization problems in ill-posed problems. *J. Comput. Appl. Math.*, 298:105–122, 2016.
- [30] Y. Nesterov et al. Gradient methods for minimizing composite objective function, 2007.
- [31] P.Q. Muoi Reconstructing conductivity coefficients based on sparsity regularization and measured data in electrical impedance tomography. *Inverse Problems in Science and Engineering*, 23(8):1366–1387, 2015.
- [32] S.K. Sahoo, D.L. Tang, and C. Dang. Enhancing security of incoherent optical cryptosystem by a simple position-multiplexing technique and ultra-broadband illumination. *Scientific reports*, 7(1):17895, 2017.
- [33] S.K. Sahoo, D.L. Tang, and C. Dang. Single-shot multispectral imaging with a monochromatic camera. *Optica*, 4(10):1209–1213, 2017.
- [34] P. Smaragdis and J.C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 177–180. IEEE, 2003.
- [35] M. Ulbrich. Semismooth Newton methods for operator equations in function spaces. *SIAM J. Optim.*, 13(3):805–842 (2003), 2002.
- [36] P. Weiss, L. Blanc-Féraud, and G. Aubert. Efficient schemes for total variation minimization under constraints in image processing. *SIAM J. Scient. Comput.*, 31(3):2047–2080, 2009.