Synthetic Aperture Radar Autofocus Based on a Bilinear Model

Kuang-Hung Liu, *Member, IEEE*, Ami Wiesel, *Member, IEEE*, and David C. Munson, Jr., *Fellow, IEEE*

Abstract—Autofocus algorithms are used to restore images in nonideal synthetic aperture radar imaging systems. In this paper, we propose a bilinear parametric model for the unknown image and the nuisance phase parameters and derive an efficient maximum-likelihood autofocus (MLA) algorithm. In the special case of a simple image model and a narrow range of look angles, MLA coincides with the successful multichannel autofocus (MCA). MLA can be interpreted as a generalization of MCA to a larger class of models with a larger range of look angles. We analyze its advantages over previous extensions of MCA in terms of identifiability conditions and noise sensitivity. As a byproduct, we also propose numerical approximations to the difficult constant modulus quadratic program that lies at the core of these algorithms. We demonstrate the superior performance of our proposed methods using computer simulations in both the correct and mismatched system models. MLA performs better than other methods, both in terms of the mean squared error and visual quality of the restored image.

Index Terms—Autofocus, Fourier-domain multichannel autofocus (FMCA), maximum-likelihood estimation, multichannel autofocus (MCA), phase gradient autofocus (PGA), semidefinite relaxation (SDR), sharpness-maximization autofocus, spotlight-mode synthetic aperture radar (SAR), successive cancellation approach (SCA).

I. INTRODUCTION

A spotlight-mode synthetic aperture radar (SAR) provides a high-resolution microwave image using an antenna of small aperture. High resolution in the range direction is achieved through traditional pulse compression, whereas high resolution in the cross-range direction is obtained by illuminating the target from many viewing angles. The collected return signals can be conveniently modeled as polar samples of the Fourier transform of the target's reflectivity function. Image reconstruction is typically accomplished via the traditional polar formatting algorithm [1], [2]. In practice, a main challenge in SAR imaging is the unavoidable demodulation timing error in the radar receiver. The timing error arises from inaccurate range measurements or unknown signal propagation effects. This error causes the reconstructed image to suffer distortion, which is sometimes so severe

that the image is completely unrecognizable. The effect of the timing error can be well approximated as an unknown phase corruption in the collected Fourier data that varies with look angle but is constant for all data collected from a fixed angle. Autofocus algorithms use signal processing techniques to restore the image in the presence of these phase errors [1], [3], [4].

SAR autofocus involves the estimation (reconstruction) of the unknown image corrupted by noise and unknown nuisance phase distortions. For this purpose, it is common to impose additional constraints, either on the underlying image or on the imaging system. Some early spotlight-mode SAR autofocus algorithms assume that the unknown autofocus phases can be described by a finite polynomial expansion [5], [6]. The first successful autofocus method to be widely applied in practice is phase gradient autofocus (PGA) [7]. PGA assumes that the underlying scene consists of isolated point reflectors and that the SAR system collects data across a narrow range of look angles. Another class of autofocus algorithms compensates the phase errors by maximizing the sharpness of the reconstructed image [8]. Popular metrics that measure image sharpness include entropy and various powers of the image intensity [9]. These sharpness-maximizing autofocus algorithms tend to favor sparse images such as collections of point scatterers. While the restoration results obtained using these approaches often are outstanding, the techniques sometimes fail to produce correct restorations when the underlying scene is poorly described by the implicitly assumed image model. Recently, a promising multichannel autofocus (MCA) has been proposed in [10] assuming a small range of look angles and prior knowledge of a region in the image with a small pixel value (low-return region). This image support constraint can be enforced by the antenna pattern found in any practical system. Due to its promise, MCA was then generalized to allow for a larger range of look angles via Fourier-domain multichannel autofocus (FMCA) [11], [12].

In this paper, we propose a novel autofocus reconstruction algorithm that is based on a bilinear parametric model. Following [13]–[16], we consider a standard linear model for the reflectivity function using a finite vector of unknown parameters $\boldsymbol{\theta}$. On the other hand, similar to [10]–[12], we propose a linear model for the SAR acquisition system involving a vector of unknown phase distortions $\boldsymbol{\gamma}$. Together, we obtain a bilinear model involving the unknown parameters $\{\boldsymbol{\theta}, \boldsymbol{\gamma}\}$ contaminated by additive noise. We analyze the conditions for identifiability and solvability of the problem in the noiseless case and then derive a novel maximum-likelihood autofocus (MLA) method to deal with noisy observations.

MLA can be interpreted as a generalization of MCA to a larger class of models and a wider range of look angles. In the

K.-H. Liu is with Schlumberger WesternGeco, Houston, TX (e-mail: khliu@umich.edu).

A. Wiesel is with the School of Computer Science and Engineering, Hebrew University of Jerusalem, Jerusalem 91904, Israel (e-mail: amiw@cs.huji.ac.il).

D. C. Munson, Jr., is with the Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2121 USA (e-mail: munson@umich.edu).

simplest setting, namely, an impulse reflectivity model with a known support constraint and a small range of look angles, MCA coincides with MLA. However, our results show that MCA's FMCA extension is suboptimal in comparison to MLA. The latter requires weaker identifiability conditions, and small error analysis reveals that it is significantly less sensitive to noise. Furthermore, MLA is more general from the standpoint that it does not require the existence of a prior low-return region. In some sense, FMCA may be considered more robust as it does not rely on any explicit parametric reflectivity model. However, the numerical simulation results presented in [11] and [12] are based on a special case of our model, and we present numerical evidence that MLA is superior to FMCA, also under mismatched models.

From a computational complexity perspective, MCA, FMCA, and MLA are very similar. At their core is the solution to a constant modulus quadratic program (CMQP). This is an NP-hard problem, which requires approximations that tradeoff accuracy with complexity [17], [18]. We review the classical approximation based on eigenvalue relaxation (EVR) and review a state-of-the-art approximation based on semidefinite relaxation (SDR) [19], [20]. Next, we propose a third approximation to CMQP, which is based on a successive cancellation approach (SCA). Successive cancellation, also known as decision feedback, was originally derived in the context of multiuser detection [17]. CMQP is very similar to maximum-likelihood (ML) multiuser detection with the only difference being that unknown phases in digital communication belong to a finite constellation set. We extend this method to allow for arbitrary unknown phases. Numerical results show that SCA provides accuracy close to SDR, while its computational complexity is similar to EVR. Thus, it is a promising approximation to MCA and FMCA, as well as to our newly proposed MLA.

A related approach was taken in [21]. There, prior knowledge of the low-return region is crucial to the problem formulation, whereas, in this work, we do not require such information.

The organization of this paper is as follows. In Section II, we introduce the proposed bilinear problem formulation. In Section III, we discuss the conditions for perfect reconstruction in the noiseless case. In Section IV, we derive MLA and discuss its implementation. In Section V, we compare MLA and FMCA through their small error analysis. Simulation results are presented in Section VI, and concluding remarks are provided in Section VII.

In this paper, a capital boldface letter $\mathbf{A}$ denotes a matrix, and a lowercase boldface letter $\mathbf{a}$ denotes a column vector. $\mathrm{conj}(\mathbf{a})$ denotes the complex conjugate of $\mathbf{a}$. $[\mathbf{a}; \mathbf{b}]$ denotes a column vector formed by concatenating vector $\mathbf{a}$ on top of vector $\mathbf{b}$. Superscript ^{H} denotes the Hermitian transpose, and † denotes the Moore–Penrose pseudo-inverse. $\mathrm{tr}(\mathbf{A})$ represents the trace of matrix $\mathbf{A}$, and $\mathrm{Diag}(\mathbf{a})$ is the diagonal matrix with elements of vector $\mathbf{a}$ on the main diagonal. $\mathrm{rank}(\mathbf{A})$ denotes rank of matrix $\mathbf{A}$, and $\mathcal{N}(\mathbf{A})$ denotes the null space of $\mathbf{A}$. For an index set s, $\mathbf{A}_s(\mathbf{A}^s)$ represents the submatrix of $\mathbf{A}$ formed by the rows(columns) of $\mathbf{A}$ indexed by s. For a complex number c, $|c|$ represents the magnitude of c, and $\angle(c)$ represents the phase of c.

II. PROBLEM FORMULATION

In this section, we present a parametric bilinear model for the SAR autofocus problem.

A. SAR Reflectivity Function Model

SAR systems image a continuous target reflectivity function denoted by $r(x, y)$. Most reconstruction methods assume a finite parametric model for $r(x, y)$. This parametric model can be mathematically described as

$$r(x,y) = \sum_{i=1}^{D} \theta_i \, h_i(x,y) \qquad (1)$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_D]^T$ is a complex-valued parameter vector of dimension D, and $h_i(x,y)$ for $i = 1, \ldots, D$ are known complex-valued functions. Common physical models that can be characterized by (1) are as follows.

(A) *Impulse model:* $r(x,y)$ is decomposed into D resolution cells, each with size $dx \times dy$, where dx and dy are two known constants determined by the SAR radar specification. Any two point reflectors within a resolution cell cannot be well resolved. With this framework, one way of approximating a SAR system is to use a discrete model, where a single complex number is assigned to each resolution cell. This complex number, which is denoted by θ_i, represents the sum of the point reflectors' reflectivities within a resolution cell and can be simply viewed as an impulse located at the center of the cell. This leads to the following definition:

$$h_i(x,y) = \delta(x - u_i \, dx, \, y - v_i \, dy) \qquad (2)$$

where $\delta(\cdot)$ denotes the standard Dirac delta function, and u_i and v_i are the spatial indexes for the ith resolution cell.

(B) *Sampling model:* In this model, $\boldsymbol{\theta}$ represents discrete samples of $r(x,y)$, and $h_i(x,y)$ represent the sampling kernel. A typical assumption in this model is a band-limited reflectivity function, which involves a finite number of parameters. Previous works that use this model include [13]–[15].

(C) *Discrete speckle model:* In this model, SAR imaging is viewed as a statistical inverse problem, and it is recognized that it is possible to reconstruct only a speckle version of $r(x,y)$. Here, model (1) can be interpreted as an approximate model for the speckle image [1], [4].

(D) *Additional prior information:* Additional linear constraints may be incorporated into the model if prior information about the underlying scene is known. For example, a previous work assumed knowledge of a zero (or low) valued region in the underlying scene [10], [11]. Such information corresponds to elements in $\boldsymbol{\theta}$ with known zero values. Alternatively, this can be translated into a linear model of reduced dimension (smaller value of D).

B. SAR Acquisition Model

Under the far-field assumption and using a narrowband transmitted waveform, the collected SAR data, which are

denoted by $G[m,n]$, can be modeled as the Fourier transform of $r(x,y)$ evaluated at nonuniform frequency locations $(F_x[m,n], F_y[m,n])$ for $m = 1,\ldots,M$ and $n = 1,\ldots,N$, i.e.,

$$G[m,n] = \int \int r(x,y) e^{-j2\pi(F_x[m,n]\,x + F_y[m,n]\,y)} dx\,dy. \quad (3)$$

Substituting (3) back into our parametric model for the reflectivity function yields

$$G[m,n] = \sum_i \theta_i \int\int h_i(x,y) e^{-j2\pi(F_x[m,n]\,x + F_y[m,n]\,y)} dx\,dy. \quad (4)$$

Using vector notation

$$\mathbf{g} = \text{vec}(\mathbf{G}) \quad (5)$$

there is a simple linear relation between $\boldsymbol{\theta}$ and $\mathbf{g}$, which can be expressed as

$$\mathbf{g} = \mathbf{L}\boldsymbol{\theta} \quad (6)$$

where L is a $MN \times D$ matrix with elements

$$L[k,l] = \int \int h_k(x,y) \, e^{-j2\pi(F_x[a,b]\,x + F_y[a,b]\,y)} dx\,dy$$
$$a = \lfloor l/N \rfloor, \; b = \text{mod}(l,N) \quad (7)$$

where $\lfloor \cdot \rfloor$ denotes "integer part of," and $\text{mod}(l,N)$ denotes "l modulo N."

Classical SAR reconstruction amounts to inversion of this linear transformation to reconstruct $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} = \mathbf{L}^\dagger \mathbf{g}. \quad (8)$$

In the special case of model *(A)* and when SAR operates across a narrow range of look angles, F_x and F_y can be well approximated as a Cartesian grid; thus, the matrix $\mathbf{L}$ is simply a discrete Fourier transform (DFT) matrix, and inversion is performed using an efficient fast Fourier transform (FFT) [10]. Otherwise, it is common to approximate this pseudo-inversion via interpolation to a uniform Cartesian grid, followed by FFT.

The above SAR model is too idealistic for practical systems. We now extend the model and introduce signal distortion and noise. Specifically, a more realistic observation model is

$$\tilde{G}[m,n] = G[m,n]\, e^{j\phi[m,n]} + N[m,n] \quad (9)$$

where $\phi[m,n]$ are autofocus phase distortions, and $N[m,n]$ represents additive noise. The phase distortions result from inaccurate range measurements or unknown signal propagation delays. The polar-format Fourier data are contaminated with unknown phase errors that cause the reconstructed image to suffer distortion. The measurements at a given look angle suffer from the same unknown delay, and, under a narrowband assumption, this corresponds to an unknown phase. The delays, and their associated phases, change between different look angles. Thus, following [1], we let

$$\phi[m,n] = \phi(m), \qquad m = 1,\ldots,M. \quad (10)$$

In addition, without loss of generality, we define

$$\phi(M) = 0 \quad (11)$$

and focus on estimating $\phi(1),\ldots,\phi(M-1)$, as we are only interested in the phase differences (PGA in [7] is also based on this approach). The additive noise samples $N[m,n]$ are assumed to be independent, zero mean, complex normal random variables with variance σ_n^2.

In vector notation, we obtain the following model:

$$\tilde{\mathbf{g}} = \mathbf{R}(\boldsymbol{\gamma})\mathbf{L}\boldsymbol{\theta} + \mathbf{n} \quad (12)$$

where $\boldsymbol{\gamma} = [e^{j\boldsymbol{\phi}};\ 1]$, $\mathbf{n} = \text{vec}(\mathbf{N})$ and

$$\mathbf{R}(\boldsymbol{\gamma}) = \text{Diag}\left(\left[\underbrace{\gamma_1,\ldots,\gamma_1}_{\text{N terms}},\ldots,\underbrace{\gamma_{M-1},\ldots,\gamma_{M-1}}_{\text{N terms}},\underbrace{1,\ldots,1}_{\text{N terms}}\right]\right). \quad (13)$$

Define Γ as the space where $\boldsymbol{\gamma}$ lies, i.e.,

$$\Gamma = \{\boldsymbol{\gamma} : \gamma_M = 1, |\gamma_i| = 1, i = 1,\ldots,M-1\}. \quad (14)$$

Then, the SAR autofocus problem can be summarized as: Find $\boldsymbol{\theta}$ (and the nuisance parameters $\boldsymbol{\gamma} \in \Gamma$) using the observations $\tilde{\mathbf{g}}$.

For completeness, we note that a more accurate problem formulation would treat the magnitudes and phases of the complex variables, $\boldsymbol{\theta}$, separately. Specifically, other work models the magnitudes and phases of the reflectivity function differently [22], but our goal is only to reconstruct the magnitude information for display [4]. For simplicity and tractability, we use a joint model and estimate both magnitude and phase together. Future work will pursue the more advanced formulation.

III. NOISELESS CASE

The autofocus problem is difficult due to the nonlinear coupling between the unknown reflectivity parameters $\boldsymbol{\theta}$ and the unknown autofocus phases $\boldsymbol{\phi}$. Even if we parameterize the phases via $\boldsymbol{\gamma}$ (rather than the more complicated $\boldsymbol{\phi}$ characterization), there is a bilinear coupling between $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$. Therefore, in order to understand when can this coupling be resolved, we begin by considering the problem in the noiseless case, i.e., $\mathbf{n} = \mathbf{0}$.

A. Perfect Reconstruction

Theorem 3.1: In the noiseless case and when $\mathbf{L}$ has full column rank, perfect reconstruction of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ from $\tilde{\mathbf{g}}$ is possible if and only if there is a unique vector $\hat{\boldsymbol{\gamma}}$ such that

$$\hat{\boldsymbol{\gamma}} \in \Gamma \cap \mathcal{N}\left((\mathbf{I} - \mathbf{L}\mathbf{L}^\dagger)\tilde{\mathbf{Y}}\right) \quad (15)$$

where

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{g}}_{1\ldots N} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{g}}_{N+1\ldots 2N} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tilde{\mathbf{g}}_{MN-N+1\ldots MN} \end{bmatrix} \quad (16)$$

and $\tilde{\mathbf{g}}_{1...N} = [\tilde{g}(1), \ldots, \tilde{g}(N)]^T$. In this case, $\hat{\boldsymbol{\gamma}} = \mathrm{conj}(\boldsymbol{\gamma})$ and

$$\boldsymbol{\theta} = \mathbf{L}^\dagger \mathbf{R}(\hat{\boldsymbol{\gamma}})\tilde{\mathbf{g}}. \tag{17}$$

Proof: Our first step is to decouple $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ from (12). This is done by first recognizing that, when $\mathbf{L}$ has full column rank, we can write

$$\boldsymbol{\theta} = \mathbf{L}^\dagger \mathbf{R}(\boldsymbol{\gamma})^{-1}\tilde{\mathbf{g}}. \tag{18}$$

Note that, in this case, $\mathbf{R}(\boldsymbol{\gamma})^{-1} = \mathbf{R}(\boldsymbol{\gamma})^H$. Substituting (18) back into (12), we have

$$\left(\mathbf{I} - \mathbf{R}(\boldsymbol{\gamma})\mathbf{L}\mathbf{L}^\dagger \mathbf{R}(\boldsymbol{\gamma})^{-1}\right)\tilde{\mathbf{g}} = \mathbf{0} \tag{19}$$

or

$$(\mathbf{I} - \mathbf{L}\mathbf{L}^\dagger)\mathbf{R}(\boldsymbol{\gamma})^{-1}\tilde{\mathbf{g}} = \mathbf{0}. \tag{20}$$

We can rewrite (20) as

$$(\mathbf{I} - \mathbf{L}\mathbf{L}^\dagger)\tilde{\mathbf{Y}}\hat{\boldsymbol{\gamma}} = \mathbf{0} \tag{21}$$

where $\hat{\boldsymbol{\gamma}} = \mathrm{conj}(\boldsymbol{\gamma})$, and $\tilde{\mathbf{Y}}$ is defined in (16). Together with the requirement $\hat{\gamma} \in \Gamma$, this yields the required condition. ∎

In practice, our noiseless autofocus algorithm takes $\tilde{\mathbf{Y}}$ as an input and searches for $\hat{\gamma}$ that satisfies (15). This search is difficult due to the nonconvex set Γ. We propose to relax it, and choose $\hat{\gamma}$ as any properly normalized vector in $\mathcal{N}((\mathbf{I} - \mathbf{L}\mathbf{L}^\dagger)\tilde{\mathbf{Y}})$ rounded to Γ.

B. Comparison to Previous Methods

We now compare Theorem 3.1 with its competing MCA and FMCA methods. Both MCA and FMCA are also based on the noiseless model

$$\tilde{\mathbf{g}} = \mathbf{R}(\boldsymbol{\gamma})\mathbf{L}\boldsymbol{\theta} \tag{22}$$

and it is *a priori* known that some of the elements of $\boldsymbol{\theta}$ are zero valued, i.e.,

$$\boldsymbol{\theta}_a = \mathbf{0} \tag{23}$$

where a is a known set of indices. FMCA searches for a vector $\hat{\boldsymbol{\gamma}} \in \Gamma$ such that the reconstruction satisfies (23), i.e.,

$$\left[\mathbf{L}^\dagger \mathbf{R}(\hat{\boldsymbol{\gamma}})^{-1}\tilde{\mathbf{g}}\right]_a = \mathbf{0}. \tag{24}$$

In terms of $\tilde{\mathbf{Y}}$ and Γ, the condition for the perfect reconstruction of (22) is equivalent to requiring a unique vector $\hat{\boldsymbol{\gamma}}$ such that

$$\hat{\boldsymbol{\gamma}} \in \Gamma \cap \mathcal{N}\left([\mathbf{L}^\dagger]_a \tilde{\mathbf{Y}}\right). \tag{25}$$

In our proposed parametric model, the prior information of the set a corresponds to a reduced model. More specifically, let $\bar{a}$ denote the complement of a so that $a \cap \bar{a} = \emptyset$, $a \cup \bar{a} = \{1, \ldots, D\}$, and we can partition $\mathbf{L}$ into

$$\mathbf{L} = [\mathbf{L}^{\bar{a}} \ \mathbf{L}^a]. \tag{26}$$

This gives us the effective model

$$\tilde{\mathbf{g}} = \mathbf{R}(\boldsymbol{\gamma})\mathbf{L}^{\bar{a}}\boldsymbol{\theta}_{\bar{a}}. \tag{27}$$

Its necessary and sufficient reconstruction condition is a unique vector satisfying

$$\hat{\boldsymbol{\gamma}} \in \Gamma \cap \mathcal{N}\left(\left(\mathbf{I} - [\mathbf{L}^{\bar{a}}][\mathbf{L}^{\bar{a}}]^\dagger\right)\tilde{\mathbf{Y}}\right). \tag{28}$$

The following theorem compares these conditions in the noiseless case.

Theorem 3.2: If $\mathbf{L}$ is full column rank (as required for reconstruction in the focused case), then

$$\mathcal{N}\left(\left(\mathbf{I} - [\mathbf{L}^{\bar{a}}][\mathbf{L}^{\bar{a}}]^\dagger\right)\tilde{\mathbf{Y}}\right) \subset \mathcal{N}\left([\mathbf{L}^\dagger]_a \tilde{\mathbf{Y}}\right). \tag{29}$$

If $\mathbf{L}$ is also invertible, then conditions (25) and (28) are equivalent since

$$\mathcal{N}\left(\left(\mathbf{I} - [\mathbf{L}^{\bar{a}}][\mathbf{L}^{\bar{a}}]^\dagger\right)\tilde{\mathbf{Y}}\right) = \mathcal{N}\left([\mathbf{L}^\dagger]_a \tilde{\mathbf{Y}}\right). \tag{30}$$

Proof: First, let $\mathbf{s} \in \mathcal{N}((\mathbf{I} - [\mathbf{L}^{\bar{a}}][\mathbf{L}^{\bar{a}}]^\dagger)\tilde{\mathbf{Y}})$ so that $(\mathbf{I} - \mathbf{L}^{\bar{a}}[\mathbf{L}^{\bar{a}}]^\dagger)\tilde{\mathbf{Y}}\mathbf{s} = \mathbf{0}$. This implies that $\tilde{\mathbf{Y}}\mathbf{s}$ is in the column space of $\mathbf{L}^{\bar{a}}$, and therefore, there exists a vector $\mathbf{z}$ such that $\tilde{\mathbf{Y}}\mathbf{s} = \mathbf{L}^{\bar{a}}\mathbf{z} = \mathbf{L}[\mathbf{z}; \mathbf{0}]$. We have

$$[\mathbf{L}^\dagger]_a \tilde{\mathbf{Y}}\mathbf{s} = [\mathbf{L}^\dagger]_a \mathbf{L}[\mathbf{z}; \mathbf{0}] = \left[\mathbf{L}^\dagger \mathbf{L}[\mathbf{z}; \mathbf{0}]\right]_a = [\mathbf{z}; \mathbf{0}]_a = \mathbf{0} \tag{31}$$

hence, $\mathbf{s} \in \mathcal{N}([\mathbf{L}^\dagger]_a \tilde{\mathbf{Y}})$, and we have the desired result.

On the other hand, assume $\mathbf{L}$ is also invertible, so that $\mathbf{L}^\dagger = \mathbf{L}^{-1}$. Let $\mathbf{v} \in \mathcal{N}([\mathbf{L}^{-1}]_a \tilde{\mathbf{Y}})$ and $\boldsymbol{\theta}' = \mathbf{L}^{-1}\tilde{\mathbf{Y}}\mathbf{v}$. Since $[\mathbf{L}^{-1}]_a \tilde{\mathbf{Y}}\mathbf{v} = \mathbf{0}$, we have $\boldsymbol{\theta}'_a = \mathbf{0}$ and $\tilde{\mathbf{Y}}\mathbf{v} = \mathbf{L}^{\bar{a}}\boldsymbol{\theta}'_{\bar{a}}$. Now, we have

$$\begin{aligned}\left(\mathbf{I} - [\mathbf{L}^{\bar{a}}][\mathbf{L}^{\bar{a}}]^\dagger\right)\tilde{\mathbf{Y}}\mathbf{v} &= \tilde{\mathbf{Y}}\mathbf{v} - \mathbf{L}^{\bar{a}}\left([\mathbf{L}^{\bar{a}}]^H[\mathbf{L}^{\bar{a}}]\right)^{-1}[\mathbf{L}^{\bar{a}}]^H \tilde{\mathbf{Y}}\mathbf{v} \\ &= \mathbf{L}^{\bar{a}}\boldsymbol{\theta}'_{\bar{a}} - \mathbf{L}^{\bar{a}}\left([\mathbf{L}^{\bar{a}}]^H \mathbf{L}^{\bar{a}}\right)^{-1}[\mathbf{L}^{\bar{a}}]^H \mathbf{L}^{\bar{a}}\boldsymbol{\theta}'_{\bar{a}} \\ &= \mathbf{0}\end{aligned} \tag{32}$$

hence, $\mathbf{v} \in \mathcal{N}((\mathbf{I} - [\mathbf{L}^{\bar{a}}][\mathbf{L}^{\bar{a}}]^\dagger)\tilde{\mathbf{Y}})$. ∎

Under the MCA framework, the matrix $\mathbf{L}$ is a square, invertible, and unitary DFT matrix; therefore, (30) is satisfied, and we conclude that MCA is optimal in the noiseless case in the sense that it is equivalent to Theorem 3.1 and that perfect reconstruction is achieved. On the other hand, its generalization to nonsquare matrices via FMCA is suboptimal and requires a condition that is too strong. In summary, we interpret Theorem 3.1 as the correct extension of MCA to nonsquare matrices and generalized parametric models with or without low-return regions.

IV. MAXIMUM-LIKELIHOOD ESTIMATION

The previous section addressed conditions and methods for perfect reconstruction in the noiseless case. In practice, the measurements are usually corrupted by additive noise, as expressed in (12). We now assume that the conditions in Theorem 3.1 hold so that the problem is solvable, and we extend the previous results to the noisy case based on an ML framework.

ML is the classical statistical method for estimating deterministic unknown parameters. In the presence of Gaussian noise,

the method reduces to nonlinear least-squares estimation. Its main advantage is that it is known to minimize the mean squared error among all unbiased estimators in low noise conditions, i.e., small error analysis shows that MLA attains the Cramer–Rao bound (CRB) on the estimation error. Specifically, we define the MLA estimator as the solution to

$$[\hat{\boldsymbol{\gamma}}_{\text{ML}}, \hat{\boldsymbol{\theta}}_{\text{ML}}] = \underset{\boldsymbol{\gamma} \in \Gamma, \boldsymbol{\theta}}{\arg\min} \|\tilde{\mathbf{g}} - \mathbf{R}(\boldsymbol{\gamma})\mathbf{L}\boldsymbol{\theta}\|^2$$

$$= \underset{\boldsymbol{\gamma} \in \Gamma, \boldsymbol{\theta}}{\arg\min} \left\|\mathbf{R}^{-1}(\boldsymbol{\gamma})\tilde{\mathbf{g}} - \mathbf{L}\boldsymbol{\theta}\right\|^2 \qquad (33)$$

where we have used the invariance of the norm to unitary transformation. Now, we can easily solve for $\hat{\boldsymbol{\theta}}_{\text{ML}}$

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \mathbf{L}^{\dagger}\mathbf{R}^{-1}(\boldsymbol{\gamma})\tilde{\mathbf{g}}. \qquad (34)$$

Substituting (34) back into (33), we have

$$\hat{\boldsymbol{\gamma}}_{\text{ML}} = \underset{\boldsymbol{\gamma} \in \Gamma}{\arg\min} \left\|(\mathbf{I} - \mathbf{L}\mathbf{L}^{\dagger})\mathbf{R}^{-1}(\boldsymbol{\gamma})\tilde{\mathbf{g}}\right\|^2 \qquad (35)$$

$$= \underset{\text{conj}(\boldsymbol{\gamma}) \in \Gamma}{\arg\min} \left\|(\mathbf{I} - \mathbf{L}\mathbf{L}^{\dagger})\tilde{\mathbf{Y}}\boldsymbol{\gamma}\right\|^2 \qquad (36)$$

where we have used the notation in (16). As will be detailed shortly, this problem is a CMQP, which is generally NP-hard [17], [18] but can be approximated well under suitable conditions.

A. Comparison to Previous Methods

The previous MCA and FMCA methods also recognized that the measurements may be noisy and proposed to approximate (24) via minimizing the energy in the low-return region of the reconstructed image, i.e.,

$$\hat{\boldsymbol{\gamma}}_{\text{FMCA}} = \underset{\boldsymbol{\gamma} \in \Gamma}{\arg\min} \left\|[\mathbf{L}^{\dagger}]_a \mathbf{R}^{-1}(\boldsymbol{\gamma})\tilde{\mathbf{g}}\right\|^2 \qquad (37)$$

$$= \underset{\text{conj}(\boldsymbol{\gamma}) \in \Gamma}{\arg\min} \left\|[\mathbf{L}^{\dagger}]_a \tilde{\mathbf{Y}}\boldsymbol{\gamma}\right\|^2. \qquad (38)$$

Similar to MLA, this is an NP-hard CMQP optimization problem, which will be discussed next.

Comparing (36) with the reduced model $\mathbf{L}^{\bar{a}}$ to (38), it is easy to see the similarity between MLA and FMCA. However, it is interesting to note the different interpretations: FMCA uses the Euclidean norm in an attempt to minimize energy, whereas MLA chooses this norm due to the Gaussian noise. In this sense, MLA is easier to generalize to other scenarios involving different noise characteristics, e.g., correlated noise or non-Gaussian noise.

B. Numerical Approximations for CMQP

We have shown that both MLA and its preceding MCA methods reduce to the solution of a CMQP. We now address the numerical approximation of this difficult optimization problem. For this purpose, we define a general CMQP as

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\gamma} \in \Gamma}{\arg\min} \|\mathbf{A}\boldsymbol{\gamma}\|^2. \qquad (39)$$

Both (36) and (38) reduce to this model by choosing

$$\mathbf{A} = \text{conj}((\mathbf{I} - \mathbf{L}\mathbf{L}^{\dagger})\tilde{\mathbf{Y}}) \qquad (40)$$

and

$$\mathbf{A} = \text{conj}([\mathbf{L}^{\dagger}]_a \tilde{\mathbf{Y}}) \qquad (41)$$

respectively.

The first and simplest approach to CMQP is based on eigendecomposition or EVR. This is the technique used in the original MCA and FMCA methods. EVR replaces the difficult feasible set Γ with

$$\left\{\boldsymbol{\gamma} : |\boldsymbol{\gamma}|^2 = M\right\}. \qquad (42)$$

The solution to this relaxation has a simple closed-form solution, namely, the properly scaled right singular vector $\boldsymbol{u}_{\min}(\mathbf{A})$ associated with the minimum singular value of $\mathbf{A}$. The approximate $\hat{\boldsymbol{\gamma}}$ is

$$\hat{\boldsymbol{\gamma}} = e^{\angle \boldsymbol{u}_{\min}(\mathbf{A})}. \qquad (43)$$

EVR requires low computational complexity but is very sensitive to noise.

A second and more accurate approximation to CMQP is SDR. This technique was recently applied to MCA and FMCA in [20]. SDR reformulates CMQP as a quadratically constrained quadratic program. Thus

$$\begin{aligned} \underset{\mathbf{x} \in \mathbb{C}^M}{\min} \quad & \mathbf{x}^H \mathbf{C} \mathbf{x} \\ s.t \quad & |\mathbf{x}_i| = 1, \quad i = 1, \ldots, M \end{aligned} \qquad (44)$$

where $\mathbf{C} = \mathbf{A}^H \mathbf{A}$. It then lifts the vector variable into the space of positive semidefinite matrices, i.e.,

$$\begin{aligned} \underset{\mathbf{X} \in \mathbb{C}^{M \times M}}{\min} \quad & \text{tr}(\mathbf{C}\mathbf{X}) \\ s.t \quad & \mathbf{X}_{ii} = 1, \quad i = 1, \ldots, M \\ & \mathbf{X} \succeq 0 \\ & \text{rank}(\mathbf{X}) = 1 \end{aligned} \qquad (45)$$

where $\mathbf{X} = \mathbf{x}\mathbf{x}^H$. Finally, it relaxes the problematic nonconvex rank-1 constraint to obtain

$$\begin{aligned} \underset{\mathbf{X} \in \mathbb{C}^{M \times M}}{\min} \quad & \text{tr}(\mathbf{C}\mathbf{X}) \\ s.t \quad & \mathbf{X}_{ii} = 1, \quad i = 1, \ldots, M \\ & \mathbf{X} \succeq 0. \end{aligned} \qquad (46)$$

This last problem is a convex semidefinite program that can be efficiently solved using standard optimization methods [23], [24]. Once the optimal solution, $\mathbf{X}_{\text{opt}}$, to (46) is found, its principal eigenvector $\boldsymbol{v}_{\min}(\mathbf{X}_{\text{opt}})$ is computed, and $\hat{\boldsymbol{\gamma}}$ is approximated as

$$\hat{\boldsymbol{\gamma}} = e^{\angle \boldsymbol{v}_{\min}(\mathbf{X}_{\text{opt}})}. \qquad (47)$$

More advanced SDR approximations based on randomization are described in [19] and [20]. SDR is known to provide a tighter

approximation to CMQP than does EVR, but its computational complexity is significantly higher.

Finally, we propose a third numerical approximation to CMQP based on an SCA. To our knowledge, this method has not been applied before in the context of autofocus. Successive cancellation, also known as decision feedback, was originally derived in the context of multiuser detection [17]. CMQP is very similar to ML multiuser detection, with the only difference being that unknown phases in digital communication belong to a finite constellation set. We now extend this method to allow for arbitrary unknown phases.

We use the QR decomposition to factorize $\mathbf{A} = \mathbf{QR}$, where $\mathbf{Q}$ is a unitary matrix, and $\mathbf{R}$ is an upper triangular matrix, with positive and real-valued diagonal elements [25]. Due to the invariance of the norm to unitary transformations, we obtain

$$\|\mathbf{A}\boldsymbol{\gamma}\|^2 = \|\mathbf{R}\boldsymbol{\gamma}\|^2. \tag{48}$$

Thus, the CMQP in (44) is equivalent to

$$\min_{\boldsymbol{\gamma} \in \Gamma} \sum_{i=1}^{M} \left| R_{i,i}\gamma_i + \sum_{k>i}^{M} R_{i,k}\gamma_k \right|^2. \tag{49}$$

The ith term in the objective depends only on γ_k for $k \geq i$. Therefore, we propose to approximate (49) by minimizing each term of the sum sequentially, starting from $i = M$ and stepping down to $i = 1$. As defined in Γ, we let $\gamma_M = 1$. Each subsequent γ_i is then given by the solution to

$$\begin{aligned} \gamma_i &= \underset{\{\gamma_i : |\gamma_i| = 1\}}{\arg\min} \left| \gamma_i + \sum_{k=i+1}^{M} \frac{R_{i,k}}{R_{i,i}}\gamma_k \right|^2 \\ &= e^{j \angle \left[-\sum\limits_{k=i+1}^{M} \frac{R[i,k]}{R[i,i]}\gamma_k \right]} \end{aligned} \tag{50}$$

where γ_k for $k > i$ are already *known* from previous *decisions*.

SCA can be represented in pseudo-code as

Input: matrix $\mathbf{A}$ with size $D \times M$

Output: autofocus phase estimates $\boldsymbol{\gamma} \in \Gamma$

1 $\mathbf{R} = \text{QR}(\mathbf{A})$;

2 $\gamma_M = 1$;

3 for $i \leftarrow M - 1$ to 1 do

4     $\gamma_i = e^{j \angle [-\sum_{k=i+1}^{M}(R[i,k]/R[i,i])\gamma_k]}$;

5 end.

V. SMALL ERROR ANALYSIS

In this section, we provide small error analysis for the FMCA and MLA methods. We define the real-valued parameter vector as

$$\boldsymbol{\xi} = \left[\boldsymbol{\phi}, \text{Re}\{\boldsymbol{\theta}^T\}, \text{Im}\{\boldsymbol{\theta}^T\} \right]^T \tag{51}$$

where $\boldsymbol{\phi} = [\phi(1), \ldots, \phi(M-1)]^T$ are the autofocus phases (recall that we assumed $\phi(M) = 0$). We denote by subscripts TRUE, ML, and FMCA the true parameters, their MLA estimates, and their FMCA estimates, respectively. For simplicity, we parameterize the unknown phases using $\boldsymbol{\phi}$ instead of $\boldsymbol{\gamma}$. Furthermore, we concentrate on estimating the autofocus phase errors, $\boldsymbol{\phi}$, while treating $\boldsymbol{\theta}$ as nuisance parameters.

Small error analysis of ML estimators in Gaussian distributions has a simple closed-form solution that is known to coincide with the well known CRB on the mean squared error [26]. Let $\hat{\boldsymbol{\phi}}_{\text{ML}} \in \mathbb{R}^{M-1}$ denote the MLA phase estimates for $\phi(1), \ldots, \phi(M-1)$. Then, the asymptotic error satisfies

$$E\left[(\hat{\boldsymbol{\phi}}_{\text{ML}} - \boldsymbol{\phi}_{\text{TRUE}})(\hat{\boldsymbol{\phi}}_{\text{ML}} - \boldsymbol{\phi}_{\text{TRUE}})^T \right] \approx \left[\mathbf{F}^{-1}(\boldsymbol{\xi}) \right]_{1:M-1, 1:M-1} \tag{52}$$

where $\mathbf{F}(\boldsymbol{\xi})$ is the Fisher information matrix defined by

$$\mathbf{F}(\boldsymbol{\xi}) = \frac{2}{\sigma_n^2} \text{Re}\left\{ \mathbf{J}(\boldsymbol{\xi}_{\text{TRUE}})^H \mathbf{J}(\boldsymbol{\xi}_{\text{TRUE}}) \right\} \tag{53}$$

where the Jacobian matrix is given by

$$\mathbf{J}(\boldsymbol{\xi}) = \left[j\mathbf{Y}_2 \text{Diag}(e^{j\boldsymbol{\phi}_{\text{TRUE}}}), \ \mathbf{R}\left([e^{j\boldsymbol{\phi}_{\text{TRUE}}}; 1]\right)\mathbf{L}, \right.$$
$$\left. j\mathbf{R}\left([e^{j\boldsymbol{\phi}_{\text{TRUE}}}; 1]\right)\mathbf{L} \right] \tag{54}$$

with

$$\begin{aligned} \mathbf{Y} &= \mathbf{R}\left([e^{j\boldsymbol{\phi}_{\text{TRUE}}}; 1]\right) \\ &\cdot \begin{bmatrix} \mathbf{g}_{1\ldots N} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{g}_{N+1\ldots 2N} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{g}_{MN-N+1,\ldots,MN} \end{bmatrix} \end{aligned} \tag{55}$$

and $\mathbf{Y}_2$ is a matrix comprising the first to the $(M-1)$th columns of $\mathbf{Y}$. Comparing $\mathbf{Y}$ with $\tilde{\mathbf{Y}}$ ($\mathbf{Y}_2$ with $\tilde{\mathbf{Y}}_2$), $\mathbf{Y}$ ($\mathbf{Y}_2$) represents noise-free data, where $\tilde{\mathbf{Y}}$ ($\tilde{\mathbf{Y}}_2$) is noise corrupted, as shown in (12).

FMCA involves solving a nonlinear least-squares problem

$$\hat{\boldsymbol{\phi}}_{\text{FMCA}} = \underset{\boldsymbol{\phi} \in \mathbb{R}^{M-1}}{\arg\min} \|\mathbf{f}(\boldsymbol{\phi})\|^2 \tag{56}$$

where

$$\mathbf{f}(\boldsymbol{\phi}) = [\mathbf{L}^{\dagger}]_a \left(\tilde{\mathbf{Y}} [e^{-j\boldsymbol{\phi}}; 1] \right). \tag{57}$$

Its small error analysis consists of approximating $\mathbf{f}(\boldsymbol{\phi})$ by a first-order Taylor series expansion about the true parameters

$$\mathbf{f}(\boldsymbol{\phi}) \approx \mathbf{f}(\boldsymbol{\phi}_{\text{TRUE}}) + \tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}})(\boldsymbol{\phi} - \boldsymbol{\phi}_{\text{TRUE}}) \tag{58}$$

where

$$\tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) = -j[\mathbf{L}^{\dagger}]_a \tilde{\mathbf{Y}}_2 \text{Diag}(e^{-j\boldsymbol{\phi}_{\text{TRUE}}}) \tag{59}$$

is the Jacobian matrix, and $\check{\mathbf{Y}}_2$ is a matrix comprising the first to the $(M-1)$th columns of $\check{\mathbf{Y}}$. Consequently, the estimate (56) is approximated by

$$\hat{\boldsymbol{\phi}}_{\text{FMCA}} \approx \underset{\boldsymbol{\phi} \in \mathbb{R}^M}{\arg\min} \; \left\| \mathbf{q} + \tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}})\boldsymbol{\phi} \right\|^2 \qquad (60)$$

where

$$\mathbf{q} = [\mathbf{L}^\dagger]_a \tilde{\mathbf{Y}} \left[e^{-j\boldsymbol{\phi}_{\text{TRUE}}}; 1 \right] - \tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}})\boldsymbol{\phi}_{\text{TRUE}}$$
$$= [\mathbf{L}^\dagger]_a \mathbf{R} \left(\left[e^{-j\boldsymbol{\phi}_{\text{TRUE}}}; 1 \right] \right) \tilde{\mathbf{g}} - \tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}})\boldsymbol{\phi}_{\text{TRUE}}.$$
$$(61)$$

Since $\mathbf{q}$ and $\tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}})$ are complex-valued and $\boldsymbol{\phi}$ is real-valued, the solution is

$$\hat{\boldsymbol{\phi}}_{\text{FMCA}} \approx - \left[\text{Re} \left\{ \tilde{\mathbf{J}}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})\tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) \right\} \right]^{-1}$$
$$\times \text{Re} \left\{ \tilde{\mathbf{J}}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})\mathbf{q} \right\}. \qquad (62)$$

Substituting (61) into (62) yields

$$\hat{\boldsymbol{\phi}}_{\text{FMCA}} - \hat{\boldsymbol{\phi}}_{\text{TRUE}} = -\mathbf{A}^{-1}\mathbf{b} \qquad (63)$$

where

$$\mathbf{A} = \text{Re} \left\{ \tilde{\mathbf{J}}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})\tilde{\mathbf{J}}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) \right\} \qquad (64)$$
$$\mathbf{b} = \text{Re} \left\{ \tilde{\mathbf{J}}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})[\mathbf{L}^\dagger]_a \mathbf{R} \left(\left[e^{-j\boldsymbol{\phi}_{\text{TRUE}}}; 1 \right] \right) \tilde{\mathbf{g}} \right\}. \qquad (65)$$

Using the low-return region constraint

$$[\mathbf{L}^\dagger]_a \mathbf{R} \left(\left[e^{-j\boldsymbol{\phi}_{\text{TRUE}}}; 1 \right] \right) \mathbf{g} = \mathbf{0} \qquad (66)$$

and explicitly writing out the signal term and the random noise term in $\tilde{\mathbf{g}}$, we get

$$\mathbf{b} = \text{Re} \left\{ \tilde{\mathbf{J}}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})[\mathbf{L}^\dagger]_a \mathbf{R} \left(\left[e^{-j\boldsymbol{\phi}_{\text{TRUE}}}; 1 \right] \right) \mathbf{n} \right\}. \qquad (67)$$

Using the linearity of the Jacobian in $\mathbf{Y}_2$ and the noise term, we find

$$\tilde{\mathbf{J}}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}}) = \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}}) + \mathbf{Q}(\mathbf{n}) \qquad (68)$$

where $\mathbf{Q}(\mathbf{n})$ is a matrix that linearly depends on $\mathbf{n}$ and

$$\mathbf{J}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) = -j[\mathbf{L}^\dagger]_a \mathbf{Y}_2 \, \text{Diag}(e^{-j\boldsymbol{\phi}_{\text{TRUE}}}). \qquad (69)$$

Neglecting second and higher order noise terms, we get

$$\mathbf{b} \approx \text{Re} \left\{ \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})[\mathbf{L}^\dagger]_a \mathbf{R} \left(\left[e^{-j\boldsymbol{\phi}_{\text{TRUE}}}; 1 \right] \right) \mathbf{n} \right\}. \qquad (70)$$

Similarly, linearizing the inverse about the true parameters yields

$$\mathbf{A}^{-1} \approx \left[\text{Re} \left\{ \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})\mathbf{J}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) \right\} \right]^{-1} + \bar{\mathbf{Q}}(\mathbf{n}) \qquad (71)$$

where $\bar{\mathbf{Q}}(\mathbf{n})$ is a matrix that linearly depends on $\mathbf{n}$. Neglecting second and higher order noise terms, we get

$$\hat{\boldsymbol{\phi}}_{\text{FMCA}} - \hat{\boldsymbol{\phi}}_{\text{TRUE}} \approx - \left[\text{Re} \left\{ \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})\mathbf{J}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) \right\} \right]^{-1}$$
$$\times \text{Re} \left\{ \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})[\mathbf{L}^\dagger]_a \mathbf{R} \left(\left[e^{-j\boldsymbol{\phi}_{\text{TRUE}}}; 1 \right] \right) \mathbf{n} \right\}. \qquad (72)$$

Finally, by using the identity

$$E \left[\text{Re}\{\mathbf{Cn}\} \text{Re}\{\mathbf{Cn}\}^T \right] = \frac{\sigma_n^2}{2} \text{Re}\{\mathbf{CC}^H\} \qquad (73)$$

where $\mathbf{C}$ is any matrix with MN columns, we have

$$E \left[(\hat{\boldsymbol{\phi}}_{\text{FMCA}} - \boldsymbol{\phi}_{\text{TRUE}})(\hat{\boldsymbol{\phi}}_{\text{FMCA}} - \boldsymbol{\phi}_{\text{TRUE}})^T \right]$$
$$\approx \frac{\sigma_n^2}{2} \left[\text{Re} \left\{ \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})\mathbf{J}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) \right\} \right]^{-1}$$
$$\times \text{Re} \left\{ \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})[\mathbf{L}^\dagger]_a [\mathbf{L}^\dagger]_a^H \mathbf{J}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) \right\}$$
$$\times \left[\text{Re} \left\{ \mathbf{J}_{\mathbf{f}}^H(\boldsymbol{\phi}_{\text{TRUE}})\mathbf{J}_{\mathbf{f}}(\boldsymbol{\phi}_{\text{TRUE}}) \right\} \right]^{-1}. \qquad (74)$$

Now, we can compare the MSE of MLA phase estimation in (52), which is also the CRB, with the MSE of FMCA phase estimation in (74). The errors are plotted in Fig. 1, where it is easy to see that FMCA is significantly more sensitive to noise. We validated this small error analysis using Monte Carlo simulation of the estimators. The setup for this simulation was: $\boldsymbol{\theta}$ was a $D = 100$ dimensional random uniformly distributed complex vector, and $\boldsymbol{\phi}$ was independent and uniformly distributed between $-\pi$ and π, except for $\phi_M = 0$. The sampling matrix $\mathbf{L}$ was generated by using an imaging scenario where the SAR operated across a $1°$ viewing angle and transmitted 15 pulses ($M = 15$). The receiver provided 15 samples per single pulse ($N = 15$). We applied a rectangular antenna pattern on $\boldsymbol{\theta}$ so that it was known *a priori* that $\theta_a = 0$ for $a = \{1, \ldots, 10, 90, \ldots, 100\}$. The additive noise was complex Gaussian, with a signal-to-noise ratio (SNR) defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{1}{\sigma_n KL} \sum_{k,l} \left| \tilde{G}[k,l] \right| \right). \qquad (75)$$

The experimental results plotted in Fig. 1 used SDR to solve both the MLA and FMCA phase estimation problems described by (36) and (38), respectively.

VI. SIMULATION RESULTS

In this section, we present numerical results to illustrate the advantages of our proposed methods in realistic SAR systems.

A. SAR Simulator

In order to test the different algorithms, we built a SAR simulator based on the bilinear model described in Section II. The amplitudes of $\boldsymbol{\theta}$ were taken from actual SAR images, whereas the phases of $\boldsymbol{\theta}$ were independently generated according to a uniform distribution between $-\pi$ and π (not to confuse with the autofocus phase error). We adopted the impulse model (2) with $dx = dy = 1$ for our simulation. The nuisance autofocus
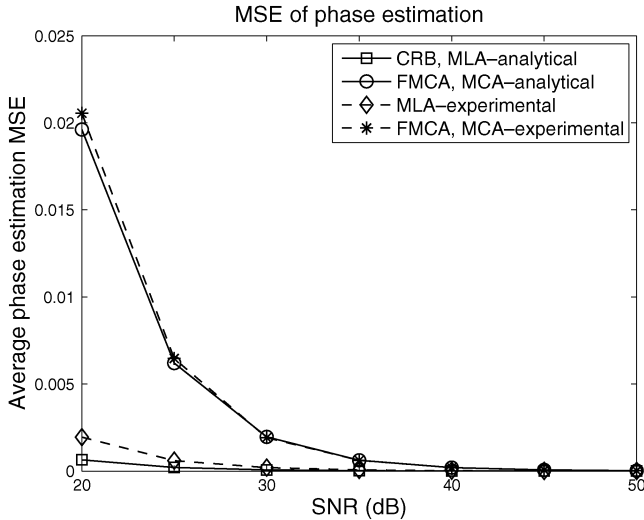
Fig. 1. MSE for MLA phase estimation compared with FMCA and MCA phase estimation under small noise (y axis shows $(1/M - 1)\sum_i (\hat{\phi}(i) - \phi_{\text{TRUE}}(i))^2$).

phases were also independently generated according to a uniform distribution. Note that this distribution is known to be the most challenging phase corruption. Subsequent figures present the magnitude of the complex reconstructed reflectivity functions (which are the magnitudes of $\boldsymbol{\theta}$ in the impulse model).

B. Numerical Approximation to CMQP

In the first set of experiments, we compare the accuracy and computational complexity of the three numerical approximations to CMQP. First, we synthetically construct random matrices $\mathbf{A}$ where we can change the dimensions and noise level in a systematic and flexible manner. Specifically, we define

$$\boldsymbol{\gamma} = e^{j\boldsymbol{\phi}}$$
$$\mathbf{P} = \mathbf{I} - \frac{\boldsymbol{\gamma}\boldsymbol{\gamma}^H}{\|\boldsymbol{\gamma}\|_2}$$
$$\mathbf{A} = \boldsymbol{\Psi}\mathbf{P} + \mathbf{W} \tag{76}$$

where $\boldsymbol{\phi}$ is a uniform *i.i.d.* phase vector, $\boldsymbol{\Psi}$ is a matrix of random uniformly distributed complex numbers, and $\mathbf{W}$ is a complex additive white Gaussian noise matrix. Note that this construction guarantees that $\boldsymbol{\gamma}$ will be the noiseless solution to (44) with a zero objective value.

Fig. 2 compares performance of SCA with SDR and EVR. Here, SDR was implemented using the SeDuMi optimization toolbox [27]. Fig. 2(a) shows the MSE of the phase estimates (phase MSE) produced by SCA, SDR, and EVR under different SNRs. It is obvious that SCA performs much closer to SDR than does EVR. Fig. 2(b) shows the minimal objective function value of (44) obtained by each of the methods under different SNRs. We can see that SDR has the best performance; however, SCA tends to close the gap, particularly at higher SNRs. Fig. 2(c) shows the performance dependence on the dimension of $\boldsymbol{\phi}$. As expected, SDR and SCA provide better scalability than EVR, which shows a rapid increase in error when the dimension of $\boldsymbol{\phi}$ increases. Fig. 2(d) illustrates the computational complexity of the three methods by measuring the amount of time

for each method to solve the minimization problem under a constant SNR and rising dimensionality of $\boldsymbol{\phi}$. The time (vertical) axis of the graph is logarithmic. We can see that the complexity of SDR grows more rapidly than for SCA and EVR. This simulation suggests that SCA provides a good compromise between EVR and SDR.

Next, in order to demonstrate the advantages in terms of visual image quality, we test the performance of SCA when used to implement the MLA phase estimator. We used an artificial image shown in Fig. 3(a) and the following imaging scenario: the SAR radar operated across a $2°$ range of look angles and transmitted 50 pulses ($M = 50$). The receiver provided 50 samples per single pulse ($N = 50$). Complex Gaussian noise, with SNR = 15 dB, was added to the perfectly focused image. Here, SDR was implemented using the algorithm in [28]. Fig. 3 shows the image reconstructed by various methods. Fig. 3(b) shows the reconstruction by MLA, using SDR (MLA-SDR), whereas the reconstructions by MLA-SCA and MLA-EVR are shown in Fig. 3(c) and (d), respectively. We can easily see the improved image enhancement of MLA-SDR and MLA-SCA over MLA-EVR. The improved performance is also apparent in the MSEs of the autofocus phase estimates, which decreased from 0.7364 for MLA-EVR to 0.2103 for MLA-SCA and 0.1096 for MLA-SDR. It took 0.1263 s using MATLAB on a standard PC for MLA-SDR to find the phase estimates and only 0.0054 s for MLA-SCA and MLA-EVR.

C. Performance of MLA

In the second set of experiments, we start with showing the full strength of MLA by comparing it with PGA and sharpness-maximization methods. Hereafter, MLA was implemented with SDR using the interior-point method developed in [28]. A SAR image, shown in Fig. 4(a), was obtained from Sandia National Laboratory. We adopted a scenario where the radar transmitted 50 pulses ($M = 50$) and the receiver provided 50 samples per single pulse ($N = 50$). The hypothetical SAR operated across a $2°$ range of look angles. No prior information about the image was used by the MLA estimator, i.e., MLA considered the image to be arbitrary. For the sharpness-maximization method, we used negative entropy as the sharpness metric. The perfectly focused image with additive noise of SNR = 10 dB is shown in Fig. 4(a). Because no low-return region can be found, FMCA cannot be applied here. The MLA reconstructed image is shown in Fig. 4(c) with a phase MSE of 0.0148. The MLA reconstructed image for SNR = 5 dB is shown in Fig. 4(d) with a phase MSE of 0.1728. The PGA image restoration for SNR = 10 dB and SNR = 5 dB are shown in Fig. 4(e) and (f), respectively. The sharpness-maximization method image restoration for SNR = 10 dB and SNR = 5 dB are shown in Fig. 4(g) and (h) with phase MSE 0.7815 and 0.9346, respectively. Note that, since PGA only produces a reconstructed image, no phase MSE is computed.

D. Comparison of MLA With FMCA

In the third set of experiments, we compared MLA with FMCA. We adopted a SAR scenario with a wide range of look angles, which is more challenging than the narrow range of look angles considered earlier. We used a SAR image obtained from the MSTAR SAR database [29]. The focused image is
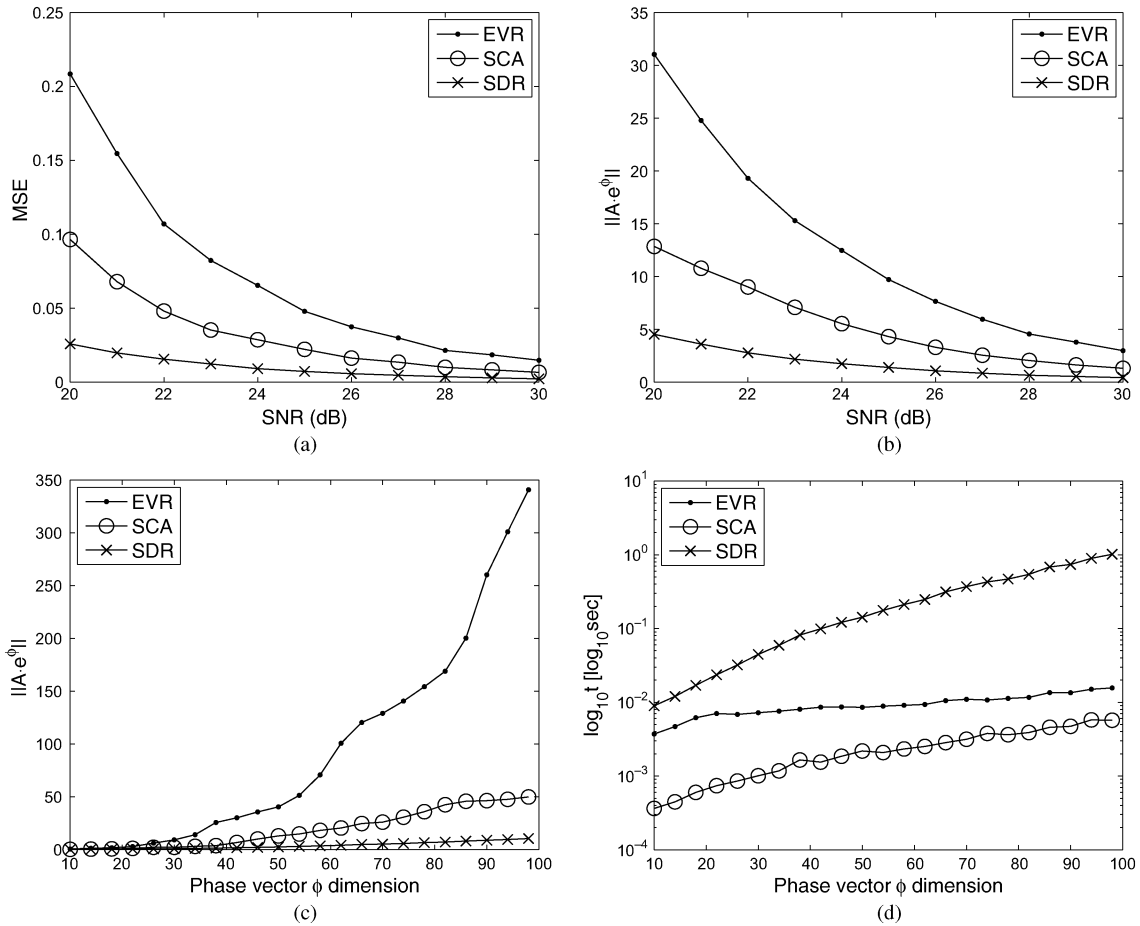
Fig. 2.   Performance of SCA compared with SDR and EVR: (a) MSE of optimal solution found by SCA, SDR, and EVR as a function of SNR; (b) optimal objective function value found by SCA, SDR, and EVR as a function of SNR; (c) optimal objective function value found by SCA, SDR, and EVR as a function of problem size; and (d) computational time required by SCA, SDR, and EVR as a function of problem size.

shown in Fig. 5(a). We applied a rectangular antenna pattern to the image so that the first and last columns were zero, i.e., a known index set a that indexes the low-return region was known *a priori* to FMCA so that $\theta_a = 0$. For MLA, the set a effectively reduced the dimension for $\boldsymbol{\theta}$. Note that knowledge of the antenna pattern (low-return region) is required only by FMCA, not MLA. We adopted an imaging scenario where the radar was collecting data across $6°$ and $M = N = 80$. The phase-corrupted image is shown in Fig. 5(b). The images restored by MLA for SNR $= 10$ dB and SNR $= 5$ dB are shown in Fig. 5(c) and (d) with phase MSE equal to 0.3697 and 0.5521, respectively. The images restored by FMCA for SNR $= 10$ dB and SNR $= 5$ dB are shown in Fig. 5(e) and (f) with phase MSE equal to 0.6690 and 0.8723, respectively.

E. Robustness Against Model Mismatching

In this last set of experiments, we examine the robustness of MLA against model mismatching. MLA is derived from the proposed parametric model, and we have analytically and experimentally demonstrated its advantages under the proposed model. However, FMCA does not explicitly assume such a model, and it is important to investigate the performance of MLA when there is a model mismatch. We continue with a similar setup, and the estimators are implemented as before.
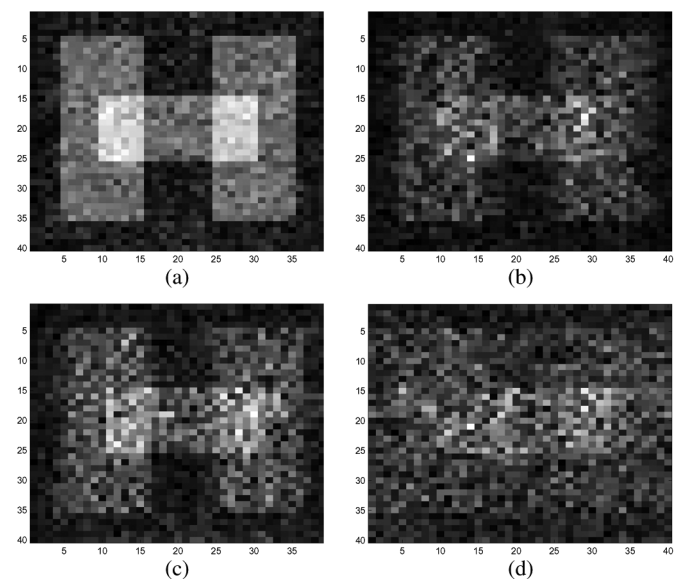

Fig. 3.   Restored image using MLA-SDR, MLA-SCA, and MLA-EVR: (a) Focused image with SNR $= 15$ dB; (b) restoration using MLA-SDR (phase MSE $= 0.1096$, time $= 0.126304$ s); (c) restoration using MLA-SCA (phase MSE $= 0.2103$, time $= 0.005472$ s); and (d) restoration using MLA-EVR (phase MSE $= 0.7364$, time $= 0.005481$ s).
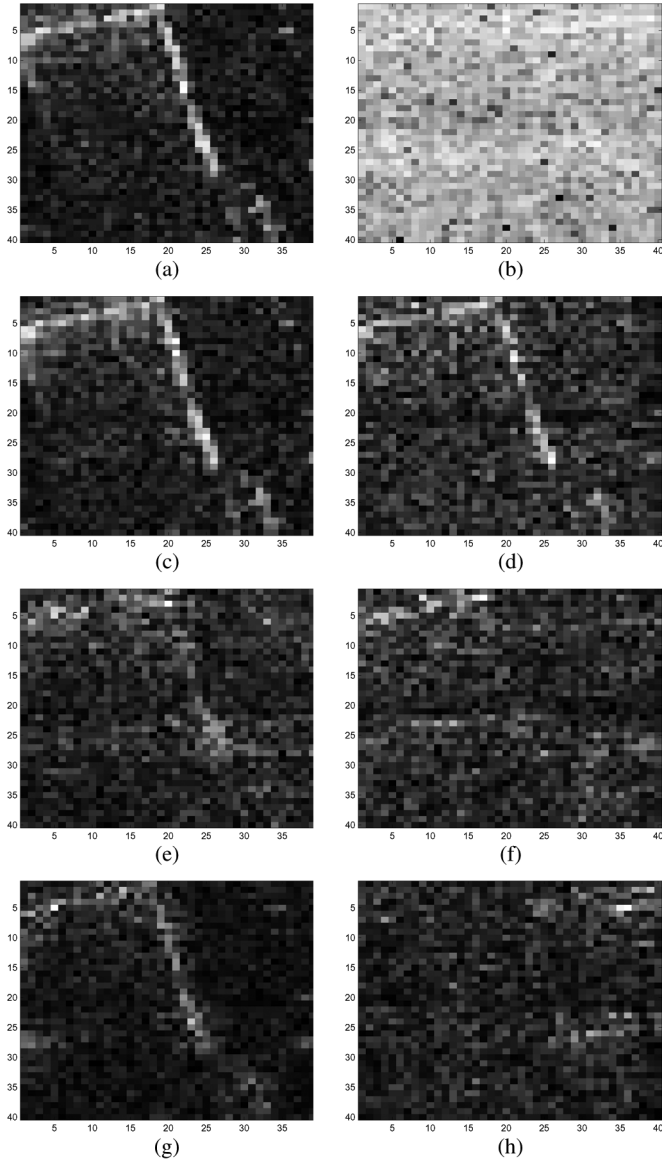
Fig. 4. Image restoration using MLA, PGA, and sharpness-maximization method: (a) Focused image with SNR = 10 dB; (b) corrupted image using an *i.i.d.* phase error function and SNR = 10 dB; (c) MLA restoration for SNR = 10 dB (phase MSE = 0.0148); (d) MLA restoration for SNR = 5 dB (phase MSE = 0.1728); (e) PGA restoration for SNR = 10 dB; (f) PGA restoration for SNR = 5 dB; (g) sharpness-maximization restoration for SNR = 10 dB (phase MSE = 0.7815); and (h) sharpness-maximization restoration for SNR = 5 dB (phase MSE = 0.9346).

However, the SAR simulator no longer uses the naive impulse basis functions in (2) but replaces them with more realistic functions

$$h_i(x, y) = k(x - u_i, y - v_i). \qquad (77)$$

This has the effect of further corrupting the collected Fourier data in the following way:

$$G'[m, n] = G[m, n] \cdot K(F_x[m, n], F_y[m, n]) \qquad (78)$$

where G' denote the actual collected data, and K denotes the Fourier transform of the kernel function k. The autofocus algorithms are not aware of this setup and mistakenly compute $\mathbf{L}$ using the model in (2).
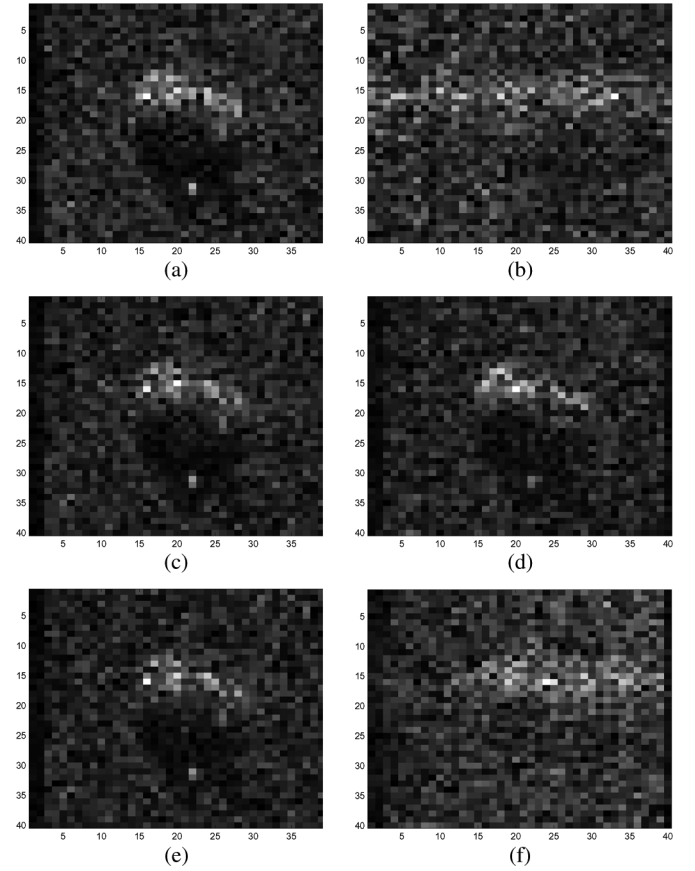

Fig. 5. Image restoration using MLA and FMCA: (a) Focused image with SNR = 5 dB; (b) corrupted image using an *i.i.d.* phase error function and SNR = 5 dB; (c) MLA restoration for SNR = 10 dB (phase MSE = 0.3697); (d) MLA restoration for SNR = 5 dB (phase MSE = 0.5521); (e) FMCA restoration for SNR = 10 dB (phase MSE = 0.6690); and (f) FMCA restoration for SNR = 5 dB (phase MSE = 0.8723).

First, we adopted a rectangular kernel

$$k(x, y) = \begin{cases} 1, & |x| < 1 \text{ and } |y| < 1 \\ 0, & \text{otherwise.} \end{cases} \qquad (79)$$

This illustrates a model where there is a constant reflectivity within a resolution cell instead of an impulse located at the center of the cell. The focused images formed using the exact and mismatched models are shown in Fig. 6(a) and (b), respectively. The images restored by MLA and FMCA for the model mismatched image are shown in Fig. 6(c) and (d), respectively. The autofocus phase error was an *i.i.d.* function and SNR = 5 dB.

Second, we adopted a 2-D Gaussian kernel with

$$k(x, y) = e^{-\left(\frac{x^2}{2} + \frac{y^2}{2}\right)}. \qquad (80)$$

This illustrates a model where the sum of reflectivity within a resolution cell also affects neighboring cells. The focused images formed by the exact and mismatched models are shown in Fig. 7(a) and (b), respectively. The images restored by MLA and FMCA for the model mismatched data are shown in Fig. 7(c) and (d), respectively. The autofocus phase error was an *i.i.d.* function and SNR = 5 dB.
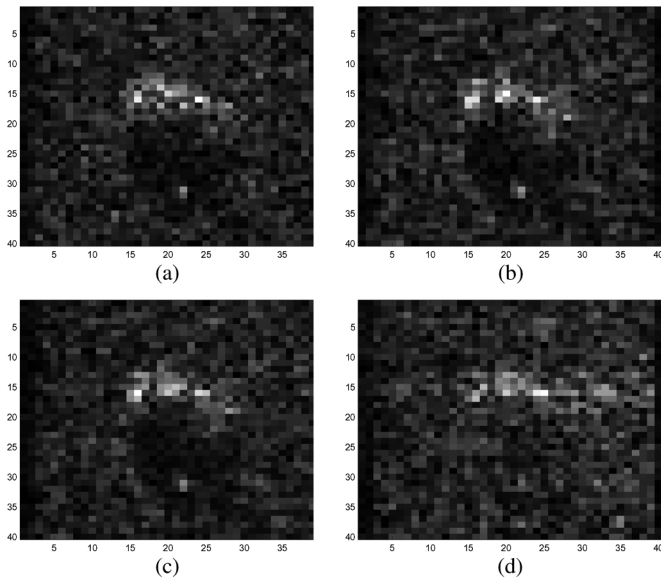
Fig. 6. Model mismatching using rectangular kernel (SNR = 5 dB): (a) Focused image using exact model; (b) focused image using mismatched model; (c) MLA restoration (phase MSE = 0.2973); and (d) FMCA restoration (phase MSE = 0.8572).
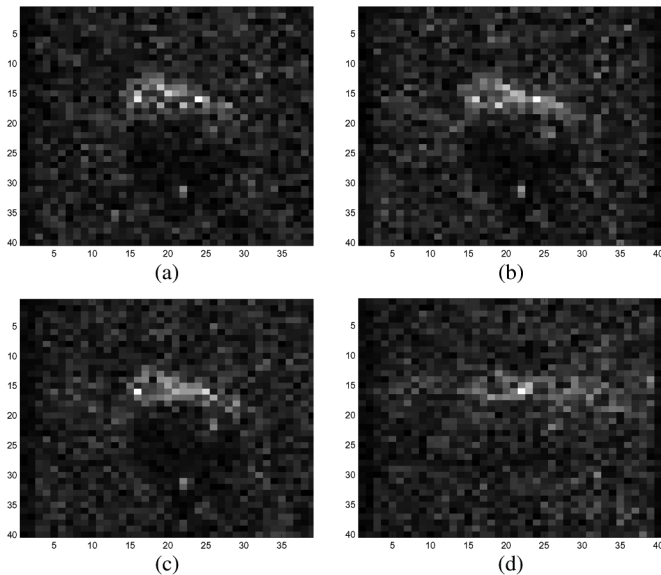

Fig. 7. Model mismatching using Gaussian kernel (SNR = 5 dB): (a) Focused image using exact model; (b) focused image using mismatched model; (c) MLA restoration (phase MSE = 0.3205); and (d) FMCA restoration (phase MSE = 0.7237).

Based on these simulations, we conclude that MLA can outperform FMCA even when there is model mismatching.

VII. DISCUSSION

In this paper, we considered the problem of SAR autofocus based on a bilinear parametric model. We derived the MLA framework and compared it with previous methods. Under simplistic conditions, MLA coincides with the successful MCA technique. In more realistic conditions, MLA outperforms FMCA and is applicable to a broader class of scenarios. As a byproduct, we also considered efficient numerical approximations to the CMQP problem, which lies at the core of all these algorithms.

An important direction for future research involves the reduction of computational complexity. Typical SAR images involve a huge number of pixels and require efficient numerical methods for their reconstruction. In practice, pseudo-inversion of the matrix **L** is typically implemented via linear interpolations and FFTs. Similar ideas should be applied to the autofocus problem and the CMQP approximations for successful application to large images.

An additional direction for future work concerns the validity of the linear reflectivity function model. A more realistic formulation might model the amplitudes and the phases of this function separately as done in [22].
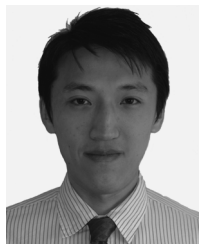
ACKNOWLEDGMENT

REFERENCES

[1] C. V. Jakowatz, Jr., D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach.*. Boston, MA: Kluwer, 1996.

[2] W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*. Boston, MA: Arthech House, 1995.

[3] D. C. Munson, Jr., J. D. O'Brien, and W. K. Jenkins, "A tomographic formulation of spotlight-mode synthetic aperture radar," *Proc. IEEE*, vol. 71, no. 8, pp. 917–925, Aug. 1983.

[4] D. C. Munson, Jr. and J. L. C. Sanz, "Image reconstruction from frequency-offset Fourier data," *Proc. IEEE*, vol. 72, no. 6, pp. 661–669, Jun. 1984.

[5] G. A. Bendor and T. W. Gedra, "Single-pass fine-resolution SAR autofocus," in *Proc. IEEE NAE-CON*, Dayton, OH, May 1983, pp. 482–488.

[6] T. C. Calloway and G. Donohoe, "Subaperture autofocus for synthetic aperture radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 2, pp. 617–621, Apr. 1994.

[7] P. Eichel and C. V. Jakowatz, Jr., "Phase-gradient algorithm as an optimal estimator of the phase derivative," *Opt. Lett.*, vol. 14, no. 20, pp. 1101–1103, Oct. 1989.

[8] L. Xi, L. Guosui, and J. Ni, "Autofocusing of ISAR images based on entropy minimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 4, pp. 1240–1252, Oct. 1999.

[9] J. R. Fienup and J. J. Miller, "Aberration correction by maximizing generalized sharpness metrics," *Opt. Soc. Amer.*, vol. 20, no. 4, pp. 609–620, Apr. 2003.

[10] R. L. Morrison, Jr., M. N. Do, and D. C. Munson, Jr., "MCA: A multichannel approach to SAR autofocus," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 840–853, Apr. 2009.

[11] K. Liu and D. C. Munson, Jr., "Fourier-domain multichannel autofocus for synthetic aperture radar," in *Proc. 42nd Asilomar Conf. Signals, Syst. Comput.*, Oct. 2008, pp. 848–852.

[12] K. Liu and D. C. Munson, Jr., "Fourier-domain multichannel autofocus for synthetic aperture radar," *IEEE Trans. Image Process*, vol. 20, no. 12, pp. 3544–3552, Dec. 2011, 10.1109/TIP.2011.2156421.

[13] V. Patel, G. R. Easley, D. M. Healy, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 244–254, Apr. 2010.

[14] I. Stojanovic and W. Karl, "Imaging of moving targets with multi-static SAR using an overcomplete dictionary," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 164–176, Feb. 2010.

[15] M. Cetin, W. Karl, and D. Castanon, "Feature enhancement and ATR performance using nonquadratic optimization-based SAR imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1375–1395, Oct. 2003.

[16] M. Cetin and W. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 623–631, Apr. 2001.

[17] S. Verdu, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, pp. 303–312, 1989.

[18] A. M. So, J. Zhang, and Y. Ye, "On approximating complex quadratic optimization problems via semidefinite programming relaxations," *J. Math. Programming, Ser. A & B*, vol. 110, no. 1, pp. 93–110, Mar. 2007.

[19] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problem using semi-definite programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995.

[20] K. Liu, A. Wiesel, and D. C. Munson, Jr., "Synthetic aperture radar autofocus via semidefinite relaxation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 1342–1345.

[21] K. Liu, A. Wiesel, and D. C. Munson, Jr., "Maximum likelihood SAR autofocus with low-return region," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2011, pp. 1377–1380.

[22] S. N. Madsen, "Spectral properties of homogeneous and nonhomogeneous radar images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 4, pp. 583–588, Jul. 1987.

[23] C. Helmberg and F. Rendl, "A spectral bundle method for semidefinite programming," *SIAM J. Optim.*, vol. 10, no. 3, pp. 673–696, 1997.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2004.

[25] G. Strang, *Linear Algebra and its Applications.*. Orlando, FL: Brace Jovanovich College Publ., 1988.

[26] S. M. Kay, *Fundamentals of statistical signal processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[27] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, no. 1, pp. 625–653, 1999.

[28] W. Ma, P. Ching, and Z. Ding, "Semidefinite relaxation based multiuser detection for M-ary PSK multiuser systems," *IEEE J. Selected Areas Commun.*, vol. 52, no. 10, pp. 2862–2872, Oct. 2004.

[29] Center for imaging science image database [Online]. Available: http://cis.jhu.edu/data.sets/MSTAR

Kuang-Hung Liu (M'08) received the B.S. degree in computer science and information engineering from National Taipei University of Technology, Taipei, Taiwan, in 2003 and the M.S. and Ph.D. degrees in electrical engineering from the University of Michigan at Ann Arbor, in 2007 and 2011, respectively.

He is currently a Research Scientist with Schlumberger WesternGeco, Houston, TX.

Ami Wiesel (M'07) received the B.Sc. and M.Sc. degrees in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2007.

He was a Postdoctoral Fellow with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, in 2007–2009. Since January 2010, he has been a Senior Lecturer with the School of Computer Science and Engineering, Hebrew University of Jerusalem, Israel.

Dr. Wiesel was a recipient of the Young Author Best Paper Award for a 2006 paper in the IEEE Transactions on Signal Processing and a Student Paper Award for the 2005 Workshop on Signal Processing Advances in Wireless Communications (SPAWC) paper. He was awarded the Weinstein Study Prize in 2002, the Intel Award in 2005, the Viterbi Fellowship in 2005 and 2007, and the Marie Curie Fellowship in 2008.

David C. Munson, Jr., (S'74–M'79–SM'84–F'91) received the B.S. degree from the University of Delaware, Newark, Delaware, in 1975 and the M.S., M.A., and Ph.D. degrees from Princeton University, Princeton, NJ, in 1978, 1978, and 1979, respectively.

He is the Robert J. Vlasic Dean of Engineering and Professor of Electrical Engineering and Computer Science at the University of Michigan, Ann Arbor. He has served as dean since 2006. He served as Chair of Electrical Engineering and Computer Science from 2003–2006. He served on the faculty of the University of Illinois from 1979–2003, where he was the Robert MacClinchie Distinguished Professor of Electrical and Computer Engineering.

Dr. Munson conducts research in the area of signal and image processing, with a specialty in synthetic aperture radar. He is a Fellow of the IEEE, a past president of the IEEE Signal Processing Society, founding editor-in-chief of the IEEE Transactions on Image Processing, and co-founder of the IEEE International Conference on Image Processing. In addition to multiple teaching awards and other honors, he was presented the Society Award of the IEEE Signal Processing Society, and he was the Texas Instruments Distinguished Visiting Professor at Rice University. He is coauthor of multiple textbooks, including "Engineering Our Digital Future," which is introducing engineering to hundreds of high schools nationwide via the Infinity Project. He is co-founder of InstaRecon, Inc., which is commercializing fast algorithms for image formation in computer tomography.