A thorough study on genetic algorithms in feedback-based wavefront shaping

Daixuan Wu, Jiawei Luo, Zhaohui Li and Yuecheng Shen*

Key Laboratory of Optoelectronic Materials and Technologies
School of Electronics and Information Technology
Sun Yat-sen University, Guangzhou 510275, P. R. China
*shenyuecheng@mail.sysu.edu.cn

Received 24 March 2019
Accepted 5 May 2019
Published 4 June 2019

Feedback-based wavefront shaping focuses light through scattering media by employing phase optimization algorithms. Genetic algorithms (GAs), inspired by the process of natural selection, are well suited for phase optimization in wavefront shaping problems. In 2012, Conkey et al. first introduced a GA into feedback-based wavefront shaping to find the optimum phase map. Since then, due to its superior performance in noisy environment, the GA has been widely adopted by lots of implementations. However, there have been limited studies discussing and optimizing the detailed procedures of the GA. To fill this blank, in this study, we performed a thorough study on the performance of the GA for focusing light through scattering media. Using numerical tools, we evaluated certain procedures that can be potentially improved and provided guidance on how to choose certain parameters appropriately. This study is beneficial in improving the performance of wavefront shaping systems with GAs.

Keywords: Genetic algorithm; wavefront shaping; scattering media; adaptive optics.

1. Introduction
Optical focusing is essentially important in many applications of biophotonics, such as optical imaging, laser surgery, phototherapy, and optical manipulation. However, microscopic inhomogeneity inherent to biological tissue causes optical scattering, which prevents optical focusing from being achieved beyond 1 mm in soft tissue.1,2 These scattering effects have been long considered as the major obstacles that limit all these applications to a shallow depth.

To overcome this challenge, wavefront shaping technology has been developed with the capability to control and refocus scattered light. By treating optical scattering as a deterministic process, wavefront shaping technology finds a special wavefront, called optimum wavefront, to compensate scattering...
effects. There are three general approaches to obtain the optimum wavefront: optical phase conjugation (OPC),\textsuperscript{3–9} the inversion of transmission matrix (TM),\textsuperscript{10–14} and feedback-based wavefront shaping.\textsuperscript{15–18} Compared to the other two approaches, feedback-based wavefront shaping employs the simplest implementation. In a typical experiment, incident light is first modulated by a spatial light modulator (SLM) at the front side of a scattering medium. The modulated light is then directed by the scattering medium. At the backside of the scattering medium, a camera or a photodiode measures the light intensity at a targeted location, providing feedback signals. To form an optical focus at the targeted position through the scattering medium, the phase map displayed on the SLM is iteratively optimized to maximize the feedback signal.

In feedback-based wavefront shaping, various algorithms have been proposed and demonstrated to search for the optimum phase map, including a stepwise sequential algorithm (SSA),\textsuperscript{19} a continuous sequential algorithm (CSA),\textsuperscript{20} a partitioning algorithm (PA),\textsuperscript{21} a simulated annealing algorithm (SA),\textsuperscript{22} a genetic algorithm (GA),\textsuperscript{18,22,23} a particle swarm,\textsuperscript{22,24} a four-element division,\textsuperscript{25} a harmony search algorithms,\textsuperscript{26} and a Hadamard encoding algorithm (HEA).\textsuperscript{27} Each algorithm has its own pros and cons, which has been well discussed in many literatures.\textsuperscript{22,28,29} Among all these algorithms, the GA has been recognized to be particularly advantageous in noisy environment. As the name suggests, the GA is an iterative optimization algorithm that originates from Darwin’s theory of evolution and follows the basic scheme of “survival of the fittest”. The GA has been widely used in solving various complex problems through iterative fitting, leading to an optimum individual with the highest fitness. In 2012, Conkey et al. first introduced the GA into feedback-based wavefront shaping to focus light through scattering media. In that work, the crossover procedure was slightly modified by bringing in a random binary mask breeding system, which exhibits a higher convergence rate compared to the conventional GA. Due to its superior performance in noisy environment, the GA has been widely adopted. The detailed procedures, as well as the choice of parameters, are largely followed by successors. However, till to date, no studies have been reported on optimizing neither the detailed procedures nor the choices of parameters of the GA. To fill this blank, in this study, we perform a thorough study on optimizing the performance of the GA. In particular, we propose several new schemes to form new generations and select parents. These new schemes can facilitate the initial increasing rate of the enhancement. We also investigate the choices of certain parameters in the GA, including the parental mask ratio and the mutation rate. By comparing the performance of the GA with various combinations of parameters, we provide certain guidance on how to choose these parameters appropriately under different noise levels.

2. Genetic Algorithm

The flowchart of the GA used in feedback-based wavefront shaping is schematically shown in Fig. 1, which can be divided into eight different steps. The GA begins by generating an initial population of $NP$ phase masks. The fitness of each phase mask is evaluated as the light intensity at the targeted position. Then, two parents, $pa$ and $ma$, are selected from the population to breed offspring. There are several possible selection schemes, but the rule of thumb is that the individual with a higher fitness has a higher probability to be selected. With two parents being selected, a new offspring is created as $pa \cdot T + ma \cdot (1 - T)$, where $T$ is a random binary template generated with a parental mask ratio $pc$. The way of generating offspring indicates that information is inherited from two parents. For this offspring, a mutation operation is then performed. The mutation rate $pm$ can be either a fixed number or an adaptive variable. After mutation, the fitness of the new offspring is measured and recorded. The above steps of generating new offspring are repeated with a certain number of times, until a generational condition is fulfilled, which means that the new generation is complete. The GA is usually executed for many generations, and this generational process is repeated for a given number of times.

Among all these steps, at least a few of them (steps 3, 4, 6, and 8), which are enclosed in the purple boxes, need further discussion and can be potentially improved. For example, in step 3, Ref. 18 selects only two parents based on the roulette algorithm without providing further explanations. Similarly, in step 8, Ref. 18 directly supplements the new generation with $NP/2$ new offspring, regardless of their fitness. Given that there exist multiple schemes for both steps, we should at least implement some of them and
compare their performances. Moreover, in step 4, the random binary template $T$ in Ref. 18 is generated using a fixed parental mask ratio $pc = 0.5$. However, the reason of this choice has never been explained. Similarly, in step 6, the choices of the parameters in the mutation operation are not explained in Ref. 18 as well. As a result, we will test different choices of these parameters in the GA and provide certain guidance on how to properly choose them.

### 3. Simulation Environment

Throughout this study, numerical tools are used to evaluate the performance of this algorithm. In this section, we describe the simulation environment used in this study. Unless otherwise stated, the following parameters will be used throughout the entire study. For each phase mask, the number of independent elements is $16 	imes 16 = 256$, which is also the number of genes in the GA. To mimic a phase-only modulation scheme, all the elements have the same amplitude of 1. For the initialization process, the phase of each element in the phase mask is randomly distributed between 0 and $2\pi$. Using random matrix theory, the scattering medium is modeled as a TM with a dimension of $256 \times 256$. Each element of this matrix is drawn from a circular Gaussian distribution. The scattered light is then computed by acting the TM on the input light field (the phase mask). During simulation, we have properly normalized the incident light field and the TM, so that the average intensities for both the incident light and the scattered light equal 1. Therefore, the enhancement, defined as the ratio between the light intensity at the targeted position and the initial average intensity, can be calculated by simply taking the absolute square of the target element in the scattered light. In this study, the element in the central field of the scattered light is chosen as the target for simplicity.

For the GA, the fitness of a given input phase mask is the same as the enhancement defined above. The population of each generation is fixed at $NP = 50$. The generational process is repeated until the number of measurements $n$ reaches 25,000.
For simplicity, when discussing one specific parameter, other parameters are set to be the same as the ones used in Ref. 18. In the mutation operation, the initial mutation rate \(pm_{\text{max}}\), the final mutation rate \(pm_{\text{min}}\), and the decay factor \(\lambda\) are set to be 0.1, 0.0025, and 200, respectively. In the breeding operation, the parental mask ratio \(pc\) is set to be 0.5. Since the GA is particularly advantageous in noisy environment, we follow the same strategy in Ref. 18 to add white Gaussian noise. Several levels of Gaussian noises with different standard deviations 0.1\(\langle I_0 \rangle\), 0.3\(\langle I_0 \rangle\), and 0.6\(\langle I_0 \rangle\), representing low-noise level, medium-noise level, and high-noise level, were added into each measurement. Here, \(\langle I_0 \rangle\) is the average intensity of the scattered light, which equals 1 due to normalization. In a typical experiment, the noise of the measurement may come from the instability of the laser and the inaccuracy of the SLM. We also note that linear interpolation was used to make plots smooth.

The numerical simulations were carried out on a personal computer with a central processing unit of Intel(R) Core(TM) i5-8600K CPU @ 3.60 GHz and a random access memory of 16.0 GB. With this computational resource, it takes about 7 s for each independent execution.

Due to statistical reasons, even for the same set of optimization parameters, the enhancement \(\eta\) shows considerable fluctuations. To minimize these statistical errors and make reliable conclusions, each plot shown below is obtained by averaging the results from 1000 independent executions. For example, the standard deviation of the enhancement \(\eta\) around 5000 measurements is about 10. By averaging the results for 1000 independent executions, the standard error, which estimates the deviation, reduces to only 0.3.

4. Results and Discussions

4.1. Schemes to form new generations (step 8)

We first discuss the scheme to form new generations, which turns out to have considerable effects on the performance of the GA. In Ref. 18, based on the rankings of fitness, the prior half of the individuals are added into the population of the next generation, while the latter half \((NP/2)\) of the individuals in the current generation are dropped. Then, we generate \(NP/2\) offspring and supplement them into the population of the next generation, regardless of their fitness. In this way, the new generation contains \(NP/2\) individuals from the current generation and \(NP/2\) offspring. This scheme is adopted in Ref. 18, and is labeled as scheme 1. For scheme 1, it is likely that some of the newly added offspring may even have a lower fitness compared to those dropped ones, which may slow down the increasing rate of the enhancement. Since the enhancement increases monotonically during the optimization process from the theoretical perspective, we can consider other schemes that are slightly aggressive. One suggested modification is that when dropping the latter half of the individuals in the current generation, we record the largest fitness of them as a threshold. When a new offspring is created, we immediately compare its fitness to the threshold. The new offspring can be added into the population of the next generation only if its fitness is larger than the threshold. In this way, we guarantee that all the dropped individuals are replaced by offspring that have larger fitness. This scheme is labeled as scheme 2. Another possible scheme is that we only generate \(NP/2\) offspring and combine them with the current generation that has \(NP\) individuals. Then, we rank all of them based on the fitness and drop the last \(NP/2\) individuals. This scheme is labeled as scheme 3.

The comparison of three schemes are shown in Fig. 2. It can be directly seen that the increasing rates at the very beginning for all three schemes are almost the same, regardless of noise levels. Observable differences start to appear after 3000 measurements and become quite prominent around 7000 measurements. For the noise-free case shown in Fig. 2(a), scheme 1 has the lowest enhancement throughout the entire plot. Such an observation is in consistent with our speculation that scheme 1 is the most conservative one. Moreover, scheme 3 always outperforms scheme 2, possibly because under the same number of measurements, scheme 3 has a greater number of generations than scheme 2. This observation also indicates that it might be better to proceed to the next generation rather than spending lots of measurements within a certain generation.

When noises are added, Figs. 2(b)–2(d) show the comparisons of these three schemes. Although scheme 3 still outperforms the other two schemes, scheme 1 gradually exhibits its advantages in handling noises. With low- and medium-noise levels, although the enhancement obtained using scheme 1
is still lower than that of the other two schemes at around 7000 measurements, the gaps become smaller compared to the noise-free case. Moreover, at around 25,000 measurements, the enhancement obtained using schemes 1 and 3 becomes close and nondifferentiable. When the noise level becomes high, as shown in Fig. 2(d), the enhancement obtained using scheme 1 has already become larger than the one obtained using scheme 2 at around 7000 measurements. Although scheme 1 still performs worse than scheme 3 at around 7000 measurements, the enhancement obtained using these two schemes at around 25,000 measurements is very close.

These results demonstrate that in a noisy environment, an aggressive scheme can increase the initial increasing rate at the expense of the final convergence rate. In the applications of biophotonics and fiber optics, our system does not have enough time to make too many measurements due to the relatively short correlation time of the scattering events. Therefore, a fast initial increase rate is preferable in the GA. For this reason, we consider scheme 3 as the most appropriate one for the GA, which will be adopted in the following simulations.

4.2. Schemes to select parents (step 3)

The GA used in Ref. 18 employs a roulette algorithm to choose two individuals from the population as two parents \( p_a \) and \( m_a \). For these two parents, the one with a higher fitness becomes \( p_a \) and the other one becomes \( m_a \). With this scheme, it is possible that both parents have low fitness, thereby breeding low-fitness offspring that is unlikely to be added into the next generation. In order to avoid the above-mentioned situation, we propose to select several individuals using the roulette algorithm. For convenience, we introduce a variable "selection
number” $S_n$ to represent the number of selected individuals. After selecting $S_n$ individuals, the ones with the highest two fitness are selected as parents. This method of selecting $pa$ and $ma$ using $S_n > 2$ can be considered as a nonlinear roulette algorithm, where high-fitness individuals are a lot easier to be selected. $S_n = 2$ corresponds the original scheme and $S_n \to \infty$ means directly taking the two individuals with the highest two fitness in the generation. Figure 3 shows the performance of the GA using different $S_n$ under different noise conditions. For noise-free case, $S_n = 2$ shows the worst performance regarding both the initial increase rate and the final convergence rate. $S_n = 30$ has the best performance but the enhancement curve is close to that of $S_n = 20$. This observation indicates that further increase of $S_n$ will not cause a big improvement. Nevertheless, one would expect that $S_n \to \infty$ should have the best performance. Surprisingly, it is shown that $S_n \to \infty$ performs even worse than $S_n = 10$. This counterintuitive effect may result from the “genetic drift”, which corresponds to nonergodic in nature. In this situation, the increase of the enhancement is largely owing to mutations.

When different levels of noises are added, Figs. 3(b)–3(d) show the performance of the GA with various $S_n$. For low- and medium-noise levels, $S_n = 2$ always has a relatively slow initial increase rate but gradually surpasses other situations in the end. This observation also indicates a counterbalance between the initial increasing rate and the final convergence rate. When the noise level becomes high, $S_n = 2$ outperforms other situations even in a very early stage. In general, we conclude that in a noisy environment, a larger $S_n$ leads to a faster initial increasing rate but a slower final convergence rate. Moreover, the the noise level is higher, the earlier $S_n = 2$ surpasses other schemes.

Based on the above analysis, we propose to use an adaptive $S_n$, which is a large number in the beginning and gradually decrease to 2 in the end. Such

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3}
\caption{Comparisons of different schemes to select parents. $S_n = 2$ is the one used in Ref. 18, which is marked with an asterisk. (a)-(d) The enhancement as a function of the number of measurements under different conditions: noise free, low-noise level, medium-noise level, and high-noise level. Insets are provided for visualization purpose.}
\end{figure}
an adaptive $S_n$ is expected to achieve a fast initial increasing rate and a fast final convergence rate simultaneously. Here, we use the following formula:

$$S_n = (S_{n, ini} - 2) \times \exp(-\beta \times n/NP) + 2,$$

where $S_{n, ini}$ is the initial selection number and $\beta$ is a constant number. Numerically, we found that GA performs quite well when $S_{n, ini} = 15$ and $\beta = 125$. Figure 4 shows the comparison of $S_n = 2$, $S_n = 30$, and the adaptive $S_n$, under different noise levels. As can be seen from the figures that the adaptive $S_n$ performs well throughout almost the entire measurement range under different noise levels. The only unsatisfactory performance of adaptive $S_n$ shows up in Fig. 4(d), possibly due to the fact that the parameters in the adaptive $S_n$ were not optimized to the best condition. Further optimizations on the adaptive $S_n$ can be performed by scanning different combinations of the parameters including $S_{n, ini}$ and $\beta$ or considering other transition formulas.

4.3. Choices of the parental mask ratio $pc$ (step 4)

A key parameter in generating the random binary template $T$ is the parental mask ratio $pc$. During generation, each element of $T$ will be randomly assigned to either “1” or “0”, with a probability of $pc$ and $(1 - pc)$, respectively. A new offspring is generated according to the following formula:

$$\text{offspring} = pa \cdot T + ma \cdot (1 - T).$$

In Ref. 18, $pc$ is fixed to 0.5, which means that the offspring inherits equal information from $pa$ and $ma$. Given that $pa$ always has a higher fitness compared with $ma$, one would expect that more information should be inherited from $pa$. To validate this assumption,
we consider several different choices of \( pc \) ranging from 0.5 to 0.65 with a step size of 0.05. Besides using a fixed \( pc \), we also consider a tunable \( pc \) that is determined as the fitness ratio of the parents: 
\[
pc = \frac{pa_{\text{fitness}}}{pa_{\text{fitness}} + ma_{\text{fitness}}},
\]
where \( pa_{\text{fitness}} \) and \( ma_{\text{fitness}} \) are the fitnesses of the selected \( pa \) and \( ma \).

Figure 5 shows the performance of the GA with different choices of \( pc \). For all noise levels, we found that \( pc = 0.5 \) always shows the best performance but is only slightly better compared to other choices. Therefore, we conclude that varying the parental mask ratio \( pc \) has little effect on the performance of the GA. This result is actually in consistent with the Darwin’s theory of evolution, where the chromosomes of offspring equally inherit from their parents (\( pc = 0.5 \)).

4.4. Choices of the mutation rate \( pm \) (step 6)

Mutation is a very important operation in the GA to maintain genetic diversity as well as to jump out of a local maximum. In general, the GA with a very small mutation rate may suffer from “genetic drift” while a very large mutation rate can lead to the loss of good solutions. Given the nature of wavefront shaping problems, Ref. 18 used an adaptive mutation rate that is high in the beginning and gradually vanishes in the end: 
\[
pm = (pm_{\text{max}} - pm_{\text{min}}) \times \exp\left(-\frac{n}{\lambda}\right) + pm_{\text{min}}.
\]
To find reasonable settings, it is worth tuning these parameters. However, a detailed analysis on how to choose these parameters has never been discussed.

To fill this blank, we perform an ergodic scanning on the parameters of \( pm_{\text{max}}, pm_{\text{min}}, \) and \( \lambda \). We
consider eight choices of $pm_{\text{max}}$ (0.08, 0.09, 0.1, 0.11, 0.12, 0.2, 0.3, 0.5), five choices of $pm_{\text{min}}$ (0.001, 0.0025, 0.005, 0.01, 0.015), and six choices of $\lambda$ (150, 250, 350, 450, 550, 650), which leads to a total number of 240 optional combinations. To save computational resources, we decrease the number of average times from 1000 to 100 during this ergodic search.

As a typical example, Fig. 6 plots the enhancement as a function of the number of measurements under the high-noise level. As can be seen from the Fig. 6(a) that different combinations of these parameters show significant variations, which indicates that choices of these parameters play significant roles in the performance of the GA. To quantify the performance, we define a quantity: $F(N) \equiv \sum_{n=1}^{N} \eta(n)$. Pictorially, $F(N)$ is the area enclosed by the enhancement curve and the horizontal axis before $N$ measurement. In general, a larger area indicates a better performance. Since the most rapid increase happens before 5000 measurements, we choose $F(5000)$ as the figure of merit.

For all the 240 possible combinations, it is found that for high-noise level, the combination of $pm_{\text{max}} = 0.1$, $pm_{\text{min}} = 0.01$, and $\lambda = 350$ leads to the largest $F(5000)$. For medium-noise level, the combination of $pm_{\text{max}} = 0.1$, $pm_{\text{min}} = 0.005$, and $\lambda = 450$ leads to the largest $F(5000)$. For low-noise level, the combination of $pm_{\text{max}} = 0.09$, $pm_{\text{min}} = 0.0025$, and $\lambda = 650$ leads to the largest $F(5000)$. From the above results, we note that $pm_{\text{max}}$ is always around 0.1, regardless of the noise level. This value may be related to certain parameters that were fixed during this simulation, such as the number of independent controls (number of genes) and the population number.

Moreover, as the operation of mutation is helpful in dealing with environmental noises, a relatively
larger $pm_{\min}$ is helpful when a higher noise level exists. For the same reason, a higher noise level requires a slower decay of mutation rates, which leads to a smaller $\lambda$.

To further illustrate the functions of each parameter separately, we compare the performance of the GA by varying only one parameter at a time. At high-noise level, $pm_{\max} = 0.1$, $pm_{\min} = 0.01$, and $\lambda = 350$ are chosen as the baseline. Figure 6(b) shows the performance of the GA when only $pm_{\max}$ is varied. As can be seen from the figure that $pm_{\max}$ significantly affects the early increasing rate of the enhancement but has little effects on the final convergence rate. Among all different choices of $pm_{\max}$, the best performance is obtained when $pm_{\max} = 0.1$. In contrast, as shown in Fig. 6(c), $pm_{\min}$ does not have significant effect on the first few thousands measurements. However, an improper choice of $pm_{\min}$ significantly degrades the enhancement after 5000 measurements. Among all different choices of $pm_{\min}$, the best performance is obtained when $pm_{\min} = 0.01$. Figure 6(d) shows the performance of the GA with different $\lambda$. Although $\lambda = 350$ is found to have the largest $F(5000)$, there is no big difference among the curves obtained with different $\lambda$. Therefore, we conclude that the performance of the GA is not sensitive to the choice of $\lambda$.

Conflict of Interest

We declare that there is no conflicts of interest in this paper.

5. Conclusion

In conclusion, we have done a thorough study on the performance of the GA in feedback-based wavefront shaping. Using numerical tools, we investigate the performance of the GA with several different schemes, including the formation of new generations and the selection of parents. With new schemes, the initial increasing rate of the enhancement can be improved. Moreover, we also discuss different choices of the parental mask ratio $pc$ and the mutation rate $pm$. Guidance on how to properly choose these parameters is provided. We anticipate that this work can be beneficial to future studies related to feedback-based wavefront shaping with GAs.

References


