Shape design of a reflecting surface using Bayesian Optimization

IOP Conf. Series: Journal of Physics: Conf. Series **963** (2018) 012003 doi:10.1088/1742-6596/963/1/012003

Shape design of a reflecting surface using Bayesian Optimization

X. Garcia-Santiago[1,2], P.-I. Schneider[1], C. Rockstuhl[2,3], S. Burger[1,4]

[1] JCMwave GmbH., 12163 Berlin, Germany
[2] Institute of Nanotechnology, Karlsruhe Institute of Technology, 76021 Karlsruhe, Germany
[3] Institute of Theoretical Solid State Physics, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
[4] Zuse Institute Berlin, 14195 Berlin, Germany

E-mail: xavi.garcia@jcmwave.com

Abstract. Optimizing free-from shapes of photonic nanostructures is a high-dimensional problem. We present a Bayesian optimization algorithm with a hyper-parameter learning routine. We apply the algorithm for optimizing the shape of a reflecting meta-surface. The algorithm exhibits efficient performance.

1. Introduction

Nanooptical technologies have drastically evolved in recent years with applications in fields such as silicon photonics, energy harvesting, or quantum technologies. This improvement is mainly due to modern fabrication technologies such as laser writing [1] or electron-beam lithography [2] which allow for the manufacturing of micro and nano optical structures with an increasing degree of accuracy and flexibility. However, the best shape of the photonic devices for achieving a good performance is not always easy to find. The most efficient designs can be sometimes counter-intuitive. In these cases inverse engineering methods become an indispensable tool and the choice of the algorithm used to derive the shape of the device becomes a critical aspect.

One important aspect of the design process is the parametrization used in order to describe the shape of the device, this parametrization should be flexible enough in order to be able to describe a high number of possible shapes. However, this flexibility presents two main drawbacks. First, it normally implies working in high dimensional parameter spaces, which directly increases the computational time needed to find the optimal point. Second, the more general the shape is allowed to be, the more likely it is that the optimization problem is not convex, leading to problems where the objective function presents many local minima.

Many algorithms have been used in order to globally optimize different nanophotonic structures. From genetic [3] or evolution [4] algorithms, direct-binary search (DBS) [5], particle swarm [6] or more recently via the use of machine learning algorithms [7]. Into this last group we could include Bayesian optimization. An algorithm based on a stochastic model where the fundamental parameters of the algorithm are updated during computation time based on the information acquired from the previous simulations. This group has recently presented a study on the use and performance of this algorithm in order to optimize different optical nanostructures [8]. We now update this work presenting an improvement in the learning process.

This improvement leads to better convergence ratios, thereby allowing the access to parameter spaces of higher dimensions and therefore, to more complex structures. As an example, we optimize the shape of a periodic meta-mirror in order to increase the power reflected into the first diffraction order.

2. Bayesian Optimization with Gaussian Processes

The goal of every global optimization algorithm is to identify the global minimum (or maximum) of an unknown objective function f in a certain design space $\mathcal{X}$,

$$\mathbf{x_{min}} = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}), \tag{1}$$

In order to achieve that, Bayesian optimization (BO) uses a stochastic model of the objective function f. The model is updated with information from the evaluations of f and then it is used to drive the optimization strategy.

Gaussian processes (GP) are frequently used as the stochastic model in Bayesian optimization. Our objective function is modeled as a Gaussian process if for any N points $\mathbf{x}_1, \cdots \mathbf{x}_N$ in the parameter space $\mathcal{X}$, the probability of the objective function to be equal to $\mathbf{Y} = (y_1, \cdots, y_N)$ at these points is following a multivariate Gaussian random distribution defined by a mean function $\mu : \mathcal{X} \to \mathbb{R}$ and a positive definite covariance function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$,

$$P(\mathbf{Y}) = \frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{Y}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mu)\right] \tag{2}$$

or more compactly $\mathbf{Y} \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Here, $\mu = [\mu(\mathbf{x}_1), \cdots, \mu(\mathbf{x}_k)]^T$ is the vector of mean values and $\mathbf{\Sigma} = [k(x_i, x_j)]_{i,j}$ the covariance matrix.

GP can describe a large class of random functions using different covariance functions $k(\mathbf{x}, \mathbf{x}')$. For our optimization purposes and based on the results obtained in [8], we will use the Matérn 5/2 kernel,

$$k_{\mathrm{M52}}(\mathbf{x}, \mathbf{x}') = \sigma^2\left(1 + \sqrt{5r^2(\mathbf{x}, \mathbf{x}')} + \frac{5}{3}r^2(\mathbf{x}, \mathbf{x}')\right)\exp\left(-\sqrt{5r^2(\mathbf{x}, \mathbf{x}')}\right) \tag{3}$$

$$r^2(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{d}(x_i - x_i')^2/l_i^2 \tag{4}$$

The hyper-parameters σ and $l_1, l_2, \cdots$ determine the standard deviation and the length scales of the random distributions. With respect to the mean vector μ, we will assume all its elements have the same constant value m. The different choices for μ do not influence the performance of the method as much as the kernel function does.

The stochastic model is then updated after the evaluations of the objective function using Bayes' theorem. In the points where the problem has been solved, the value of the objective is no longer an unknown but rather a well known value without any uncertainty. Also in the regions close to the evaluated points the uncertainty will be reduced. The posterior probability distribution is computed based on the computed objective evaluations and on the prior distribution of the geometry parameters.

Suppose M values $\mathbf{Y_{ev}} = (y_{ev1}, \cdots y_{evM}) = (f(\mathbf{x_{ev}}_1), \cdots, f(\mathbf{x_{ev}}_M))$ of the unknown objective function are known. Then the posterior distribution will also be a Gaussian process G'. The new mean vector μ' and the covariance matrix $\mathbf{\Sigma}'$ are functions of the previous mean vector μ, of the covariance matrix $\mathbf{\Sigma}$ and also of its values in the evaluation points $\mathbf{x_{ev}}_i$

$$\mu'(\mathbf{x}) = \mu(\mathbf{x})^T + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{Y_{ev}} - \mu(\mathbf{x_{ev}})^T) \tag{5}$$

$$\mathbf{\Sigma}' = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}) \tag{6}$$

with $(\mathbf{\Sigma}_{11})_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$, $(\mathbf{\Sigma}_{12})_{i,j} = k(\mathbf{x_{ev}}_i, \mathbf{x}_j)$, and $(\mathbf{\Sigma}_{22})_{i,j} = k(\mathbf{x_{ev}}_i, \mathbf{x_{ev}}_j)$.

In Fig. 1 we show a stochastic model of an objective function before (left) and after (right) being updated with five evaluations.

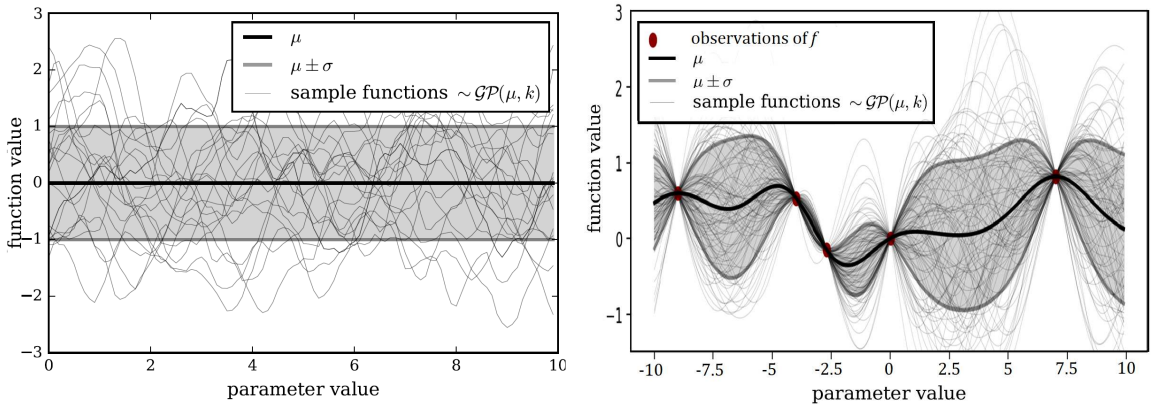

Figure 1. a. Different samples of a Gaussian process with the Matérn 5/2 kernel defined in Eq. (3) **b**. Different samples of a Gaussian process with the Matérn 5/2 kernel updated after five evaluations of the sample objective function $f(x) = 0.5\sin(x) + 0.01x^2$.

Finally, the optimization strategy i.e., the decision where to make the next evaluation is determined by an *acquisition function*. In our case we use the expected improvement function α_{EI},

$$\begin{aligned}\alpha_{\mathrm{EI}}(x, y_{\min}) &= \mathbb{E}[\max(0, y_{\min} - f(x))] \\ &= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\mathrm{y_{min}} - \mu_{\mathrm{N}}(\mathbf{x})}{\sqrt{2}\sigma_{\mathrm{N}}(\mathbf{x})}\right)\right](y_{\min} - \mu_N(\mathbf{x})) + \frac{\sigma_N(\mathbf{x})}{\sqrt{2\pi}}\exp\left(\frac{(y_{\min} - \mu_N(\mathbf{x}))^2}{2\sigma_N(\mathbf{x})^2}\right)\end{aligned} \tag{7}$$

In Fig. 2 an example of the expected improvement function is shown. With this function, the algorithm selects the next point to evaluate based on where the stochastic model expects the largest reduction in the currently computed minimum.

3. The hyper-parameters learning process

Initially, the hyper-parameters of the GP $\omega = (\sigma, l_1, l_2, \cdots)$ are unknown and are usually initialized with some reasonable values. Having drawn some samples of the objective function one can optimize the hyper-parameters by maximizing the probability of drawing these samples with respect to the values of the hyper-parameters:

$$\omega_{\mathrm{opt}} = \arg\max_{\omega}\left(\log[P_\omega(\mathbf{Y_{ev}})] + \log[P(\omega)]\right) \tag{8}$$

Here, $P(\omega)$ is the prior distribution of the hyper-parameters (the hyper-prior) and $P_\omega(\mathbf{Y_{ev}})$ the likelihood of observing $\mathbf{Y_{ev}}$ given ω.

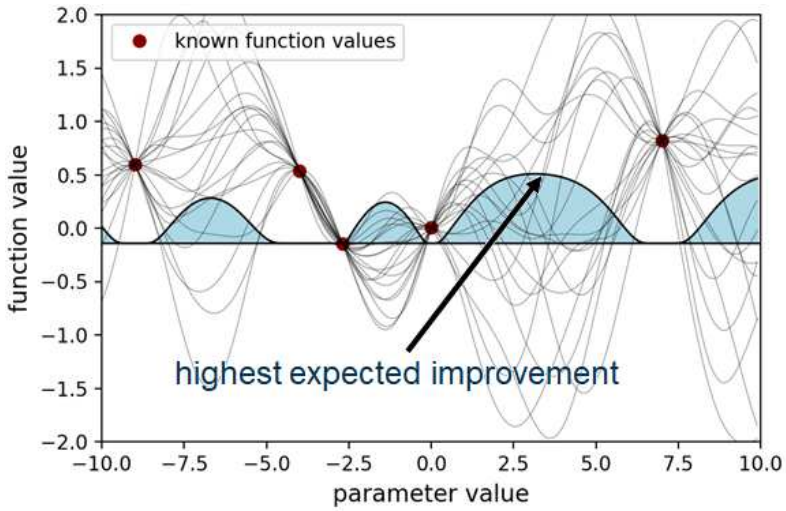

Figure 2. The optimization strategy evaluates the objective function in the point with the highest expected improvement (eq. 7)

The value of the hyper-parameters is crucial for an efficient optimization. A too small length scale, for example, would overestimate the uncertainty in points close to the evaluated points. However, if it is too large, the acquisition function would tend to point to the boundaries of the parameter space and to under-sample the central regions. However, every time the hyper-parameters are updated the covariance matrix has to be inverted, an expensive task which slows down the optimization algorithm considerably, specially for high-dimensional parameter spaces. One solution to overcome this problem is to not update the hyper-parameters after each evaluation but rather to use some measure for estimating the possible improvement (I). Then the hyper-parameters would be updated if and only if they are expected to improve the likelihood $P_\omega(\mathbf{Y_{ev}})$.

However, not all the hyper-parameters force the inversion of the covariance matrix. This is the case for μ and σ where, furthermore, analytical expressions for their optimal values can be obtained. Therefore, we can optimize these hyper-parameters after every evaluation and just leave the length scales l_i to be updated based on the possible improvement estimator I.

The log likelyhood can be written as

$$\log[P_\omega(\mathbf{Y_{ev}})] = -\frac{N}{2} log(2\pi) - \frac{1}{2}\log(|\mathbf{\Sigma}|) - \frac{1}{2}(\mathbf{Y_{ev}} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1}(\mathbf{Y_{ev}} - \boldsymbol{\mu}). \tag{9}$$

If we define a matrix $\mathbf{K}$ such that $\mathbf{\Sigma} = \sigma^2\mathbf{K}$, the optimal value of the hyperparameter σ can be derived analytically from

$$\frac{\partial \log[P_\omega(\mathbf{Y_{ev}})]}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}(\mathbf{Y_{ev}} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{Y_{ev}} - \boldsymbol{\mu}) = 0. \tag{10}$$

Hence the optimal value of σ is given by the Mahalanobis distance divided by $\sqrt{N}$

$$\sigma^2_{\text{opt}} = \frac{1}{N}(\mathbf{Y_{ev}} - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{Y_{ev}} - \boldsymbol{\mu}). \tag{11}$$

Repeating the procedure for the optimal mean value m we obtain,

$$m_{\text{opt}} = \frac{\sum_{ij}(\mathbf{K}^{-1})_{ij} Y_{evj}}{\sum_{ij}(\mathbf{K}^{-1})_{ij}} \tag{12}$$

The optimization of these hyper-parameters after every evaluation is a fast calculation and it improves the convergence rate of the algorithm. Higher convergence rates allow us to investigate more complex structures as geometry parametrization spaces with higher dimensions become feasible to solve.

For the length scales l_i we use a first order approximation of the log likelihood as measure of the possible improvement,

$$I = \max_i (l_i \frac{\partial \log[P_\omega(\mathbf{Y_{ev}})]}{\partial l_i}), \tag{13}$$

The optimization of the l_i's is done provided

$$I > \epsilon \log[P_\omega(\mathbf{Y_{ev}})], \tag{14}$$

with $\epsilon = 0.1$.

4. Numerical results

In order to demonstrate the performance of BO using GP with optimized hyper-parameters we revisit the example of optimization of a reflective surface [8]. The structure consists on a periodic corrugated silver surface as shown in Fig. 3. The mirror has a periodic length of 1350 nm, 3 times the wavelength of the incident plane wave ($\lambda = 450$ nm). The optimization problem consists on finding the mirror shape which increases the power flux reflected into the first diffraction order when the mirror is illuminated with a normal incident TE polarized plane wave.

$$f_{ob} = \frac{S_{z,+1}}{S_{z,inc}} \tag{15}$$

where $S_{z,inc}$ is the z-component of the Poynting vector of the incident wave and $S_{z,+1}$ of the first diffraction order.

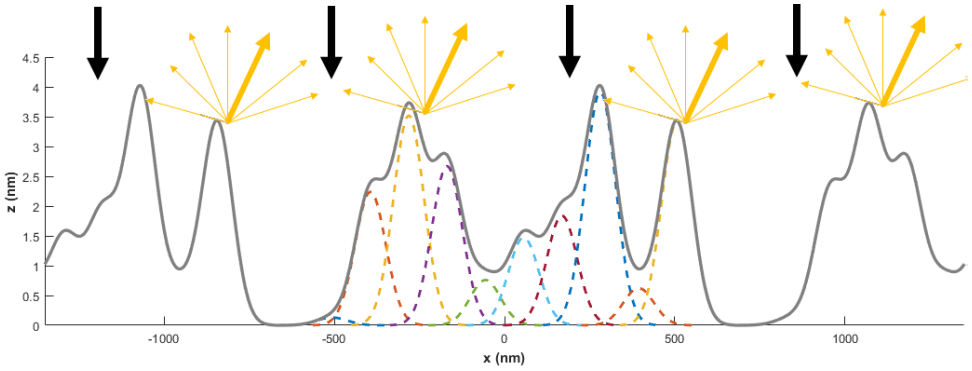

Figure 3. Schematic of the system. A plane wave illuminates the periodic mirror meta-surface (solid line) under normal incidence, which produces several diffraction orders. The objective is to find the shape of the mirror that maximizes the reflectance into the first diffraction order. The shape of the mirror is described with a parametrization of its periodic unit cell using 10 third order b-splines (dotted lines).

The mirror is parametrized with 3rd order b-splines with uniformly distributed knots as shown in Fig. 3.

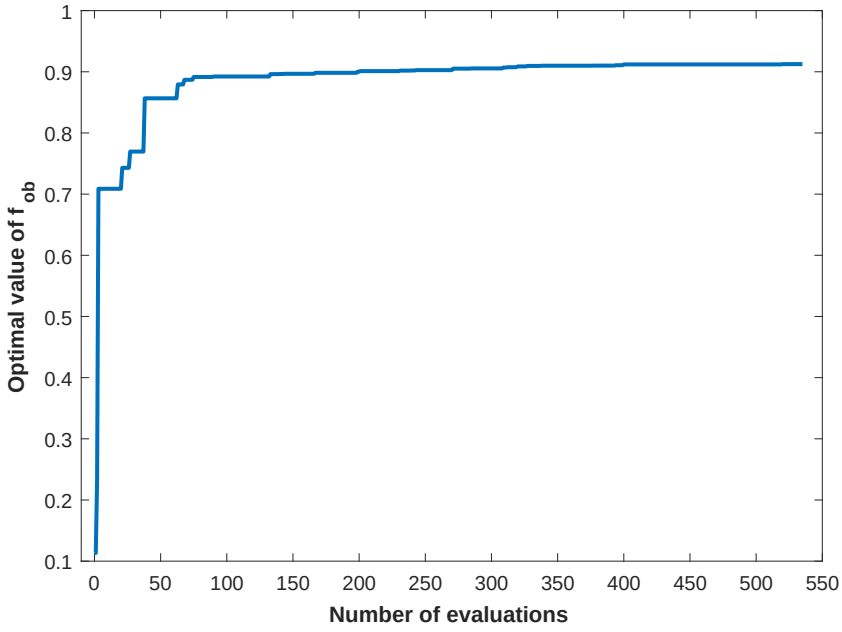

Figure 4. Convergence plot of the optimization.

The simulations of the electromagnetic field were done with the FEM solver JCMsuite [9]. We use elements with a polynomial degree of 3 and a mesh size of $\lambda/(10n(\mathbf{r}))$, where $n(\mathbf{r})$ is the refractive index of the material. The convergence results are shown in Fig. 4. The best shape deviates 91.3% of the incident power flux into the first diffraction order. The optimum shape is shown in Fig. 5. After 80 evaluations the algorithm had already found a shape with a performance of 89% and of 91% after 330 evaluations. Note that these 530 evaluations correspond to less than two points per dimension in the case of a uniformly distributed parameter scan.

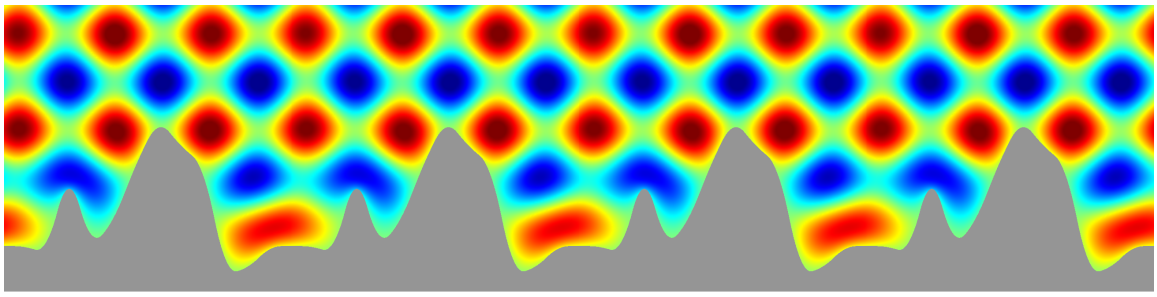

Figure 5. Electric field profile created by the optimal mirror.

5. Conclusions

We have implemented a new hyper-parameter learning routine. With the calculation of the optimum hyper-parameters σ and μ after every evaluation we get an improvement in the convergence ratio. This improvement allows us to look for shapes in higher dimensional parameter spaces, which potentially leads to a design with a better performance. We have improved the performance of a meta-surface mirror with regards to previously reported results [8] by a factor of around 25 %. This improvement is obtained thanks to the increment of the dimensions of the parameter space, going from 4 to 10 dimensions. The number of equivalent

points-per-dimension needed to obtain the optimum shape was less than two. This suggests the enormous potential of BO for globally optimizing nanophotonic structures.

Acknowledgments
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675745 (MSCA-ITN-EID NOLOSS). This project has received funding from the Senate of Berlin (IBB, ProFIT grant agreement No 10160385, FI-SEQUR), co-financed by the European Fund for Regional Development (EFRE).

References
[1] Fischer J and Wegener M 2013 *Laser Photon. Rev.* **7** 22–44
[2] Chen Y 2015 *Microelectron. Eng.* **135** 57 – 72
[3] Goh J, Fushman I, Englund D and Vučković J 2007 *Opt. Express* **15** 8218–8230
[4] Preble S, Lipson M and Lipson H 2005 *Applied Physics Letters* **86** 061111
[5] Shen B, Wang P, Polson R and Menon R 2015 *Nature Photonics* **9** 378 – 382
[6] Forestiere C, Donelli M, Walsh G F, Zeni E, Miano G and Negro L D 2010 *Opt. Lett.* **35** 133–135
[7] Malkiel I, Nagler A, Mrejen M, Arieli U, Wolf L and Suchowski H 2017 *ArXiv e-prints* (*Preprint* `1702.07949`)
[8] Schneider P I, Garcia-Santiago X, Rockstuhl C and Burger S 2017 *Proc. SPIE* **10335** 103350O
[9] Pomplun J, Burger S, Zschiedrich L and Schmidt F 2007 *physica status solidi (b)* **244** 3419–3434