

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Sparse ACEKF for phase reconstruction.

### Permalink

<https://escholarship.org/uc/item/9n40x9b1>

### Journal

Optics Express, 21(15)

### ISSN

1094-4087

### Authors

Jingshan, Zhong  
Dauwels, Justin  
Vázquez, Manuel A  
[et al.](#)

### Publication Date

2013-07-29

### DOI

10.1364/oe.21.018125

Peer reviewed

# Sparse ACEKF for phase reconstruction

Zhong Jingshan,<sup>1,\*</sup> Justin Dauwels,<sup>1</sup> Manuel A. Vázquez,<sup>2</sup> and Laura Waller<sup>3</sup>

<sup>1</sup>Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore 639798

<sup>2</sup>Universidad Carlos III de Madrid, Depto. de Teoría de la Señal y Comunicaciones, Madrid, Spain.

<sup>3</sup>Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

\*[zhongjingshan@hotmail.com](mailto:zhongjingshan@hotmail.com)

**Abstract:** We propose a novel low-complexity recursive filter to efficiently recover quantitative phase from a series of noisy intensity images taken through focus. We first transform the wave propagation equation and nonlinear observation model (intensity measurement) into a complex augmented state space model. From the state space model, we derive a sparse augmented complex extended Kalman filter (ACEKF) to infer the complex optical field (amplitude and phase), and find that it converges under mild conditions. Our proposed method has a computational complexity of  $N_z N \log N$  and storage requirement of  $\mathcal{O}(N)$ , compared with the original ACEKF method, which has a computational complexity of  $\mathcal{O}(N_z N^3)$  and storage requirement of  $\mathcal{O}(N^2)$ , where  $N_z$  is the number of images and  $N$  is the number of pixels in each image. Thus, it is efficient, robust and recursive, and may be feasible for real-time phase recovery applications with high resolution images.

© 2013 Optical Society of America

**OCIS codes:** (100.5070) Phase retrieval; (100.3010) Image reconstruction techniques.

---

## References and links

1. L. Waller, M. Tsang, S. Ponda, S. Yang, and G. Barbastathis, "Phase and amplitude imaging from noisy images by Kalman filtering," *Opt. Express* **19**, 2805–2814 (2011).
2. L. Waller, Y. Luo, S. Y. Yang, and G. Barbastathis, "Transport of intensity phase imaging in a volume holographic microscope," *Opt. Lett.* **35**, 2961–2963 (2010).
3. S. S. Gorthi and E. Schonbrun, "Phase imaging flow cytometry using a focus-stack collecting microscope," *Opt. Lett.* **37**, 707–709 (2012).
4. R. Gerchberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane picture," *Optik* **35**, 237–246 (1972).
5. J. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* **21**, 2758–2769 (1982).
6. G. Pedrini, W. Osten, and Y. Zhang, "Wave-front reconstruction from a sequence of interferograms recorded at different planes," *Opt. Lett.* **30**, 833–835 (2005).
7. N. Streibl, "Phase imaging by the transport equation of intensity," *Opt. Commun.* **49**, 6–10 (1984).
8. M. Soto and E. Acosta, "Improved phase imaging from intensity measurements in multiple planes," *Appl. Opt.* **46**, 7978–7981 (2007).
9. L. Waller, L. Tian, and G. Barbastathis, "Transport of intensity phase-amplitude imaging with higher order intensity derivatives," *Opt. Express* **18**, 12552–12561 (2010).
10. R. Paxman, T. Schulz, and J. Fienup, "Joint estimation of object and aberrations by using phase diversity," *J. Opt. Soc. Am. A* **9**, 1072–1085 (1992).
11. R. Paxman and J. Fienup, "Optical misalignment sensing and image reconstruction using phase diversity," *J. Opt. Soc. Am. A* **5**, 914–923 (1988).

12. Z. Jingshan, J. Dauwels, M. A. Vázquez, and L. Waller, "Efficient Gaussian inference algorithms for phase imaging," Proc. IEEE ICASSP 2012 pp. 25–30.
  13. J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 1996).
  14. D. H. Dini and D. P. Mandic, "Class of widely linear complex Kalman filters," IEEE Trans. Neural Netw. Learn. Syst. **23**, 775–786 (2012).
  15. R. E. Kalman, "A new approach to linear filtering and prediction problems," Journal of basic Engineering **82**, 35–45 (1960).
  16. K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the discrete-time extended Kalman filter," IEEE Trans. Autom. Control **44**, 714–728 (1999).
  17. T. Lefebvre, H. Bruyninckx, and J. D. Schutter, "Kalman filters for nonlinear systems: a comparison of performance," Internat. J. Control **77**, 639–653 (2004).
  18. A. J. Krener, "The convergence of the extended Kalman filter," in "Directions in mathematical systems theory and optimization," (Springer, 2003), pp. 173–182.
  19. D. J. Lee, M. C. Roggemann, and B. M. Welsh, "Cramér-Rao analysis of phase-diverse wave-front sensing," J. Opt. Soc. Am. A **16**, 1005–1015 (1999).
- 

## 1. Introduction

When light propagates, its amplitude and phase evolve according to the wave equation. These amplitude and phase perturbations contain important information about the complex-field in the original (focal) plane. Intensity can be measured easily; however, phase cannot be measured directly and it needs to be reconstructed from intensity images with computational methods. Most solutions for phase recovery do not explicitly account for noise in the measurement, and thus typical methods are not applicable in high noise situations, such as in X-ray imaging or photon-starved situations. Recently, Kalman filters have been used to estimate phase from very noisy intensity images (signal-to-noise ratio less than one) [1]; however the method is very computationally intensive and impractical for images with large pixel counts (e.g. 1 megapixel or greater). In this paper, we solve these problems by developing efficient inference methods for recovering the phase from a set of defocus intensity images, which can be captured either sequentially by moving camera along the optical axis, or all at once by using various focal stack collection methods [2, 3]. Such methods can be useful for applications in medical imaging, neuroscience, and materials science, where low noise images are difficult to obtain. Furthermore, since our proposed method is recursive (not iterative), it has the ability to estimate phase in real-time during the measurement sequence, progressively improving the estimate as each new image is captured.

Traditional methods for recovering phase involve complicated interferometric setups, so there is a significant experimental advantage to methods such as ours, which only require a set of intensity images taken with different amounts of defocus. One of the originally proposed algorithms, the Gerchberg-Saxton(GS) method [4, 5], uses only two images and treats the problem as convex (which it is generally not), iterating back and forth between two domains to reduce error. In [6], a generalization of the GS method is given for cases of multiple defocus images. These methods tend to be strongly sensitive to the noise in the last image, although techniques exist for averaging at each iteration. Direct methods generally linearize the problem, either by assuming that the object is weak or that the propagation distance is small. It allows one to exploit the Transport of Intensity Equation (TIE) [7], which reveals relationship between phase and the first derivative of intensity with respect to the optical axis. The first derivative of intensity is usually approximated by finite difference method [8, 9], and it is not robust to severe noise. A few statistical approaches have been proposed as well. An approximation to the maximum likelihood estimator is derived in [10, 11]. However, it easily gets stuck in local maxima, and sometimes leads to poor results.

In [1], an augmented complex extended Kalman filter (ACEKF) was used to solve for phase in the presence of significant noise corruption. The Kalman filter is a well-known recursive

algorithm for estimating dynamically changing quantities from a set of noisy measurements. In the linear case, with Gaussian noise statistics, the Kalman filter is the optimal estimation method. However, intensity is a nonlinear measure of complex field, and noise in images generally follows Poisson statistics, so the solution will not be provably optimal. Extended Kalman filters add a linearization step in order to account for nonlinear measurements. Further, we use a complex EKF in order to deal with complex numbers. The resulting ACEKF outperforms many previous algorithms, at the cost of very high computational complexity of  $\mathcal{O}(N_z N^3)$  and storage requirement of  $\mathcal{O}(N^2)$ . The memory storage requirements are particularly limiting - for example, an image with pixel counts on the order of megapixels will require terabytes of data to be stored and processed - thus, the algorithm is infeasible for large images and impractical for real-time applications. In [12], a diagonalized complex extended Kalman filter (diagonalized CEKF) was proposed to alleviate those issues, without jeopardizing the reconstruction accuracy. The diagonalized CEKF is iterative: it needs to cycle through the set of intensity images repeatedly, yielding a more accurate phase reconstruction after each cycle. The computational complexity increases with each cycle.

The sparse ACEKF proposed here uses the same augmented state space model as ACEKF. With the assumption that the phase is small, we derive theorems to simplify the update equations of ACEKF. It eventually results in a low-complexity noise-robust phase reconstruction algorithm. The proposed method has a computational complexity of  $N_z N \log N$ , while ACEKF requires  $N_z N^3$ . However, our method still achieves better phase reconstruction than both ACEKF and the diagonalized CEKF.

This paper is organized as follows. In the next section, we turn the propagation and observation models from scalar wave optics into an augmented state space model in signal processing. In Section 3, we derive the proposed sparse ACEKF algorithm. In Section 4, we analyze the convergence and stability of the new ACEKF. In Section 5, we compare results of different methods using synthetic and experimental data. We offer concluding remarks in Section 6.

## 2. Problem description and state space model of the optical field

We aim to estimate the 2D complex-field  $A(x, y, z_0)$  at the focal plane  $z_0$ , from a sequence of noisy intensity images  $I(x, y, z)$  captured at various distance  $z_0, z_1, z_2, \dots, z_{N_z}$ . We assume a linear medium with homogeneous refractive index and coherent (laser) illumination, such that the complex-field at  $z_0$  fully determines the complex-field at all other planes. The complex optical field at  $z$  is  $A(x, y, z) = |A(x, y, z)|e^{i\phi(x, y, z)}$ , where  $|A(x, y, z)|$  is the amplitude, and  $\phi(x, y, z)$  is the phase. Propagation is modeled by the homogeneous paraxial wave equation:

$$\frac{\partial A(x, y, z)}{\partial z} = \frac{i\lambda}{4\pi} \nabla_{\perp}^2 A(x, y, z), \quad (1)$$

where  $\lambda$  is the wavelength of the illumination, and  $\nabla_{\perp}$  is the gradient operator in the lateral  $(x, y)$  dimensions. The noisy measurements  $I(x, y, z)$  usually adhere to a (continuous) Poisson distribution:

$$p[I(x, y, z) | A(x, y, z)] = e^{-\gamma |A(x, y, z)|^2} \frac{(\gamma |A(x, y, z)|^2)^{I(x, y, z)}}{I(x, y, z)!}, \quad (2)$$

where  $\gamma$  is the photon count detected by the camera. The measurement at each pixel  $I(x, y, z)$  is assumed statistically independent of any other pixel (conditioned on the optical field  $A(x, y, z)$ ).

We discretize the optical field  $A(x, y, z)$  as a raster-scanned complex column vector  $\mathbf{a}_n$ , and similarly discretize the intensity measurement  $I(x, y, z)$  as column vector  $\mathbf{I}_n$ . We denote by  $b(u, v, z)$  the 2-D Fourier transform of  $A(x, y, z)$ . The column vector  $\mathbf{b}_n$  is again raster-scanned from  $b(u, v, z)$ , and hence can be expressed as  $\mathbf{b}_n = \mathbf{K}\mathbf{a}_n$ , where  $\mathbf{K}$  is the discrete Fourier trans-

form matrix. Since  $\mathbf{K}$  is unitary, we can write  $\mathbf{K}\mathbf{K}^H = \mathbf{K}^H\mathbf{K} = \mathbf{U}$  (with normalization), where  $\mathbf{U}$  is the identity matrix or unit matrix and  $\mathbf{K}^H$  denotes the hermitian of  $\mathbf{K}$ .

Let us assume the distance between two consecutive planes is a constant  $\Delta z$ . Then, we can define the propagation matrix from  $z_{n-1}$  to  $z_n$  as [13]:

$$\mathbf{H} = \text{diag} \left( \exp \left[ -i\lambda\pi \left( \frac{u_1^2}{L_x^2} + \frac{v_1^2}{L_y^2} \right) \Delta z \right], \dots, \exp \left[ -i\lambda\pi \left( \frac{u_M^2}{L_x^2} + \frac{v_N^2}{L_y^2} \right) \Delta z \right] \right), \quad (3)$$

where  $L_x$  and  $L_y$  are the width and height of the image, respectively.

The relation between two images with distance  $\Delta z$  in Fourier domain can be written as:

$$\mathbf{b}_n = \mathbf{H}\mathbf{b}_{n-1}. \quad (4)$$

In order to fit within the Kalman filter framework, we approximate the Poisson observation model as in Eq. (2) with a Gaussian distribution of the same mean and covariance. In particular, we consider the approximate observation model:

$$\mathbf{I}_n = \gamma |\mathbf{a}_n|^2 + \mathbf{v}, \quad (5)$$

where  $\mathbf{v}$  is a Gaussian vector with zero mean and covariance  $\mathbf{R} = \gamma \text{diag}(\mathbf{a}_n^*) \text{diag}(\mathbf{a}_n)$ . Here  $\mathbf{a}_n^*$  denotes the complex conjugate of  $\mathbf{a}_n$ , and  $\text{diag}(\mathbf{a}_n^*)$  is a diagonal matrix with its corresponding diagonal entries equal to the elements in the vector  $\mathbf{a}_n^*$ .

The nonlinear observation model in Eq. (5) is linearized as [14]:

$$\mathbf{I}_n = \gamma |\hat{\mathbf{a}}_n|^2 + \gamma \text{diag}(\hat{\mathbf{a}}_n^*)(\mathbf{a}_n - \hat{\mathbf{a}}_n) + \gamma \text{diag}(\hat{\mathbf{a}}_n)(\mathbf{a}_n^* - \hat{\mathbf{a}}_n^*) + \mathbf{v}, \quad (6)$$

where  $\hat{\mathbf{a}}_n$  is the state predicted from the previous  $n-1$  observations, and Eq. (6) is the first order Taylor series expansion of Eq. (5) with respect to  $\hat{\mathbf{a}}_n$ .

Summarizing, the augmented state space model is given as:

$$\text{state: } \begin{bmatrix} \mathbf{b}_n \\ \mathbf{b}_n^* \end{bmatrix} = \begin{bmatrix} \mathbf{H} & 0 \\ 0 & \mathbf{H}^* \end{bmatrix} \begin{bmatrix} \mathbf{b}_{n-1} \\ \mathbf{b}_{n-1}^* \end{bmatrix} \quad (7)$$

$$\text{observation: } \mathbf{I}_n = \begin{bmatrix} \mathbf{J}_n & \mathbf{J}_n^* \end{bmatrix} \begin{bmatrix} \mathbf{b}_n \\ \mathbf{b}_n^* \end{bmatrix} - \gamma |\hat{\mathbf{a}}_n|^2 + \mathbf{v}, \text{ with } \mathbf{v} \sim (\mathbf{0}, \mathbf{R}), \quad (8)$$

where

$$\mathbf{R} = \gamma \text{diag}(\hat{\mathbf{a}}_n^*) \text{diag}(\hat{\mathbf{a}}_n) \text{ and } \mathbf{J}_n = \gamma \text{diag}(\hat{\mathbf{a}}_n^*) \mathbf{K}^H. \quad (9)$$

We adopt the augmented state space model because the ACEKF with the state augmented outperforms CEKF in both estimation error and convergence [14]. If the state is not augmented, such as in CEKF algorithm, the covariance  $\mathbf{E}[\mathbf{b}_n \mathbf{b}_n^H]$  is not sufficient for a complete description of the complex state. Since the state in ACEKF is augmented from  $\mathbf{b}_n$  to  $[\mathbf{b}_n \ \mathbf{b}_n^*]^T$ , it introduces both the covariance  $\mathbf{E}[\mathbf{b}_n \mathbf{b}_n^H]$  and the pseudo-covariance  $\mathbf{E}[\mathbf{b}_n \mathbf{b}_n^T]$ . Intuitively,  $\mathbf{E}[\mathbf{b}_n \mathbf{b}_n^H]$  and  $\mathbf{E}[\mathbf{b}_n \mathbf{b}_n^T]$  together provide a more comprehensive statistical description of the complex state.

### 3. State estimation by sparse augmented complex extended Kalman filter

The state covariance matrix of the augmented state has the form:

$$\mathbf{S}_n = \begin{bmatrix} \mathbf{S}_n^Q & \mathbf{S}_n^P \\ (\mathbf{S}_n^P)^* & (\mathbf{S}_n^Q)^* \end{bmatrix} \quad (10)$$

From the update equations of ACEKF [1, 15], we have the following steps:

1. Initialize:  $b_0$ ,  $\mathbf{S}_0^Q$ , and  $\mathbf{S}_0^P$ .
2. Predict:

$$\hat{\mathbf{S}}_n^Q = \mathbf{H}\mathbf{S}_{n-1}^Q\mathbf{H}^H, \quad (11)$$

$$\hat{\mathbf{S}}_n^P = \mathbf{H}\mathbf{S}_{n-1}^P\mathbf{H}. \quad (12)$$

3. Update:

$$\mathbf{S}_n^Q = \hat{\mathbf{S}}_n^Q - (\hat{\mathbf{S}}_n^Q\mathbf{J}_n^H + \hat{\mathbf{S}}_n^P\mathbf{J}_n^T)(\mathbf{J}_n\hat{\mathbf{S}}_n^Q\mathbf{J}_n^H + \mathbf{J}_n\hat{\mathbf{S}}_n^P\mathbf{J}_n^T + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^Q)^*\mathbf{J}_n^T + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^P)^*\mathbf{J}_n^H + \mathbf{R})^{-1}(\mathbf{J}_n\hat{\mathbf{S}}_n^Q + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^P)^*), \quad (13)$$

$$\mathbf{S}_n^P = \hat{\mathbf{S}}_n^P - (\hat{\mathbf{S}}_n^Q\mathbf{J}_n^H + \hat{\mathbf{S}}_n^P\mathbf{J}_n^T)(\mathbf{J}_n\hat{\mathbf{S}}_n^Q\mathbf{J}_n^H + \mathbf{J}_n\hat{\mathbf{S}}_n^P\mathbf{J}_n^T + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^Q)^*\mathbf{J}_n^T + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^P)^*\mathbf{J}_n^H + \mathbf{R})^{-1}(\mathbf{J}_n\hat{\mathbf{S}}_n^P + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^Q)^*), \quad (14)$$

$$\mathbf{G}_n = (\mathbf{S}_n^Q\mathbf{J}_n^H + \mathbf{S}_n^P\mathbf{J}_n^T)\mathbf{R}^{-1}, \quad (15)$$

$$\mathbf{b}_n = \hat{\mathbf{b}}_n + \mathbf{G}_n(\mathbf{I}_n - \gamma|\hat{\mathbf{a}}_n|^2). \quad (16)$$

The size of  $\mathbf{S}_n^Q$  or  $\mathbf{S}_n^P$  is  $N^2$ , where  $N$  is the total number of the pixels in the image. Thus, the covariance matrix dominates the memory requirements. The inversion of the covariance matrix has a computational complexity of  $\mathcal{O}(N^3)$  in each step. Both the storage requirement and computational burden make the above update algorithm impractical for real applications. Here we make some mild assumptions and derivations, resulting in a low-complexity algorithm with reduced storage requirement. After some derivations (see more in Appendices A and B), we prove Lemma 1 as well as Theorems 1 and 2. Theorem 1 is derived from the ACEKF update Eqs. (11)-(14). It shows that if the covariance matrix at step  $n-1$  has the form  $\mathbf{S}_{n-1}^Q = \mathbf{Q}_{n-1}$  and  $\mathbf{S}_{n-1}^P = \mathbf{P}_{n-1}\mathbf{E}$ , where  $\mathbf{Q}_{n-1}$  and  $\mathbf{P}_{n-1}$  are diagonal, then the covariance matrix at step  $n$  has the same form as that of step  $n-1$  (see Eqs. (23)-(24)). Therefore, once the first covariance matrix is initialized as  $\mathbf{S}_0^Q = \mathbf{Q}_0$  and  $\mathbf{S}_0^P = \mathbf{P}_0\mathbf{E}$ , then the subsequent matrices will keep the same form. Since the matrices in Theorem 1 are diagonal, the computational complexity in Theorem 1 is  $\mathcal{O}(N)$ . Theorem 2, derived from the Eqs. (15)-(16), provides a new Kalman gain and update formula using the diagonal matrices  $\mathbf{Q}_n$  and  $\mathbf{P}_n$ .

**Lemma 1.** *If a matrix  $\mathbf{M}$  is diagonal and its diagonal entries are rotationally symmetric in 2-D, then  $\mathbf{E}\mathbf{M}\mathbf{E} = \mathbf{M}$ , where  $\mathbf{E} = \mathbf{K}\mathbf{K}^T$ , and  $\mathbf{K}$  is the Discrete Fourier Transform Matrix. ■*

**Theorem 1.** *If*

$$\mathbf{S}_{n-1}^Q = \mathbf{Q}_{n-1} \quad (17)$$

$$\mathbf{S}_{n-1}^P = \mathbf{P}_{n-1}\mathbf{E}, \quad (18)$$

where  $\mathbf{Q}_{n-1}$  and  $\mathbf{P}_{n-1}$  are diagonal, then we derive from the ACEKF update Eqs. (11)-(16) that

the covariance matrix can be updated as follows

**Predict:**

$$\hat{\mathbf{Q}}_n = \mathbf{Q}_{n-1} \quad (19)$$

$$\hat{\mathbf{P}}_n = \mathbf{H}\mathbf{P}_{n-1}\mathbf{H} \quad (20)$$

**Update:**

$$\mathbf{Q}_n = \hat{\mathbf{Q}}_n - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n)(\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^* + (\hat{\mathbf{P}}_n)^* + q\mathbf{I})^{-1}(\hat{\mathbf{Q}}_n + (\hat{\mathbf{P}}_n)^*) \quad (21)$$

$$\mathbf{P}_n = \hat{\mathbf{P}}_n - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n)(\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^* + (\hat{\mathbf{P}}_n)^* + q\mathbf{I})^{-1}(\hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^*) \quad (22)$$

$$\mathbf{S}_n^Q = \mathbf{Q}_n \quad (23)$$

$$\mathbf{S}_n^P = \mathbf{P}_n\mathbf{E}, \quad (24)$$

where  $\mathbf{Q}_n$  and  $\mathbf{P}_n$  are diagonal, and  $q = \frac{1}{\gamma}$ . ■

**Theorem 2.** The Kalman gain and update formula for the state are

$$\mathbf{G}_n = (\mathbf{S}_n^Q \mathbf{J}_n^H + \mathbf{S}_n^P \mathbf{J}_n^T) \mathbf{R}^{-1} = (\mathbf{Q}_n + \mathbf{P}_n)(\mathbf{J}_n)^{-1} q^{-1}. \quad (25)$$

$$\mathbf{b}_n = \hat{\mathbf{b}}_n + \mathbf{G}_n(\mathbf{I}_n - \gamma |\hat{\mathbf{a}}_n|^2). \quad (26)$$

■

Table 1. Sparse augmented complex extended Kalman filter for estimating a wave field.

<b>(1) Initialization <math>\mathbf{b}_0</math>, <math>\mathbf{Q}_0</math> and <math>\mathbf{P}_0</math>.</b>	
<b>(2) Prediction:</b>	
	$\hat{\mathbf{b}}_n = \mathbf{H}\mathbf{b}_{n-1} \quad (27)$
	$\hat{\mathbf{Q}}_n = \mathbf{Q}_{n-1} \quad (28)$
	$\hat{\mathbf{P}}_n = \mathbf{H}\mathbf{P}_{n-1}\mathbf{H} \quad (29)$
<b>(3) Update:</b>	
	$\hat{\mathbf{a}}_n = \mathbf{K}^H \hat{\mathbf{b}}_n \quad (30)$
	$\mathbf{Q}_n = \hat{\mathbf{Q}}_n - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n)(\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^* + (\hat{\mathbf{P}}_n)^* + q\mathbf{I})^{-1}(\hat{\mathbf{Q}}_n + (\hat{\mathbf{P}}_n)^*) \quad (31)$
	$\mathbf{P}_n = \hat{\mathbf{P}}_n - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n)(\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^* + (\hat{\mathbf{P}}_n)^* + q\mathbf{I})^{-1}(\hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^*) \quad (32)$
	$\mathbf{b}_n = \hat{\mathbf{b}}_n + (\mathbf{Q}_n + \mathbf{P}_n)(\mathbf{J}_n)^{-1} q^{-1}(\mathbf{I}_n - \gamma  \hat{\mathbf{a}}_n ^2). \quad (33)$

The resulting algorithm, referred to as the sparse ACEKF, is summarized in Table 1. Matrices  $\mathbf{Q}_n$  and  $\mathbf{P}_n$  are diagonal, and hence they can be stored as two vectors. The storage burden of Eqs. (13)-(14) in the update step is reduced from  $N^2$  to  $N$ . The inverse of  $\mathbf{J}_n$  in Eq. (33) can be computed by Fast Fourier Transform (FFT), which has a computational complexity of  $\mathcal{O}(N \log(N))$ . Since  $\mathbf{Q}_n$  and  $\mathbf{P}_n$  are diagonal, the matrix multiplications and inversions in Eqs. (31)-(32) have a computational complexity of  $\mathcal{O}(N)$ . The overall computational complexity of the sparse ACEKF is  $\mathcal{O}(N_z N \log(N))$ , limited by the complexity of FFT.

#### 4. Analysis of the convergence of the augmented complex extended Kalman filter

The convergence of ACEKF, for the complex state space model, has not been well analyzed in the literature. However, the convergence and stability of the extended Kalman filter (EKF) for the real state is well studied [16–18]. Thus we first convert the ACEKF state space model into its equivalent dual form as a real state space model (EKF), then we use existing methods to analyze the convergence of the dual form.

We follow [14] to derive the dual real form of the ACEKF state space model (7)-(8). Define

$$\mathbf{M} = \frac{1}{2} \begin{bmatrix} \mathbf{U} & \mathbf{U} \\ -j\mathbf{U} & j\mathbf{U} \end{bmatrix}, \quad (34)$$

where  $\mathbf{U}$  is identity matrix. When multiplied by  $\mathbf{M}$ , a complex variable is converted to its corresponding real form:

$$\begin{bmatrix} \text{Re}(\mathbf{b}_n) \\ \text{Im}(\mathbf{b}_n) \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathbf{b}_n \\ \mathbf{b}_n^* \end{bmatrix}, \quad (35)$$

where  $\text{Re}(\mathbf{b}_n)$  and  $\text{Im}(\mathbf{b}_n)$  are the real and imaginary parts of  $\mathbf{b}_n$ , respectively. The complex model (7)-(8) becomes a real value model:

$$\text{state: } \begin{bmatrix} \text{Re}(\mathbf{b}_n) \\ \text{Im}(\mathbf{b}_n) \end{bmatrix} = \mathbf{A}_n \begin{bmatrix} \text{Re}(\mathbf{b}_{n-1}) \\ \text{Im}(\mathbf{b}_{n-1}) \end{bmatrix}, \quad (36)$$

$$\text{observation: } \mathbf{I}_n = \mathbf{C}_n \begin{bmatrix} \text{Re}(\mathbf{b}_n) \\ \text{Im}(\mathbf{b}_n) \end{bmatrix} - \gamma |\hat{\mathbf{a}}_n|^2 + \mathbf{v}, \quad (37)$$

where

$$\mathbf{A}_n = \mathbf{M} \begin{bmatrix} \mathbf{H} & 0 \\ 0 & \mathbf{H}^* \end{bmatrix} \mathbf{M}^{-1}, \quad (38)$$

$$\mathbf{C}_n = [\mathbf{J}_n \quad \mathbf{J}_n^*] \mathbf{M}^{-1}. \quad (39)$$

By applying the theory of EKF convergence in [16] to our real value model, we obtain Theorem 3 which provides the convergence and stability of the model.

**Theorem 3.** *The estimation error of the EKF algorithm in Eqs. (36)-(37) is bounded, and hence also the estimation error of ACEKF algorithm for Eqs. (7)-(8) is bounded, when the initial estimation error and the noise are small enough, and the EKF model satisfies the following conditions [16]:*

1. There exist positive numbers  $\bar{a}, \bar{c}, \bar{s}, \underline{s}, \underline{v} > 0$ , such that for every  $n \geq 0$ :

$$\|\mathbf{A}_n\| \leq \bar{a}, \quad (40)$$

$$\|\mathbf{C}_n\| \leq \bar{c}, \quad (41)$$

$$\underline{s}\mathbf{U} \leq \mathbf{S}_n \leq \bar{s}\mathbf{U}, \quad (42)$$

$$\underline{v}\mathbf{U} \leq \mathbf{v}. \quad (43)$$

2.  $\mathbf{A}_n$  is non-singular for every  $n$ .

3. The linearization at Eq. (6) has a bounded error: ■



In Theorem 3,  $\|\cdot\|$  denotes the spectral norm of matrices. Assuming the initial estimation error and the noise are small enough, let us check each condition of Theorem 3. From the definition in Eq. (3), the propagation matrix  $\mathbf{H}$  is non-singular and determined by the propagation distance  $\Delta z$ . Since  $\mathbf{M}$  is full rank, the matrix  $\mathbf{A}_n$  given in Eq. (38) is bounded and non-singular. Therefore, the inequality (40) and the second condition of Theorem 3 are satisfied. The linearization in Eq. (6) is a first-order truncation of Taylor series expansion of a quadratic function. Therefore, the linearization has a bounded error and the real observation matrix  $\mathbf{C}_n$  of Eq. (37) is bounded as well, as required by the inequality (41) and the third condition of Theorem 3. The matrix  $\mathbf{R}$  is the covariance matrix of the Poisson noise in Eq. (5), and it is bounded as the amplitude of the state is bounded.

The inequality (42) requires the covariance matrix to be bounded for every  $n$ . In [16], it is shown that the boundedness of the covariance matrix is closely related to observability and detectability properties of the EKF model. However, it is difficult to prove the observability of the model directly, because  $\mathbf{A}_n$  and  $\mathbf{C}_n$  are based on the observed data at each step. We calculate the rank of the observability matrix and the bound on the covariance matrix by testing our algorithm on the synthetic data (see Data Set 2 in section 5). We initialize the covariance matrix  $\mathbf{Q}_0$  and  $\mathbf{P}_0$  in Table 1 with 30 different values. The matrix  $\mathbf{Q}_0$  is initialized as  $q\mathbf{U}$  ( $\mathbf{U}$  is identity matrix), with  $q$  ranging from 70 to 7000, and  $\mathbf{P}_0$  is initialized as  $p\mathbf{U}$ , with  $p$  ranging from 50 to 5000. Figure 1 shows that the minimum singular value of the observability matrix is larger than zero, so the observability matrix is full rank at each step. Figures 2 and 3 show the maximum and minimum singular value of the covariance matrix at each step with different initializations. The covariance matrix is bounded, as required by inequality (42).

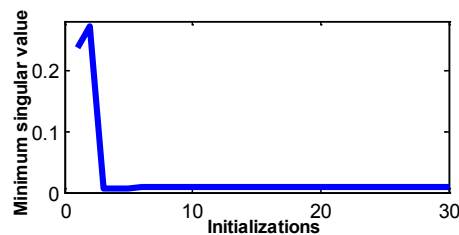


Fig. 1. Minimum singular value of observability matrix with 30 different initializations of the covariance matrix  $\mathbf{Q}_0$  and  $\mathbf{P}_0$ . The matrix  $\mathbf{Q}_0$  is initialized as  $q\mathbf{U}$ , with  $q$  ranging from 70 to 7000, and  $\mathbf{P}_0$  is initialized as  $p\mathbf{U}$ , with  $p$  ranging from 50 to 5000.

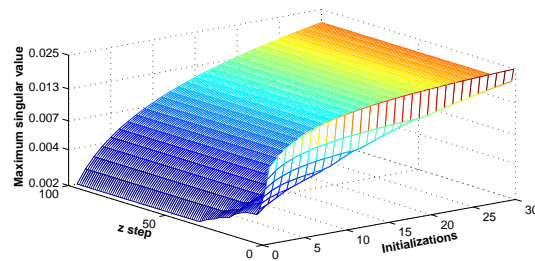


Fig. 2. Maximum singular value of covariance matrix with 30 different initializations of  $p$  and  $q$  in the covariance matrix.

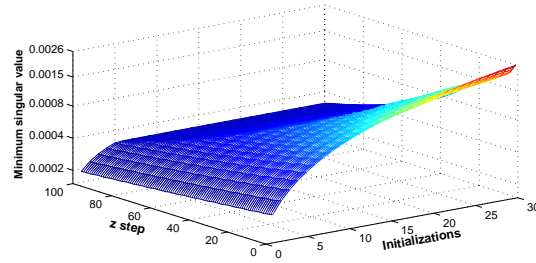


Fig. 3. Minimum singular value of covariance matrix with 30 different initializations of  $p$  and  $q$  in the covariance matrix.

## 5. Experimental results

We consider three sets of data to assess the performance the augmented Kalman filter. **Data Set 1** consists of 100 images of size  $100 \times 100$  pixels artificially generated to simulate a complex field propagating from focus in  $0.5 \mu\text{m}$  steps over a distance of  $50 \mu\text{m}$  with illumination wavelength of  $532 \text{ nm}$ . Pixels are corrupted by Poisson noise so that, on average, each pixel detects  $\gamma = 0.998$  photons. **Data Set 2** comprises 50 images of size  $150 \times 150$  pixels acquired by a microscope. The wavelength was again  $532 \text{ nm}$ , and the defocused intensity images were captured by moving the camera axially with a step size of  $2 \mu\text{m}$  over a distance of  $100 \mu\text{m}$ . **Data Set 3** has 101 images of size  $492 \times 656$  pixels acquired by a microscope. The light source is partially coherent, and filtered by a narrow-band color filter of center wavelength  $633 \text{ nm}$ . The images were captured by moving the camera axially with a step size of  $2 \mu\text{m}$ . Figure 4 shows the images of synthetic data (Data Set 1) and experimental data (Data Set 2 and Data Set 3).

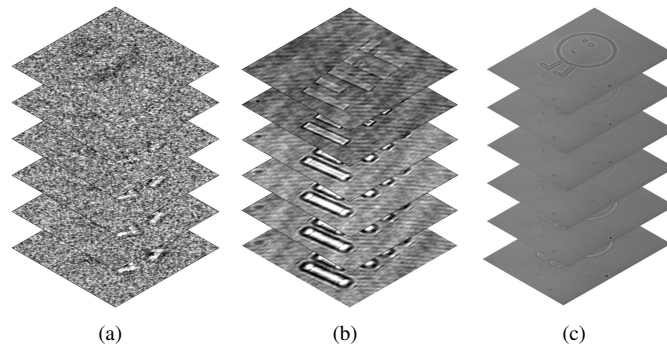


Fig. 4. (a) Data Set 1: synthetic images with strong noise ( $100 \times 100$  pixels with size of  $1 \mu\text{m} \times 1 \mu\text{m}$ ). (b) Data Set 2: experimental data acquired by a microscope ( $150 \times 150$  pixels with size of  $2 \mu\text{m} \times 2 \mu\text{m}$ ). (c) Data Set 3: experimental data of large size ( $492 \times 656$  pixels with size of  $0.74 \mu\text{m} \times 0.74 \mu\text{m}$ ) obtained by a microscope.

### 5.1. Synthetic data

Table 2 summarizes the results of Data Set 1 using three methods: ACEKF (augmented complex extended Kalman filter) [1], diagonalized CEKF [12], and the proposed method sparse ACEKF. The ACEKF method has a high computational complexity of  $\mathcal{O}(N_z N^3)$  and storage requirement of  $\mathcal{O}(N^2)$ . In order to alleviate the computational burden of ACEKF, the images are divided into

independent blocks of size  $50 \times 50$ , but it still takes  $1.4 \times 10^4$  seconds by a standard CPU. On the other hand, the computational complexity of the sparse ACEKF is  $\mathcal{O}(N_z N \log N)$ , and it takes 0.40 seconds to process the 100 (full) images, thus giving a speedup factor of 35000x.

As illustrated in Table 2, the computational complexity of the diagonalized CEKF is lower than that of ACEKF. However, the latter yields better results in terms of phase error. In order to reduce the error of the diagonalized CEKF, forward and backward sweeps (iterations) are applied in [12]. However, the iteration increases the computational complexity linearly, and makes the method no longer recursive. The sparse ACEKF method has an intensity error of 0.0071, and a phase error of 0.0143 (radian). Compared with the diagonalized CEKF, the sparse ACEKF has the same computational complexity and storage requirement, but returns more accurate images.

We assess the quality of reconstruction by root mean square error (RMSE). The proposed sparse ACEKF has an error near to that of ACEKF, while the recovered phase and intensity images of the sparse ACEKF in Fig. 5 might look better. The images recovered by ACEKF exhibit a block effect as straight lines crossing the images, whereas the result of sparse ACEKF is free of such block effect. The sparse ACEKF has a much lower complexity, which avoids the need of dividing the images into independent blocks. The images recovered by ACEKF and the diagonalized CEKF contain traces of phase in the intensity images due to errors. However, the trace of phase is mostly removed in the estimated intensity image from the sparse ACEKF.

Table 2. Comparison of different methods. Each image is divided into 4 blocks of size  $50 \times 50$  for ACEKF, while the proposed sparse ACEKF and diagonalized CEKF processes the images without separating into blocks.

	Complexity	Time[s]	Storage	Intensity error	Phase error [radians]
ACEKF	$\mathcal{O}(N_z N^3)$	$1.3 \times 10^4$	$\mathcal{O}(N^2)$	0.0091	0.0139
Diagonalized CEKF	$\mathcal{O}(N_z N \log N)$	0.30	$\mathcal{O}(N)$	0.0079	0.0166
<b>Sparse ACEKF</b>	$\mathcal{O}(N_z N \log N)$	<b>0.40</b>	$\mathcal{O}(N)$	<b>0.0071</b>	<b>0.0143</b>

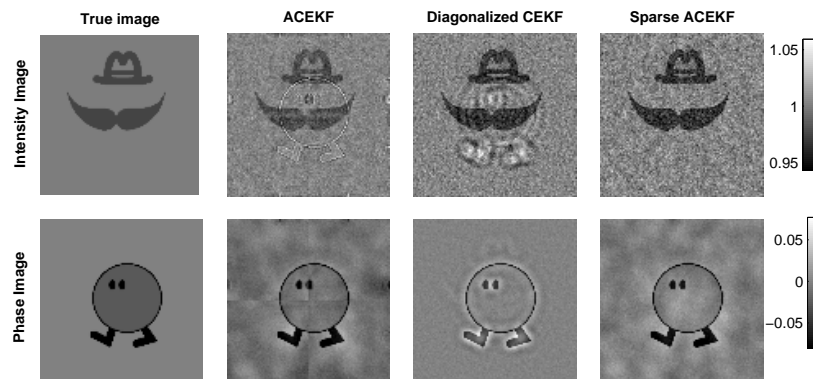


Fig. 5. Recovered intensity and phase [radian] image from synthetic Data Set 1 by ACEKF, diagonalized CEKF, and the proposed sparse ACEKF.

## 5.2. Experimental data

In Fig. 6, we compare the estimated intensity and phase images of Data Set 2 using the ACEKF, the diagonalized CEKF, and the sparse ACEKF. Stripes in the phase image recovered by the diagonalized CEKF look darker, while the stripes in the recovered phase image of the sparse

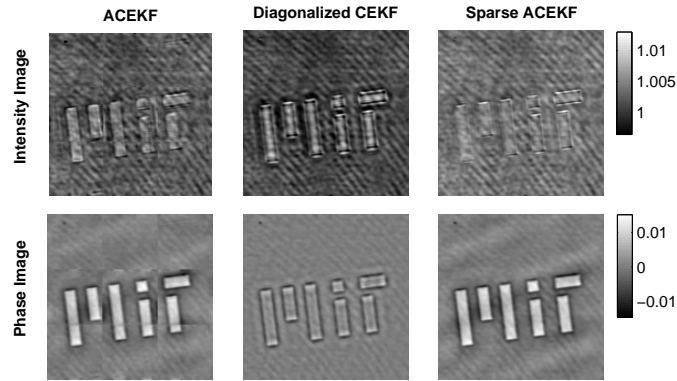


Fig. 6. Estimated intensity and phase [radians] of Data Set 2 by ACEKF, diagonalized CEKF, and the proposed sparse ACEKF.

ACEKF method have stronger contrast. In Fig. 7(a), we show the recovered phase of the Data Set 3 by ACEKF, the diagonalized CEKF, and the sparse ACEKF. The real depth of the sample in Data Set 3 is around  $75 \pm 5nm$ . The proposed method takes 20.24 seconds to process 101 images of size  $492 \times 656$ . However, the ACEKF method takes 54.15 hours and each image is separated into 117 pieces of  $50 \times 50$  blocks. In Fig. 7(b), we compare the depth across the black line of the recovered phase in Fig. 7(a). The sparse ACEKF method shows a result much closer to the true value, compared to the ACEKF and the diagonalized CEKF. However, the reconstructed phase image obtained by sparse ACEKF contains low-frequency noise, especially at the edges. This shadow effect may be attributed to the partial coherence of the light, which is not incorporated into our model (we assume full coherence). The low frequency issue may be alleviated through a nonlinear diffusion filter, which we will explore in future work.

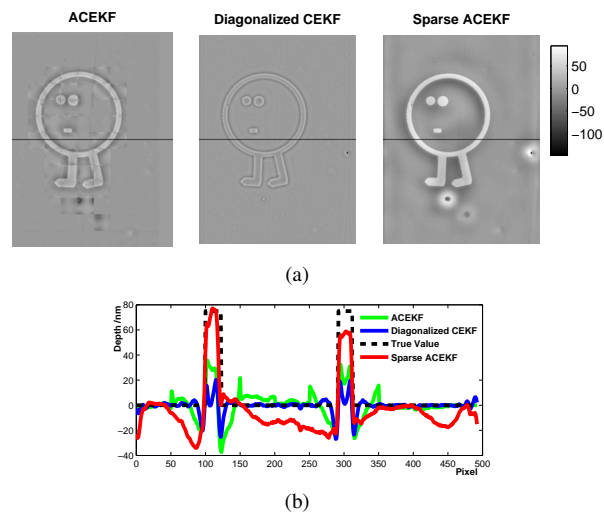


Fig. 7. (a) Estimated height [nm] for Data Set 3 by ACEKF, diagonalized CEKF, and the proposed sparse ACEKF. (b) Depth along the black line in (a).

## 6. Conclusions

The proposed method efficiently recovers phase and amplitude from a series of noisy defocused images. Although a constant defocus step size is used for the through-focal series in this paper, it can be generalized to cases of non-constant step-size. The method can directly deal with large data sets and high-resolution images because of its low computational complexity and storage requirement. It can process the images recursively in real-time after the data has been captured, or even during the capture sequence. For example, with further work, this method could form the basis for an adaptive phase imaging method, in which the current estimate of the phase informs the choice of the next measurement plane (defocus distance). Since the optimal measurement planes will depend on the object itself [19], real-time estimation during capture could lead to optimization not only of the processing, but also optimization of the measurement scheme itself. Furthermore, due to the scalability of the wave equations and the simplicity of the measurement technique, this method could find use in phase imaging beyond optical wavelengths (for example, X-ray or neutron imaging), where high-quality images are difficult to obtain and noise is significant and unavoidable.

### Appendix A: proof of theorem 1

#### Predict

If  $\mathbf{Q}_{n-1}$  and  $\mathbf{P}_{n-1}$  are diagonal, from the definition of  $\mathbf{H}$  in Eq. (3) it follows:

$$\hat{\mathbf{P}}_n = \mathbf{H}\mathbf{P}_{n-1}\mathbf{H} \quad (44)$$

$$\hat{\mathbf{Q}}_n = \mathbf{H}\mathbf{Q}_{n-1}\mathbf{H}^H = \mathbf{Q}_{n-1}. \quad (45)$$

Since the propagation matrix  $\mathbf{H}$  is diagonal and its diagonal entries are rotationally symmetric, from Eq. (11)-(12) and Lemma 1, it follows:

$$\hat{\mathbf{S}}_n^Q = \mathbf{H}\mathbf{S}_{n-1}^Q\mathbf{H}^H = \mathbf{H}\mathbf{Q}_{n-1}\mathbf{H}^H = \hat{\mathbf{Q}}_n \quad (46)$$

$$\hat{\mathbf{S}}_n^P = \mathbf{H}\mathbf{S}_{n-1}^P\mathbf{H} = \mathbf{H}\mathbf{P}_{n-1}\mathbf{H} = \hat{\mathbf{P}}_n\mathbf{E}. \quad (47)$$

#### Update

From the update formula in Eqs. (13)-(14), we have

$$\mathbf{S}_n^Q = \hat{\mathbf{S}}_n^Q - (\hat{\mathbf{S}}_n^Q\mathbf{J}_n^H + \hat{\mathbf{S}}_n^P\mathbf{J}_n^T)(\mathbf{J}_n\hat{\mathbf{S}}_n^Q\mathbf{J}_n^H + \mathbf{J}_n\hat{\mathbf{S}}_n^P\mathbf{J}_n^T + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^Q)^*\mathbf{J}_n^T + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^P)^*\mathbf{J}_n^H + \mathbf{R})^{-1}(\mathbf{J}_n\hat{\mathbf{S}}_n^Q + \mathbf{J}_n^*(\hat{\mathbf{S}}_n^P)^*). \quad (48)$$

From Eq. (9),  $\mathbf{J}_n = \gamma \text{diag}(\hat{\mathbf{a}}_n^*)\mathbf{K}^H$ . Let us define  $\mathbf{D} = \gamma \text{diag}(\hat{\mathbf{a}}_n^*)$ , so that  $\mathbf{J}_n = \mathbf{D}\mathbf{K}^H$ . It follows:

$$\begin{aligned} \mathbf{S}_n^Q &= \hat{\mathbf{S}}_n^Q - (\hat{\mathbf{S}}_n^Q\mathbf{K}\mathbf{D}^H + \hat{\mathbf{S}}_n^P\mathbf{K}^*\mathbf{D}^T)(\mathbf{D}\mathbf{K}^H\hat{\mathbf{S}}_n^Q\mathbf{K}\mathbf{D}^H + \mathbf{D}\mathbf{K}^H\hat{\mathbf{S}}_n^P\mathbf{K}^*\mathbf{D}^T + \mathbf{D}^*\mathbf{K}^T(\hat{\mathbf{S}}_n^Q)^*\mathbf{K}^*\mathbf{D}^T \\ &\quad + \mathbf{D}^*\mathbf{K}^T(\hat{\mathbf{S}}_n^P)^*\mathbf{K}\mathbf{D}^H + q\mathbf{D}\mathbf{K}^H\mathbf{K}\mathbf{D}^H)^{-1}(\mathbf{D}\mathbf{K}^H\hat{\mathbf{S}}_n^Q + \mathbf{D}^*\mathbf{K}^T(\hat{\mathbf{S}}_n^P)^*), \end{aligned} \quad (49)$$

where  $q = \frac{1}{\gamma}$ .

Because  $\mathbf{D}$  is diagonal matrix, it has  $\mathbf{D}^T = \mathbf{D}$  and  $\mathbf{D}^H = \mathbf{D}^*$ .

$$\begin{aligned} \mathbf{S}_n^Q &= \hat{\mathbf{S}}_n^Q - (\hat{\mathbf{S}}_n^Q\mathbf{K} + \hat{\mathbf{S}}_n^P\mathbf{K}^*\mathbf{D}(\mathbf{D}^*)^{-1})(\mathbf{K}^H\hat{\mathbf{S}}_n^Q\mathbf{K} + \mathbf{K}^H\hat{\mathbf{S}}_n^P\mathbf{K}^*\mathbf{D}(\mathbf{D}^*)^{-1} + \mathbf{D}^{-1}\mathbf{D}^*\mathbf{K}^T(\hat{\mathbf{S}}_n^Q)^*\mathbf{K}^*\mathbf{D}(\mathbf{D}^*)^{-1} \\ &\quad + \mathbf{D}^{-1}\mathbf{D}^*\mathbf{K}^T(\hat{\mathbf{S}}_n^P)^*\mathbf{K} + q\mathbf{K}^H\mathbf{K})^{-1}(\mathbf{K}^H\hat{\mathbf{S}}_n^Q + \mathbf{D}^{-1}\mathbf{D}^*\mathbf{K}^T(\hat{\mathbf{S}}_n^P)^*). \end{aligned} \quad (50)$$

Since in our phase recovery problem the value of phase is small, here we consider the approximation  $\mathbf{D}^*\mathbf{D}^{-1} \approx \mathbf{I}$ . Future research will explore the influence of large phase values.

$$\begin{aligned} \mathbf{S}_n^Q &= \hat{\mathbf{S}}_n^Q - (\hat{\mathbf{S}}_n^Q\mathbf{K} + \hat{\mathbf{S}}_n^P\mathbf{K}^*)(\mathbf{K}^H\hat{\mathbf{S}}_n^Q\mathbf{K} + \mathbf{K}^H\hat{\mathbf{S}}_n^P\mathbf{K}^* + \mathbf{K}^T(\hat{\mathbf{S}}_n^Q)^*\mathbf{K}^* \\ &\quad + \mathbf{K}^T(\hat{\mathbf{S}}_n^P)^*\mathbf{K} + q\mathbf{K}^H\mathbf{K})^{-1}(\mathbf{K}^H\hat{\mathbf{S}}_n^Q + \mathbf{K}^T(\hat{\mathbf{S}}_n^P)^*). \end{aligned} \quad (51)$$

Taking  $\mathbf{K}^H$  and  $\mathbf{K}$  out of the inverse term yields

$$\mathbf{S}_n^Q = \hat{\mathbf{S}}_n^Q - (\hat{\mathbf{S}}_n^Q \mathbf{K} + \hat{\mathbf{S}}_n^P \mathbf{K}^*) (\mathbf{K}^H (\hat{\mathbf{S}}_n^Q + \hat{\mathbf{S}}_n^P \mathbf{K}^* \mathbf{K}^H + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^Q)^* \mathbf{K}^* \mathbf{K}^H + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^P)^* + qI) \mathbf{K})^{-1} (\mathbf{K}^H \hat{\mathbf{S}}_n^Q + \mathbf{K}^T (\hat{\mathbf{S}}_n^P)^*). \quad (52)$$

It follows:

$$\mathbf{S}_n^Q = \hat{\mathbf{S}}_n^Q - (\hat{\mathbf{S}}_n^Q + \hat{\mathbf{S}}_n^P \mathbf{K}^* \mathbf{K}^H) (\hat{\mathbf{S}}_n^Q + \hat{\mathbf{S}}_n^P \mathbf{K}^* \mathbf{K}^H + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^Q)^* \mathbf{K}^* \mathbf{K}^H + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^P)^* + qI)^{-1} (\hat{\mathbf{S}}_n^Q + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^P)^*). \quad (53)$$

Likewise, we can derive a similar expression for the update of  $\mathbf{S}_n^P$ :

$$\mathbf{S}_n^P = \hat{\mathbf{S}}_n^P - (\hat{\mathbf{S}}_n^Q + \hat{\mathbf{S}}_n^P \mathbf{K}^* \mathbf{K}^H) (\hat{\mathbf{S}}_n^Q + \hat{\mathbf{S}}_n^P \mathbf{K}^* \mathbf{K}^H + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^Q)^* \mathbf{K}^* \mathbf{K}^H + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^P)^* + qI)^{-1} (\hat{\mathbf{S}}_n^P + \mathbf{K} \mathbf{K}^T (\hat{\mathbf{S}}_n^Q)^*). \quad (54)$$

From Lemma 1 and Eqs. (46)-(47), we obtain:

$$\mathbf{S}_n^Q = \hat{\mathbf{Q}}_n - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n \mathbf{E} \mathbf{E}) (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n \mathbf{E} \mathbf{E} + \mathbf{E} (\hat{\mathbf{Q}}_n)^* \mathbf{E} + \mathbf{E} (\hat{\mathbf{P}}_n)^* \mathbf{E} + qI)^{-1} (\hat{\mathbf{Q}}_n + \mathbf{E} (\hat{\mathbf{P}}_n)^* \mathbf{E}) \quad (55)$$

$$\mathbf{S}_n^P = \hat{\mathbf{P}}_n \mathbf{E} - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n \mathbf{E} \mathbf{E}) (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n \mathbf{E} \mathbf{E} + \mathbf{E} (\hat{\mathbf{Q}}_n)^* \mathbf{E} + \mathbf{E} (\hat{\mathbf{P}}_n)^* \mathbf{E} + qI)^{-1} (\hat{\mathbf{P}}_n \mathbf{E} + \mathbf{E} (\hat{\mathbf{Q}}_n)^* \mathbf{E} \mathbf{E}). \quad (56)$$

We initialize  $\mathbf{Q}_0$  and  $\mathbf{P}_0$  as a scalar times the identity matrix. From Eqs. (44)-(45) and (55)-(56), the covariance matrices remain rotationally symmetric during the update. From Lemma 1, it follows:

$$\mathbf{S}_n^Q = \hat{\mathbf{Q}}_n - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n) (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^* + (\hat{\mathbf{P}}_n)^* + qI)^{-1} (\hat{\mathbf{Q}}_n + (\hat{\mathbf{P}}_n)^*) \quad (57)$$

$$\begin{aligned} \mathbf{S}_n^P &= \hat{\mathbf{P}}_n \mathbf{E} - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n) (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^* + (\hat{\mathbf{P}}_n)^* + qI)^{-1} (\hat{\mathbf{P}}_n \mathbf{E} + (\hat{\mathbf{Q}}_n)^* \mathbf{E}) \\ &= (\hat{\mathbf{P}}_n - (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n) (\hat{\mathbf{Q}}_n + \hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^* + (\hat{\mathbf{P}}_n)^* + qI)^{-1} (\hat{\mathbf{P}}_n + (\hat{\mathbf{Q}}_n)^*)) \mathbf{E}. \end{aligned} \quad (58)$$

## Appendix B: proof of theorem 2

In Eq. (15), the Kalman gain for the augmented Kalman filter is given as:

$$\begin{aligned} G_n &= (\mathbf{S}_n^Q \mathbf{J}_n^H + \mathbf{S}_n^P \mathbf{J}_n^T) \mathbf{R}^{-1} \\ &= (\mathbf{S}_n^Q \mathbf{J}_n^H + \mathbf{S}_n^P \mathbf{J}_n^T) (\mathbf{J}_n^H)^{-1} (\mathbf{J}_n)^{-1} q^{-1} \\ &= (\mathbf{S}_n^Q + \mathbf{S}_n^P \mathbf{E}) (\mathbf{J}_n)^{-1} q^{-1} \\ &= (\mathbf{Q}_n + \mathbf{P}_n) (\mathbf{J}_n)^{-1} q^{-1}. \end{aligned} \quad (59)$$

## Acknowledgments

The authors appreciate the help from George Barbastathis's group at MIT. Specifically, the authors thank George Barbastathis, Justin Lee, Nikhil Vadhavkar, Adam Pan and other colleagues for helpful discussions and providing the experimental data. Manuel A. Vázquez acknowledges the support of the Ministry of Education of Spain (Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i 2008-2011).