



Adaptive stochastic parallel gradient descent approach for efficient fiber coupling

QINTAO HU,^{1,2,3} LIANGLI ZHEN,⁴ YAO MAO,^{1,2,3,*} SHIWEI ZHU,³ XI ZHOU,^{1,2} AND GUOZHONG ZHOU^{1,2}

¹Key Laboratory of Optical Engineering, Chinese Academy of Sciences, Chengdu, China

²Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China

³University of Chinese Academy of Sciences, Beijing, China

⁴Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore

*maoyao@ioe.ac.cn

Abstract: In high-speed free-space optical communication systems, the received laser beam must be coupled into a single-mode fiber at the input of the receiver module. However, propagation through atmospheric turbulence degrades the spatial coherence of a laser beam and poses challenges for fiber coupling. In this paper, we propose a novel method, called as adaptive stochastic parallel gradient descent (ASPGD), to achieve efficient fiber coupling. To be specific, we formulate the fiber coupling problem as a model-free optimization problem and solve it using ASPGD in parallel. To avoid converging to the extremum points and accelerate its convergence speed, we integrate the momentum and the adaptive gain coefficient estimation to the original stochastic parallel gradient descent (SPGD) method. Simulation and experimental results demonstrate that the proposed method reduces 50% of iterations, while keeping the stability by comparing it with the original SPGD method.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Free space optical communication (FSOC), which is a high-speed alternative communication technology between satellites, has attracted increasing attention of researchers [1–5]. In FSOC system, the remote distance between satellites and the tiny shake occurring on the transmitter result in severe jitter of the beam and degrade the spatial coherence of a laser beam, thus making the quality of the link decrease dramatically [6]. Ideally, the received laser beam must be coupled into single-mode fiber (SMF) at the input of the receiver module. If the beam fluctuates owing to outer turbulence, the wavefront is introduced with tip/tilt aberration and mismatch with the field of SMF. In consequence, the power of beam coupled into SMF, *i.e.*, the coupling efficiency (CE), is decreased [7–9]. Generally, adaptive optics is an effective method to compensate for the wavefront aberration. The fast steering mirror (FSM) is the primary control unit for steering the beam from the laser to improve CE in the fiber coupling system [10–13].

Due to the complexity of the system, it is challenging to formulate the fiber coupling system explicitly. As a result, researchers usually treat it as a red-box system and formulate the fiber coupling as a model-free optimization problem. Various approaches have been proposed to perform fiber coupling. For instance, stochastic gradient descent (SGD) [14], hill climbing [15], and random search methods [15]. However, these methods all optimize the controlling variables sequentially, which dramatically limits its efficiency on fiber coupling.

In order to accelerate the optimization process, the stochastic parallel gradient descent (SPGD) method is adopted to achieve fiber coupling in parallel. SPGD is firstly adopted by Vorontsov *et al.* for adaptive optical problems in 1997 [14]. Since then, many applications of the SPGD method have been presented [16–23]. However, the SPGD method may converge to local extremum points and its converge speed can be extremely slow [22], which limits its application in real-world applications, especially in complex systems. In recent years, a few attempts have been conducted

to speed-up the convergence and/or avoid converging to the local extremum points. For example, in 2012, Chen *et al.* improved the SPGD method for satellite-to-ground laser communication links [18]. In 2013, Geng *et al.* proposed the divergence cost function method, where divergence cost function was proposed as a merit function for SPGD method [19]. In 2015, Wu *et al.* proposed the multi-perturbation SPGD method with faster convergence than the original SPGD method [20]. In 2017, Yang *et al.* improved the SPGD method to avoid local extremum points for incoherent beam combination [21]. In 2018, Huang *et al.* deployed the precisely-delayed SPGD method for adaptive SMF coupling in the free space optical communication [22]. Although these methods have achieved promising results, most of them were proposed for specific optical problems and cannot be adopted to achieve efficient fiber coupling directly.

In this paper, we propose a novel method, called adaptive stochastic parallel gradient descent (ASPGD), to achieve efficient fiber coupling. Specifically, inspired by the Adam optimizer [24,25], which is widely used to optimize the connection weights of deep neural networks, we integrate the momentum and the adaptive gain coefficient estimation to the original SPGD method. The novelty and the main contribution of this work are two-fold: 1) An improved SPGD method is proposed to solve the model-free optimization problem in parallel. It is capable of escaping local extremum points and accelerating convergence. At the same time, it sets the corresponding gain coefficients for different controlling variables adaptively, which makes ASPGD more robust to the learning rate; and 2) we apply the proposed ASPGD method to achieve efficient fiber coupling in a real-world system, which can further advance the FSOC research. Extensive simulation and experiments have been conducted. The simulation and experimental results demonstrate that the proposed method reduces not only 50% of iterations but also keeps the stability by comparing it with the original SPGD method, which verifies the effectiveness and efficiency of our proposed method.

2. Our proposed approach

2.1. Problem formulation

In FSOC between satellites, the vibration of satellite platform where FSO terminals are mounted induces wavefront tip-tilt aberration into the beam, degrading the beam coupling efficiency (CE) into the single-mode fiber. Fortunately, the optical fiber coupling has proven to be a significant technique for adaptive optical tasks [14], which can effectively improve the fiber CE of the system. As shown in Fig. 1, after the reflection of the mirror and the disturbance of disturbing fast steering mirror (FSM), the laser enters the energy meter after the correction of coupling FSM. The disturbing FSM is used to simulate atmospheric turbulence and satellite, the vibration of the satellite platform. The coupling FSM is the primary control unit for steering the beam from the laser into SMF to improve CE in the fiber coupling system, power meter as the sensor measures the coupling energy of optical fiber. The goal of fiber coupling is to control the FSM to reach the maximal coupling energy by adjusting the controlling variables. To formulate this fiber coupling system, we take the power meter measurement as the objective function J , which is associated with the FSM voltage parameters u_1 and u_2 as $J = g(u_1, u_2)$. Even the function of g is not explicitly defined, its result can be obtained by reading the power meter, and it is assumed to be differentiable w.r.t the FSM voltage parameters u_1 and u_2 [23]. Thus, the fiber coupling can be achieved by searching the optimal FSM voltage parameters to maximize J .

2.2. ASPGD

The original SPGD method is widely used in AO for correcting the spot jitter error caused by atmospheric turbulence and mechanical jitter at the receiving equipment to maximize the power meter reading result. It is also can be used to solve our formulated fiber coupling problem. In the original SPGD method, the gradient estimation of the objective function is realized by

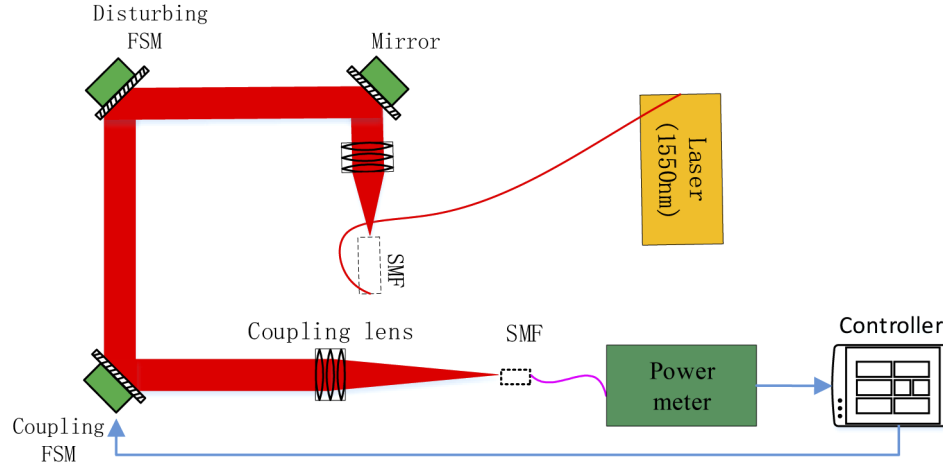


Fig. 1. A typical fiber coupling system. The laser enters the power meter after being reflected by FSM, and the controller controls FSM to maximize the power meter measurement value.

applying random disturbances $\Delta u_1, \Delta u_2, \dots, \Delta u_m$ to the controlling variables of the function, u_1, u_2, \dots, u_m , simultaneously. The disturbances $\Delta u_1, \Delta u_2, \dots, \Delta u_m$ have fixed amplitude, *i.e.*, $|\Delta u_k| = \Delta u$ for $k \in \{1, 2, \dots, m\}$, where m denotes the number of the controlling variables.

Following [26], we define the change in objective function as

$$\Delta J = J(u_1 + \Delta u_1, u_2 + \Delta u_2, \dots, u_m + \Delta u_m) - J(u_1, u_2, \dots, u_m). \quad (1)$$

By using the Taylor expansion, we can rewrite Eq. (1) as follows:

$$\Delta J = \sum_{i=1}^m \frac{\partial J}{\partial u_k} \Delta u_k + O([\Delta u_k]), \quad (2)$$

which yields the following approximation if we ignore the higher order terms:

$$\Delta J \cdot \Delta u_k = \frac{\partial J}{\partial u_k} (\Delta u_k)^2 + \sum_{i>0, i \neq k}^m \left(\frac{\partial J}{\partial u_k} \cdot \frac{\Delta u_i}{\Delta u_k} \right). \quad (3)$$

Reference [26] points out that the last term of Eq. (3) has expectation value of zero for random and independently distributed since the term $(\Delta u_k)^2$ is equally Δu^2 . Thus, we can approximate the gradient by disturbing all variables simultaneously as

$$g_k = \frac{\partial J}{\partial u_k} = \frac{\Delta J \cdot \Delta u_k}{(\Delta u)^2}. \quad (4)$$

We note that the learning scheme of the original SPGD method can be very slow when there is a long and narrow valley in the objective function surface. In such a situation, the direction of the gradient is almost perpendicular to the long axis of the valley. Thus, the optimizer would oscillate forth and back in the direction of the short axis and moves very slowly along the long axis of the valley. Inspired by [27] and [24], we first introduce the momentum term into the SPGD method to accelerate its convergence. Mathematically, we compute the first momentum of the current time step as:

$$m_k^t = \beta_1 m_k^{t-1} + (1 - \beta_1) \cdot g_k^t, \quad (5)$$

where m_k^{t-1} stands for the momentum of the last time step, and β_1 is a scalar hyper-parameter

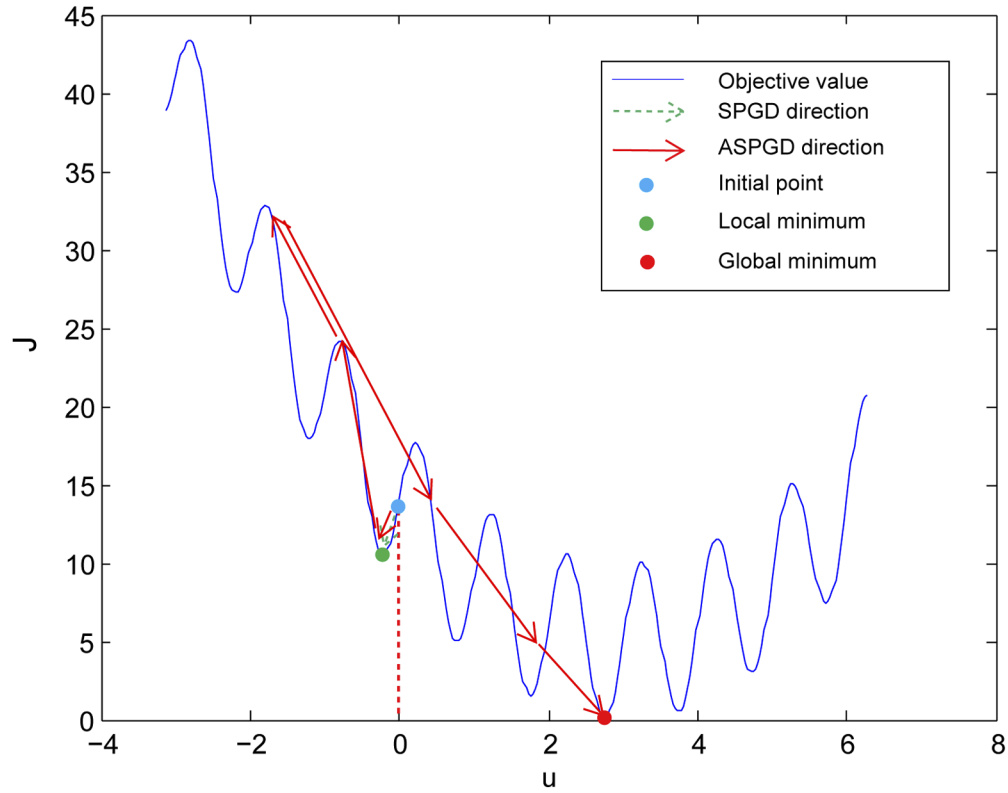


Fig. 2. The optimization trace of SPGD and our ASPGD for minimizing the objective function in Eq. (12), which has many local minimum. SPGD falls into local minimum and our method can find the global minimal (Best viewed in color).

controlling the decay rates of the past momentum. The momentum depends on both the current gradient and the previous gradients. This manner helps average out the oscillation along the short axis while adds up contributions along the long axis [27].

Furthermore, the original SPGD method adopts a united gain rate for all the optimizing parameters. It would be difficult to search for a suitable gain rate value in the real-world fiber coupling systems. By following [25], we adjust the gain rate for different parameters in SPGD by involving a second momentum term as follows:

$$v_k^t = \beta_2 v_k^{t-1} + (1 - \beta_2) \cdot (g_k^t)^2, \quad (6)$$

where v_k^t stands for the second momentum of the past time step and β_2 is a scalar hyper-parameter controlling the decay rates of the second momentum in the last step. This term sums up the weighted square results of the past gradients, which indicates the uncentered variance of the gradients. In the learning process, we adjust the learning step by dividing the second momentum term. In consequence, we update the parameters as follows:

$$u_k^t = u_k^{t-1} - \alpha m_k^t / (\sqrt{v_k^t} + \varepsilon), \quad (7)$$

where ε is a small number to avoid numerical problems, and we typically set it as 10^{-8} .

It can be seen that the updating rule in Eq. (7) makes the momentum biased towards the initial value of the momentum at $t = 0$, especially when β_1 and β_2 close to 1. To address this issue, [25]

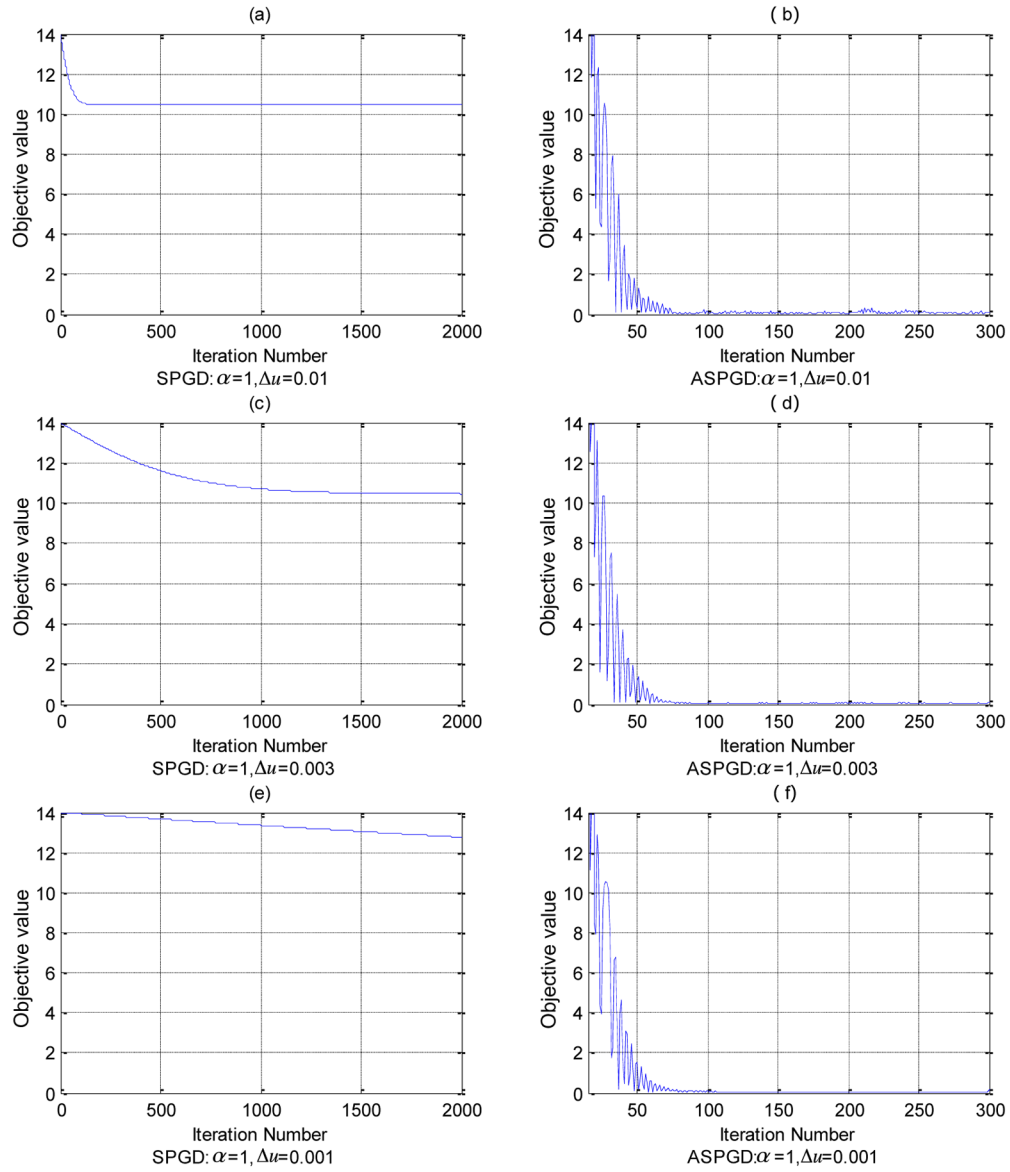


Fig. 3. Comparison between SPGD and ASPGD under different parameter settings. The ASPGD method can quickly converge to the global minimum under different parameters. However, the SPGD method converges to the local extrema or does not converge under all the three settings.

has proposed a correction strategy to estimate the bias-corrected estimates of the momentum values as:

$$\begin{aligned}\hat{m}_k^t &= m_k^t / [1 - (\beta_1)^t], \\ \hat{v}_k^t &= v_k^t / [1 - (\beta_2)^t].\end{aligned}\tag{8}$$

The derivation of Eq. (8) can be find in [25]. Let us initialize the momentum value as zero. Then, the first momentum at time step t can be written as:

$$\begin{aligned} m_k^t &= (1 - \beta_1) \sum_{i=1}^t \beta_1 \cdot g_k^i, \\ v_k^t &= (1 - \beta_2) \sum_{i=1}^t \beta_2 \cdot (g_k^i)^2. \end{aligned} \quad (9)$$

Taking expectations of the both sides of Eq. (9), we have

$$\begin{aligned} \mathbb{E}(m_k^t) &= \mathbb{E} \left[(1 - \beta_1) \sum_{i=1}^t \beta_1 \cdot g_k^i \right]; \\ \mathbb{E}(v_k^t) &= \mathbb{E} \left[(1 - \beta_2) \sum_{i=1}^t \beta_2 \cdot (g_k^i)^2 \right], \end{aligned} \quad (10)$$

which gives

$$\begin{aligned} \mathbb{E}(m_k^t) &= \mathbb{E}(g_k^t) \cdot [1 - (\beta_1)^t] + \eta_1; \\ \mathbb{E}(v_k^t) &= \mathbb{E}[(g_k^t)^2] \cdot [1 - (\beta_2)^t] + \eta_2, \end{aligned} \quad (11)$$

where $\eta_1 = 0$ if the true first momentum is stationary; otherwise η_1 can be kept small [25]. Similar result can be obtained for η_2 . To correct the discrepancy between $\mathbb{E}(m_k^t)$, $\mathbb{E}(v_k^t)$ and $\mathbb{E}(g_k^t)$, $\mathbb{E}[(g_k^t)^2]$, we need to conduct the initialization bias correction via Eq. (8).

Algorithm 1 The procedure of our proposed ASPGD method

Require: The gain rate α , the exponential decay rates for the moment estimates β_1 and β_2 , the stochastic objective function $J(u_1, u_2, \dots, u_m)$, the amplitude Δu , and the maximal number of learning iterations T .

Ensure: The optimal parameters u_1, u_2, \dots, u_m .

```

1:  $m_0 \leftarrow 0$                                 ▶ Initialize the first moment vector
2:  $v_0 \leftarrow 0$                                 ▶ Initialize the second moment vector
3:  $\varepsilon \leftarrow 10^{-8}$ 
4: Randomly initialize the parameters  $u_1^0, u_2^0, \dots, u_m^0$ .
5: for  $t = 1, \dots, T$  do                                ▶ Different time steps
6:   Randomly generate the values of  $\Delta u_1, \Delta u_2, \dots, \Delta u_m$  based on the value of  $\Delta u$ .
7:    $\Delta J \leftarrow J(u_1^{t-1} + \Delta u_1, u_2^{t-1} + \Delta u_2, \dots, u_m^{t-1} + \Delta u_m) - J(u_1^{t-1}, u_2^{t-1}, \dots, u_m^{t-1})$ 
8:   parfor  $k = 1, 2, \dots, m$  do                                ▶ Parallel for loop
9:      $g_k^t \leftarrow \frac{\Delta J \cdot \Delta u_k^t}{(\Delta u)^2}$ 
10:     $m_k^t \leftarrow \beta_1 m_k^{t-1} + (1 - \beta_1) \cdot g_k^t$ 
11:     $v_k^t \leftarrow \beta_2 v_k^{t-1} + (1 - \beta_2) \cdot (g_k^t)^2$ 
12:     $\hat{m}_k^t \leftarrow m_k^t / [1 - (\beta_1)^t]$ 
13:     $\hat{v}_k^t \leftarrow v_k^t / [1 - (\beta_2)^t]$ 
14:     $u_k^t \leftarrow u_k^{t-1} + \alpha \hat{m}_k^t / (\hat{v}_k^t + \varepsilon)$ 
15:   end parfor
16: end for

```

The details of the learning procedure of ASPGD are summarized in Algorithm 1. It is notable that the code block in Lines 9 - 14 is executed in parallel for different values of k . The maximal

number of learning iterations is taken as the termination condition in this work and is typically set as 100.

To intuitively illustrate the effectiveness of the ASPGD method, we apply it to minimize the objective function:

$$J = (u - 3)^2 + \sin(2\pi u) + 5. \quad (12)$$

As shown in Fig. 2, it can be seen that it has many local minimum points, and its global minimum value is 0. For comparison, we use SPGD to optimize the objective function J as well, and three sets of parameters are evaluated for each method.

The values of the objective function obtained by the two methods during the optimization process are shown in Fig. 3, where the left column shows the results of SPGD and the right column shows the results of ASPGD, and different rows display the results under different parameter settings. From the simulation results, we can see that the SPGD method can converge quickly when the parameters are appropriately provided, but it falls into the local minimum (Fig. 3(a)). When the parameters are changed from $(\Delta u = 0.01$ to $\Delta u = 0.003$, $\Delta u = 0.001$), its convergence speed is reduced (Fig. 3(c) and Fig. 3(e)). In contrast, the ASPGD method can converge quickly within 100 iterations and reach the global minimum under all the three parameter settings (Fig. 3(b), Fig. 3(d) and Fig. 3(f)). The comparison demonstrates that ASPGD can accelerate the convergence speed and improve the capability to reach the global minimum. Also, it shows that the proposed method is robust to the hyper-parameter Δu .

3. Simulation

3.1. SMF coupling efficiency

The scheme of SMF coupling is shown in Fig. 4. A beam propagates through an aperture with a diameter of d located at plane A, and is focused via an optical lens with a focal length of f . The tip of the stationary SMF is mounted at the focal plane signed as plane B. The SMF mode field at plane B can be approximated as a Gaussian beam with 1% error. The symbol of λ is the wavelength of the laser beam and ω_0 is the radius of SMF field. For convenience we consider the calculation of coupling efficiency η in plane A, which is defined as follows [28]:

$$\begin{aligned} \eta &= \left| \iint_A \sqrt{\frac{2}{\pi\omega_\alpha^2}} \exp\left(-\frac{r^2}{\omega_\alpha^2} - j\phi(r, \theta)\right) dr d\theta \right|^2 \\ &= \frac{2}{\pi\omega_\alpha^2} (a_r^2 + a_i^2), \end{aligned} \quad (13)$$

where $a_r = \iint_A \exp\left(-\frac{r^2}{\omega_\alpha^2}\right) \cos[\phi(r, \theta)] dr d\theta$, $a_i = \iint_A \exp\left(-\frac{r^2}{\omega_\alpha^2}\right) \sin[\phi(r, \theta)] dr d\theta$ and $\omega_\alpha = \frac{\lambda f}{\pi\omega_0}$.

In adaptive optical systems, Zernike polynomial is generally adopted to decompose the wavefront phase with distortion to the sum of weighted orthogonal polynomials, which represent various types of aberrations. The wavefront phase $\phi(r, \theta)$ can be expanded as [28]:

$$\phi(r, \theta) = a_0 + a_1 Z_1(r, \theta) + a_2 Z_2(r, \theta) + \sum_{i=3}^{\infty} a_i Z_i(r, \theta) \quad (14)$$

where $Z_i(r, \theta)$ denotes the i^{th} Zernike polynomial and a_i is the corresponding coefficient of polynomials. In the Zernike polynomials, the 0^{th} term with coefficient a_0 represents piston that is insignificant to SMF coupling, while $Z_1(r, \theta)$ and $Z_2(r, \theta)$ represent the tilt aberrations along x and y directions, respectively.

Tip/tilt error accounts for 87% of the total wavefront aberrations caused by the atmosphere turbulence [28]. In addition, the tracking system is based on the optical communication link in space with a thin atmosphere. Thus, in this work, we ignore the high-order aberrations and compensate only tip/tilt error caused by vibration and atmospheric turbulence.

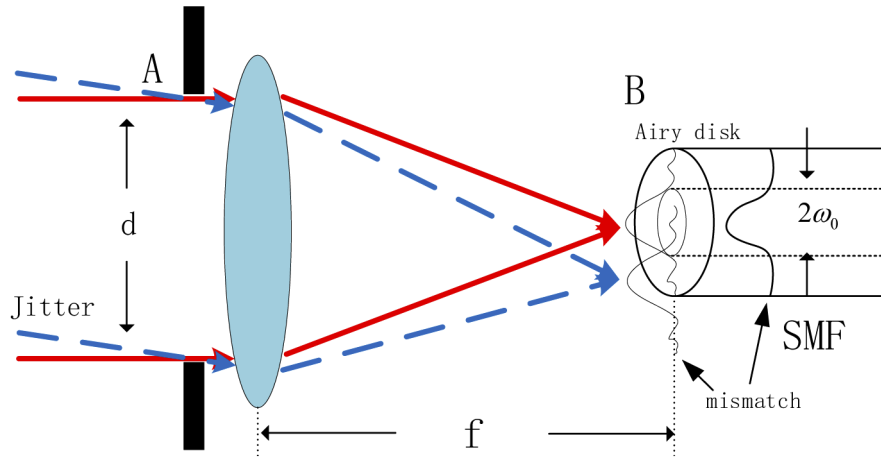


Fig. 4. The scheme of the SMF coupling system. The vibration of satellite platform and atmospheric turbulence affect the CE of the optical fiber.

3.2. Simulation analysis

In order to imitate slight atmosphere turbulence and inherent aberrations of the lens, Zernike polynomials with 10 terms is fabricated as the distorted wavefront. The initial coefficients for a_1 to a_{10} are given as 2, 2, 0.34, 0.2, 0.15, 0.12, 0.13, 0.16, 0.08 and 0.09, respectively. In the simulation, λ is set to 1550nm , f is 0.71m , ω_0 is $5.2\mu\text{m}$ and d is set to 0.15m . Since the control voltages of FSM have an approximately linear relationship with the coefficients a_1 and a_2 , we regulate a_1 and a_2 to equivalently simulate tip/tilt control of FSM. The normalized CE is used as the index rather than the absolute value of CE to observe the feature of the method more intuitively, and x-label is set as the motion times of FSM because of the fixed control frequency. The wavefront before the compensation is shown in Fig. 5(a) and the wavefront after the compensation is shown in Fig. 5(b). The normalized CE value obtained by using the compensation is 67.8%, which is much larger than the value of 3×10^{-4} before the compensation. PV means the peak value, and RMS is the root mean square. Clearly, most of the distortion has been well compensated. Note that to facilitate our observation, the simulation results treat the optimization objective as normalized coupling efficiency.

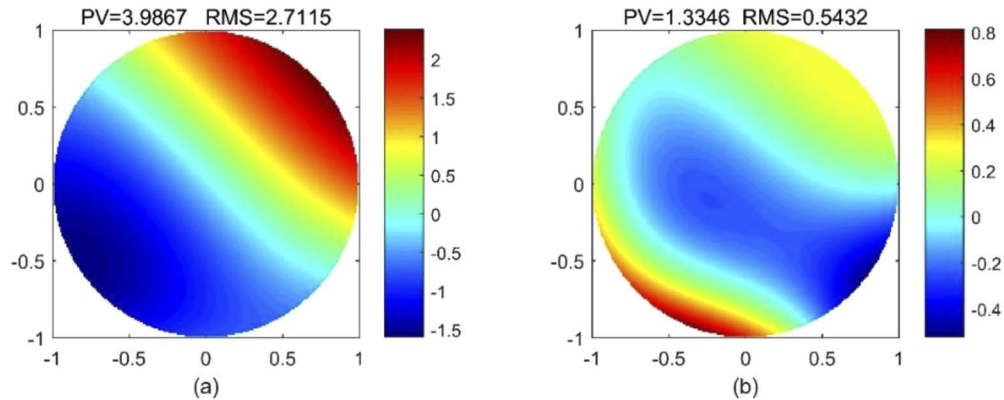


Fig. 5. The results of the wavefronts with/without compensation. (a) The wavefront before the compensation. (b) The wavefront after the compensation.

In the simulation, we use Eq. (13) as the optimization goal of SPGD and ASPGD, and control FSM by optimizing a_1 and a_2 of the Zernike coefficients. By considering the randomness of the method, we execute each method 200 times. First of all, we do experiments on two parameters β_1 and β_2 introduced by SPGD to find the optimal parameters. As shown in Fig. 6(a) and Fig. 6(b), we can see that the minimum convergent numbers under $\beta_1 = 0.2$ and $\beta_2 = 0.999$. Then, we use the setting of $\beta_1 = 0.2$ and $\beta_2 = 0.999$ for ASPGD and compare it with the SPGD in the simulation. Figure 7(a) and Fig. 7(b) show the optimization curves of SPGD and ASPGD under their optimal parameters, respectively.

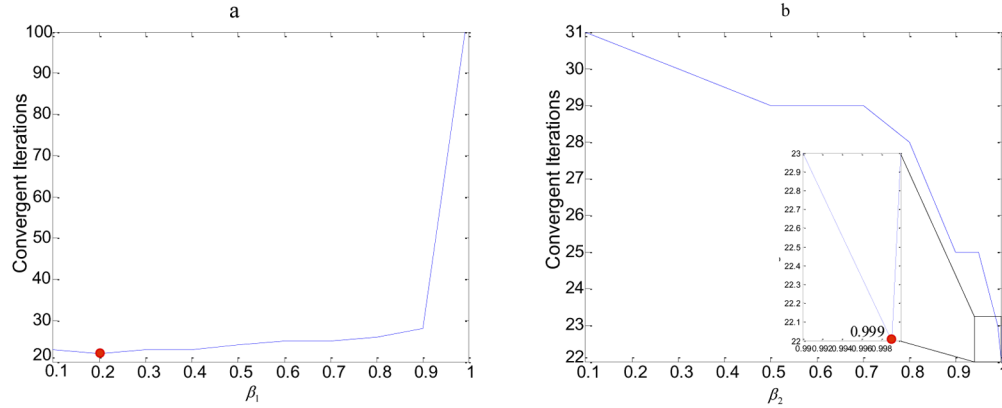


Fig. 6. The simulation results of convergent iteration and convergent normalized CE obtained by ASPGD under different β_1 and β_2 values.

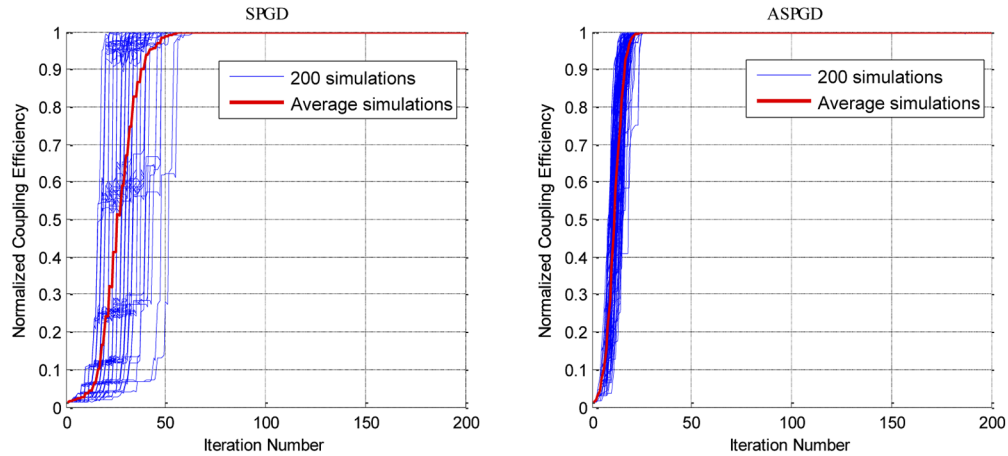


Fig. 7. Simulation comparison of SPGD and ASPGD on fiber coupling.

From Fig. 7, the SPGD method converges after at least 20 iterations, and in the worst case, it converges after up to 65 iterations, averaging at the number around 52 iterations. ASPGD converges to a fixed point after at least 11 iterations, and maximally 27 iterations. The average number of iterations for the convergence of ASPGD is about 22, which is less than half of the SPGD method. In addition, the results of SPGD merely depend on the random disturbance at each iteration and the current gradient, which fluctuates greatly. While the ASPGD method not only considers the current gradient information in the iteration process, but also the historical

gradient, thus effectively reducing the impact of randomness. Overall, ASPGD converges faster than SPGD, and it is more robust to the randomness of the disturbance.

To further compare the robustness of the two methods to the hyper-parameter Δu , we evaluate the two methods under the same setting as previous simulation except change the value of Δu from 0.01 to 0.015 and 0.005. The results are shown in Fig. 8, from which we can see that the SPGD method is extremely sensitive to Δu . When Δu becomes 0.015, SPGD almost diverges (Fig. 8(a)), while when Δu is reduced to 0.005, the convergence speed of SPGD is reduced twice (Fig. 8(c)). Differently, the ASPGD method still works well under $\Delta u \in 0.015, 0.005$.

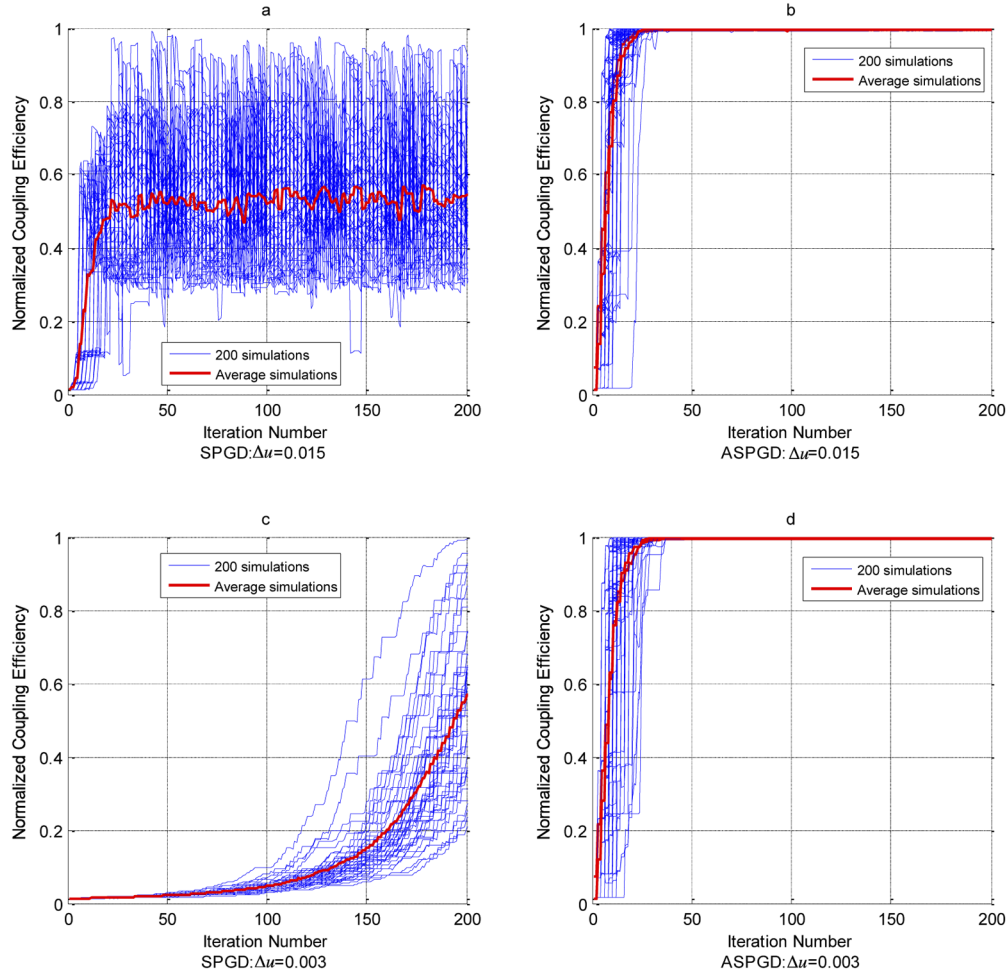


Fig. 8. Simulation comparison of SPGD and ASPGD under different values of Δu .

To explore the limitation of the ASPGD method, we adjust Δu from 0.00005 to 0.5. The results of convergent iteration and the normalized CE obtained by ASPGD are shown in Fig. 9, from which we can see that our ASPGD is able to converge to the optimal CE under $\Delta u \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1\}$ within 120 iterations, and it obtains the normalized CE of 80% under $\Delta u = 1$. The results show that ASPGD works well under a large range of Δu values, which makes it easily be applied to real-world applications for the users.

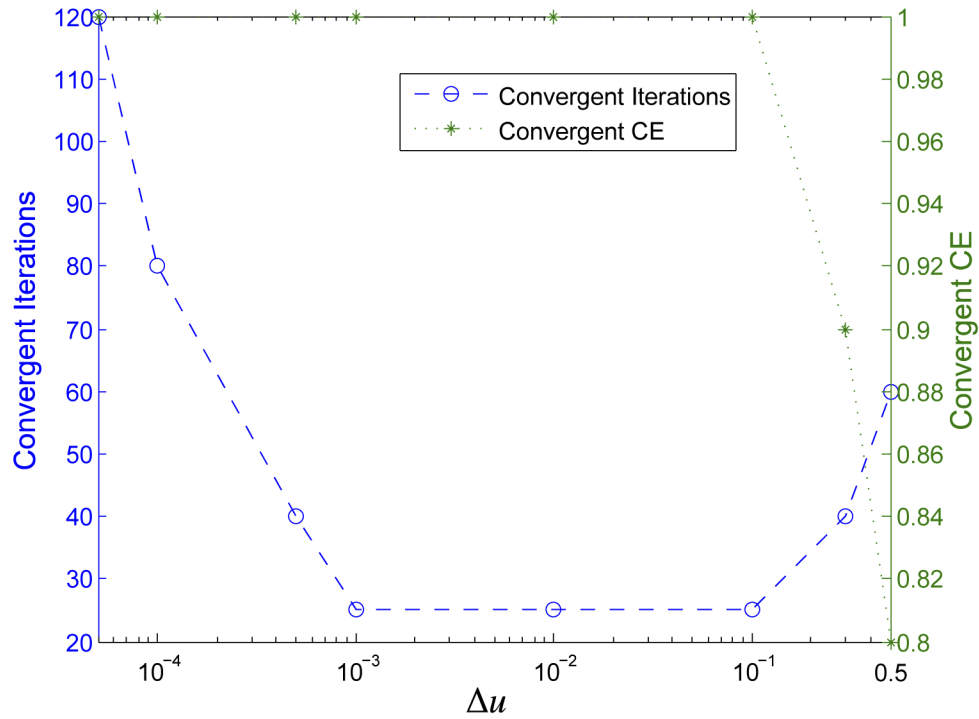


Fig. 9. The simulation results of convergent iteration and convergent normalized CE obtained by ASPGD under different Δu values.

4. Experiment

To further investigate the performance of ASPGD for fiber coupling, and verify the performance in real-world application systems, we compare the SPGD method with our ASPGD on a fiber coupling platform. It consists of a laser, an SMF, an FSM, and an optical power meter. The scheme and the experimental setup are shown in Fig. 1 and Fig. 10, respectively.

As shown in Fig. 1, the power meter is designed for receiving a light beam from the laser. The wavelength of the laser beam is 1550nm , the conversion coefficient of the optical power meter's output (voltage) and input (optical power) is measured to be 39.475V/mW , the diameter of fiber core is $9\mu\text{m}$ and the sampling frequency of the controller is 500Hz . The beam is reflected by FSM and enters the optical power meter. According to the variation of optical power, FSM is controlled to move tinily so as to calculate the gradient [4]. For a fair comparison, we evaluate the performance of SPGD and ASPGD under the same initial conditions.

We set the same initial point for both tested methods, and report the results with their corresponding optimal parameter values. The optimal setting for SPGD is $\Delta u = 1$, $\alpha = 7000$, and the optimal setting for ASPGD is $\Delta u = 1$, $\alpha = 50$, $\beta_1 = 0.2$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$. The experimental results are shown in Fig. 11, from which we find that the curve of the SPGD method rises slowly at the beginning and reaches the maximum upon about 130 iterations. However, the ASPGD method can dynamically adjust the gain according to the gradient value at the beginning to achieve rapid convergence. Finally, it reaches the maximum after about 50 iterations, which is much faster than SPGD.

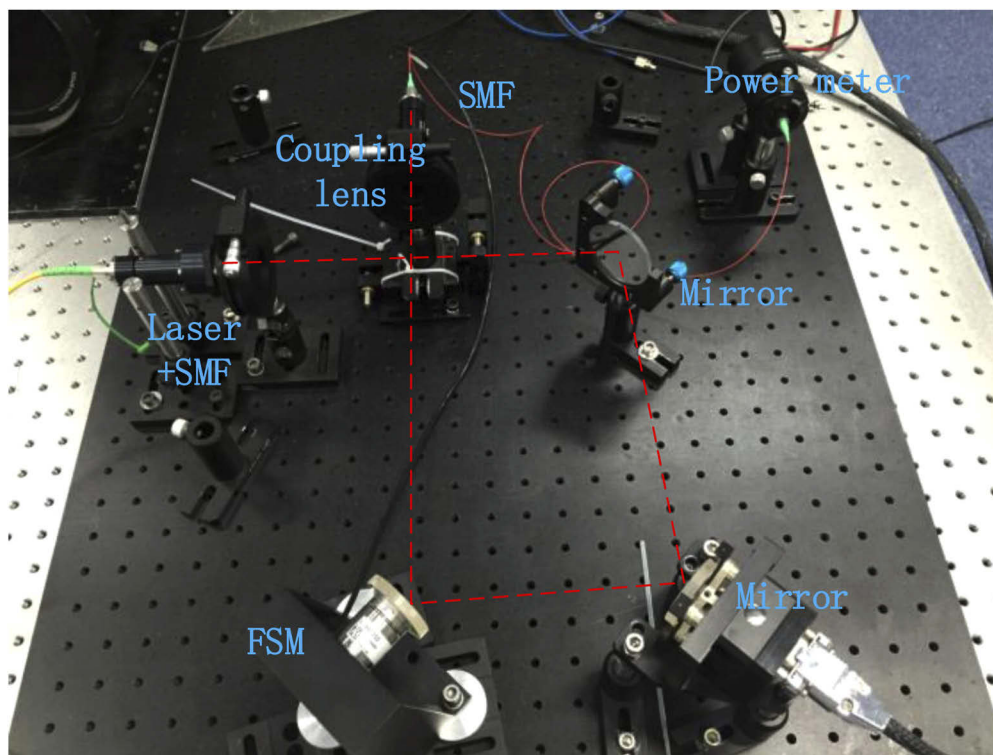


Fig. 10. A real-world fiber coupling system. It is constructed based on the fiber coupling scheme shown in Fig. 1.

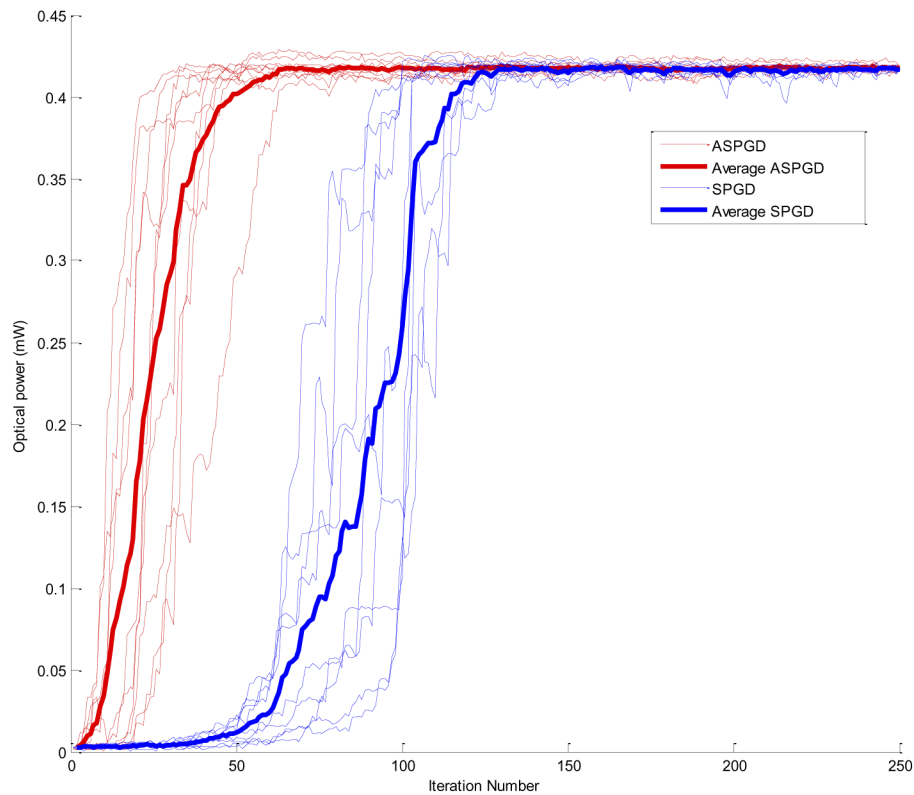


Fig. 11. Comparison of SPGD and ASPGD on the real-world fiber coupling system. The red curves indicate the results of SPGD, and the red curves indicate the results of ASPGD.

5. Conclusion

In this paper, an improved SPGD method (ASPGD) is proposed to achieve efficient fiber coupling. By integrating the momentum and adaptive gain coefficient estimation into the original SPGD, our proposed method is able to avoid converging to the local extremum points and accelerate the convergence speed. The simulation results show that the ASPGD method can improve the stability of the method and accelerate the convergence speed. Specifically, compared with SPGD, the iteration number of ASPGD is reduced by 50%. At the same time, the method is robust to parameter uncertainties and can converge for a wide range of parameters ($\Delta u = 0.00005 - 0.5$). At last, the effectiveness of the method is also evaluated on a real-world fiber coupling system. The experimental results show that our ASPGD converges much faster than the original SPGD method as well.

In the future, as a general optimization method, we will investigate how to apply the ASPGD method to more complex optical problems.

Funding

National Natural Science Foundation of China (NO.61905253).

Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions.

Disclosures

The authors declare no conflicts of interest.

References

1. X. Yi, Z. Liu, and P. Yue, "Optical scintillations and fade statistics for fso communications through moderate-to-strong non-kolmogorov turbulence," *Opt. Laser Technol.* **47**, 199–207 (2013).
2. A. v. Eekeren, K. Schutte, J. Dijk, P. Schwering, M. v. Iersel, and N. Doelman, "Turbulence compensation: an overview," *Proc. SPIE* **8355**, 83550Q (2012).
3. G. Huang, C. Geng, F. Li, Y. Yang, and X. Li, "Adaptive smf coupling based on precise-delayed spgd algorithm and its application in free space optical communication," *IEEE Photonics J.* **10**(3), 1–12 (2018).
4. J. Cao, X. Zhao, Z. Li, W. Liu, and Y. Song, "Stochastic parallel gradient descent laser beam control algorithm for atmospheric compensation in free space optical communication," *Optik* **125**(20), 6142–6147 (2014).
5. H. Endo, M. Fujiwara, M. Kitamura, O. Tsuzuki, T. Ito, R. Shimizu, M. Takeoka, and M. Sasaki, "Free space optical secret key agreement," *Opt. Express* **26**(18), 23305–23332 (2018).
6. N. Werth, M. S. Müller, J. Meier, and A. W. Koch, "Diffraction errors in micromirror-array based wavefront generation," *Opt. Commun.* **284**(9), 2317–2322 (2011).
7. H. Takenaka, M. Toyoshima, and Y. Takayama, "Experimental verification of fiber-coupling efficiency for satellite-to-ground atmospheric laser downlinks," *Opt. Express* **20**(14), 15301–15308 (2012).
8. W. Liu, W. Shi, J. Cao, Y. Lv, K. Yao, S. Wang, J. Wang, and X. Chi, "Bit error rate analysis with real-time pointing errors correction in free space optical communication systems," *Optik* **125**(1), 324–328 (2014).
9. Z. Guang and Y. Zhang, "Coupling ultrafast laser pulses into few-mode optical fibers: a numerical study of the spatiotemporal field coupling efficiency," *Appl. Opt.* **57**(33), 9835–9844 (2018).
10. D. Zheng, Y. Li, E. Chen, B. Li, D. Kong, W. Li, and J. Wu, "Free-space to few-mode-fiber coupling under atmospheric turbulence," *Opt. Express* **24**(16), 18739–18744 (2016).
11. N. Martínez, L. F. R. Ramos, and Z. Sodnik, "Simulating the performance of adaptive optics techniques on fso communications through the atmosphere," in *Laser Communication and Propagation through the Atmosphere and Oceans VI*, vol. 10408 (International Society for Optics and Photonics, 2017), p. 1040808.
12. A. E. Willner, Y. Ren, G. Xie, Y. Yan, L. Li, Z. Zhao, J. Wang, M. Tur, A. F. Molisch, and S. Ashrafi, "Recent advances in high-capacity free-space optical and radio-frequency communications using orbital angular momentum multiplexing," *Phil. Trans. R. Soc. A* **375**(2087), 20150439 (2017).
13. M. F. Coughlan and A. V. Goncharov, "Nonpupil adaptive optics for visual simulation of a customized contact lens," *Appl. Opt.* **57**(22), E57–E63 (2018).
14. M. Vorontsov, G. Carhart, and J. Ricklin, "Adaptive phase-distortion correction based on parallel gradient-descent optimization," *Opt. Lett.* **22**(12), 907–909 (1997).
15. A. J. Wright, D. Burns, B. A. Patterson, S. P. Poland, G. J. Valentine, and J. M. Girkin, "Exploration of the optimisation algorithms used in the implementation of adaptive optics in confocal and multiphoton microscopy," *Microsc. Res. Tech.* **67**(1), 36–44 (2005).
16. W. Xiong, W. Xiaolin, Z. Pu, X. Xiaojun, and S. Bohong, "Numerical simulation of tilt-tip control in coherent beam combining using spgd algorithm," *Opt. Laser Technol.* **48**, 343–350 (2013).
17. M. A. Vorontsov and V. Sivokon, "Stochastic parallel-gradient-descent technique for high-resolution wave-front phase-distortion correction," *J. Opt. Soc. Am. A* **15**(10), 2745–2758 (1998).
18. E. Chen, H. Cheng, Y. An, and X. Li, "The improvement of spgd algorithm convergence in satellite-to-ground laser communication links," *Procedia Eng.* **29**, 409–414 (2012).
19. C. Geng, W. Luo, Y. Tan, H. Liu, J. Mu, and X. Li, "Experimental demonstration of using divergence cost-function in spgd algorithm for coherent beam combining with tip/tilt control," *Opt. Express* **21**(21), 25045–25055 (2013).
20. K. Wu, Y. Sun, Y. Huai, S. Jia, X. Chen, and Y. Jin, "Multi-perturbation stochastic parallel gradient descent method for wavefront correction," *Opt. Express* **23**(3), 2933–2944 (2015).
21. G. Yang, L. Liu, Z. Jiang, T. Wang, and J. Guo, "Improved spgd algorithm to avoid local extremum for incoherent beam combining," *Opt. Commun.* **382**, 547–555 (2017).
22. G. Yang, L. Liu, Z. Jiang, J. Guo, and T. Wang, "Incoherent beam combining based on the momentum spgd algorithm," *Opt. Laser Technol.* **101**, 372–378 (2018).
23. V. I. Polejaev and M. A. Vorontsov, "Adaptive active imaging system based on radiation focusing for extended targets," in *Adaptive Optics and Applications*, vol. 3126 (International Society for Optics and Photonics, 1997), pp. 216–220.
24. N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.* **12**(1), 145–151 (1999).
25. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
26. J. Alspector, R. Meir, B. Yuhua, A. Jayakumar, and D. Lippe, "A parallel gradient descent method for learning in analog vlsi neural networks," in *Advances in neural information processing systems*, (1993), pp. 836–844.
27. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature* **323**(6088), 533–536 (1986).
28. R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**(3), 207–211 (1976).