A Nonlocal-Gradient Descent Method for Inverse Design in Nanophotonics

Sirui Bi
Computational Sciences and Engineering Division
Oak Ridge National Laboratory, Oak Ridge, TN 37830
bis1@ornl.gov

Jiaxin Zhang
Computer Science and Mathematics Division
Oak Ridge National Laboratory, Oak Ridge, TN 37830
zhangj@ornl.gov

Guannan Zhang
Computer Science and Mathematics Division
Oak Ridge National Laboratory, Oak Ridge, TN 37830
zhangg@ornl.gov

Abstract

Local-gradient-based optimization approaches lack nonlocal exploration ability required for escaping from local minima when searching non-convex landscapes. A directional Gaussian smoothing (DGS) approach was recently proposed in [29] and used to define a truly nonlocal gradient, referred to as the DGS gradient, in order to enable nonlocal exploration in high-dimensional black-box optimization. Promising results show that replacing the traditional local gradient with the nonlocal DGS gradient can significantly improve the performance of gradient-based methods in optimizing highly multi-modal loss functions. However, the current DGS method is designed for unbounded and uncontrained optimization problems, making it inapplicable to real-world engineering optimization problems where the tuning parameters are often bounded and the loss function is usually constrained by physical processes. In this work, we propose to extend to the DGS approach to the constrained inverse design framework in order to find better optima of multi-modal loss functions. A series of adaptive strategies for smoothing radius and learning rate updating are developed to improve the computational efficiency and robustness. Our methodology is demonstrated by an example of designing a nanoscale wavelength demultiplexer, and shows superior performance compared to the state-of-the-art approaches. By incorporating volume constraints, the optimized design achieves an equivalently high performance but significantly reduces the amount of material usage.

1 Introduction

There recently have been significant interests in using *computational inverse design* approaches to explore the full design space of novel nanophotonic devices with a broad variety of applications [24, 17, 2, 18, 10, 28, 23, 13]. Much of this progress is made by the gradient-based optimization that is a promising method to efficiently search the enormous degrees of freedom in high-dimensional design spaces. The gradient-based optimization typically relies on the adjoint method [11, 3],

Workshop on machine learning for engineering modeling, simulation and design @ NeurIPS 2020

which calculates the local gradient of a loss function with respect to the design variables by solving differential equations, i.e., the adjoint equation [12, 8]. However, the adjoint method basically depends on an efficient estimation or analytical formulation of the adjoint equation. Typically, some assumptions are often made for simplicity to utilize gradient-based optimization. For example, electromagnetism is modeled using Maxwell's equations assuming statics, linear, homogeneous, and isotropic materials as well as time-harmonic behavior of the field [6]. All of these assumptions lead to a large challenge in real-world nanophotonic design with dynamic, nonlinear, and dispersive material properties in multiphysics conditions [9]. In addition, for complicated objective functions and constraints, gradient estimation with adjoint method may be either not easily accessible or unreliable. Sometimes, additional efforts are required to derive the gradient if unusual objectives or constraints are incorporated into the optimization formulation [24, 8].

Another important challenge is that, up to now, most studies use local gradient-based approaches for inverse design so that the optimized devices converge to a *local minimum*. In many electromagnetic design problems, their landscapes have been proven to be highly nonlinear and non-convex such that many possible local minima exist [15, 24]. These local minima depend on the initialization and vary largely as the initial guess changes. These challenges in gradient-based approaches have attracted much attention [24, 14, 7, 4]. Several gradient-free algorithms such as Bayesian optimization are used to explore the global minima but finding the optimal solution to complex high-dimensional, multimodal problems often converges very slowly and requires very expensive loss function evaluations [21]. Additionally, it is important to impose fabrication ((also material usage) constraints into design optimization workflow because a fundamental challenge in nanophotonics is that arbitrary permittivity distribution including tiny feature and grey-scale value, can not be fabricated in practice [3, 27, 19]. Although stochastic algorithms, e.g., SGD, Adam, and RMSProp, have widely used in the machine learning community, it is still a non-trivial task to incorporate multiple equality or inequality constraints into these stochastic methods because they mainly focus on unconstrained optimization tasks like neural network training.

To address these challenges, we propose a nonlocal inverse design workflow by incorporating the nonlocal gradient that was recently developed in [29]. The nonlocal gradient was defined by directional Gaussian smoothing, thus it is referred to as *the DGS gradient* hereinafter. The DGS gradient conducts 1D nonlocal explorations along with d orthogonal directions and each of which defines a nonlocal directional derivative as a 1D integral. The d directional derivatives are assembled to the DGS gradient. The Gauss-Hermite (GH) quadrature is used to approximate the 1D integrals (i.e., the directionsl derivatives) to achieve a higher accuracy than Monte Carlo (MC) sampling. We improved the existing DGS approach from two perspectives in the context of inverse design. First, we established a workflow in which the DGS gradient can be combined with a variety of constraints, e.g., fabrication constraints and materials usage constraint in the practical material design. Second, we developed a series of adaptive strategies for the smoothing radius and the learning rate in order to improve computational efficiency and robustness. Compared to the local gradient method, the directional smoothing allows for a large smoothing radius to capture the global structure of loss landscapes and thus provide a strong nonlocal exploration capability for escaping from local minima in non-convex landscapes. Furthermore, our workflow h multiple assumptions for simplicity so that it has wider feasibility to nonlinear, dynamic, and non-isotropic materials under complex physical conditions. In the meantime, our method having the benefits of gradient-based optimization can be easily scaled to high-dimensional design space, which alleviates the challenges in derivative-free global optimization, such as Bayesian optimization.

2 The nonlocal DGS gradient for optimization

Considering a general black-box optimization problem: $\min_{\boldsymbol{x}\in\mathbb{R}^d} F(\boldsymbol{x})$, where $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ consists of d inputs, and $F : \mathbb{R}^d \to \mathbb{R}$ is a d-dimensional loss function, we assume that the gradient $\nabla F(\boldsymbol{x})$ is unavailable, and $F(\boldsymbol{x})$ is only accessible via function evaluations.

The traditional Gaussian smoothing To better explain the DGS strategy, we briefly recall the standard Gaussian smoothing [16] for estimating local gradients. Specifically, it starts by defining a smoothed loss function $F_\sigma(\mathbf{x}) = \mathbb{E}_{\boldsymbol{u}\sim\mathcal{N}(0,\mathbf{I}_d)}\left[F(\mathbf{x}+\sigma\boldsymbol{u})\right]$, where $\mathcal{N}(0,\mathbf{I}_d)$ is the d-dimensional standard Gaussian distribution, and $\sigma > 0$ is the smoothing radius. Then, the gradient $\nabla F_\sigma(\mathbf{x})$ can

be represented as an expectation and estimated by drawing M samples $\{\boldsymbol{u}_m\}_{m=1}^M$ from $\mathcal{N}(0, \mathbf{I}_d)$

$$\nabla F_\sigma(\mathbf{x}) = \frac{1}{\sigma}\mathbb{E}_{\boldsymbol{u}\sim\mathcal{N}(0,\mathbf{I}_d)}\left[F(\mathbf{x}+\sigma\boldsymbol{u})\,\boldsymbol{u}\right] \approx \frac{1}{M\sigma}\sum_{m=1}^{M} F(\mathbf{x}+\sigma\boldsymbol{u}_m)\boldsymbol{u}_m. \tag{1}$$

The Monte Carlo (MC) estimator in Eq. (1) is substituted into any gradient-based algorithm to update the state $\mathbf{x}$. *The major drawback is that the error of the MC estimator in Eq. (1) is on the order of* $\varepsilon \sim \mathcal{O}(d\sigma/\sqrt{M})$. When the dimension d is large (e.g., on the order of thousands) and the computing budget (the upper bound of M) is given, practitioners often have to sacrifice a nonlocal smoothing effect (with a relatively big σ) that helps skipping local mimina to achieve a required accuracy. In other words, Eq. (1) is mostly used in the local regime with a small value for σ.

The nonlocal DGS gradient The DGS gradient was developed to alleviate the above challenge with the standard Gaussian smoothing. The key idea behind the DGS gradient is to conduct 1D nonlocal explorations along d orthogonal directions in $\mathbb{R}^d$, each of which defines a nonlocal directional derivative as a 1D integral. The Gauss-Hermite quadrature is used to estimate the d 1D integrals to achieve high accuracy. Specifically, we first define a 1D cross section of $F(\mathbf{x})$ as $G(y\,|\,\mathbf{x},\boldsymbol{\xi}) = F(\mathbf{x}+y\,\boldsymbol{\xi}),\ \ y\in\mathbb{R}$, where $\mathbf{x}$ is the current state of $F(\mathbf{x})$ and $\boldsymbol{\xi}$ is a unit vector in $\mathbb{R}^d$. The Gaussian smoothing of $G(y)$, denoted by $G_\sigma(y)$, is defined by

$$G_\sigma(y\,|\,\mathbf{x},\boldsymbol{\xi}) := \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} G(y+\sigma v\,|\,\mathbf{x},\boldsymbol{\xi})\,\mathrm{e}^{-\frac{v^2}{2}}\,dv = \mathbb{E}_{v\sim\mathcal{N}(0,1)}\left[G(y+\sigma v\,|\,\mathbf{x},\boldsymbol{\xi})\right]. \tag{2}$$

This is also the Gaussian smoothing of $F(\mathbf{x})$ along the direction $\boldsymbol{\xi}$ in the neighbourhood of $\mathbf{x}$. The derivative of $G_\sigma(y|\mathbf{x},\boldsymbol{\xi})$ at $y=0$ can be represented by a 1D integral

$$\mathscr{D}[G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi})] = \frac{1}{\sigma}\,\mathbb{E}_{v\sim\mathcal{N}(0,1)}\left[G(\sigma v\,|\,\mathbf{x},\boldsymbol{\xi})\,v\right], \tag{3}$$

where $\mathscr{D}[\cdot]$ denotes the differential operator. We emphasize that Eq. (3) is fundamentally different from the directional derivative of $F_\sigma(\mathbf{x})$, because $G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi})$ only conducts the directional smoothing along $\boldsymbol{\xi}$. For a matrix $\boldsymbol{\Xi} := (\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_d)$ consisting of d orthonormal vectors, we can define d directional derivatives like those in Eq. (3) and assemble our DGS gradient as

$$\nabla_{\sigma,\boldsymbol{\Xi}}[F](\mathbf{x}) := \left[\mathscr{D}[G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi}_1)],\cdots,\mathscr{D}[G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi}_d)]\right]\boldsymbol{\Xi}, \tag{4}$$

where the orthogonal system $\boldsymbol{\Xi}$ and the smoothing radius σ can be adjusted during an optimization process. The next step is to develop an accurate DGS estimator. Since each component of $\nabla_{\sigma,\boldsymbol{\Xi}}[F](\mathbf{x})$ only involves a 1D integral, such that the Gauss-Hermite quadrature rule [20, 1] can be used to approximate the integrals with high accuracy (shown in Eq. (6)). By doing a simple change of variable in Eq. (3), the GH rule can be directly used to obtain the following estimator for each directional derivative $\mathscr{D}[G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi})]$ in Eq. (3)

$$\mathscr{D}^M[G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi})] = \frac{1}{\sqrt{\pi}\sigma}\sum_{m=1}^{M} w_m\,F(\mathbf{x}+\sqrt{2}\sigma v_m\boldsymbol{\xi})\sqrt{2}v_m, \tag{5}$$

where $\{v_m\}_{m=1}^M$ are the roots of the M-th order Hermite polynomial and $\{w_m\}_{m=1}^M$ are quadrature weights. Both v_m and w_m can be found online[1] or in [1]. Compared with MC sampling, the error of Eq. (5) can be bounded by

$$\left|(\,\mathscr{D}^M-\mathscr{D})[G_\sigma]\right| \le C\frac{M\,!\sqrt{\pi}}{2^M(2M)\,!}\sigma^{2M-1}, \tag{6}$$

where $M!$ is the factorial of M and the constant $C>0$ is independent of M and σ. Applying the GH quadrature rule $\mathscr{D}^M$ to each component of $\nabla_{\sigma,\boldsymbol{\Xi}}[F](\mathbf{x})$ in Eq. (4), we define the following estimator:

$$\nabla^M_{\sigma,\boldsymbol{\Xi}}[F](\mathbf{x}) = \left[\,\mathscr{D}^M[G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi}_1)],\cdots,\,\mathscr{D}^M[G_\sigma(0\,|\,\mathbf{x},\boldsymbol{\xi}_d)]\right]\boldsymbol{\Xi}. \tag{7}$$

In illustration of the difference between the DGS gradient and local gradient is given in Figure 1. We herein discuss some important features of the DGS gradient and the GH quadrature estimator.

[1]Nodes and weights for GH quadrature: https://keisan.casio.com/exec/system/1281195844

- *Nonlocality*: The directional smoothing allows for a large radius σ to capture global structures of loss landscapes and help escape from local minima.
- *Accuracy*: The GH quadrature with the error bounded in Eq. (6) provides an estimator having much higher accuracy than MC, even when a large smoothing radius σ is used.
- *Portability*: The DGS gradient can be integrated into a majority of gradient-based algorithms, e.g., gradient descent, Adam, and those with constraints.
- *Scalability*: The DGS estimator in Eq. (7) requires $M \times d$ evaluations of $F(\mathbf{x})$, and these evaluations are completely parallelizable as those in random sampling.

A mathematical example To illustrate the performance of the DGS gradient, we combine the DGS gradient with the standard gradient descent algorithm to optimize the 1000D Ackley function, which is one of the benchmark functions used to test non-convex optimization algorithms. As shown in Figure 1 (Middle), the Ackley function has many local minima which pose significant challenges for optimization. Figure 1 (Right) shows that the DGS gradient exploited its nonlocal exploration ability to skip the local minima and converge to the global minimum $F(\mathbf{x}) = 0$. DGS gradient successfully found the global minimum while the other baselines are trapped in local minima.

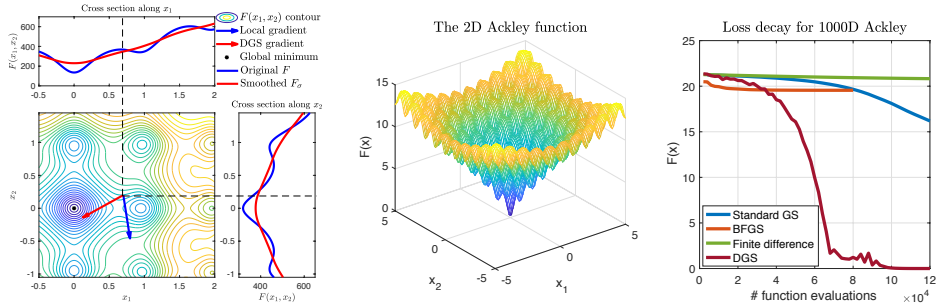

Figure 1: (Left) Illustration of the nonlocal exploration capability of our DGS gradient. In the central plot, the blue arrow points to the local gradient direction and the red arrow points to the DGS gradient direction. The top and right plots show the directionally smoothed functions along the two axes. Because the DGS gradient captures the nonlocal features of F, it can point to a direction much closer to the global minimum than the local gradient. (Middle) The landscape of the 2D Ackley function that possesses many local minima. (Right) Comparison of the loss decay w.r.t. # function evaluations for the 1000D Ackley function.

3 Inverse design workflow with nonlocal DGS gradient

Optimization problem formulation A general electromagnetic design problem can be cast into the following optimization formulation:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{E}_1, ..., \mathbf{E}_n, \boldsymbol{\epsilon}_1, ..., \boldsymbol{\epsilon}_n, \mathbf{x}) \\ \text{subject to} \quad & g_j(\mathbf{x}) = 0, \quad j = 1, ..., m \\ & h_k(\mathbf{x}) \leq 0, \quad k = 1, ..., l \end{aligned} \tag{8}$$

where $\mathbf{E}_i$ is the electric field corresponding to the permittivity distribution $\boldsymbol{\epsilon}_i$, which depends on a parameterization vector $\mathbf{x} \in \mathbb{R}$ which is the computational design domain, and f is the objective function that defines the target of the optimization. A typical objective is to maximize the transmission, which is equivalent to minimize the negative

$$f_{obj}(\mathbf{x}) = -|\mathbf{c}^\dagger \mathbf{E}(\boldsymbol{\epsilon}(\mathbf{x}))|^2, \quad \mathbf{c}^\dagger \mathbf{E}(\boldsymbol{\epsilon}(\mathbf{x})) = \iint_S \zeta \cdot \mathbf{E}(\boldsymbol{\epsilon}(\mathbf{x})) dS \tag{9}$$

where $\mathbf{c}^\dagger \mathbf{E}$ means the overlap integrals to compute the model coupling efficiency of the electric field $\mathbf{E}$ with the target mode at the output ζ. $h_k(\mathbf{x})$ and $g_j(\mathbf{x})$ in Eq. (8) are inequality and equality constraints on $\mathbf{x}$, particularly fabrication and volume of materials constraints. For the optimization

problem in Eq. (8), the electric fields $\mathbf{E}_i$ generated by the input permittivity distribution $\boldsymbol{\epsilon}(\mathbf{x})$ should satisfy the Maxwell's equations in the frequency domain,

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{E}_i - \omega_i^2 \boldsymbol{\epsilon}(\mathbf{x}) \mathbf{E}_i = -i\omega_i \mathbf{J}_i \tag{10}$$

where $i = 1, ..., n$ is the input modes, ω_i is the angular frequency, μ is the magnetic permeability of free space and $\mathbf{J}_i$ is the input source which injects the current mode into the input waveguide. Eq. (10) is often solved by electromagnetic simulation using the finite-difference frequency-domain (FDFD) method [5] or finite-difference time-domain (FDTD) method [25].

Parameterization and constraints in optimization Solving the optimization problem defined in Eq. (8) led to continuously varying features of $\boldsymbol{\epsilon}(\mathbf{x})$, which is difficult for fabricating devices in practice. It is therefore critical to describe the permittivity distribution through a *parameterization* that addresses the fabrication challenges in device design [24]. Parameterization basically consists of two key components: *projection* operator and *filtering* operator. Projection operator aims to convert the continuous features to a binary feature that better captures a clear "0-1" design, where "0" represents a *background* material and "1" represents a *foreground* material in permittivity distribution. This can be achieved by defining an operator through the equation $\boldsymbol{\epsilon}(\mathbf{x}) = \boldsymbol{\epsilon}_b(\mathbf{x}) + \mathbb{H}(\varphi(\mathbf{x}))$,where $\boldsymbol{\epsilon}_b(\mathbf{x})$ is a permittivity background (constant) and $\varphi(\mathbf{x})$ is a 2D slice of the permittivity distribution and ranges from 0 to 1. A possible projection operator $\mathbb{H}$ is using nonlinear penalty methods [11, 12]. Filtering operator is often used to eliminate very tiny features and avoid to the formation of checker-board pattern in material layout [22]. Another common constraint in optimizing material layout is the volume fraction of material usage, $V(\mathbf{x})/V_0 \leq \gamma$, where V and V_0 are the expected material volume and design volume respectively, and γ is the specific volume fraction.

Methodology workflow Figure 2 shows the workflow to implement DGS-based nonlocal optimization method for inverse design with four core components, that are parameterization, physics simulation, objective formulation and optimization.

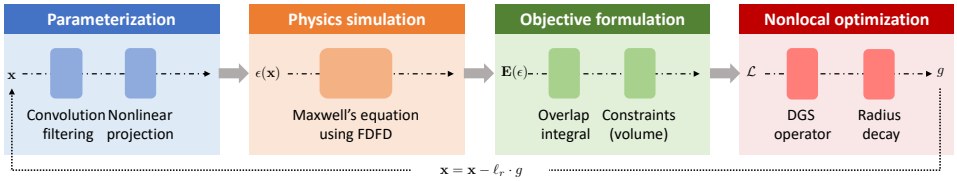

Figure 2: The proposed inverse design workflow with nonlocal DGS gradient-based optimization

4 Wavelength demultiplexer example

As shown in Figure 3, we choose a three-port structure with 500 nm input waveguide and output waveguides and a square 2.5 μm × 2.5 μm design region. We design a device for the 220 nm silicon-on-insulator (SOI) platform where the structure is constrained to a single fully etched Si layer on a SiO_2 substrate with air cladding. For illustration, the refractive indices of $n_{\text{air}} = 1$, $n_{SiO_2} = 1.45$ and $n_{\text{Si}} = 3.5$ are used. The purpose of inverse design is to separate 1300 nm signal to the upper waveguide and 1500 nm signal to the bottom waveguide.

The final optimized design and the corresponding electric field intensity are diagrammed in Figure 4. Note that the device designed by nonlocal optimization method using the DGS gradient, shown in Figure 4, displays a nonintuitive geometry while retaining relatively large features and a clear permittivity distribution with ideal binarization. The light takes a relatively confined path through the structure at both wavelengths. The optimization history, shown in Figure 3, provides iterative changes of the permittivity distribution during the optimization process. It is clear to note that the local gradient shows a fast convergence but it quickly traps into a local minimum and difficult to escape even though a large number of evaluations are performed. On the contrary, the DGS outperformance the local method and converges to a better solution, $f^*_{\text{DGS}} = 13.58$, that performs $\sim 10\%$ improvement compared with the local gradient method that is $f^*_{\text{local}} = 12.40$.

Robustness performance In this case, we repeatably run 100 times optimization with three different levels of randomness described by Gaussian noise: $N(0, 0.1)$, $N(0, 0.05)$ and $N(0, 0.01)$.

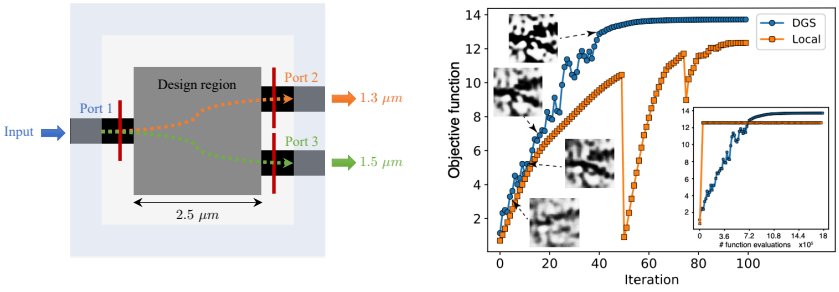

Figure 3: (Left) Illustration of wavelength demultiplexer design. The structure consists of one input waveguide (port 1) and two output waveguides (port 2 and port 3). The outer hatched light blue frame represents the simulation domain, specifically, the perfectly matched layer (PML) boundaries. The goal of inverse design is route to 1.3 μm through the top waveguide and 1.5 μm through the bottom waveguide. (Right) Optimization iteration history.

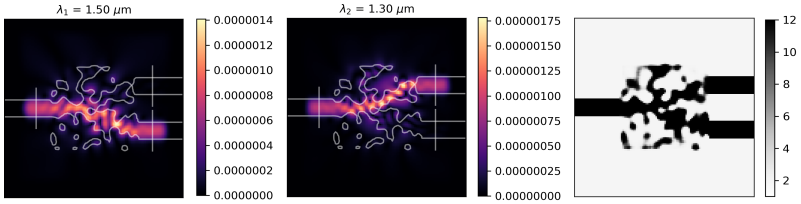

Figure 4: Electric field intensity of the optimized device at 1500 nm (Left) and 1300 nm (Middle), as well as the optimized permittivity distribution (Right) using optimization method with DGS gradient

The histogram of final objective values, as shown in Figure 5 (a-c), shows the distribution of the final objective values using the nonlocal optimization method with DGS gradient and local gradient. The nonlocal method outperforms the local gradient in terms of the objective values and shows a smaller variation in all three levels of randomness. In other words, the nonlocal method demonstrates stronger robustness and reliability to resist the local minima caused by random initialization.

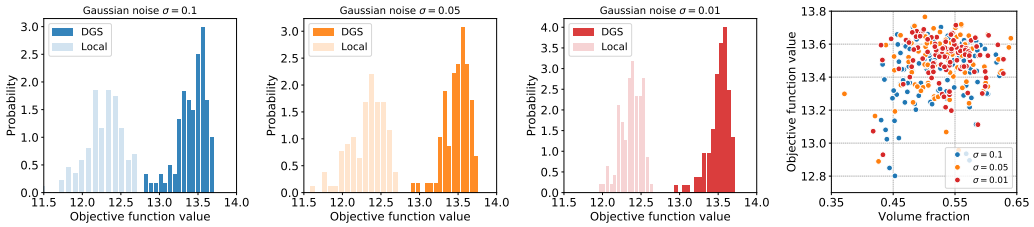

Figure 5: Objective function values of 100 local minmia given different noise levels. Dark color represents the distribution using the nonlocal optimization method with DGS gradient and light color represents the distribution using local gradient algorithm. The first three columns represent the three noise levels and the right column displays a fact that there is no obvious correlation between the volume fraction and final objective values.

Inverse design optimization with volume constraint We conduct a study to investigate the influence of material volume constraint. Most cases of volume fraction γ in Figure 5 (d) concentrates in 0.45−0.65 and no cases are lower than 0.35. We therefore reformulate the optimization by adding a volume constraint and solve this constrained optimization problem using the Method of Moving Asymptotes (MMA) [26]. However, MMA is limited to seek optima using local gradients information, either via adjoint method or finite difference. We address this challenge by inserting the DGS gradient into the MMA optimizer so that we can exploit the nonlocal exploration of the DGS operator to search for a better design. The final designed devices with three different amounts of material usage are shown in Figure 6.

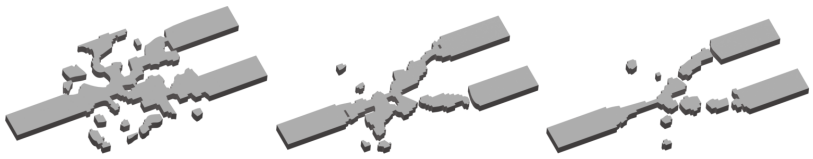

Figure 6: A three-dimensional rendering of the optimized design. Silicon is shown in grey and light enters the optimized device from the input waveguide on the left-hand side (port 1) and exits via one of the two output waveguides (port 2 and port 3) on the right side. There are three optimized devices with different volume fraction: (left) $\gamma = 0.474$, (middel) $\gamma = 0.3$ and (right) $\gamma = 0.2$.

5 Conclusion

This work focuses on the development of a nonlocal method for computational inverse design in nanophotonics. A novel DGS gradient is introduced into inverse design to improve the nonlocal exploration required for escaping from local minima in high-dimensional non-convex landscapes. Our approach has advantages in portability and flexibility so that it is naturally incorporated with parameterization, physics simulation, and objective formulation to build up an effective optimization workflow for inverse design. The proposed method is demonstrated on benchmark mathematical functions and a real-world wavelength demultiplexer design problem.

6 Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR), Applied Mathematics program under contract ERKJ352, ERKJ369; and by the Artificial Intelligence Initiative at the Oak Ridge National Laboratory (ORNL). ORNL is operated by UT-Battelle, LLC., for the U.S. Department of Energy under Contract DEAC05-00OR22725.

References

[1] M. Abramowitz and I. Stegun, editors. *Handbook of Mathematical Functions*. Dover, New York, 1972.

[2] Guillermo Angeris, Jelena Vučković, and Stephen P. Boyd. Computational Bounds for Photonic Design. *ACS Photonics*, 6(5):1232–1239, 2019.

[3] Peter Ingo Borel, Anders Harpøth, Lars Hagedorn Frandsen, Martin Kristensen, Peixiong Shi, Jakob Søndergaard Jensen, and Ole Sigmund. Topology optimization and fabrication of photonic crystal structures. *Optics express*, 12(9):1996–2001, 2004.

[4] Sawyer D Campbell, David Sell, Ronald P Jenkins, Eric B Whiting, Jonathan A Fan, and Douglas H Werner. Review of numerical optimization techniques for meta-device design. *Optical Materials Express*, 9(4):1842–1863, 2019.

[5] Andreas Christ and Hans L Hartnagel. Three-dimensional finite-difference method for the analysis of microwave-device embedding. *IEEE Transactions on Microwave Theory and Techniques*, 35(8):688–696, 1987.

[6] Rasmus E Christiansen and Ole Sigmund. A tutorial for inverse design in photonics by topology optimization. *arXiv preprint arXiv:2008.11816*, 2020.

[7] Yuriy Elesin, Boyan Stefanov Lazarov, Jakob Søndergaard Jensen, and Ole Sigmund. Design of robust and efficient photonic switches using topology optimization. *Photonics and nanostructures-Fundamentals and Applications*, 10(1):153–165, 2012.

[8] Lars H Frandsen, Yuriy Elesin, Louise F Frellsen, Miranda Mitrovic, Yunhong Ding, Ole Sigmund, and Kresten Yvind. Topology optimized mode conversion in a photonic crystal waveguide fabricated in silicon-on-insulator material. *Optics express*, 22(7):8525–8532, 2014.

[9] Louise F Frellsen, Yunhong Ding, Ole Sigmund, and Lars H Frandsen. Topology optimized mode multiplexing in silicon-on-insulator photonic wire waveguides. *Optics express*, 24(15):16866–16873, 2016.

[10] Tyler W. Hughes, Momchil Minkov, Ian A.D. Williamson, and Shanhui Fan. Adjoint Method and Inverse Design for Nonlinear Nanophotonic Devices. *ACS Photonics*, 5(12):4781–4787, 2018.

[11] Jakob S Jensen and Ole Sigmund. Systematic design of photonic crystal structures using topology optimization: Low-loss waveguide bends. *Applied physics letters*, 84(12):2022–2024, 2004.

[12] Jakob Søndergaard Jensen and Ole Sigmund. Topology optimization for nano-photonics. *Laser & Photonics Reviews*, 5(2):308–321, 2011.

[13] Jiaqi Jiang and Jonathan A Fan. Simulator-based training of generative neural networks for the inverse design of metasurfaces. *Nanophotonics*, 1(ahead-of-print), 2019.

[14] Yang Jiao, Shanhui Fan, and David AB Miller. Systematic photonic crystal device design: global and local optimization and sensitivity analysis. *IEEE journal of quantum electronics*, 42(3):266–279, 2006.

[15] Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vucković, and Alejandro W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670, 2018.

[16] Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

[17] John Peurifoy, Yichen Shen, Li Jing, Yi Yang, Fidel Cano-Renteria, Brendan G. DeLacy, John D. Joannopoulos, Max Tegmark, and Marin Soljačić. Nanophotonic particle simulation and inverse design using artificial neural networks. *Science Advances*, 4(6):1–8, 2018.

[18] Thaibao Phan, David Sell, Evan W Wang, Sage Doshay, Kofi Edee, Jianji Yang, and Jonathan A Fan. High-efficiency, large-area, topology-optimized metasurfaces. *Light: Science & Applications*, 8(1):1–9, 2019.

[19] Alexander Y Piggott, Jan Petykiewicz, Logan Su, and Jelena Vučković. Fabrication-constrained nanophotonic inverse design. *Scientific reports*, 7(1):1–7, 2017.

[20] Alfio Quarteroni, Riccardo Sacco, and Fausto Saleri. *Numerical Mathematics*, volume 332. Springer Science Business Media &, 2007.

[21] Philipp-Immanuel Schneider, Xavier Garcia Santiago, Victor Soltwisch, Martin Hammerschmidt, Sven Burger, and Carsten Rockstuhl. Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction. *ACS Photonics*, 6(11):2726–2733, 2019.

[22] Ole Sigmund and Joakim Petersson. Numerical instabilities in topology optimization: a survey on procedures dealing with checkerboards, mesh-dependencies and local minima. *Structural optimization*, 16(1):68–75, 1998.

[23] Logan Su, Alexander Y. Piggott, Neil V. Sapra, Jan Petykiewicz, and Jelena Vučković. Inverse Design and Demonstration of a Compact on-Chip Narrowband Three-Channel Wavelength Demultiplexer. *ACS Photonics*, 5(2):301–305, 2018.

[24] Logan Su, Dries Vercruysse, Jinhie Skarda, Neil V. Sapra, Jan A. Petykiewicz, and Jelena Vučković. Nanophotonic inverse design with SPINS: Software architecture and practical considerations. *Applied Physics Reviews*, 7(1):1–26, 2020.

[25] Dennis M Sullivan. *Electromagnetic simulation using the FDTD method*. John Wiley & Sons, 2013.

[26] Krister Svanberg. The method of moving asymptotes—a new method for structural optimization. *International journal for numerical methods in engineering*, 24(2):359–373, 1987.

[27] Dries Vercruysse, Neil V Sapra, Logan Su, Rahul Trivedi, and Jelena Vučković. Analytical level set fabrication constraints for inverse design. *Scientific reports*, 9(1):1–7, 2019.

[28] Dan Yan, Lili Qiu, Min Xue, Zihui Meng, and Yunfeng Wang. A flexible surface-enhanced raman substrates based on cellulose photonic crystal/ag-nanoparticles composite. *Materials & Design*, 165:107601, 2019.

[29] Jiaxin Zhang, Hoang Tran, Dan Lu, and Guannan Zhang. A Novel Evolution Strategy with Directional Gaussian Smoothing for Blackbox Optimization. *arXiv e-prints*, page arXiv:2002.03001, February 2020.