



Article

Sparse Regularization with a Non-Convex Penalty for SAR Imaging and Autofocusing

Zi-Yao Zhang *, Odysseas Pappas , Igor G. Rizaev and Alin Achim

Visual Information Laboratory, University of Bristol, Bristol BS1 5TE, UK; o.pappas@bristol.ac.uk (O.P.); i.g.rizaev@bristol.ac.uk (I.G.R.); alin.achim@bristol.ac.uk (A.A.)

* Correspondence: ziyao.zhang@bristol.ac.uk

Abstract: In this paper, SAR image reconstruction with joint phase error estimation (autofocusing) is formulated as an inverse problem. An optimization model utilising a sparsity-enforcing Cauchy regularizer is proposed, and an alternating minimization framework is used to solve it, in which the desired image and the phase errors are estimated alternatively. For the image reconstruction sub-problem (f -sub-problem), two methods are presented that are capable of handling the problem's complex nature. Firstly, we design a complex version of the forward-backward splitting algorithm to solve the f -sub-problem iteratively, leading to a complex forward-backward autofocusing method (CFBA). For the second variant, techniques of Wirtinger calculus are utilized to minimize the cost function involving complex variables in the f -sub-problem in a direct fashion, leading to Wirtinger alternating minimization autofocusing (WAMA) method. For both methods, the phase error estimation sub-problem is solved by simply expanding and observing its cost function. Moreover, the convergence of both algorithms is discussed in detail. Experiments are conducted on both simulated and real SAR images. In addition to the synthetic scene employed, the other SAR images focus on the sea surface, with two being real images with ship targets, and another two being simulations of the sea surface (one of them containing ship wakes). The proposed method is demonstrated to give impressive autofocusing results on these datasets compared to state-of-the-art methods.

Keywords: SAR autofocusing; Cauchy regularization; Wirtinger calculus; forward-backward splitting; KL property; sea surface



Citation: Zhang, Z.-Y.; Pappas, O.; Rizaev, I.G.; Achim, A. Sparse Regularization with a Non-Convex Penalty for SAR Imaging and Autofocusing. *Remote Sens.* **2022**, *14*, 2190. <https://doi.org/10.3390/rs14092190>

Academic Editor: Dusan Gleich

Received: 25 March 2022

Accepted: 28 April 2022

Published: 4 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Synthetic aperture radar (SAR) has become one of the most-employed modalities in the field of remote sensing, largely due to its capability to collect data in all kinds of weather and lighting conditions. As a coherent imaging radar system often mounted on an airplane or satellite platform, it can transmit signals after frequency modulation to a certain scene on the ground and record the return radar echoes in flight. As with other coherent radar systems, these raw radar returns must then be processed to form an image suitable for visual interpretation. A detailed introduction on the working mechanisms of SAR can be found in [1,2].

The SAR data acquisition process is unfortunately frequently plagued by phase errors. Due to inaccuracies in the SAR platform trajectory measurement as well as the possible existence of moving targets in the observed scene, the acquired data (radar returns) will contain phase errors. These phase errors, in turn, result in a defocusing effect in the formed SAR images. Techniques aiming at the direct estimation of these phase errors from the raw SAR data and the removal of them so as to improve the quality of the reconstructed SAR images are called autofocusing techniques. Additionally, extending autofocusing to the case of moving targets, such as ships in motion, is also of particular interest. It can not only address the additional blurring effects caused by target motion, but also address the direct target return displacement effect, where a ship appears disjointed from its wake,

making localisation and velocity/heading estimation problematic [3]. An example of such displacement can be seen in Figure 1.

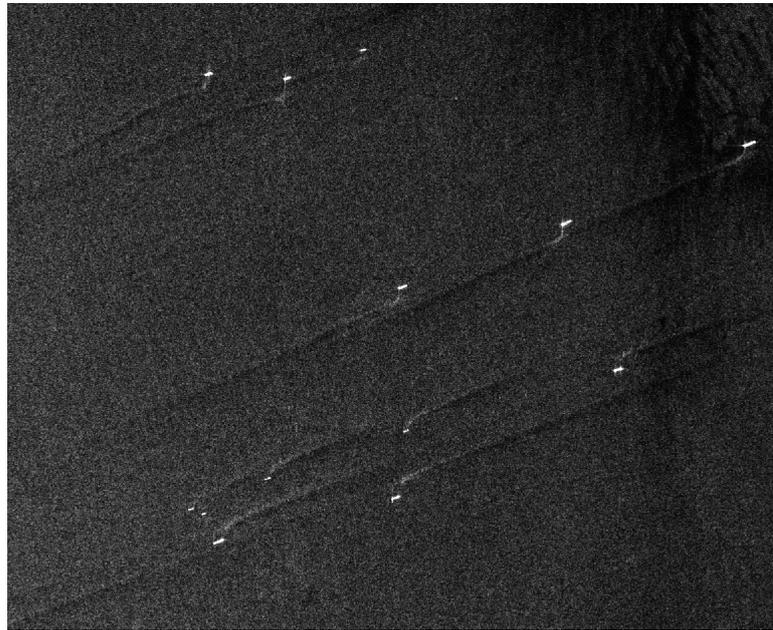


Figure 1. Example of moving ship targets appearing displaced from their wakes in Sentinel-1 image (IW intensity image, VV polarisation).

Among the earliest autofocusing techniques, phase gradient autofocus (PGA) [4] is a very well-known method. It first circularly shifts and windows the data, then uses these processed data to estimate the gradient of the phase errors, and finally integrates the estimations to obtain the phase errors themselves. Mapdrift autofocus is another classical technique for SAR autofocusing [5]. It partitions the whole aperture into several sub-apertures from which the images (maps) are reconstructed, and measures the drift between each pair of maps to obtain the phase errors. These classic methods often serve as the basis for newer methods, such as the variants of the mapdrift and phase gradient autofocus algorithms presented in [6], which employ fractional lower-order moments of the alpha-stable distribution for modelling the phase history data.

Differing from them, various methods based on optimization techniques have been proposed in recent years. Many of these methods belong to one of two categories. The first involves the construction of a sharpness metric to be maximized. For instance, a power function is chosen as the sharpness metric in [7,8], and it is further demonstrated in [8] that when the image is of multiple columns, and under certain statistical assumptions, the perfectly focused image can be closely approximated according to the strong law of large numbers. Image entropy is another popular alternative [9–11], though in this case it is minimized (rather than maximized) to enforce sharpness.

The second category of methods adopts an inverse problem approach. Based on a forward observation model associating the corrupted phase history with the underlying SAR image, SAR autofocusing is formulated as an inverse problem, and variational models with a variety of regularizers have been designed to obtain its solution. For instance, Onhon et al. uses the p th power of the approximate l_p norm as the regularization term and an alternating minimization framework to solve the problem [12]. There are also various methods addressing the problem in a compressive sensing context, such as the majorization-minimization-based method [13,14], iteratively re-weighted augmented Lagrangian-based method [15,16] and conjugate gradient-based method with a cost function involving hybrid regularization terms (approximate l_1 norm and approximate total variation regularization) [17].

Besides these, there are also SAR autofocus approaches built by directly strengthening traditional SAR imaging methods. For example, an autofocus method which maximizes a sharpness metric for each pulse in the imaging process of back-projection is proposed in [18] and further extended to the case of moving ship targets [19]. A polar format algorithm-based autofocus approach [20] which combines [12] with a classical autofocus method like PGA has been proposed recently as well.

Moreover, with the development of deep learning, deep neural networks have also recently been considered in SAR autofocus. A recurrent auto-encoder-based SAR imaging network mimicking the behavior of an iterative shrinkage thresholding algorithm (ISTA) is proposed [21], in which the removal of phase errors is achieved by learning the forward model (observation matrix). Another auto-encoder and decoder-based neural network is built in [22], and the motion compensation is achieved by adding an updating step of the observation matrix in their alternating minimization framework.

In this paper, inspired by [12], we formulate the SAR autofocus problem as an inverse problem. This enables us to jointly estimate the desired SAR image and the unknown phase errors instead of treating autofocus as a post-processing step after image reconstruction like traditional autofocus methods do, thus leading to better performance. In the proposed cost function, we use a non-convex Cauchy regularization on the magnitude of the desired image (thus, we call it “magnitude Cauchy”), exploiting its sparsity-enforcing ability and differentiability. However, this inverse problem poses an interesting challenge in comparison to many other inverse problems in imaging in that it deals with real-valued functions with complex variables. To handle its complex nature, we present two methods for SAR image formation and autofocus under a framework of alternating minimization. The first method is based on a complex-domain adaptation of the forward-backward splitting algorithm, and is called CFBA. The second method is based on Wirtinger calculus [23–26] which is designed to deal with differentiable (in the sense of [23]) real-valued functions with complex variables. The second method has been introduced in our previous work [27], referred to as the Wirtinger alternating minimization autofocus (WAMA) method, but we further give a thorough discussion of its convergence and show that it can be extended to the cases of several other regularizers.

When validating the performance of our methods, we mainly focus on datasets representing the sea surface, i.e., real SAR images containing ships and simulated images of sea surface containing sea wave and ship wakes. The autofocus of this kind of SAR image can be beneficial for many practical applications, including ship detection [28], the characterization of ship wakes [29], and so on. The source code used to produce the presented results has been made available on GitHub at <https://github.com/zy-zhangc/SAR-autofocusing-with-a-non-convex-penalty> (accessed on 3 April 2022).

The rest of this paper is organized as follows. In Section 2, a brief introduction is given to the data acquisition model for SAR and the associated problem formulation. In Section 3, the proposed forward-backward splitting-based SAR autofocus method is described in detail, then its convergence is analyzed. In Section 4, the WAMA method is reviewed, and some more discussion is added, including its extension to the cases of other regularizers and its convergence. In Section 5, experimental results on both simulated scenes and real SAR images are shown to demonstrate the effectiveness of the proposed method. Finally, conclusions are presented in Section 6.

2. SAR Data Acquisition Model

In this paper, a SAR platform operating in spotlight mode is considered, whose transmitted signal at each azimuth position can be expressed as:

$$s(t) = \text{Re}\{e^{j(\omega_0 t + \alpha t^2)}\}, \quad (1)$$

where ω_0 is the carrier frequency, 2α is the chirp rate, and t is fast time.

The obtained data $r_m(t)$ at the m th aperture position and the latent SAR image $F(x, y)$ can be associated by:

$$r_m(t) = \iint F(x, y)e^{-jU(x\cos\theta+y\sin\theta)} dx dy. \tag{2}$$

The region over which the integral is computed is $x^2 + y^2 \leq L^2$, with L being the radius of the circular patch on the ground to be imaged. θ is the look angle, and U is defined by

$$U = \frac{2}{c}(\omega_0 + 2\alpha(t - \tau_0)), \tag{3}$$

with τ_0 being the demodulation time. Let M be the number of aperture positions, K be the number of range samples in each range line, and N be the number of pixels of the desired SAR image; then, the discretized version of this model is

$$r_m = C_m f, \quad m = 1, \dots, M, \tag{4}$$

where r_m and C_m are the $K \times 1$ column vector of the phase history and the $K \times N$ observation matrix for the m th aperture position, respectively. f is the $N \times 1$ column vector of the underlying SAR image. Stacking (4) with respect to all the aperture positions, and considering phase errors as well as possible noise, the model becomes

$$g = C(\phi) f + n, \tag{5}$$

with g being the $MK \times 1$ vector of the corrupted phase history, ϕ being the $MK \times 1$ vector of phase errors, and n being the $N \times 1$ vector of Gaussian white noise. $C(\phi)$ is the $MK \times N$ corrupted observation matrix. For more details on this signal model, please see [30,31]. In this paper, only the case of 1D phase errors varying in the azimuth direction is dealt with, which leads to [12]:

$$C_m(\phi) = e^{j\phi_m} C_m, \tag{6}$$

where $C_m(\phi)$ and ϕ_m are the corrupted observation sub-matrix for the m th aperture position and the phase error for the m th aperture position, respectively. The case of 2D separable phase errors as well as the case of 2D non-separable phase errors can be formulated similarly; see [12]. For simplicity, we set $MK = N$ in this paper.

3. The Proposed CFBA Method

3.1. The Optimization Model

We formulate SAR autofocus as an inverse problem and minimize the following cost function:

$$J(f, \phi) = \|g - C(\phi) f\|_2^2 - \lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + |f_i|^2}, \tag{7}$$

where λ is the regularization parameter and γ is the scale parameter for a Cauchy distribution.

The penalty term in (7) is a Cauchy regularization merely imposed on the magnitude of the latent SAR image. Consequently, we refer to it as ‘‘magnitude Cauchy regularization’’. Like the l_p norm, it too is a regularization term enforcing statistical sparsity [32], whose effectiveness has already been validated in SAR autofocus [27], as well as SAR imaging and other inverse problems [33,34].

In (7), the desired SAR image f and the phase errors ϕ are both unknown. By jointly estimating them, SAR image reconstruction and the removal of phase errors are accomplished simultaneously. To do this, we adopt an alternating minimization autofocus framework similar to [12]. Specifically, f and ϕ are updated alternatively by fixing one of them while optimizing the other. This iterative process will terminate when the relative error between $f^{(n)}$ and $f^{(n+1)}$ is smaller than 10^{-3} .

3.2. Complex Forward-Backward Splitting-Based Method

Overall, in our alternating minimization framework, we iteratively implement the following two steps:

(1) Image reconstruction step: solve the f -sub-problem by

$$f^{(n+1)} = \arg \min_{f \in \mathbb{C}^N} J(f, \phi^{(n)}); \tag{8}$$

(2) Phase error estimation step: solve the ϕ -sub-problem by

$$\phi^{(n+1)} = \arg \min_{\phi \in \mathbb{C}^N} J(f^{(n+1)}, \phi). \tag{9}$$

The first step will be detailed in Section 3.2.1, while the second step will be introduced in Section 3.2.2.

3.2.1. Image Reconstruction Step

Under the framework of alternating minimization, each f -sub-problem to be solved is

$$f^{(n+1)} = \arg \min_{f \in \mathbb{C}^N} \|g - C(\phi^{(n)})f\|_2^2 - \lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + |f_i|^2}. \tag{10}$$

Unlike many other inverse problem formulations in computational imaging, Equation (10) is an optimization problem involving a complex unknown vector, and it thus needs to be handled with appropriate mathematical tools. To this end, we design an iterative solution, which is a complex version of the well-known forward-backward splitting algorithm, and thus we call it complex forward-backward autofocusing (CFBA). It holds the benefit that the computation of the involved complex proximity operators can be converted to the computation of real proximity operators related to the magnitudes of the original complex vector’s components. Therefore, the techniques for real optimization regarding proximal operators can be leveraged. As a result, it is also possible to solve (7) with these techniques when the magnitude Cauchy regularization is replaced by certain non-smooth regularizers. Note that this is not always the case with the WAMA algorithm—the ability to be easily generalised to a number of non-smooth regularizers is a distinct advantage of the CFBA method.

Similar to the forward-backward splitting algorithm for a real case, we recast (10) as

$$f^{(n+1)} = \arg \min_{f \in \mathbb{C}^N} H(f) + G(f), \tag{11}$$

where $H(f) = \|g - C(\phi^{(n)})f\|_2^2$ and $G(f) = \sum_{i=1}^N r(f_i) = -\lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + |f_i|^2}$.

To minimize (11), with a given initial $\mathbf{o}^{(0)}$, we iteratively implement the following step:

$$\mathbf{o}^{(k+1)} = \text{prox}_{\mu G}(\mathbf{o}^{(k)} - 2\mu \nabla_f H(\mathbf{o}^{(k)})), \tag{12}$$

with the proximity operator defined by

$$\text{prox}_{\mu G}(x) = \arg \min_{y \in \mathbb{C}^N} \frac{1}{2} \|x - y\|_2^2 + \mu G(y), \tag{13}$$

and the function $\frac{1}{2} \|x - y\|_2^2 + \mu G(y)$ on its right-hand side is called a Moreau envelope [35]. $f^{(n+1)}$ is given by the final output $\mathbf{o}^{(K)}$ of this inner iterative loop.

In (12), $H(f)$ is a real-valued function with a complex vector variable, which is much discussed in Wirtinger calculus. Therefore, instead of using an ordinary gradient operator

defined in a real case ∇ , here we use the complex gradient operator ∇_f defined by Wirtinger calculus. Its definition is first proposed in [23] and further extended in [25]:

$$\nabla_f h = \Omega_f^{-1} \left(\frac{\partial h}{\partial \bar{f}} \right)^H. \tag{14}$$

where Ω_f^{-1} is a metric tensor. Using Brandwood’s setting, i.e., letting it be equal to the identity matrix, we have

$$\nabla_f h = \left(\frac{\partial h}{\partial \bar{f}} \right)^H. \tag{15}$$

As in (14) and (15), in the rest of this paper, we will use ∇_f to denote the complex gradient operator defined by Wirtinger calculus, and ∇ to denote the ordinary gradient operator defined in the real case.

Since the cost function is real-valued, according to [25], we have:

$$\left(\frac{\partial h}{\partial \bar{f}} \right)^H = \overline{\left(\frac{\partial h}{\partial \bar{f}} \right)^T} = \left(\frac{\partial h}{\partial \bar{f}} \right)^T = \left(\frac{\partial h}{\partial \bar{f}_1}, \dots, \frac{\partial h}{\partial \bar{f}_N} \right)^T. \tag{16}$$

The last term in the right side of (16) can be computed using the chain rule and the definition of the conjugate \mathbb{R} -derivative, i.e., $\frac{\partial h}{\partial \bar{f}_i} = \frac{1}{2} \left(\frac{\partial h}{\partial x_i} + i \frac{\partial h}{\partial y_i} \right)$, with x_i and y_i being the real and imaginary part of f_i , respectively [25].

It should also be noted that in (12), the step size is written as 2μ rather than the commonly used μ in the real-case forward-backward splitting algorithm. This choice is implied by the relationship between the complex gradient ∇_f and real gradient ∇ ; see e.g., [23]. Furthermore, μ should satisfy $\mu \leq \frac{1}{L}$, as will be further discussed in Section 3.2.

As a result of Wirtinger calculus, it can be shown that

$$\nabla_f H(f) = C(\boldsymbol{\phi}^{(n)})^H (C(\boldsymbol{\phi}^{(n)})f - \mathbf{g}). \tag{17}$$

As for the computation of (13), we first expand it as

$$\text{prox}_{\mu G}(x) = \arg \min_{\mathbf{y} \in \mathbb{C}^N} \frac{1}{2} \sum_{i=1}^N |x_i - y_i|^2 - \mu \lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + |y_i|^2}. \tag{18}$$

Then we solve (18) by independently solving

$$\text{prox}_{\mu \lambda r}(x_i) = \arg \min_{y_i \in \mathbb{C}} \frac{1}{2} |x_i - y_i|^2 - \mu \lambda \ln \frac{\gamma}{\gamma^2 + |y_i|^2} \tag{19}$$

for each i ($i = 1, \dots, N$).

For (19), observe that the logarithm term in the Moreau envelope merely depends on $|y_i|$; thus the solution y_i^* must lie on the line passing through the origin and the input x_i , as long as x_i is not 0. When searching among all the y_i for y_i^* , if $|y_i|$ is fixed, then only the first term of (19) needs to be minimized, since the second term therein is a constant in this case. To minimize the former, we just need to find the point on a circle with radius $|y_i|$ in the complex plane which is closest to the fixed point x_i . Obviously, this point would be the one which also lies on the line passing through the origin and x_i , if x_i is not 0. Therefore, the desired point will have the same argument as x_i . Since the choice of $|y_i|$ is arbitrary, the argument of y_i^* must be the same as that of x_i .

However, if $x_i = 0$, every y_i on a certain circle of the complex plane will be the solution of (19). In this case, we set the argument of y_i as 0, as in [35].

Therefore, the solution of (19) can now be split into two steps. The first is to solve the corresponding real optimization problem which gives $|y_i^*|$:

$$|y_i^*| = \arg \min_{y \in \mathbb{R}} \frac{1}{2}(|x_i| - y)^2 - \mu\lambda \ln \frac{\gamma}{\gamma^2 + y^2}. \quad (20)$$

The second step is to let

$$y_i^* = \begin{cases} |y_i^*| e^{j\phi_{x_i}} & x_i \neq 0 \\ |y_i^*| & x_i = 0 \end{cases} \quad (21)$$

where $e^{j\phi_{x_i}} = x_i/|x_i|$. Similar techniques can also be seen in phase retrieval [35] and SAR imaging [36].

Note that (20) is a non-convex optimization problem. If we compute the gradient of the Moreau envelope in the right-hand side and set it to 0, we will get a cubic equation. It may have three real roots, which stands for three stationary points. Here, however, we can add some constraints to simplify the problem. By restricting $\gamma > \frac{\sqrt{\mu\lambda}}{2}$, this Moreau envelope becomes strictly convex and thus implies the existence of only one, real stationary point [34]. In this case, the corresponding cubic equation must have a single real root and a pair of complex roots, and since our desired solution is a magnitude, then the solution we seek must be the real root. This real root is given by [33]:

$$|y_i^*| = \frac{|x_i|}{3} + s + t, \quad (22)$$

where

$$s = \sqrt[3]{\frac{q}{2} + \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}}, \quad (23)$$

$$t = \sqrt[3]{\frac{q}{2} - \sqrt{\frac{p^3}{27} + \frac{q^2}{4}}}, \quad (24)$$

$$p = \gamma^2 + 2\mu\lambda - \frac{|x_i|^2}{3}, \quad (25)$$

$$q = \gamma^2|x_i| + \frac{2|x_i|^3}{27} - \frac{\gamma^2 + 2\mu\lambda}{3}|x_i|. \quad (26)$$

3.2.2. Optimization of the Phase Errors

After obtaining each $f^{(n+1)}$, we use it to compute $\phi^{(n+1)}$. For 1D phase errors varying along the azimuth direction, the vector of phase errors can be updated by solving the following sequence of sub-problems concerning its components:

$$\phi_m^{(n+1)} = \arg \min_{\phi_m} \|g_m - e^{j\phi_m} C_m f^{(n+1)}\|_2^2, m = 1, \dots, N, \quad (27)$$

with g_m and C_m being the parts of g and C corresponding to the m th aperture position. According to [12], the solution is

$$\phi_m^{(n+1)} = \arctan\left(\frac{\operatorname{Re}\{[f^{(n+1)}]^H C_m g_m\}}{\operatorname{Im}\{[f^{(n+1)}]^H C_m g_m\}}\right). \quad (28)$$

The corrupted observation matrix can then be estimated by:

$$C_m(\phi_m^{(n+1)}) = e^{j\phi_m^{(n+1)}} C_m. \quad (29)$$

The cases of 2D phase errors varying in both range direction and cross-range direction can also be solved by similar methods; see [12] for more details. The whole process of the proposed CFBA method is summarized in Algorithm 1.

We point out here that CFBA differs from [33] mainly in two aspects. First, the authors of the other study do not deal with the same problem. The method in [33] is applied to SAR imaging, while CFBA addresses a different and more challenging problem, i.e., SAR autofocusing, which removes phase errors and reconstructs SAR image simultaneously. Second, only part of the solving process in CFBA is inspired by [33]. When facing each f -sub-problem of CFBA, techniques similar to [35,36] are used to tackle the complex nature of the problem, which makes it possible to handle the magnitudes and the arguments of the components of the complex solution separately. On the basis of this, Formulas (22)–(26) in [33] are utilized to facilitate the computation of the magnitudes.

Algorithm 1: CFBA.

Initialize $n = 0$, $f^{(0)} = C^H g$, $\phi^{(0)} = 0$, $C(\phi^0) = C$, and set the values of γ , λ , and μ according to $\mu \in (0, \frac{1}{L})$ and $\gamma \geq \frac{\sqrt{\mu\lambda}}{2}$
while $n < 300$ or $\|f^{(n+1)} - f^{(n)}\| / \|f^{(n)}\| > 0.001$ **do**
 1. Compute $f^{(n+1)}$ by complex forward-backward splitting method, i.e.,
 while $k < 500$ or $\|o^{(k+1)} - o^{(k)}\| / \|o^{(k)}\| > 0.001$ **do**
 Find $o^{(k+1)} = \text{prox}_{\lambda\mu R}(o^{(k)} - \mu C(\phi^{(n)})^H (C(\phi^{(n)})o^{(k)} - g))$ by (20)–(26)
 $k = k + 1$
 end while
 2. Compute $\phi_m^{(n+1)}$ by (28)
 3. Compute $C(\phi_m^{(n+1)})$ by (29)
 4. $n = n + 1$
end while

3.3. Convergence Analysis

For the proposed CFBA method, the issue of convergence is twofold. That is to say, the discussion needs to cover the convergence of the inner complex forward-backward splitting algorithm as well as the convergence of the outer alternating minimization algorithm.

3.3.1. Convergence of the Inner Complex Forward-Backward Splitting Algorithm

In the n th image reconstruction step, we find

$$f^{(n+1)} = \arg \min_{f \in \mathbb{C}^N} \|g - C(\phi^{(n)})f\|_2^2 - \lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + |f_i|^2}, \tag{30}$$

and in each step of the complex FB splitting algorithm which solves (30) iteratively, we compute

$$o^{(k+1)} = \arg \min_{o \in \mathbb{C}^N} \frac{1}{2} \|o - z^{(k)}\|_2^2 - \mu\lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + |o_i|^2}, \tag{31}$$

where

$$z^{(k)} = o^{(k)} - 2\mu C(\phi^{(n)})^H (C(\phi^{(n)})o^{(k)} - g). \tag{32}$$

Since $\phi^{(n)}$ is fixed for the n th f -sub-problem, for simplicity of notation, we will denote $C(\phi^{(n)})$ by $C^{(n)}$ in this subsection. Using the notations of (11), we denote the cost function in (30) as $J_n(f) = H(f) + G(f)$, with $H(f) = \|g - C^{(n)}f\|_2^2$ and $G(f) = -\lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + |f_i|^2}$.

Now, note that for an arbitrary N -dimensional $f = (f_1, \dots, f_N)^T = (x_1 + iy_1, \dots, x_N + iy_N)^T \in \mathbb{C}^N$, we can obtain a $2N$ -dimensional real vector $\tilde{f} = (x_1, \dots, x_N, y_1, \dots, y_N)^T \in \mathbb{R}^{2N}$. Conversely, for an arbitrary $2N$ -dimensional $\tilde{f} = (f_1, \dots, f_N, \tilde{f}_{N+1}, \dots, \tilde{f}_{2N})^T \in \mathbb{R}^{2N}$, we can obtain a N -dimensional complex vector $f = (f_1, \dots, f_N)^T = (f_1 + if_{N+1}, \dots, f_N + if_{2N})^T$.

$i\tilde{\mathbf{f}}_{2N}^T \in \mathbb{C}^N$. For simplicity, we denote $\tilde{\mathbf{f}}_R = (\tilde{\mathbf{f}}_1, \dots, \tilde{\mathbf{f}}_N)^T$ and $\tilde{\mathbf{f}}_I = (\tilde{\mathbf{f}}_{N+1}, \dots, \tilde{\mathbf{f}}_{2N})^T$ for $\tilde{\mathbf{f}}$. With these notations, we give the following lemma.

Lemma 1. $J_n(\mathbf{f})$ can be rewritten as $\tilde{J}_n(\tilde{\mathbf{f}})$, which is a real analytic function.

The proof of this lemma is put in Appendix A.

Now, since $\tilde{J}_n(\tilde{\mathbf{f}})$ is real analytic, it satisfies the Kurdyka–Lojasiewicz (KL) property [37], which means that for every $\tilde{\mathbf{f}} \in \mathbb{R}^{2N}$ and every bounded neighborhood U of $\tilde{\mathbf{f}}$, there exists $\kappa \in (0, +\infty)$, $\eta \in (0, +\infty)$ and $\theta \in [0, 1)$ such that

$$\|\nabla \tilde{J}_n(\tilde{\mathbf{f}})\| \geq \kappa |\tilde{J}_n(\tilde{\mathbf{f}}) - \tilde{J}_n(\tilde{\mathbf{f}}')|^\theta \tag{33}$$

for every $\tilde{\mathbf{f}} \in U \cap \{\tilde{\mathbf{f}} \mid |\tilde{J}_n(\tilde{\mathbf{f}}) - \tilde{J}_n(\tilde{\mathbf{f}}')| \leq \eta\}$.

The proof of Lemma 1 also implies the proof of the convergence of the CFBA algorithm from a perspective of real vector variables, because $J_n(\mathbf{f})$ (a function of the complex \mathbf{f}) can now be viewed as $\tilde{J}_n(\tilde{\mathbf{f}})$ (a function of real vector variable $\tilde{\mathbf{f}}$). If the real forward-backward splitting algorithm minimizing $\tilde{J}_n(\tilde{\mathbf{f}})$ can be connected with the proposed complex forward-backward splitting algorithm minimizing $J_n(\mathbf{f})$, then the convergence analysis of the latter will benefit from the convergence analysis of the former.

Theorem 1. Note that $k \rightarrow +\infty$, $\mathbf{o}^{(k)}$ generated by (31) will converge to some critical point of $J_n(\mathbf{f})$. Moreover, denote by \mathbf{f}^* the global minimizer of $J_n(\mathbf{f})$; then for each $r > 0$, there exist $u \in (0, r)$, $\delta > 0$ such that when the inequalities $\min J_n(\mathbf{f}) < J_n(\mathbf{o}^{(0)}) < \delta + \min J_n(\mathbf{f})$ and $\|\mathbf{o}^{(0)} - \mathbf{f}^*\| < u$ hold, the sequence $\mathbf{o}^{(k)}$ generated for each $J_n(\mathbf{f})$ will converge to some \mathbf{o}^* with $\mathbf{o}^{(k)} \in B(\mathbf{f}^*, r)$ for an arbitrary k and $J_n(\mathbf{u}^*) = \min J_n(\mathbf{f})$.

Proof. Let us first formulate the real forward-backward splitting algorithm minimizing $\tilde{J}_n(\tilde{\mathbf{f}})$ with respect to $\tilde{\mathbf{f}}$. First denote $\mathbf{u}^{(0)} = \mathbf{o}^{(0)}$. Then in each step, this algorithm finds

$$\mathbf{u}^{(k+1)} = \text{prox}_{\mu\tilde{\mathbf{G}}}(\mathbf{w}^{(k)}) = \arg \min_{\mathbf{u} \in \mathbb{R}^{2N}} \frac{1}{2} \|\mathbf{u} - \mathbf{w}^{(k)}\|_2^2 - \mu\lambda \sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + \|\mathbf{s}_{(i)}\mathbf{u}\|^2}, \tag{34}$$

where

$$\mathbf{w}^{(k)} = \mathbf{u}^{(k)} - 2\mu\widetilde{\mathbf{C}}^{(n)T} (\widetilde{\mathbf{C}}^{(n)}\mathbf{u}^{(k)} - \tilde{\mathbf{g}}). \tag{35}$$

Observe that

$$(\mathbf{C}^{(n)})^H \mathbf{f} = (\overline{c_{11}}\mathbf{f}_1 + \dots + \overline{c_{N1}}\mathbf{f}_N, \dots, \overline{c_{1N}}\mathbf{f}_1 + \dots + \overline{c_{NN}}\mathbf{f}_N)^T, \tag{36}$$

and

$$(\widetilde{\mathbf{C}}^{(n)})^H \mathbf{f} = \begin{pmatrix} (c_{11})_R \mathbf{x}_1 + (c_{11})_I \mathbf{y}_1 + \dots + (c_{N1})_R \mathbf{x}_N + (c_{N1})_I \mathbf{y}_N \\ \vdots \\ (c_{1N})_R \mathbf{x}_1 + (c_{1N})_I \mathbf{y}_1 + \dots + (c_{NN})_R \mathbf{x}_N + (c_{NN})_I \mathbf{y}_N \\ (-c_{11})_I \mathbf{x}_1 + (c_{11})_R \mathbf{y}_1 + \dots + (-c_{N1})_I \mathbf{x}_N + (c_{N1})_R \mathbf{y}_N \\ \vdots \\ (-c_{1N})_I \mathbf{x}_1 + (c_{1N})_R \mathbf{y}_1 + \dots + (-c_{NN})_I \mathbf{x}_N + (c_{NN})_R \mathbf{y}_N \end{pmatrix}. \tag{37}$$

This can be exactly decomposed as

$$(\widetilde{\mathbf{C}}^{(n)})^H \mathbf{f} = \widetilde{\mathbf{C}}^{(n)T} \tilde{\mathbf{f}}, \tag{38}$$

where $\widetilde{\mathbf{C}}^{(n)}$ is as in (A5).

Therefore, we have

$$\mathbf{w}^{(0)} = \widetilde{\mathbf{o}}^{(0)} - 2\mu \widetilde{\mathbf{C}}^{(n)T} (\widetilde{\mathbf{C}}^{(n)} \widetilde{\mathbf{o}}^{(0)} - \widetilde{\mathbf{g}}) = \widetilde{\mathbf{o}}^{(0)} - 2\mu \widetilde{\nabla h}(\widetilde{\mathbf{o}}^{(0)}) = \widetilde{\mathbf{z}}^{(0)}, \tag{39}$$

with $\mathbf{z}^{(0)}$ defined by (32) for CFBA.

If now we rewrite the \mathbf{u} in (34) by $\mathbf{u} = \widetilde{\mathbf{o}}$, then the Moreau envelope therein becomes a function of $\widetilde{\mathbf{o}}$. Additionally, if this function is rewritten as a function of \mathbf{o} , the result will exactly take the form of the Moreau envelope in (31). Moreover, we have $\nabla \widetilde{J}_n(\widetilde{\mathbf{f}}) = ((\frac{\partial \widetilde{J}_n}{\partial \widetilde{\mathbf{f}}_R})^T, (\frac{\partial \widetilde{J}_n}{\partial \widetilde{\mathbf{f}}_I})^T)^T$ by the definition of the real gradient and $\nabla_f J_n(\mathbf{f}) = \frac{1}{2}(\frac{\partial \widetilde{J}_n}{\partial \widetilde{\mathbf{f}}_R} + i\frac{\partial \widetilde{J}_n}{\partial \widetilde{\mathbf{f}}_I})$ by the definition in Wirtinger calculus. Therefore, if \mathbf{f} is a stationary point of $J_n(\mathbf{f})$, then the corresponding $\widetilde{\mathbf{f}}$ is a stationary point of $\widetilde{J}_n(\widetilde{\mathbf{f}})$. Since we have forced convexity of the Moreau envelopes in (34) and (31) by restricting the range of parameters, they will each have only one stationary point. Due to these three conclusions, $\mathbf{u}^{(1)} = \widetilde{\mathbf{o}}^{(1)}$ holds.

By induction, following similar deduction, a sequence $\mathbf{o}^{(k)}$ complying with (31) in the CFBA algorithm and meanwhile satisfying $\mathbf{u}^{(k)} = \widetilde{\mathbf{o}}^{(k)}$ for all k can be obtained. Therefore, the convergence of the CFBA algorithm can be analyzed equivalently by discussing this real FB splitting algorithm.

The analysis above also implies that the n th f -sub-problem can be solved by finding the corresponding real solution for $\widetilde{J}_n(\widetilde{\mathbf{f}})$ and transforming it back to the desired complex solution. However, the proposed CFBA algorithm is more compact in form, since it deals with N dimensional vectors instead of $2N$ dimensional vectors, and it does not require the construction of $\widetilde{\mathbf{C}}$ based on \mathbf{C} .

Now, $\widetilde{J}_n(\widetilde{\mathbf{f}}) = \widetilde{G}(\widetilde{\mathbf{f}}) + \widetilde{H}(\widetilde{\mathbf{f}})$ is proper and lower semicontinuous, bounded from below if hyper-parameters are chosen appropriately (by selecting $\lambda > 0$ and $\gamma < 1$, we have $\widetilde{J}_n(\widetilde{\mathbf{f}}) > \lambda N \ln \gamma$), and it satisfies the KL property. $\widetilde{H}(\widetilde{\mathbf{f}})$ is finite-valued, differentiable, and has a Lipschitz continuous gradient. Moreover, $\widetilde{G}(\widetilde{\mathbf{f}})$ is continuous on its domain. That is to say, all the conditions in Theorem 5.1 of [38] are satisfied. Therefore, according to that theorem, with the conditions aforementioned and the setting $\mu \leq \frac{1}{L}$ (which guarantees the monotonic decreasing nature of $\widetilde{J}_n(\mathbf{u}^{(k)})$), we come to the conclusion that the iterates $\mathbf{u}^{(k)}$ will converge to some critical point of $\widetilde{J}_n(\widetilde{\mathbf{f}})$. Therefore, equivalently, the iterates $\mathbf{o}^{(k)}$ produced by the proposed CFBA method will converge to some critical point of $J_n(\mathbf{f})$.

Moreover, let us denote by $\widetilde{\mathbf{f}}^*$ the global minimizer of $\widetilde{J}_n(\widetilde{\mathbf{f}})$. According to Theorem 2.12 in [38], for each $r > 0$, there exist $u \in (0, r), \delta > 0$ such that when the inequalities $\|\mathbf{u}^{(0)} - \widetilde{\mathbf{f}}^*\| < u$ and $\min \widetilde{J}_n(\widetilde{\mathbf{f}}) < \widetilde{J}_n(\mathbf{u}^{(0)}) < \delta + \min \widetilde{J}_n(\widetilde{\mathbf{f}})$ hold, the sequence $\mathbf{u}^{(k)}$ generated for each $\widetilde{J}_n(\widetilde{\mathbf{f}})$ will converge to some \mathbf{u}^* with $\mathbf{u}^{(k)} \in B(\widetilde{\mathbf{f}}^*, r)$ for arbitrary k and $\widetilde{J}_n(\mathbf{u}^*) = \min \widetilde{J}_n(\widetilde{\mathbf{f}})$. That is to say, under some assumptions about the initial value $\mathbf{u}^{(0)}$, convergence of the sequence $\mathbf{u}^{(k)}$ to a global minimizer of $\widetilde{J}_n(\widetilde{\mathbf{f}})$ can be obtained. Equivalently, for each $r > 0$, there exist $u \in (0, r), \delta > 0$ such that when the inequalities $\|\mathbf{o}^{(0)} - \mathbf{f}^*\| < u$ and $\min J_n(\mathbf{f}) < J_n(\mathbf{o}^{(0)}) < \delta + \min J_n(\mathbf{f})$ hold, the sequence $\mathbf{o}^{(k)}$ generated for each $J_n(\mathbf{f})$ will converge to some \mathbf{o}^* with $\mathbf{o}^{(k)} \in B(\mathbf{f}^*, r)$ for arbitrary k and $J_n(\mathbf{o}^*) = \min J_n(\mathbf{f})$, i.e., the iterates $\mathbf{o}^{(k)}$ produced by the proposed CFBA method will converge to a global minimizer of $J_n(\mathbf{f})$. \square

In fact, the above proof, which is elaborated from a real perspective, can also be done alternatively by working on the complex iterates $\mathbf{o}^{(k)}$ themselves. However, this process is more complicated (see Appendix A).

3.3.2. Convergence of the Outer Alternating Minimization Method

Based on Theorem 1, the following result can be obtained:

Theorem 2. For each f -sub-problem, if the assumptions related to the initial value $\mathbf{o}^{(0)}$ stated in Theorem 1 are satisfied, with $\mathbf{f}^{(n)}$ and $\boldsymbol{\phi}^{(n)}$ computed by CFBA, $J(\mathbf{f}^{(n)}, \boldsymbol{\phi}^{(n)})$ will converge to a certain value (not necessarily equal to $\inf J(\mathbf{f}, \boldsymbol{\phi})$).

Proof. For each f -sub-problem, if the assumptions related to the initial value $\mathbf{o}^{(0)}$ stated in the last section are satisfied, the sequence $\mathbf{o}^{(k)}$ will converge to a global minimizer of $J_n(f)$. In this case, $J(f^{(n+1)}, \boldsymbol{\phi}^{(n)}) \leq J(f^{(n)}, \boldsymbol{\phi}^{(n)})$. Additionally, since each ϕ -sub-problem has a closed form solution, we have $J(f^{(n+1)}, \boldsymbol{\phi}^{(n+1)}) \leq J(f^{(n+1)}, \boldsymbol{\phi}^{(n)})$. As a result, $J(f^{(n+1)}, \boldsymbol{\phi}^{(n+1)}) \leq J(f^{(n)}, \boldsymbol{\phi}^{(n)})$ holds for every n , i.e., $J(f^{(n)}, \boldsymbol{\phi}^{(n)})$ is a monotonically decreasing sequence. Since it is also bounded below, it will converge to a certain value, though not necessarily equal to $\inf J(f, \boldsymbol{\phi})$. \square

If stronger assumptions are satisfied, better results of convergence can be obtained. For instance, if the five-point property [39,40] holds, i.e., if

$$J(f, \boldsymbol{\phi}) + J(f, \boldsymbol{\phi}^{(n)}) \geq J(f, \boldsymbol{\phi}^{(n+1)}) + J(f^{(n+1)}, \boldsymbol{\phi}^{(n)}) \tag{40}$$

holds for every $f, \boldsymbol{\phi}$, and n , then

$$\lim_{n \rightarrow +\infty} J(f^{(n)}, \boldsymbol{\phi}^{(n)}) = \inf J(f, \boldsymbol{\phi}). \tag{41}$$

4. Wirtinger Alternating Minimization Autofocusing

4.1. The Original Method

In this section, we review the Wirtinger alternating minimization autofocusing (WAMA) method originally proposed in [27]. After that, we briefly expand on how to extend this method to several other cost functions with the same fidelity term but with different regularizers. Finally, we will discuss the convergence of this method, a topic not covered in previous publications.

The WAMA method also adopts the framework of alternating minimization, including two types of sub-problems to be solved. For each f -sub-problem formulated in (10), Wirtinger calculus is used to solve it. On the one hand, Wirtinger calculus is a powerful theory covering the analysis of real-valued functions of complex variables, and within which many real optimization problems can have their complex counterparts defined. On the other hand, it is also a rather elegant approach, due to its ability to address the problem in a concise way. Namely, there is no need to expand the complex variables as real vectors in the computational process.

Specifically, to solve (10), we compute the complex gradient of the cost function therein directly using Wirtinger calculus. Using the notation in Section 3.2.1, we have

$$(\nabla_f G(f))_i = \frac{\lambda f_i}{\gamma^2 + |f_i|^2}, i = 1, \dots, N. \tag{42}$$

Since $\nabla_f H(f)$ is already given in (17), the complex gradient of (10) can be written as:

$$\nabla_f J(f, \boldsymbol{\phi}) = \mathbf{C}(\boldsymbol{\phi}^{(n)})^H (\mathbf{C}(\boldsymbol{\phi}^{(n)})f - \mathbf{g}) + \lambda \mathbf{W}(f)f, \tag{43}$$

where

$$\mathbf{W}(f) = \text{diag}(\mathbf{s}), \tag{44}$$

$$s_i = \frac{1}{\gamma^2 + |f_i|^2}, i = 1, \dots, N. \tag{45}$$

Now we set (43) to zero, according to the necessary and sufficient condition for a stationary point of a real-valued complex function [23,25], which leads to

$$[\mathbf{C}(\boldsymbol{\phi}^{(n)})^H \mathbf{C}(\boldsymbol{\phi}^{(n)}) + \lambda \mathbf{W}(f)]f = \mathbf{C}(\boldsymbol{\phi}^{(n)})^H \mathbf{g}. \tag{46}$$

It is worth pointing out that even though the exact solution of (46) can be obtained, that solution is not necessarily the global minimum of (10) due to the non-convexity of the Cauchy penalty.

Now we rewrite (46) as $Af = b$, where $A = \lambda W(f) + C(\phi^{(n)})^H C(\phi^{(n)})$ and $b = C(\phi^{(n)})^H g$. Since $W(f)$ depends on f , so does A . This makes (46) nonlinear with respect to f , and it is difficult to find its closed-form solution. However, if we sacrifice some accuracy and approximate the f in $W(f)$ with the f computed during the last iteration of the alternating minimization framework, A is converted into a constant matrix, and thus $Af = b$ becomes a linear system of equations which can be solved efficiently.

That is to say, when computing an unknown $f^{(n+1)}$, the actually solved equation is

$$[C(\phi^{(n)})^H C(\phi^{(n)}) + \lambda W(f^{(n)})]f^{(n+1)} = C(\phi^{(n)})^H g. \tag{47}$$

This equation can be viewed as a fixed-point algorithm with a single iterative step, and its solution can be efficiently obtained by using the conjugate gradient (CG) algorithm [41]. The experimental results in Section 5 imply that the obtained solution is sufficiently good.

As for the ϕ -sub-problems, the solutions are the same as that of CFBA introduced in Section 3.1, so they are not presented here. Now, the whole process of the WAMA method can be summarized as Algorithm 2 as follows:

Algorithm 2: WAMA.

Initialize $n = 0, f^{(0)} = C^H g, \phi^{(0)} = 0, C(\phi^0) = C$, and set the values of γ and λ
while $n < 300$ or $\|f^{(n+1)} - f^{(n)}\| / \|f^{(n)}\| > 0.001$ **do**
 1. Compute $f^{(n+1)}$ by finding the solution of (47) via CG
 2. Compute $\phi_m^{(n+1)}$ by (28)
 3. Compute $C(\phi_m^{(n+1)})$ by (29)
 4. $n = n + 1$
end while

4.2. Extension to Several Other Regularizers

As an extension, the same computational processes can also be followed to handle the cases where the magnitude Cauchy regularization in (7) is replaced by some other \mathbb{R} -differentiable regularizers [25]. For those cases, Equation (47) will also be obtained, despite the fact that the involved s is different. We give several examples as follows:

- (1) The p th power of an approximate l_p norm

In this case,

$$G(f) = \lambda \sum_{i=1}^N (|f_i|^2 + \beta)^{\frac{p}{2}}, \tag{48}$$

and now

$$s_i = \frac{p}{2(|f_i|^2 + \beta)^{1-\frac{p}{2}}}, i = 1, \dots, N. \tag{49}$$

This yields the same algorithm as in this case of [12], but no reference to the literature of Wirtinger calculus is made therein. Therefore, this deduction can be viewed as an alternative perspective to interpret SDA;

- (2) Approximate total variation

For approximate total variation, the situation is more complicated. However, the result can still be incorporated in the form of (47). Let F be the 2D $a \times a$ matrix form of the N -dimensional vector $f(N = a \times a)$; then

$$G(f) = \lambda \sum_{i=1}^a \sum_{j=1}^a \sqrt{|(\nabla_i F)_{i,j}|^2 + |(\nabla_j F)_{i,j}|^2 + \beta}, \tag{50}$$

with

$$(\nabla_i F)_{i,j} = \begin{cases} F_{i,j} - F_{i-1,j} & i > 1 \\ 0 & i = 1 \end{cases} \tag{51}$$

$$(\nabla_j \mathbf{F})_{i,j} = \begin{cases} \mathbf{F}_{i,j} - \mathbf{F}_{i,j-1} & j > 1 \\ 0 & j = 1 \end{cases} \tag{52}$$

Now

$$\mathbf{W}(\mathbf{f}) = \mathbf{W}' \mathbf{D}' + \mathbf{W}'' \mathbf{D}'' + \mathbf{W}''' \mathbf{D}''' + \mathbf{W}'''' \mathbf{D}'''' , \tag{53}$$

$$\mathbf{W}' = \text{diag}(\text{vec}(\mathbf{S}')), \mathbf{W}'' = \text{diag}(\text{vec}(\mathbf{S}'')), \mathbf{W}''' = \text{diag}(\text{vec}(\mathbf{S}''')), \tag{54}$$

where vec is the operation which turns a matrix into a column vector by stacking its columns in order. Additionally,

$$(\mathbf{S}')_{i,j} = \frac{1}{2\sqrt{|(\nabla_i \mathbf{F})_{i,j}|^2 + |(\nabla_j \mathbf{F})_{i,j}|^2 + \beta}} , \tag{55}$$

$$(\mathbf{S}'')_{i,j} = \frac{1}{2\sqrt{|(\nabla_i \mathbf{F})_{i,j+1}|^2 + |(\nabla_j \mathbf{F})_{i,j+1}|^2 + \beta}} , \tag{56}$$

$$(\mathbf{S}''')_{i,j} = \frac{1}{2\sqrt{|(\nabla_i \mathbf{F})_{i+1,j}|^2 + |(\nabla_j \mathbf{F})_{i+1,j}|^2 + \beta}} . \tag{57}$$

As for $\mathbf{D}', \mathbf{D}'', \mathbf{D}'''$, and \mathbf{D}'''' , they are matrices that contain only 0, 1, and -1 and are constructed so that they realize the following relations:

$$\begin{aligned} \mathbf{D}' \mathbf{f} &= \text{vec}((\nabla_i \mathbf{F})_{i,j}), \mathbf{D}'' \mathbf{f} = \text{vec}((\nabla_j \mathbf{F})_{i,j}), \\ \mathbf{D}''' \mathbf{f} &= -\text{vec}((\nabla_j \mathbf{F})_{i,j+1}), \mathbf{D}'''' \mathbf{f} = -\text{vec}((\nabla_i \mathbf{F})_{i+1,j}); \end{aligned} \tag{58}$$

(3) Welsh potential

In this case, a $l_2 - l_0$ regularization [42] is imposed on the magnitude of f , and we have

$$G(\mathbf{f}) = \lambda \sum_{i=1}^N (1 - e^{-\frac{|f_i|^2}{2\delta^2}}), \tag{59}$$

and

$$s_i = \frac{e^{-\frac{|f_i|^2}{2\delta^2}}}{2\delta^2}, i = 1, \dots, N; \tag{60}$$

(4) Geman–McClure potential

In this case, another variant of $l_2 - l_0$ regularization [42] is imposed on the magnitude of f , and we have

$$G(\mathbf{f}) = \lambda \sum_{i=1}^N \frac{|f_i|^2}{2\delta^2 + |f_i|^2} \tag{61}$$

and

$$s_i = \frac{2\delta^2}{(2\delta^2 + |f_i|^2)^2}, i = 1, \dots, N. \tag{62}$$

4.3. Convergence Analysis

The approximation (47) used for the solution of each image reconstruction step adds much difficulty to the discussion of the convergence of the WAMA method. However, the WAMA method can be analyzed from another perspective, which renders its convergence analysis tractable.

Theorem 3. As $n \rightarrow \infty$, with $\mathbf{f}^{(n)}$ and $\boldsymbol{\phi}^{(n)}$ computed by WAMA, $J(\mathbf{f}^{(n)}, \boldsymbol{\phi}^{(n)})$ will converge to a certain value (not necessarily equal to $\inf J(\mathbf{f}, \boldsymbol{\phi})$).

Proof. Similar to [43], the key point is the construction of a $K(\mathbf{b}, f, \boldsymbol{\phi})$ such that $\inf_{\mathbf{b}} K(\mathbf{b}, f, \boldsymbol{\phi}) = J(f, \boldsymbol{\phi})$, with $J(f, \boldsymbol{\phi})$ given by (7). Following the theories introduced in [44], this $K(\mathbf{b}, f, \boldsymbol{\phi})$ is constructed as

$$K(\mathbf{b}, f, \boldsymbol{\phi}) = \|\mathbf{g} - \mathbf{C}(\boldsymbol{\phi})f\|_2^2 - \lambda \sum_{i=1}^N [(|f_i|^2 + \gamma^2)b_i - \ln(\gamma b_i) - 1], \tag{63}$$

where \mathbf{b} is an auxiliary vector.

Now, for verification, we let $\frac{\partial K}{\partial \mathbf{b}} = 0$ to find the \mathbf{b}^* minimizing $K(\mathbf{b}, f, \boldsymbol{\phi})$ for a fixed f and $\boldsymbol{\phi}$. Consequently,

$$b_i^* = \frac{1}{\gamma^2 + |f_i|^2}. \tag{64}$$

Substituting (64) into (63), $K(\mathbf{b}^*, f, \boldsymbol{\phi}) = J(f, \boldsymbol{\phi})$ is obtained, and therefore the equality $\inf_{\mathbf{b}} K(\mathbf{b}, f, \boldsymbol{\phi}) = J(f, \boldsymbol{\phi})$ is verified.

Therefore, minimizing the original cost function (7) with respect to f and $\boldsymbol{\phi}$ is equivalent to minimizing (63) with respect to \mathbf{b} , f , and $\boldsymbol{\phi}$. If an alternating minimization scheme is imposed directly on $K(\mathbf{b}, f, \boldsymbol{\phi})$, the procedure will consist of the repetition of the following three steps:

1. Find $\mathbf{b}^{(n+1)}$ by

$$\mathbf{b}^{(n+1)} = \arg \min_{\mathbf{b}} K(\mathbf{b}, f^{(n)}, \boldsymbol{\phi}^{(n)}). \tag{65}$$

This leads to:

$$b_i^{(n+1)} = \frac{1}{\gamma^2 + |f_i^{(n)}|^2}; \tag{66}$$

2. Find $f^{(n+1)}$ by

$$f^{(n+1)} = \arg \min_f K(\mathbf{b}^{(n+1)}, f, \boldsymbol{\phi}^{(n)}). \tag{67}$$

This leads to:

$$[\mathbf{C}(\boldsymbol{\phi}^{(n)})^H \mathbf{C}(\boldsymbol{\phi}^{(n)}) + \lambda \mathbf{W}]f^{(n+1)} = \mathbf{C}(\boldsymbol{\phi}^{(n)})^H \mathbf{g}, \tag{68}$$

where

$$\mathbf{W} = \text{diag}(\mathbf{b}^{(n+1)}); \tag{69}$$

3. Find $\boldsymbol{\phi}^{(n+1)}$ by

$$\boldsymbol{\phi}^{(n+1)} = \arg \min_{\boldsymbol{\phi}} K(\mathbf{b}^{(n+1)}, f^{(n+1)}, \boldsymbol{\phi}). \tag{70}$$

This leads to:

$$\boldsymbol{\phi}_m^{(n+1)} = \arctan\left(\frac{\text{Re}\{[f^{(n+1)}]^H \mathbf{C}_m \mathbf{g}_m\}}{\text{Im}\{[f^{(n+1)}]^H \mathbf{C}_m \mathbf{g}_m\}}\right). \tag{71}$$

Notice that if we combine (66) and (68) as one step, then the Formulas (66), (68) and (71) are exactly the same as (47) and (28). Therefore, the convergence of the WAMA method can be analyzed by discussing this equivalent alternating minimization process.

Therefore, since (66), (68), and (71) all give closed-form solutions or sufficient accuracy (suggested by the convergence analysis; see (73)), we assert that:

$$K(\mathbf{b}^{(n+1)}, f^{(n+1)}, \boldsymbol{\phi}^{(n+1)}) \leq K(\mathbf{b}^{(n)}, f^{(n)}, \boldsymbol{\phi}^{(n)}). \tag{72}$$

That is to say, $K(\mathbf{b}^{(n)}, f^{(n)}, \boldsymbol{\phi}^{(n)})$ is a monotonically decreasing sequence. Since it is bounded below, it will converge to a certain value as n goes to infinity. \square

Computationally, according to WAMA, conjugate gradient (CG) method is used to obtain the solution of (68). Denote by $f^{*(n+1)}$ the exact solution of (68), and by $q^{(j)}$ ($j = 1, \dots, J$)

the iterates in the loop of CG are such that $f^{(n+1)} = q^{(l)}$. According to [45], if the matrix $A = [C(\phi^{(n)})^H C(\phi^{(n)}) + \lambda W]$ is non-singular, we can get

$$\|q^{(j)} - f^{*(n)}\|_A^2 \leq \|q^{(0)} - f^{*(n)}\|_A^2 \left(\frac{\sqrt{\text{cond}(A)} - 1}{\sqrt{\text{cond}(A)} + 1}\right)^{2j}, \tag{73}$$

with $\|x\|_A^2 = x^H A x$, and $\text{cond}(A)$ being the condition number of A . That is to say, each set of iterates $q^{(j)}$ generated by the conjugate gradient method will converge to its corresponding $f^{*(n)}$ as j goes to infinity. Therefore, $f^{(n+1)}$ is of sufficient accuracy. This is also validated by the experimental results in Section 5.

Note that similar analysis can be carried out for the two variants using $l_2 - l_0$ regularization and approximate l_p regularization, as mentioned in Section 4.2.

When the overall cost function takes the form

$$J(f, \phi) = \|g - C(\phi)f\|_2^2 + \lambda \sum_{i=1}^N (1 - e^{-\frac{|f_i|^2}{2\delta^2}}), \tag{74}$$

the corresponding $K(b, f, \phi)$ is constructed as:

$$K(b, f, \phi) = \|g - C(\phi)f\|_2^2 + \lambda \sum_{i=1}^N [(|f_i|^2 - 2\delta^2)b_i + 2\delta^2 b_i \ln(2\delta^2 b_i) + 1]. \tag{75}$$

This leads to:

$$b_i^* = \frac{e^{-\frac{|f_i|^2}{2\delta^2}}}{2\delta^2}. \tag{76}$$

When the overall cost function takes the form

$$J(f, \phi) = \|g - C(\phi)f\|_2^2 + \lambda \sum_{i=1}^N \frac{|f_i|^2}{|f_i|^2 + \delta^2}, \tag{77}$$

the corresponding $K(b, f, \phi)$ is can then be constructed as:

$$K(b, f, \phi) = \|g - C(\phi)f\|_2^2 + \lambda \sum_{i=1}^N [(|f_i|^2 + 2\delta^2)b_i - 2\sqrt{2}\delta\sqrt{b_i} + 1]. \tag{78}$$

This leads to:

$$b_i^* = \frac{2\delta^2}{(|f_i|^2 + 2\delta^2)^2}. \tag{79}$$

Whereas when the overall cost function takes the form

$$J(f, \phi) = \|g - C(\phi)f\|_2^2 + \lambda \sum_{i=1}^N (|f_i|^2 + \beta)^{\frac{p}{2}}, \tag{80}$$

the corresponding $K(b, f, \phi)$ can be constructed as:

$$K(b, f, \phi) = \|g - C(\phi)f\|_2^2 + \lambda \sum_{i=1}^N [b_i(|f_i|^2 + \beta) + \frac{2-p}{2} \left(\frac{2b_i}{p}\right)^{\frac{p}{p-2}}]. \tag{81}$$

This leads to:

$$b_i^* = \frac{p}{2(|f_i|^2 + \beta)^{1-\frac{p}{2}}}. \tag{82}$$

5. Experimental Results

For the numerical experiments in this paper, the same radar system model as in [12] is used, whose parameters are listed in Table 1.

Table 1. Parameters of the radar system.

Carrier Frequency	$2\pi \times 10^{10}$ rad/s
Chirp Rate	$2\pi \times 10^{12}$ rad/s ²
Pulse Duration	4×10^{-4} s
Angular Range	2.3°

In each experiment, this radar system model is used to generate a simulated phase history from a given reflectivity scene. This phase history is then corrupted by adding 1D random phase errors along the azimuth direction. After that, complex white Gaussian noise is added so that the SNR of the data is 25 dB. This corrupted phase history is used to reconstruct a SAR image while correcting for the phase errors.

We compare the performance of four methods in each experiment. The first method is the traditional polar format algorithm [46] which does not involve a process of autofocus and is therefore expected to result in a blurry formed image as a result of the phase errors added into the simulated phase history. The second method is the sparsity-driven autofocus (SDA) method of [12] (we choose the approximate l_1 norm as the regularizer of their cost function as an example), a state-of-the-art SAR autofocus technique operating in an inverse problem framework similar to the one proposed in this paper. The remaining two methods are the WAMA method in [27] and the proposed CFBA method.

Apart from visual comparison, two numerical metrics are also computed to better assess the performance of each method. One is the mean square error (MSE) between the reconstructed SAR image (using the corrupted phase history) and the ground truth (the reconstructed SAR image from the uncorrupted phase history). The second metric we employ is the entropy of the reconstructed SAR image as an indicator of sharpness. For both of these two metrics, smaller values indicate better performance. For all the compared methods with tunable parameters, we present the result corresponding to the setting of the parameters which gives the best MSE value for that method.

In the first experiment, we use a simulated scene measuring 32×32 pixels. The visual results are presented in Figure 1, and the numerical results are listed in Table 2. It can be observed that the visual result of the polar format suffers from a severe defocusing effect, making it impossible to discern the targets. However, the reconstructed images by SDA, WAMA, and the proposed CFBA method are all very focused and highly resemble the original scene. Since the values of MSE and entropy for WAMA and the proposed CFBA method are lower than SDA, their results are suggested to be sharper and more similar to the original scene.

In the second and the third experiment, a real SAR image from TerraSAR-X and another real SAR image from Sentinel-1 are used. These two images originally contain multiple ships scattered on sea surface. For both experiments, however, due to the high computational burden of CFBA and WAMA even for scenes of moderate size (mainly due to the need to construct the observation matrix C), a 64×64 patch is cut from the original SAR image and regarded as an input scene such that one ship target is still discernible in it. The corrupted pseudo-phase history is then generated as described above.

Figure 2 shows the reconstructed images by all 4 methods for the second experiment (in the first row) and the third experiment (in the second row). Table 2 again contains all the corresponding results of the two numerical indices for these scenes. According to Figures 2 and 3, the polar format algorithm once again gives reconstructed results with seriously smeared ship targets, which is especially notable in Figure 2. In contrast, SDA, WAMA and the proposed CFBA method can remove phase errors effectively and present focused ship targets, displaying significant improvement over the result of the polar format algorithm. Nevertheless, the results of the numerical indices in Table 2 demonstrate that WAMA and the proposed CFBA method both give results slightly better than that of SDA.

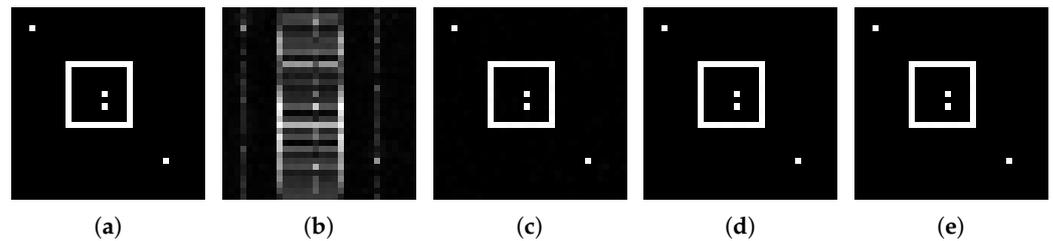


Figure 2. Visual results for Scene 1, a simulated 32×32 scene obtained by various methods. (a) Original. (b) Polar format. (c) SDA. (d) WAMA. (e) CFBA.

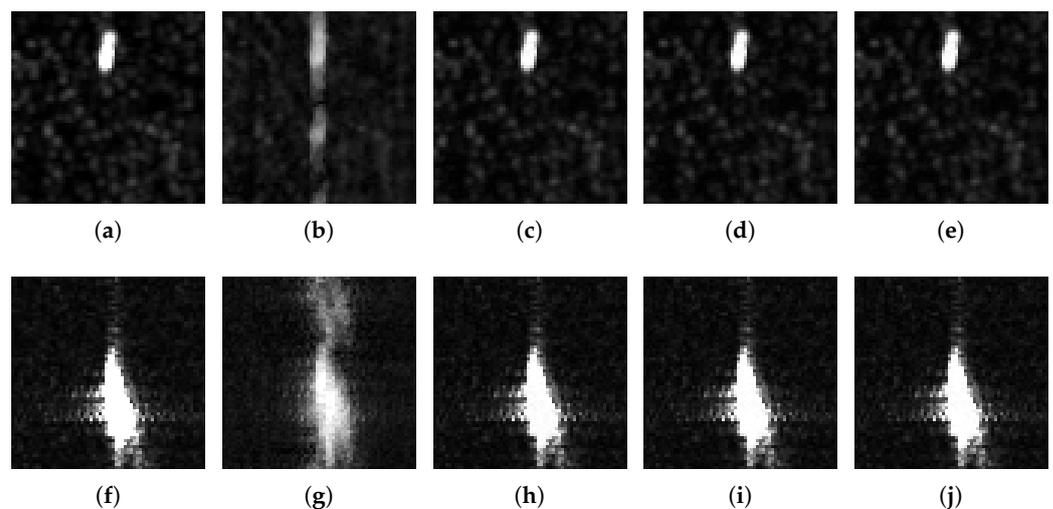


Figure 3. Visual results for Scene 2 (a 64×64 patch from TerraSAR-X) and Scene 3 (a 64×64 patch from Sentinel-1) obtained by various methods. First row for Scene 2, the second row for Scene 3. (a) Original. (b) Polar format. (c) SDA. (d) WAMA. (e) CFBA. (f) Original. (g) Polar format. (h) SDA. (i) WAMA. (j) CFBA.

Table 2. Numerical evaluation of the experimental results for all the methods.

MSE					
Method	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5
SDA	5.4310×10^{-6}	6.4964×10^{-5}	6.3576×10^{-5}	1.3997×10^{-5}	6.8909×10^{-6}
WAMA	1.2227×10^{-6}	6.3029×10^{-5}	5.3663×10^{-5}	2.2250×10^{-5}	7.8785×10^{-6}
CFBA	1.1836×10^{-6}	6.2940×10^{-5}	5.4803×10^{-5}	1.3483×10^{-5}	6.5628×10^{-6}
Entropy					
Method	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5
SDA	1.4621	5.4410	5.6918	4.5720	4.2847
WAMA	0.3327	5.4333	5.6641	4.5230	4.2782
CFBA	0.3430	5.4228	5.6602	4.5617	4.2916

In the fourth and the fifth experiment, two simulated images of the sea surface were used as the original scene, with the first one only including sea waves, and the second including a travelling ship and its wake as well. Simulated as the scenes are, they are not as simple as Scene 1, which is a mere combination of black and white regions resembling point reflectors, but are rather based on an exquisite model taking the most important SAR imaging effects into account [29]. The scenes are based on a model of the sea surface using the Pierson–Moskowitz spectrum and cosine-squared spreading function with wind speed $V_w = 8$ m/s for the first image and $V_w = 4$ m/s for the second image, with waves

traveling at 45° relative to the SAR flight direction. For the second image, the size of the ship is 55 m, with 8 m beam and 3 m draft, moving at a velocity of 8 m/s at 45° relative to the SAR flight direction. The original size of both SAR images is 1×1 km with a spatial resolution of 1.25 m, while SAR platform parameters are as follows: the platform altitude is 2.5 km, platform velocity is 125 m/s, the incidence angle is $\theta_r = 35^\circ$, and signal parameters are X-band (9.65 GHz) and VV polarization. Both scenes shown here are of 128×128 pixels, patches from the original images due to heavy computational burden, and their corresponding corrupted pseudo-phase histories are generated from them in the same way as aforementioned.

The visual results for all 4 methods for the fourth experiment and the fifth experiment are shown in Figure 4 in the first row and the second row, respectively. Since the original pixel values in the images are rather small, for visual convenience, the “imadjust” function in Matlab was used before depicting. For the fourth experiment, the results of SDA, WAMA, and CFBA are with better contrast, i.e., the bright regions in (c), (d), (e) of Figure 4 are brighter than those in (b), and their dark regions are darker. For the fifth experiment, (h), (i), (j) are sharper than (g) and display much more concentrated ship wakes. Meanwhile, the numerical results in Table 2 also demonstrate that CFBA, SDA, and WAMA give comparable performance. In terms of MSE, SDA outperforms WAMA and is outperformed by CFBA. As for entropy values, WAMA is the best.

As for the convergence of CFBA and WAMA, Figure 5 displays how $J(f^{(n)}, \phi^{(n)})$ changes with increasing n until the stopping criterion is satisfied for both the WAMA method and the proposed CFBA method, taking the first three experiments as examples. In each sub-figure, the vertical axis represents $J(f^{(n)}, \phi^{(n)})$, the value of the cost function (7) computed for $f^{(n)}$ and $\phi^{(n)}$, while the horizontal axis represents the iterative numbers n in the loop of the alternating minimization. The red line represents the case of WAMA, and the blue line represents the case of CFBA. It can be seen that in all three experiments, $J(f^{(n)}, \phi^{(n)})$ decreases monotonically for both CFBA and WAMA. This is consistent with the conclusions of our convergence analysis, and gives an experimental validation for the convergence of CFBA and WAMA in a sense. Moreover, it can be observed that CFBA converges more rapidly, since it reaches the stopping criterion with fewer iterations.

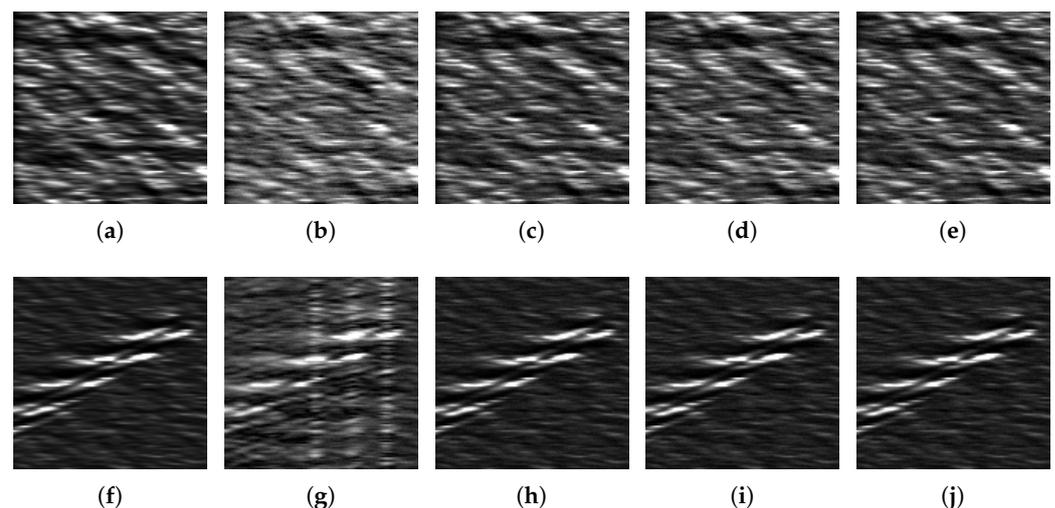
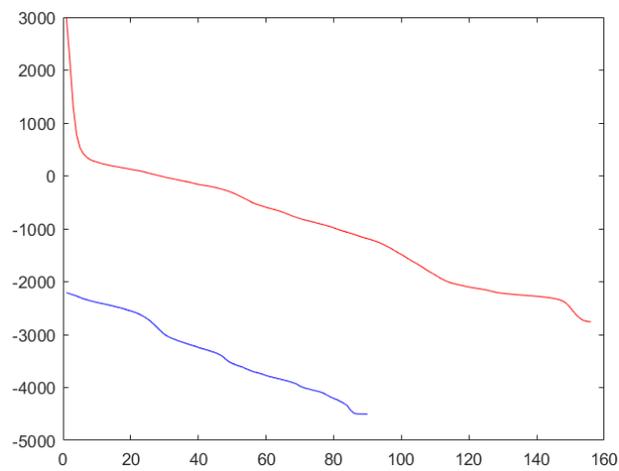
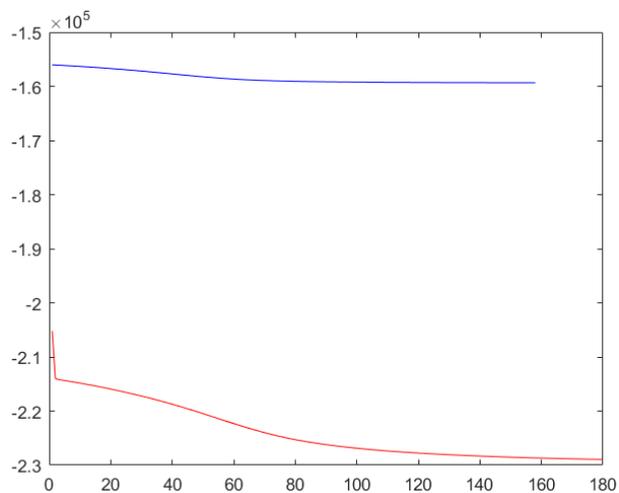


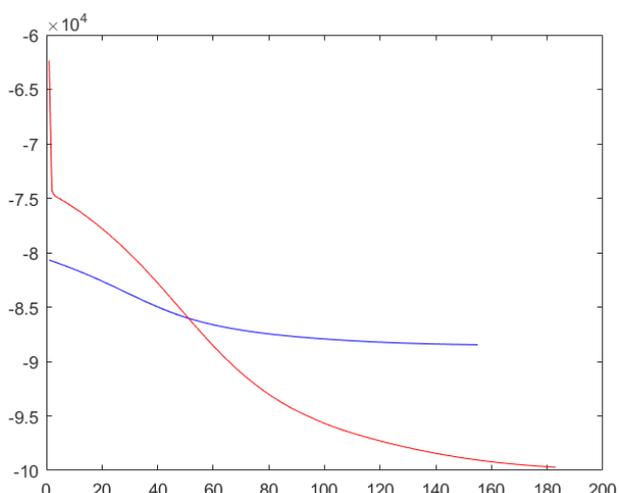
Figure 4. Visual results for Scene 4 and Scene 5, two 128×128 simulated sea surfaces, obtained by various methods. First row for Scene 4, the second row for Scene 5. (a) Original. (b) Polar format. (c) SDA. (d) WAMA. (e) CFBA. (f) Original. (g) Polar format. (h) SDA. (i) WAMA. (j) CFBA.



(a)



(b)



(c)

Figure 5. The values of $J(f^{(n)}, \phi^{(n)})$ for WAMA (red lines) and CFBA (blue lines) until the stopping criterion was satisfied. (a) Scene 1. (b) Scene 2. (c) Scene 3.

6. Conclusions

In this paper, an optimization solution regularized by the magnitude Cauchy penalty is proposed for the problem of simultaneously reconstructing and autofocusing a SAR image. An alternating minimization framework named CFBA is proposed to solve this inverse problem, in which the sub-problems related to the desired SAR image are solved by a complex forward-backward splitting method, and its convergence is analyzed. Additionally, the WAMA method based on Wirtinger calculus is revisited and further discussed, in particular in respect to its convergence. Experimental results on simulated phase history data derived from several simulated as well as real SAR images of the sea surface demonstrate that the proposed CFBA method can reconstruct highly focused SAR images and effectively remove phase errors, showing performance competitive with that of WAMA. Our current work focuses on the design of SAR autofocusing approaches in the context of compressive data acquisition, and we will aim to address the problem by combining our current model-based techniques with data-driven methods.

Author Contributions: Methodology, Z.-Y.Z., O.P., A.A.; software, Z.-Y.Z.; resources, O.P., A.A., I.G.R.; writing—original draft preparation, Z.-Y.Z.; writing—review and editing, O.P., A.A., I.G.R.; supervision, A.A., O.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by a Chinese Scholarship Council PhD studentship (to Zhang) and in part by the Engineering and Physical Sciences Research Council (EPSRC) under grant EP/R009260/1 (AssenSAR).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Proof of Lemma 1

Proof. A plain but important fact is that $\|f\|_2^2 = f^H f = (\tilde{f})^T \tilde{f} = \|\tilde{f}\|_2^2$. Therefore, for $H(f)$, we have

$$H(f) = \|g - C^{(n)} f\|_2^2 = \|\tilde{g} - \widetilde{C^{(n)} f}\|_2^2. \quad (\text{A1})$$

Since

$$C^{(n)} f = (c_{11}f_1 + \dots + c_{1N}f_N, \dots, c_{N1}f_1 + \dots + c_{NN}f_N)^T, \quad (\text{A2})$$

we have

$$\widetilde{C^{(n)} f} = \begin{pmatrix} (c_{11})_R x_1 - (c_{11})_I y_1 + \dots + (c_{1N})_R x_N - (c_{1N})_I y_N \\ \vdots \\ (c_{N1})_R x_1 - (c_{N1})_I y_1 + \dots + (c_{NN})_R x_N - (c_{NN})_I y_N \\ (c_{11})_I x_1 + (c_{11})_R y_1 + \dots + (c_{1N})_I x_N + (c_{1N})_R y_N \\ \vdots \\ (c_{N1})_I x_1 + (c_{N1})_R y_1 + \dots + (c_{NN})_I x_N + (c_{NN})_R y_N \end{pmatrix}, \quad (\text{A3})$$

where $(c_{ij}), i = 1, \dots, N, j = 1, \dots, N$ are elements of $C^{(n)}$, with $(c_{ij})_R, i = 1, \dots, N, j = 1, \dots, N$ and $(c_{ij})_I, i = 1, \dots, N, j = 1, \dots, N$ being the real and imaginary part of (c_{ij}) , respectively.

Furthermore, (A3) can be rewritten as

$$\widetilde{C^{(n)} f} = \widetilde{C^{(n)}} \tilde{f}, \quad (\text{A4})$$

where

$$\widetilde{\mathbf{C}}^{(n)} = \begin{pmatrix} (c_{11})_R & \cdots & (c_{1N})_R & (-c_{11})_I & \cdots & (-c_{1N})_I \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ (c_{N1})_R & \cdots & (c_{NN})_R & (-c_{N1})_I & \cdots & (-c_{NN})_I \\ (c_{11})_I & \cdots & (c_{1N})_I & (c_{11})_R & \cdots & (c_{1N})_R \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ (c_{N1})_I & \cdots & (c_{NN})_I & (c_{N1})_R & \cdots & (c_{NN})_R \end{pmatrix}. \tag{A5}$$

As a result, $H(\mathbf{f}) = \|\widetilde{\mathbf{g}} - \widetilde{\mathbf{C}}^{(n)}\widetilde{\mathbf{f}}\|_2^2 = \widetilde{H}(\widetilde{\mathbf{f}})$.

Now, $\widetilde{H}(\widetilde{\mathbf{f}})$ can be expressed as $\widetilde{H}(\widetilde{\mathbf{f}}) = l(H_{(1)}(\widetilde{\mathbf{f}}), \dots, H_{(2N)}(\widetilde{\mathbf{f}}))$, i.e., the composition of real analytic functions $l(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ and $H_{(i)}(\widetilde{\mathbf{f}}) = (\widetilde{\mathbf{g}})_i - (\widetilde{\mathbf{C}}^{(n)}\widetilde{\mathbf{f}})_i, i = 1, \dots, 2N$. According to [47], $\widetilde{H}(\widetilde{\mathbf{f}})$ is a real analytic function of $\widetilde{\mathbf{f}}$.

On the other hand, for $G(\mathbf{f})$, we have

$$G(\mathbf{f}) = -\sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + \|\mathbf{f}_i\|^2} = -\sum_{i=1}^N \ln \frac{\gamma}{\gamma^2 + \|\mathbf{s}_{(i)}\widetilde{\mathbf{f}}\|^2} = \widetilde{G}(\widetilde{\mathbf{f}}) \tag{A6}$$

where $\mathbf{s}_{(i)}$ is a $2 \times 2N$ matrix whose first row and second row are the i th row and the $(i + N)$ th row of a $2N \times 2N$ identity matrix, respectively. Therefore, each summed term of $\widetilde{G}(\widetilde{\mathbf{f}})$, i.e., $-\ln \frac{\gamma}{\gamma^2 + \|\mathbf{s}_{(i)}\widetilde{\mathbf{f}}\|^2}$, can be written as $-\ln \frac{\gamma}{\gamma^2 + l(G_{(1)}(\widetilde{\mathbf{f}}), G_{(2)}(\widetilde{\mathbf{f}}))}$, where $l(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ and $G_{(j)}(\widetilde{\mathbf{f}}) = (\mathbf{s}_{(i)}\widetilde{\mathbf{f}})_j, j = 1, 2$. That is to say, it is a composition of real analytic functions and thus is itself real analytic. As a result, $\widetilde{G}(\widetilde{\mathbf{f}})$ is a real analytic function of $\widetilde{\mathbf{f}}$.

Based on the above two conclusions, $J_n(\mathbf{f}) = H(\mathbf{f}) + G(\mathbf{f}) = \widetilde{H}(\widetilde{\mathbf{f}}) + \widetilde{G}(\widetilde{\mathbf{f}}) = \widetilde{J}_n(\widetilde{\mathbf{f}})$ is a real analytic function of $\widetilde{\mathbf{f}}$. □

Appendix B. Another Proof of Theorem 1

It has already been shown in the main text that $\widetilde{J}_n(\widetilde{\mathbf{f}})$ satisfies the KL property. In order to apply this conclusion to $J_n(\mathbf{f})$, just notice that the definitions of $\nabla_{\mathbf{f}} J_n(\mathbf{f})$ and $\nabla_{\widetilde{\mathbf{f}}} \widetilde{J}_n(\widetilde{\mathbf{f}})$ imply $\|\nabla_{\mathbf{f}} J_n(\mathbf{f})\| = \frac{1}{2} \|\nabla_{\widetilde{\mathbf{f}}} \widetilde{J}_n(\widetilde{\mathbf{f}})\|$.

Therefore, $J_n(\mathbf{f})$ is a KL function. It can also be shown that $J_n(\mathbf{f}) = G(\mathbf{f}) + H(\mathbf{f})$ is proper, continuous, and bounded from below; $H(\mathbf{f})$ is finite-valued, differentiable, and has a Lipschitz continuous gradient; $G(\mathbf{f})$ is continuous on its domain. Apart from differentiability, which is defined by Wirtinger calculus in a special way, the rest of these mentioned properties can be easily established in the complex case by directly replacing the real variable in the original definitions with a complex variable.

Now, we continue to show that all the three assumptions for Theorem 4.2 in [38] are satisfied. We point out that this is not a trivial task, because the inner product used in the original proof [38] takes only real values, which does not hold in our complex case.

First, by computing the optimality condition of the Moreau envelope for $J_n(\mathbf{f})$, and letting $\mathbf{v}^{(k+1)} \in \partial G(\mathbf{o}^{(k+1)})$, we have:

$$2\mu\mathbf{v}^{(k+1)} + 2\mu\nabla H(\mathbf{o}^{(k)}) + \mathbf{o}^{(k+1)} - \mathbf{o}^{(k)} = 0, \tag{A7}$$

and therefore

$$\|\mathbf{v}^{(k+1)} + \nabla H(\mathbf{o}^{(k)})\| = \frac{1}{2\mu} \|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|. \tag{A8}$$

For $H(\mathbf{f}) = \|\mathbf{g} - \mathbf{C}^{(n)}\mathbf{f}\|_2^2$, according to the convexity of $H(\mathbf{f})$ [48] and the property of $\nabla H(\mathbf{f})$ [49], we have for any \mathbf{f}_1 and \mathbf{f}_2 :

$$H(\mathbf{f}_1) - H(\mathbf{f}_2) \leq 2\text{Re}\{(\mathbf{f}_1 - \mathbf{f}_2)^H \nabla H(\mathbf{f}_1)\}. \tag{A9}$$

Therefore,

$$H(\mathbf{f}_1) - H(\mathbf{f}_2) - 2\operatorname{Re}\{(\mathbf{f}_1 - \mathbf{f}_2)^H \nabla H(\mathbf{f}_2)\} \leq 2\operatorname{Re}\{(\mathbf{f}_1 - \mathbf{f}_2)^H (\nabla H(\mathbf{f}_1) - \nabla H(\mathbf{f}_2))\}. \quad (\text{A10})$$

Since $\nabla H(\mathbf{f}) = (\mathbf{C}^{(n)})^H (\mathbf{C}^{(n)} \mathbf{f} - \mathbf{g})$, we have

$$\operatorname{Re}\{(\mathbf{f}_1 - \mathbf{f}_2)^H (\nabla H(\mathbf{f}_1) - \nabla H(\mathbf{f}_2))\} = \|\mathbf{C}^{(n)}(\mathbf{f}_1 - \mathbf{f}_2)\|_2^2. \quad (\text{A11})$$

Therefore, the right side of (A10) is real, and thus

$$\begin{aligned} H(\mathbf{f}_1) - H(\mathbf{f}_2) - 2\operatorname{Re}\{(\mathbf{f}_1 - \mathbf{f}_2)^H \nabla H(\mathbf{f}_2)\} &\leq 2(\mathbf{f}_1 - \mathbf{f}_2)^H (\nabla H(\mathbf{f}_1) - \nabla H(\mathbf{f}_2)) \\ &\leq 2\|(\mathbf{f}_1 - \mathbf{f}_2)\| \|\nabla H(\mathbf{f}_1) - \nabla H(\mathbf{f}_2)\| \leq 2L\|\mathbf{f}_1 - \mathbf{f}_2\|_2^2. \end{aligned} \quad (\text{A12})$$

As a result, we have

$$H(\mathbf{o}^{(k+1)}) \leq H(\mathbf{o}^{(k)}) + 2\operatorname{Re}\{(\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)})^H \nabla H(\mathbf{o}^{(k)})\} + 2L\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|_2^2. \quad (\text{A13})$$

On the other hand, from the definition of a proximal operator,

$$\mu G(\mathbf{o}^{(k+1)}) + \frac{1}{2}\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)} + 2\mu \nabla H(\mathbf{o}^{(k)})\|_2^2 \leq \mu G(\mathbf{o}^{(k)}) + \frac{1}{2}\|2\mu \nabla H(\mathbf{o}^{(k)})\|_2^2. \quad (\text{A14})$$

Expanding (A14) yields

$$\mu G(\mathbf{o}^{(k+1)}) + \frac{1}{2}\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|_2^2 + 2\mu \operatorname{Re}\{(\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)})^H \nabla H(\mathbf{o}^{(k)})\} \leq \mu G(\mathbf{o}^{(k)}), \quad (\text{A15})$$

and therefore

$$G(\mathbf{o}^{(k+1)}) + \frac{1}{2\mu}\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|_2^2 + 2\operatorname{Re}\{(\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)})^H \nabla H(\mathbf{o}^{(k)})\} \leq G(\mathbf{o}^{(k)}), \quad (\text{A16})$$

Combining (A13) with (A16), we have

$$\begin{aligned} &G(\mathbf{o}^{(k+1)}) + H(\mathbf{o}^{(k+1)}) + \frac{a-4L}{2}\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|_2^2 \\ &\leq G(\mathbf{o}^{(k+1)}) + H(\mathbf{o}^{(k)}) + 2\operatorname{Re}\{(\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)})^H \nabla H(\mathbf{o}^{(k)})\} + \frac{a}{2}\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|_2^2 \\ &\leq G(\mathbf{o}^{(k)}) + H(\mathbf{o}^{(k)}), \end{aligned} \quad (\text{A17})$$

with $a = \frac{1}{\mu}$. This is actually $J_n(\mathbf{o}^{(k+1)}) + \frac{a-4L}{2}\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|_2^2 \leq J_n(\mathbf{o}^{(k)})$, as long as $a > 4L$. In contrast, in the proof, from a real perspective, the corresponding requirement is just $a > L$. Therefore, if some other appropriate techniques are utilized, it may be possible to get an inequality better than (A13) and obtain $a > L$.

Second, using differential rule, we have $\mathbf{v}^{(k+1)} + \nabla h(\mathbf{o}^{(k+1)}) \in \nabla J_n(\mathbf{o}^{(k+1)})$.

At last, with (A8), it can be deduced that

$$\begin{aligned} \|\mathbf{v}^{(k+1)} + \nabla H(\mathbf{o}^{(k+1)})\| &\leq \|\mathbf{v}^{(k+1)} + \nabla H(\mathbf{o}^{(k)})\| + \|\nabla H(\mathbf{o}^{(k+1)}) - \nabla H(\mathbf{o}^{(k)})\| \\ &\leq \frac{a}{2}\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\| + L\|\mathbf{o}^{(k+1)} - \mathbf{o}^{(k)}\|. \end{aligned} \quad (\text{A18})$$

Now we are exactly in the case of Theorem 4.2 in [38] and the rest of the proof is similar (just formally substitute the real vectors therein by complex vectors). In conclusion, we can get the same result as Theorem 5.1 in [38], i.e., the convergence of the sequence $\mathbf{o}^{(k)}$ to a critical point of $J_n(\mathbf{f})$.

Moreover, denote by \mathbf{f}^* the global minimizer of $J_n(\mathbf{f})$. According to Theorem 2.12 in [38], for each $r > 0$, there exist $u \in (0, r)$, $\delta > 0$ such that the inequalities $\|\mathbf{o}^{(0)} - \mathbf{f}^*\| < u$ and $\min J_n(\mathbf{f}) < J(\mathbf{o}^{(0)}) < \delta + \min J_n(\mathbf{f})$ imply that the sequence $\mathbf{o}^{(k)}$ will converge to some \mathbf{o}^* with $\mathbf{o}^{(k)} \in B(\mathbf{f}^*, r)$ for arbitrary k and $J_n(\mathbf{o}^*) = \min J_n(\mathbf{f})$.

References

1. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A Tutorial on Synthetic Aperture Radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43.
2. Ouchi, K. Recent Trend and Advance of Synthetic Aperture Radar with Selected Topics. *Remote Sens.* **2013**, *5*, 716–807.
3. Graziano, M.D.; D’Errico, M.; Rufino, G. Wake Component Detection in X-Band SAR Images for Ship Heading and Velocity Estimation. *Remote Sens.* **2016**, *8*, 498. 10.3390/rs8060498.
4. Wahl, D.E.; Eichel, P.; Ghiglia, D.; Jakowatz, C. Phase Gradient Autofocus - a Robust Tool for High Resolution SAR Phase Correction. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 827–835.
5. Calloway, T.M.; Donohoe, G.W. Subaperture Autofocus for Synthetic Aperture Radar. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 617–621.
6. Tsakalides, P.; Nikias, C. High-resolution Autofocus Techniques for SAR Imaging Based on Fractional Lower-order Statistics. *IEE Proc.-Radar Sonar Navig.* **2001**, *148*, 267–276.
7. Fienup, J.; Miller, J. Aberration Correction by Maximizing Generalized Sharpness Metrics. *JOSA A* **2003**, *20*, 609–620.
8. Morrison, R.L.; Do, M.N.; Munson, D.C. SAR Image Autofocus by Sharpness Optimization: A Theoretical Study. *IEEE Trans. Image Process.* **2007**, *16*, 2309–2321.
9. Kragh, T.J.; Kharbouch, A.A. Monotonic Iterative Algorithm for Minimum-entropy Autofocus. *Adapt. Sens. Array Process. Workshop* **2006**, *40*, 1147–1159.
10. Zeng, T.; Wang, R.; Li, F. SAR Image Autofocus Utilizing Minimum-entropy Criterion. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1552–1556.
11. Kantor, J.M. Minimum Entropy Autofocus Correction of Residual Range Cell Migration. In Proceedings of the 2017 IEEE Radar Conference (RadarConf), Seattle, WA, USA, 8–12 May 2017; pp. 0011–0016.
12. Onhon, N.Ö.; Cetin, M. A Sparsity-driven Approach for Joint SAR Imaging and Phase Error Correction. *IEEE Trans. Image Process.* **2011**, *21*, 2075–2088.
13. Kelly, S.I.; Yaghoobi, M.; Davies, M.E. Auto-focus for Under-sampled Synthetic Aperture Radar. In Proceedings of the Sensor Signal Processing for Defence (SSPD 2012), London, UK, 25–27 September 2012.
14. Kelly, S.; Yaghoobi, M.; Davies, M. Sparsity-based Autofocus for Undersampled Synthetic Aperture Radar. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 972–986.
15. Güngör, A.; Cetin, M.; Güven, H.E. An Augmented Lagrangian Method for Autofocused Compressed SAR Imaging. In Proceedings of the 2015 3rd International Workshop on Compressed Sensing Theory and Its Applications to Radar, Sonar and Remote Sensing (CoSeRa), Pisa, Italy, 17–19 June 2015; pp. 1–5.
16. Güngör, A.; Çetin, M.; Güven, H.E. Autofocused Compressive SAR Imaging Based on the Alternating Direction Method of Multipliers. In Proceedings of the 2017 IEEE Radar Conference (RadarConf), Seattle, WA, USA, 8–12 May 2017; pp. 1573–1576.
17. Uğur, S.; Arıkan, O. SAR Image Reconstruction and Autofocus by Compressed Sensing. *Digit. Signal Process.* **2012**, *22*, 923–932.
18. Ash, J.N. An Autofocus Method for Backprojection Imagery in Synthetic Aperture Radar. *IEEE Geosci. Remote Sens. Lett.* **2011**, *9*, 104–108.
19. Sommer, A.; Ostermann, J. Backprojection Subimage Autofocus of Moving Ships for Synthetic Aperture Radar. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8383–8393.
20. Kantor, J.M. Polar Format-Based Compressive SAR Image Reconstruction With Integrated Autofocus. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3458–3468.
21. Mason, E.; Yonel, B.; Yazici, B. Deep learning for SAR image formation. In *Algorithms for Synthetic Aperture Radar Imagery XXIV*; International Society for Optics and Photonics: San Diego, CA, USA, 2017; Volume 10201, p. 1020104.
22. Pu, W. Deep SAR Imaging and Motion Compensation. *IEEE Trans. Image Process.* **2021**, *30*, 2232–2247.
23. Brandwood, D. A Complex Gradient Operator and its Application in Adaptive Array Theory. *IEE Proc. H-Microwaves Opt. Antennas* **1983**, *130*, 11–16.
24. Van Den Bos, A. Complex Gradient and Hessian. *IEE Proc.-Vis. Image Signal Process.* **1994**, *141*, 380–382.
25. Kreutz-Delgado, K. The Complex Gradient Operator and the CR-calculus. *arXiv* **2009**, arXiv:0906.4835.
26. Bouboulis, P. Wirtinger’s Calculus in General Hilbert Spaces. *arXiv* **2010**, arXiv:1005.5170.
27. Zhang, Z.Y.; Pappas, O.; Achim, A. SAR Image Autofocusing using Wirtinger calculus and Cauchy regularization. In Proceedings of the ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1455–1459.
28. Pappas, O.; Achim, A.; Bull, D. Superpixel-level CFAR detectors for ship detection in SAR imagery. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1397–1401.
29. Rizaev, I.G.; Karakuş, O.; Hogan, S.J.; Achim, A. Modeling and SAR Imaging of the Sea Surface: a Review of the State-of-the-Art with Simulations. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 120–140.
30. Çetin, M.; Karl, W.C. Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization. *IEEE Trans. Image Process.* **2001**, *10*, 623–631.
31. Carrara, W.; Goodman, R.S.; Majewski, R.M. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*; Artech House: Norwood, MA, USA, 1995.
32. McCullagh, P.; Polson, N.G. Statistical Sparsity. *Biometrika* **2018**, *105*, 797–814. <https://doi.org/10.1093/biomet/asy051>.

33. Karakuş, O.; Achim, A. On Solving SAR Imaging Inverse Problems Using Non-convex Regularization with a Cauchy-Based Penalty. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5828–5840. <https://doi.org/10.1109/TGRS.2020.3011631>.
34. Karakuş, O.; Mayo, P.; Achim, A. Convergence Guarantees for Non-convex Optimisation with Cauchy-based Penalties. *IEEE Trans. Signal Process.* **2020**, *68*, 6159–6170.
35. Soulez, F.; Thiébaud, É.; Schutz, A.; Ferrari, A.; Courbin, F.; Unser, M. Proximity Operators for Phase Retrieval. *Appl. Opt.* **2016**, *55*, 7412–7421.
36. Güven, H.E.; Güngör, A.; Cetin, M. An Augmented Lagrangian Method for Complex-valued Compressed SAR Imaging. *IEEE Trans. Comput. Imaging* **2016**, *2*, 235–250.
37. Attouch, H.; Bolte, J.; Redont, P.; Soubeyran, A. Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Łojasiewicz Inequality. *Math. Oper. Res.* **2010**, *35*, 438–457.
38. Attouch, H.; Bolte, J.; Svaiter, B.F. Convergence of Descent Methods for Semi-algebraic and Tame Problems: Proximal Algorithms, Forward-backward Splitting, and Regularized Gauss-Seidel Methods. *Math. Program.* **2013**, *137*, 91–129.
39. Csiszár, I. Information Geometry and Alternating Minimization Procedures. *Stat. Decis.* **1984**, *1*, 205–237.
40. Byrne, C.L. Alternating Minimization as Sequential Unconstrained Minimization: a Survey. *J. Optim. Theory Appl.* **2013**, *156*, 554–566.
41. Barrett, R.; Berry, M.; Chan, T.F.; Demmel, J.; Donato, J.; Dongarra, J.; Eijkhout, V.; Pozo, R.; Romine, C.; Van der Vorst, H. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*; SIAM Press: Philadelphia, PA, USA, 1994.
42. Florescu, A.; Chouzenoux, E.; Pesquet, J.C.; Ciuciu, P.; Ciochina, S. A Majorize-minimize Memory Gradient Method for Complex-valued Inverse Problems. *Signal Process.* **2014**, *103*, 285–295.
43. Çetin, M.; Karl, W.C.; Willsky, A.S. Feature-preserving Regularization Method for Complex-valued Inverse Problems with Application to Coherent Imaging. *Opt. Eng.* **2006**, *45*, 017003.
44. Geman, D.; Reynolds, G. Constrained Restoration and the Recovery of Discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 367–383.
45. Joly, P.; Meurant, G. Complex Conjugate Gradient Methods. *Numer. Algorithms* **1993**, *4*, 379–406.
46. Walker, J.L. Range-Doppler Imaging of Rotating Objects. *IEEE Trans. Aerosp. Electron. Syst.* **1980**, *AES-16*, 23–52.
47. Krantz, S.G.; Parks, H.R. *A Primer of Real Analytic Functions*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2002.
48. Zhang, S.; Xia, Y.; Zheng, W. A Complex-valued Neural Dynamical Optimization Approach and its Stability Analysis. *Neural Netw.* **2015**, *61*, 59–67.
49. Liu, S.; Jiang, H.; Zhang, L.; Mei, X. A Neurodynamic Optimization Approach for Complex-variables Programming Problem. *Neural Netw.* **2020**, *129*, 280–287.