CHIN. PHYS. LETT. Vol. 38, No. 6 (2021) 064202

Bayesian Optimization for Wavefront Sensing and Error Correction

Zhong-Hua Qian(钱忠华)[1,2], Zi-Han Ding(丁子涵)[3], Ming-Zhong Ai(艾铭忠)[1,2], Yong-Xiang Zheng(郑永祥)[1,2], Jin-Ming Cui(崔金明)[1,2*], Yun-Feng Huang(黄运锋)[1,2], Chuan-Feng Li(李传锋)[1,2], and Guang-Can Guo(郭光灿)[1,2]

[1]*CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China*
[2]*CAS Center For Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China*
[3]*Department of Electrical and Computer Engineering, Princeton University, Princeton, New Jersey 08544, USA*

(Received 21 February 2021; accepted 20 April 2021; published online 25 May 2021)

Algorithms for wavefront sensing and error correction from intensity attract great concern in many fields. Here we propose Bayesian optimization to retrieve phase and demonstrate its performance in simulation and experiment. For small aberration, this method demonstrates a convergence process with high accuracy of phase sensing, which is also verified experimentally. For large aberration, Bayesian optimization is shown to be insensitive to the initial phase while maintaining high accuracy. The approach's merits of high accuracy and robustness make it promising in being applied in optical systems with static aberration such as AMO experiments, optical testing shops, and electron or optical microscopes.

DOI: 10.1088/0256-307X/38/6/064202

Wavefront sensing (WFS) is the technology that measures the deviation of an optical wavefront from a reference wavefront such as a plane wave. WFS has broad applications in microscopy,[1,2] astronomy observation,[3,4] ultrahigh intensity lasers,[5,6] endoscopes for biomedical imaging[7] and AMO experiments.[8] Compared with wavefront sensors such as Shack–Hartmann sensors,[9] phase retrieval techniques compute phases from PSF images rather than using complex optical components.[10,11] These methods are desirable in many aspects such as easy calibration, low monetary cost, and high portability.

Phase retrieval techniques have been extensively studied in previous decades. The Gerchberg–Saxton (GS) algorithm[10,12] attempts to recover the unknown phase using a four-step iteration until the Fourier transform of the phase solution matches the target point spread function (PSF). The convergence of these methods usually relies on the estimation of the initial phase. The interferometer can deterministically measure the phase map using a multi-step phase shift method,[13,14] which is widely employed in optical shop testing and microscopes. This scheme involves the usage of the reference beam and limitations by rigorous conditions such as a temperature-stable environment and vibration-resistant platforms. Aberration correction by using phase diversity[15,16] employs data of both in-focus and defocused images, managing to restore aberration in small distortion. However, the successful rate drops severely when root-mean-square error (RMSE) of aberration is over 0.5λ.[15] Recently, machine learning techniques have been introduced to image-based wavefront sensing.[17−20] In these works, typical models are built with multi-layer perceptrons (MLPs) or convolutional neural networks (CNNs). PSF images with labels of parameter values are used for training the models. A well-trained model has a quick forward-inference process satisfying the requirement by real-time sensing. Hundreds of thousands of images need to be collected before the training phase at first. The data collection and model training consume considerable time and computing power.

In this work, we propose and demonstrate to use Bayesian optimization for wavefront sensing (BOWFS) and error correction from intensity in simulation and experiment. In the simulation part, cases of both small and large aberration are considered. In the experiment part, this method is performed on a real optical imaging system. These results show BOWFS's advantages of high accuracy, low computational cost, and insensitivity to the initial estimated phase. Our approach is promising to be used in optical systems with static aberration such as optical testing shops, electron or optical microscopes, and high-power laser facilities.

Bayesian optimization (BO)[21] as a global optimization algorithm has been extensively studied and used in computer science,[22] robotics,[23] and particle physics.[24] BO can be formulated as optimizing

Supported by the National Key Research and Development Program of China (Grant Nos. 2017YFA0304100 and 2016YFA0302700), the National Natural Science Foundation of China (Grant Nos. 11874343, 61327901, 11774335, 11474270, 11734015, and 11821404).
*Corresponding author. Email: jmcui@ustc.edu.cn
© 2021 Chinese Physical Society and IOP Publishing Ltd

an unknown function termed as black-box objective function (BBF):

$$\max_{\boldsymbol{z}\in\mathcal{A}} F(\boldsymbol{z}), \tag{1}$$

where $F(\boldsymbol{z})$ is the BBF, $\boldsymbol{z} \in \mathbb{R}^d$ is the parameter to be optimized, and d is the dimension of parameter space. $\mathcal{A}$ refers to the parameter space, typically a hyper-rectangle $\{\boldsymbol{z} \in \mathbb{R}^d : a_i \leq z_i \leq b_i\}$. $F(\boldsymbol{z})$ is required to have neither first-order derivative nor concavity structure, which makes the BO algorithm quite versatile compared with many other optimization algorithms such as gradient-descent methods. The dimension d is usually less than 20 in typical applications.

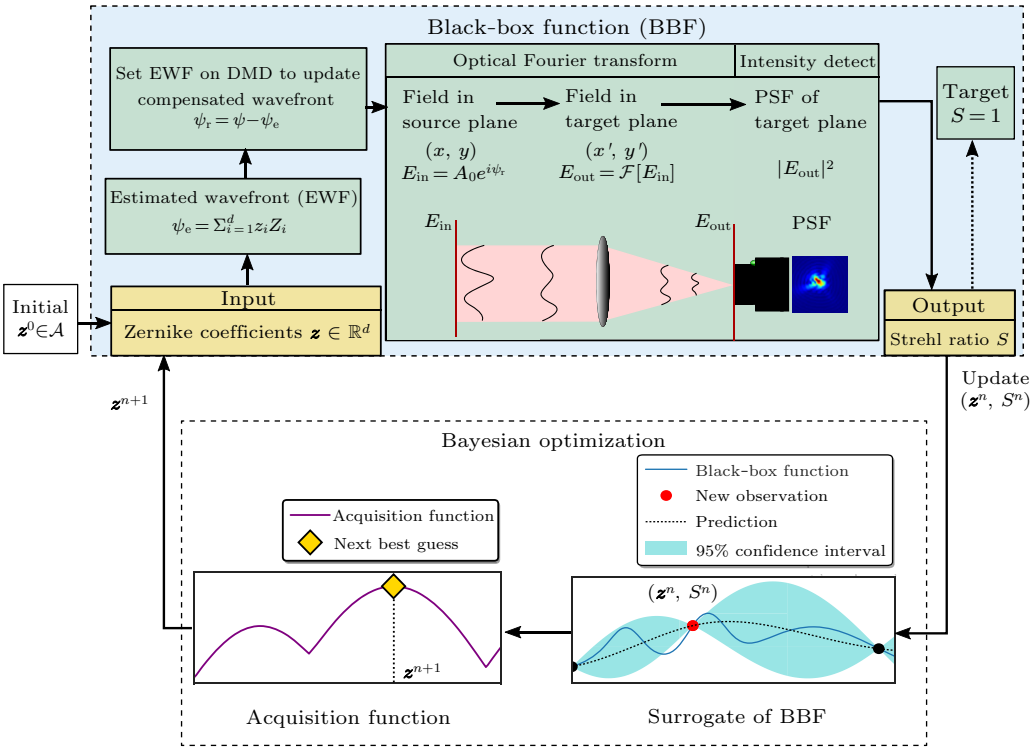

Fig. 1. Overview of the proposed BOWFS method. Top: the black-box function determining the quality of wavefront sensing given by the estimated Zernike coefficients. The wavefront of the system with raw aberration being ψ is compensated by EWF ψ_e through setting EWF on a digital micromirror device (DMD), producing new wavefront error as ψ_r. The optical Fourier transform is implemented by a Fourier lens, creating the intensity of output field as PSF. The Strehl ratio measured from the PSF image serves as the output of black-box function. Bottom: the Bayesian optimization (BO) process for updating the Zernike coefficients with respect to the Strehl ratio. The BOWFS method takes an iterative process between the two modules. In each iteration, a new dataset $(\boldsymbol{z}^n, S^n)$ produced by BBF is updated and next best guess of Zernike coefficients $\boldsymbol{z}^{n+1}$ is generated by BO. This process repeats until target Strehl ratio is found.

The proposed BOWFS aims to search the optimal Zernike coefficients which maximizes the PSF's Strehl ratio, and the concept is illustrated in Fig. 1. This method can be divided into two modules: BBF and BO process. BBF establishes a map from Zernike coefficients $\boldsymbol{z}$ as input to Strehl ratio S as output. Strehl ratio is a valid indicator to assess image quality and phase error.[25] It is measured by the ratio of the maximum intensity of the aberrated PSF to the maximum value of the diffraction limit spot. BBF $F(\boldsymbol{z})$ is described by

$$F(\boldsymbol{z}) = \frac{\max\limits_{(x',y')} |E_{\text{out}}(x', y', \boldsymbol{z})|^2}{\max\limits_{(x',y')} \text{PSF}_d(x', y')}, \tag{2}$$

where E_{out} denotes the optical field of the target plane, and $\text{PSF}_d(x', y')$ is the intensity profile of an ideal image in target plane satisfying diffraction limit. The denominator is a constant value. E_{out} is the optical Fourier transform of the field in source plane.

$$E_{\text{out}}(x', y') = \mathcal{F}[A_0 e^{i[\psi(x,y) - \psi_e(x,y,\boldsymbol{z})]}], \tag{3}$$

where $\psi(x, y)$, $\psi_e(x, y, \boldsymbol{z})$ are phases of initial wavefront and estimated wavefront (EWF) in source plane, respectively. The process of BBF is stated as follows: First the Zernike coefficients are initialized as $\boldsymbol{z}^0$ and taken as the input, with which the ψ_e can be calculated. Then the initial system aberration ψ is compensated by ψ_e by setting ψ_e on aberration-correction devices such as digital micromirror devices (DMDs), spatial light modulators (SLMs) or deformable mirrors (DMs), producing the residual phase error ψ_r. Subsequently, an optical Fourier transform on the field of source plane E_{in} is achieved with a Fourier lens or its

counterparts. In practice the inevitable aberration of the Fourier lens will be incorporated into BBF, leading to a slightly deviation from Eq. (2). However, BOWFS still works because it has no restriction on the explicit expression of BBF as long as the input and output are clearly defined. The intensity of the field in the target plane is measured by a camera. The Strehl ratio S measured from PSF images is output by BBF, which generates a new observed data point $(\boldsymbol{z}^n, S^n)$ in the nth iteration. With the newly observed data point from BBF, the new dataset integrated with this data point will be used by the BO process to update its posterior distribution and generate the next point $\boldsymbol{z}^{n+1}$ to query.[21]

In the BO process, at the beginning, a number of n_0 Zernike coefficients are initially sampled to establish an estimate of BBF $F(\boldsymbol{z})$ called the surrogate function $f(\boldsymbol{z})$. In later iterations, the posterior distribution of the surrogate will be kept updated with the new observed data points.[21] An acquisition function $\alpha(\boldsymbol{z})$ (UCB in our implementation[26]) is applied at each time with the dataset to generate the next Zernike coefficients $\boldsymbol{z}^{n+1}$, where $\alpha(\boldsymbol{z}^{n+1})$ reaches its maximum. The details of derivation of BO see reference[21,27] and the implementation of BO is available in open-source package.[28] After several iterations, BOWFS will output an estimated phase map $\psi_{\rm e}$ very close to the raw wavefront aberration ψ. In that case, the optical imaging system creates a PSF of the near diffraction limit and outputs a Strehl ratio close to 1.0.

In this work, we fit the phase up to the tenth Zernike coefficients. The zero order aberration termed piston will not affect the shape of the wavefront. First order aberrations (x-tilt and y-tilt) are defined as the tipping or tilting of the wavefront along the x-(horizontal) or y-(vertical) axis.[29] These two terms can be calculated with a centroiding algorithm[30] and have no effect on the quality of images. Thus, these three coefficients are excluded in the optimizing process thus we have the parameter dimension $d = 8$.

The simulation is implemented to verify the performance of BOWFS, in which BBF is formulated as Eq. (2). The phase maps of all planes consist of 128×128 pixels with a circle aperture. The aberrated PSF in focus is obtained through optical Fourier transform with a resolution of 128×128 pixels. Here, two types of wavefront distortion are considered: small wavefront error with RMSE 1.02 rad (0.16λ) and large wavefront error with RMSE $= 3.55$ rad (0.57λ). The corresponding Strehl ratios of initial PSFs are 0.28 and 0.05. The single iteration process is called distillation here.

For small aberration, in the first 20 steps, the algorithm samples the parameter space randomly to build the surrogate of BBF. In the following 140 steps of distillation, point to probe next is determined by the location where the acquisition function achieves its maximum. The Strehl ratio would gradually improve with fluctuation because the algorithm tries to maintain both exploitation and exploration to avoid stagnating in the local maximum. The results are illustrated in Fig. 2. The Strehl ratio improves to 0.95 within 46 steps and 0.9992 within 160 steps. The wavefront RMSE decreases to 0.027 rad (0.004λ). We sampled 100 different aberration phase maps and started each distillation with 10 different seeds. After iterating 160 steps it takes 3.64 ± 0.21 min and the average Strehl ratio improves to 0.990 ± 0.022, and wavefront RMSE reduces to 0.05 ± 0.05 rad.

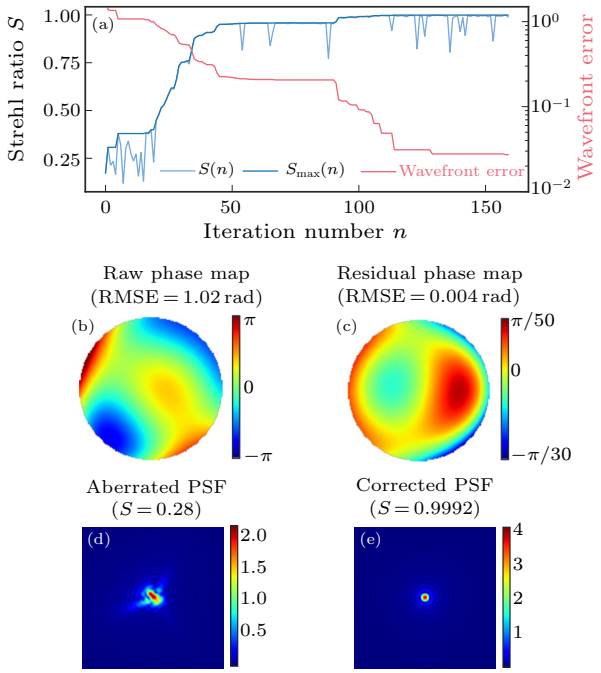

Fig. 2. Bayesian optimization results in simulation for small aberration. (a) Strehl ratio $S(n)$ and wavefront error versus iteration number n. $S_{\max}(n)$ denotes the maximum $S(n)$ among previous n iterations. Red line: the minimum wavefront error (RMSE) achieved among previous n iterations. [(b), (c)] Aberrated phase map with raw distortion (RMSE = 1.02 rad) and residual phase map (RMSE = 0.027 rad). [(d), (e)] Stretched PSF (the square root of PSF) with raw distortion and corrected PSF with Strehl ratio of 0.9992.

For large aberration, we demonstrate multiple rounds of distillation to retrieve phase with insensitivity to initial states and high accuracy. Ten different wavefront aberration cases with RMSE $= 3.6 \pm 0.5$ rad are implemented in simulation. One case result is shown in Fig. 3. The computing processor we used is an ordinary CPU (core i7-7700@3.6 GHz). For BOWFS, three rounds of distillation are run and a phase map is obtained with corrected PSF's Strehl ratio of 0.997 within 7 min. In the first distillation of BO, 100 points are first randomly sampled, with a following procedure of 150 iterations of sampling and

distribution update (one point per iteration) based on the Bayesian rule. Many points with Strehl ratios greater than 0.3 appear among the 150 iterations, which indicates the BO algorithm finds more accurate parameters than the initial random process. In the second distillation, the parameter space $\mathcal{A}$ is adjusted. For each parameter, the central value of the parameter range is chosen as the best value in the previous distillation, and the width of the range is narrowed down to 2.0. This time the BO process starts from scratch rather than continuing from the last round to avoid falling into local maximum. In the same way, a better set of parameter values is achieved with the PSF's Strehl ratio being 0.997. Then, the third round of distillation is conducted with a decreased span of 0.2 for each parameter. We finally obtain a corrected PSF's Strehl ratio of 0.998. In these three rounds of distillation, the number of initial sampling points decreases from 100, 50 to 10 in order to reduce the exploration and make use of previous results. Five sets of random initial coefficients are implemented and the Strehl ratio reaches 0.997 ± 0.002 taking 6.36 ± 0.45 min from start to end. The image loading time t_{load} and projection time t_{proj} of DMD, and PSF data acquisition time t_{acq} are also considered; $t = t_{\text{load}} + t_{\text{proj}} + t_{\text{acq}} \approx (10\,\text{ms} + 1\,\text{ms} + 1\,\text{ms}) \times 460 = 5.5\,\text{s}$. They are small compared to BOWFS running time in the computer.

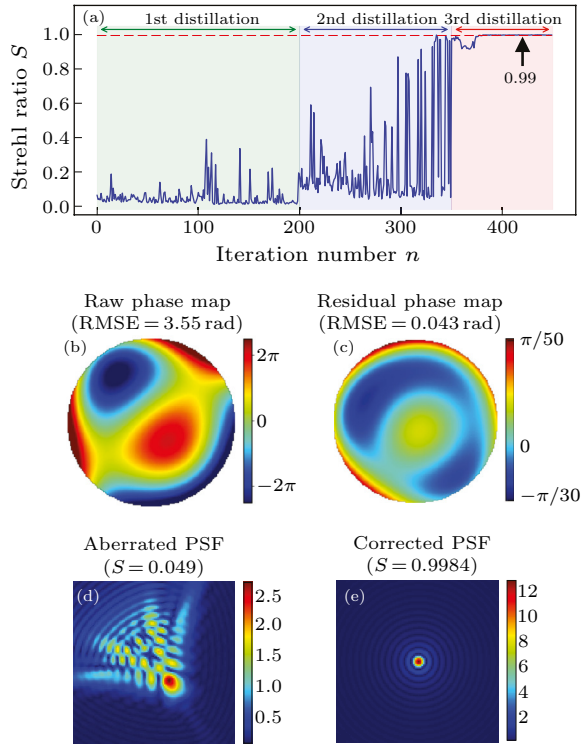

Fig. 3. Results of multi-distillation of BOWFS for large aberration. (a) Three rounds of the distillation of BOWFS. The Strehl ratio improves after each round of distillation and eventually surpasses 0.99 (red dashed line) at the 3rd distillation. [(b), (c)] Raw phase map and Residual phase map. [(d), (e)] Raw stretched PSF and corrected stretched PSF.

We also conduct an experiment to verify the performance of BOWFS. The setup is illustrated in Fig. 4(a). Phase maps are encoded into computer-generated holograph (CGH) patterns with the Lee method[31] and loaded on digital micromirror device (DMD) with a resolution of 1024×768 (DLP7000). The DMD region that illuminated by the beam is 440×440 pixels with each pixel mirror's size of $13.7\,\mu\text{m} \times 13.7\,\mu\text{m}$. The diameter of entrance pupil is about 6 mm and the effective focus of lens is 1300 mm. Only the first-order diffraction beam of CGH can be modulated.[8] It is filtered and focused on a CMOS camera (Andor Zyla 4.2). The diffraction limit spot has a diameter of $195.6\,\mu\text{m}$. The pixel size of the CMOS camera is $6.5\,\mu\text{m}$, which is sufficient to sample the Airy spot. The laser's wavelength in this experiment is 369.5 nm.

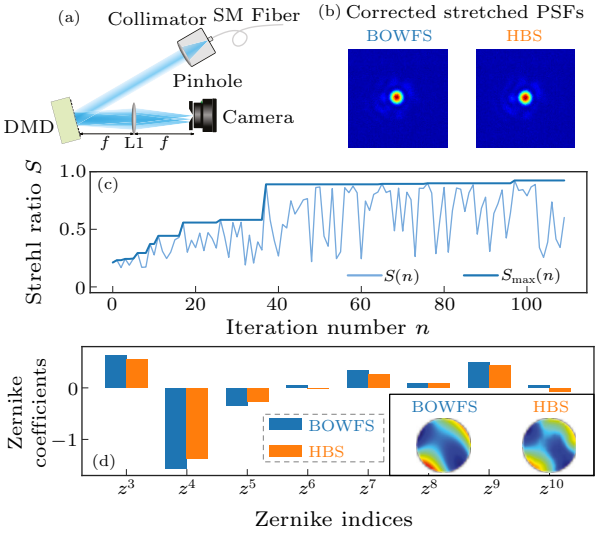

Fig. 4. Experimental results. (a) Setups of wavefront sensing. The collimated Gaussian beam is diffracted by holographic pattern on DMD. The first order beam is focused by a single lens L_1, then passes through the pinhole and is measured by a camera. (b) Corrected stretched PSFs using BOWFS and HBS (holographic beam shaping). (c) Single distillation Bayesian optimization process. (b) Zernike coefficients result obtained by BOWFS and HBS. Only up to tenth Zernike coefficients of HBS is shown. Inset: measured phase maps using BOWFS and HBS ranging from $-\pi$ to 2π.

We define the normalized energy E_{norm} as the output of BBF instead of Strehl ratio. Normalized energy is defined as follows:

$$E_{\text{norm}} = \sum_{\text{valid pixels}} \frac{I}{I_{\max}} = \frac{E}{I_{\max}}, \tag{4}$$

where I is the intensity on camera, and $I_{\max}$ is the maximum value among all valid pixels. The window on the camera is set large enough to cover most of the incident light. The pixels are preprocessed by removing bad pixels to obtain valid pixels. E_{norm} is inversely proportional to Strehl ratio. Compared with Strehl ratio defined as intensity ratio, E_{norm} would be less

affected by the power fluctuation of the light source. The finally optimized parameter is $-E_{\rm norm}$ because the algorithm only optimizes the maximum term.

The result is shown in Fig. 4. The algorithm successfully senses the phase with a Strehl ratio of 0.882 ± 0.017 within 97 steps. The whole process (iterating 110 steps) takes 141 s, where the BO computing time takes about 139 s. The blue line rises with fluctuation, which is in agreement with the previous simulation results in Fig. 2(a). The BOWFS algorithm finds an accurate answer (Strehl ratio = 0.863) in the 37th step and begins to stagnate for 60 steps until it locates a slightly better parameter in the 97th step.

We also implement the holographic beam shaping (HBS) method (refer to Ref. [8]) to verify the results of our approach. Each sample patch and reference patch are turned on to interfere on the image plane. The sample patch's phase is measured through a three-step phase shift technique. The phase map can be measured by scanning the sample patch over the DMD area. In our experiment, the patch size is 40×40 pixels. To acquire a complete phase map, $11 \times 11 \times 3 \times 2 = 726$ CGH patterns are loaded in sequence. Then the phase map is unwrapped and interpolated to make it continuous and smooth. The Zernike coefficients are calculated from the measured phase, which is consistent with the result using the BOWFS method [see Fig. 4(b)]. The Strehl ratio with this method is 0.863 ± 0.014, which is slightly lower than the BO's accuracy. However, it is 7 times slower than the BO method. Another advantage of the BO method is that it requires less light intensity compared with the HBS. The latter requires that only two small patches turn on at each time, so the light power in the camera is $11 \times 11 = 121$ smaller than that of BOWFS, which indicates a light threshold two magnitudes higher than our method. This method is applicable in systems where aberration is static or varies slowly, such as fixed optical systems in AMO experiments and microscopy, to eliminate aberration and to improve the fidelity of quantum state manipulation. On the other hand, there are also some limitations. BOWFS is not suitable to scenarios with dynamical aberration such as astronomical observations or with severely distorted amplitude distribution like laser speckle noise. In systems with very large aberrations, such as in biological tissues, the convergence speed can be increased when combining BOWFS with the Gerchberg–Saxton (GS) algorithm because the former is ideally suited for searching a fairly accurate phase estimate as the initial input for the latter.

In summary, we have demonstrated Bayesian optimization for image-based wavefront sensing and error correction from the perspective of simulation and experiment. For small aberration, BO converges within several minutes and obtains high accuracy with over 0.999 Strehl ratio in simulation. For large aberration, the BO method is also efficient to converge by adjusting the parameter space. This approach would contribute to the toolkit of image-based wavefront shaping and adaptive optics.

References

[1] Wilding D, Pozzi P, Soloviev O, Vdovin G and Verhaegen M 2016 *Opt. Express* **24** 24896
[2] Tao X, Fernandez B, Azucena O, Fu M and Garcia D 2011 *Opt. Lett.* **36** 1062
[3] Engler B, Weddell S and Clare R 2017 *International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Christchurch, New Zealand, 4–6 December 2017, pp 1–7
[4] Andersen T, Owner-Petersen M and Enmark A 2019 *Opt. Lett.* **44** 4618
[5] Yoon J W, Jeon C, Shin J, Lee S K and Lee H W 2019 *Opt. Express* **27** 20412
[6] Kudryashov A, Samarkin V, Alexandrov A, Rukosuev A and Zavalova V 2005 *Adaptive Optics for Industry and Medicine* (Berlin: Springer) p 237
[7] Zhao T, Deng L, Wang W, Elson D S and Su L 2018 *Opt. Express* **26** 20368
[8] Zupancic P, Preiss P M, Ma R, Lukin A and Tai M E 2016 *Opt. Express* **24** 13881
[9] Platt B C and Shack R 2001 *J. Refractive Surg.* **17** S573
[10] Fienup J R 1993 *Appl. Opt.* **32** 1737
[11] Song H, Fraanje R, Schitter G, Kroese H and Vdovin G 2010 *Opt. Express* **18** 24070
[12] Fienup J R 1982 *Appl. Opt.* **21** 2758
[13] Zhu W, Chen L, Gu C, Wan J and Zheng D 2015 *Appl. Opt.* **54** 6155
[14] Wang J, Léger J F, Binding J, Boccara A C and Gigan S 2012 *Biomed. Opt. Express* **3** 2510
[15] Carrara D A, Thelen B J and Paxman R G 2000 *Proc. SPIE* **4123** 56
[16] Thelen B J, Paxman R G, Carrara D A and Seldin J H 2009 *J. Opt. Soc. Am. A* **26** 206
[17] Paine S W and Fienup J R 2018 *Opt. Lett.* **43** 1235
[18] Xu Y, He D, Wang Q, Guo H and Li Q 2019 *Sensors* **19**
[19] Nishizaki Y, Valdivia M, Horisaki R, Kitaguchi K and Saito M 2019 *Opt. Express* **27** 240
[20] Allan G, Kang I, Douglas E S, Barbastathis G and Cahoy K 2020 *Opt. Express* **28** 26267
[21] Frazier P I 2018 arXiv:1807.02811 [stat.ML]
[22] Snoek J, Larochelle H and Adams R P 2012 *Advances in Neural Information Processing Systems 25* (Curran Associates, Inc.) p 2951
[23] Calandra R, Seyfarth A, Peters J and Deisenroth M P 2016 *Ann. Math. Artificial Intell.* **76** 5
[24] Ekström A, Forssén C, Dimitrakakis C, Dubhashi D and Johansson H T 2019 *J. Phys. G: Nucl. Part. Phys.* **46** 095101
[25] Ross T S 2009 *Appl. Opt.* **48** 1812
[26] Shahriari B, Swersky K, Wang Z, Adams R P and de Freitas N 2016 *Proc. IEEE* **104** 148
[27] Snoek J, Larochelle H and Adams R P 2012 *Proc. 25th International Conference on Neural Information Processing Systems* (Curran Associates Inc.) vol 2 p 2951
[28] https://github.com/fmfn/BayesianOptimization
[29] Unterhorst H A and Rubin A 2015 *Afr. Vision Eye Health* **74** 6
[30] Delabie T, Schutter J D and Vandenbussche B 2014 *J. Astronaut. Sci.* **61** 60
[31] Lee W H 1974 *Appl. Opt.* **13** 1677