

# Sensitivity analysis and optimization of sub-wavelength optical gratings using adjoints

Armand C. R. Niederberger,<sup>1,2,3,\*</sup> David A. Fattal,<sup>3,4</sup>  
Nicolas R. Gauger,<sup>5</sup> Shanhui Fan,<sup>6</sup> and Raymond G. Beausoleil<sup>3</sup>

<sup>1</sup>*SUPA, Department of Physics, University of Strathclyde, Glasgow G4 0NG, UK*

<sup>2</sup>*Edward L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA*

<sup>3</sup>*Hewlett-Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304, USA*

<sup>4</sup>*LEIA Inc., 2440 Sand Hill Rd Ste 303, Menlo Park, CA 94025, USA*

<sup>5</sup>*Department of Mathematics and Center for Computational Engineering Science (CCES),  
RWTH Aachen University, 52062 Aachen, Germany*

<sup>6</sup>*Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA*

[\\*Armand.Niederberger@hp.com](mailto:Armand.Niederberger@hp.com)

**Abstract:** Numerical optimization of photonic devices is often limited by a large design space: the finite-differences gradient method requires as many electric field computations as there are design parameters. Adjoint-based optimization can deliver the same gradients with only two electric field computations. Here, we derive the relevant adjoint formalism and illustrate its application for a waveguide slab, and for the design of optical sub-wavelength gratings.

© 2014 Optical Society of America

**OCIS codes:** (050.0050) Diffraction and gratings; (220.4830) Systems design; (230.0230) Optical devices; (310.6805) Theory and design; (350.4238) Nanophotonics and photonic crystals.

---

## References and links

1. J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations* (Springer, 1971).
2. R. Becker and R. Rannacher, "An optimal control approach to error control and mesh adaption in finite element methods," *Acta Numerica* **10**, 1–102 (2001).
3. K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, "Introduction to adaptive methods for differential equations," *Acta Numerica* **4**, 105–158 (1995).
4. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Parallel Data Processing* **1**, 318–362 (1986).
5. J. Reuther, A. Jameson, J. J. Alonso, M. J. Remlinger, and D. Saunders, "Constrained multipoint aerodynamic shape optimization using an adjoint formulation and parallel computers, part 1," *Journal of Aircraft* **36**(1), 51–60 (1999).
6. J. Reuther, A. Jameson, J. J. Alonso, M. J. Remlinger, and D. Saunders, "Constrained multipoint aerodynamic shape optimization using an adjoint formulation and parallel computers, part 2," *Journal of Aircraft* **36**(1), 61–74 (1999).
7. Y. seek Chung, Changyul-Cheon, I.-H. Park, and S.-Y. Hahn, "Optimal shape design of microwave device using fdfd and design sensitivity analysis," *Microwave Theory and Techniques, IEEE Transactions on* **48**, 2289–2296 (2000).
8. N. Georgieva, S. Glavic, M. Bakr, and J. Bandler, "Feasible adjoint sensitivity technique for em design optimization," *Microwave Theory and Techniques, IEEE Transactions on* **50**, 2751–2758 (2002).
9. N. K. Nikolova, H. W. Tam, and M. H. Bakr, "Sensitivity analysis with the fdfd method on structured grids," *Microwave Theory and Techniques, IEEE Transactions on* **52**, 1207–1216 (2004).
10. N. K. Nikolova, Y. Li, Y. Li, and M. H. Bakr, "Sensitivity analysis of scattering parameters with electromagnetic time-domain simulators," *Microwave Theory and Techniques, IEEE Transactions on* **54**, 1598–1610 (2006).

11. G. Veronis, R. W. Dutton, and S. Fan, "Method for sensitivity analysis of photonic crystal devices," *Optics Letters* **29**, 2288–2290 (2004).
12. Y. Jiao, S. Fan, and D. A. B. Miller, "Photonic crystal device sensitivity analysis with wannierbasis gradients," *Optics Letters* **30**, 302–304 (2005).
13. P. Seliger, M. Mahvash, C. Wang, and A. F. J. Levi, "Optimization of aperiodic dielectric structures," *Journal of Applied Physics* **100**, 034310 (2006).
14. C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Optics Express* **21**, 21693–21701 (2013).
15. O. D. Miller, C. W. Hsu, M. T. H. Reid, W. Qiu, B. G. DeLacy, J. D. Joannopoulos, M. Soljačić, and S. G. Johnson, "Fundamental limits to extinction by metallic nanoparticles," *Physical Review Letters* **112**, 123903 (2014).
16. D. Fattal, J. Li, Z. Peng, M. Fiorentino, and R. G. Beausoleil, "Flat dielectric grating reflectors with focusing abilities," *Nature Photonics* **4**, 466–470 (2010).
17. V. Liu, D. Miller, and S. Fan, "Highly tailored computational electromagnetics methods for nanophotonic design and discovery," *Proceedings of the IEEE* **101**, 484–493 (2013).
18. J. Lu and Vučković, "Inverse design of nanophotonic structures using complementary convex optimization," *Optics Express* **18**, 3793–3804 (2010).
19. R. Courant, K. Friedrichs, and H. Lewy, "Über die partiellen Differenzgleichungen der mathematischen Physik," *Mathematische Annalen* **100**, 32–74 (1928).
20. K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *Antennas and Propagation, IEEE Transactions on* **14**, 302–307 (1966).
21. D. Taillaert, W. Bogaerts, P. Bienstman, T. Krauss, P. van Daele, I. Moerman, S. Verstuyft, K. De Mesel, and R. Baets, "An out-of-plane grating coupler for efficient butt-coupling between compact planar waveguides and single-mode fibers," *Quantum Electronics, IEEE Journal of* **38**, 949–955 (2002).
22. G. Roelkens, D. V. Thourhout, and R. Baets, "High efficiency silicon-on-insulator grating coupler based on a poly-silicon overlay," *Optics Express* **14**, 11622–11630 (2006).
23. A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, "MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method," *Computer Physics Communications* **181**, 687–702 (2010).

## 1. Introduction

When designing optical structures via numerical simulations, we are often lead to compute the gradient of some target function with respect to our design variables: we may, for example, want to assess the sensitivity of our structure to manufacturing imprecisions, or optimize the structure's performance with respect to a given metric. Computing the gradient of a function usually requires two optical simulations per design parameter. This is fine for simple systems, but intractable for many optical systems of interest (e.g. diffractive elements or photonics circuits) which can contain millions of independent parameters.

Fortunately, if the target function is specified in advance, it is possible to exploit the mathematical structure of a large class of optical simulations to compute the multivariate gradient using only two simulations, no matter how many parameters are present. Next to the *primal solve*, one has to solve the *adjoint problem* once. The adjoint solution acts as a kind of influence function regarding the respective target function and avoids to solve the primal problem for each parameter variation. This is the principle idea behind *adjoint-based optimization and sensitivity analysis*. This approach is well-known to several fields of application as control theory, mesh adaptation, error estimation or propagation [1–4] and has been used very successfully for many years in performing aerodynamic shape optimizations [5,6]. Later, it was used by electrical engineers in the microwave regime [7–10]. In the optical arena, the approach has been used by only a few groups [11–15].

The potential of this approach to optimize complex optical structures with many design parameters can hardly be overstated. For example, optimizing optical lens gratings [16] typically requires supercomputers or very large clusters. Countless runs with slightly different parameter settings are performed to find an appropriate set of design parameters. Being able to reduce the required number of these computations significantly relaxes the computational requirements.

Thus, instead of probing a huge number of parameter settings, we can probe only a few and then perform efficient local optimization in the vicinity of the best settings. Adjoint-based optimization is an elegant method to compute the gradients that are needed for local optimization. This paper shows how to reduce the computational complexity of finding the gradient tremendously: from approximately as many field computations as there are design parameters (which is often in the hundreds or thousands) to two, independently of the number of design parameters. This makes it possible to optimize rather complex photonic systems with respect to an arbitrary number of parameters on a single desktop or laptop computer.

It is worth noticing that there are several different approaches to optimizing photonic structures and devices [17]. One particularly elegant approach is inverse design [18], where the structure is optimized to fit a particular field. In any case, when an optimization approach uses local gradients, adjoints may be used to efficiently compute them.

This paper derives the relevant equations for adjoint-based optimization and sensitivity analysis for optical systems. In particular, we present a formalism that is well suited for sensitivity analysis and optimization using a Maxwell solver. Therefore, this paper provides a detailed tutorial on the adjoint method used to solve electromagnetic problems. We provide a simple example of a slab waveguide where the adjoint method can be compared to the exact analytic model. We also show that the adjoint method can be effectively applied to the optimization of a grating coupler – a task where the adjoint method has not been applied before.

## 2. Problem description

In all generality, adjoint-based optimization is used to locally optimize a target function  $T(\mathbf{x}, \mathbf{p})$  which depends on the physical state of the system  $\mathbf{x}$ , as well as a number of design parameters  $\mathbf{p}$ . For optical gratings, for example, the physical state of the system can be described conveniently by the electric field (or sometimes just a single component thereof) at each point of the computational grid. The vector of design parameters  $\mathbf{p}$  contains, for example, geometric parameters describing the layout of our grating. In order to determine the state of our system  $\mathbf{x}(\mathbf{p})$ , we solve a given vector equation  $\mathbf{R}(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ , such as Maxwell's equations.

For our optimization, we need to know how each of the design parameters affects the value of our target function: we need to compute  $\frac{dT}{d\mathbf{p}}$ . For each design parameter  $p_k$ , this derivative is

$$\frac{dT}{dp_k} = \sum_i \frac{\partial T}{\partial x_i} \frac{dx_i}{dp_k} + \frac{\partial T}{\partial p_k} = \frac{\partial T}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dp_k} + \frac{\partial T}{\partial p_k}, \quad (1)$$

provided that the design parameters are mutually independent. What makes the evaluation of the above equation computationally costly is the term  $\frac{d\mathbf{x}}{dp_k}$ . In fact, it is important to understand that the dependency of our solution  $\mathbf{x}$  on the design parameters  $\mathbf{p}$  is *implicit* via the condition that  $\mathbf{R}(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ . Otherwise, we would be able to analytically compute the relevant derivatives with respect to our design parameters and our optimization problem would be trivial.

Computing  $\frac{d\mathbf{x}}{d\mathbf{p}}$  via the finite-differences approximation is numerically very inefficient:

$$\frac{d\mathbf{x}}{dp_k} \approx \frac{1}{\Delta p_k} (\mathbf{x}(\mathbf{p}_{\Delta p_k}) - \mathbf{x}(\mathbf{p})), \quad (2)$$

where  $\Delta p_k$  is a small variation on the  $k$ -th design variable, and  $\mathbf{p}_{\Delta p_k}$  denotes the vector of design parameters where  $\Delta p_k$  was added to the  $k$ -th component only. As we can see from the above finite-difference equation, this approach demands for one reference computation of  $\mathbf{R}(\mathbf{x}, \mathbf{p}) = \mathbf{0}$  to compute  $\mathbf{x}(\mathbf{p})$ , and then one additional  $\mathbf{R}(\mathbf{x}, \mathbf{p}_{\Delta p_k}) = \mathbf{0}$  to find the necessary  $\mathbf{x}(\mathbf{p}_{\Delta p_k})$  for each component of the gradient. To compute the full gradient with respect to  $n$  design variables, we would therefore have to solve the vector equation  $\mathbf{R} = \mathbf{0}$  a total of  $n + 1$  times. For costly  $\mathbf{R}$  and

large numbers of design parameters, this often exceeds the available resources of computational power and time.

As we show below, it is often possible to exploit the mathematical structure of the local optimization and sensitivity analysis problem: the adjoint method allows us to compute the relevant gradient  $\frac{dT}{d\mathbf{p}}$  performing only two non-trivial simulations. No matter how many design variables we are dealing with.

### 3. Derivation of adjoint approach

Using the adjoint method, we are able to determine  $\frac{dT}{d\mathbf{p}}$  using only one solution of  $\mathbf{R}(\mathbf{x}, \mathbf{p}) = \mathbf{0}$ , and one solution of the adjoint equation. As long as the latter is of similar numerical complexity as the direct equation  $\mathbf{R}$ , the adjoint-based optimization is likely to reduce the computational cost of  $\frac{dT}{d\mathbf{p}}$ , which we are interested in.

Our starting point is the fact that, for any set of design parameters  $\mathbf{p}$ , and any corresponding state  $\mathbf{x}$ , we know that the governing vector equation  $\mathbf{R}$  vanishes,

$$\mathbf{R}(\mathbf{x}, \mathbf{p}) = \mathbf{0}. \quad (3)$$

Since this condition has to hold for *any* set of parameters  $\mathbf{p}$  with corresponding  $\mathbf{x}$ , its Taylor expansion also has to vanish. Intuitively speaking, in the solution space, where we only consider design parameters and their corresponding states, the derivative of the state function  $\mathbf{R}$  with respect to a design parameter  $p_k$  also vanishes,

$$\frac{d\mathbf{R}}{dp_k} = \frac{\partial \mathbf{R}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dp_k} + \frac{\partial \mathbf{R}}{\partial p_k} = \mathbf{0}. \quad (4)$$

We can now multiply  $\frac{d\mathbf{R}}{dp_k}$  on the left by any complex vector  $\mathbf{v}$  and subtract the product from  $\frac{dT}{dp_k}$  without changing the latter gradient's value

$$\frac{dT}{dp_k} = \frac{\partial T}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dp_k} + \frac{\partial T}{\partial p_k} - \mathbf{v}^\dagger \left( \frac{\partial \mathbf{R}}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dp_k} + \frac{\partial \mathbf{R}}{\partial p_k} \right). \quad (5)$$

By regrouping the terms, we find

$$\frac{dT}{dp_k} = \frac{\partial T}{\partial p_k} - \mathbf{v}^\dagger \frac{\partial \mathbf{R}}{\partial p_k} + \left( \frac{\partial T}{\partial \mathbf{x}} - \mathbf{v}^\dagger \frac{\partial \mathbf{R}}{\partial \mathbf{x}} \right) \frac{d\mathbf{x}}{dp_k}. \quad (6)$$

This expression is the key to the adjoint method: if we are able to find a  $\mathbf{v}$  such that the pre-factor of  $\frac{d\mathbf{x}}{dp_k}$  vanishes, we do not have to perform the expensive  $n + 1$  solutions of our state function to find the gradient  $\frac{dT}{d\mathbf{p}}$ . In fact, we will simply have to solve  $\mathbf{R} = \mathbf{0}$  once to find our  $\mathbf{x}(\mathbf{p})$ . Then, we will solve our adjoint equation

$$\frac{\partial T}{\partial \mathbf{x}} = \mathbf{v}^\dagger \frac{\partial \mathbf{R}}{\partial \mathbf{x}}, \quad (7)$$

to find a convenient  $\mathbf{v}$ . And these two computations then straightforwardly allow us to determine the derivative of our target function with respect to our design parameters,

$$\frac{dT}{dp_k} = \frac{\partial T}{\partial p_k} - \mathbf{v}^\dagger \frac{\partial \mathbf{R}}{\partial p_k}. \quad (8)$$

To illustrate the use of this powerful approach, we will now show how to apply the adjoint method in two concrete situations and derive the relevant equations for optical systems.

#### 4. Optical systems

In this section, we derive the relevant equations for adjoint-based optimization and sensitivity analysis in optical systems, starting the same way as Nikolova *et al.* [9] did for the microwave regime. For mathematical simplicity, we consider the dispersion-free, and linear case. Nonetheless, adjoint-based optimization can also be used for non-uniform, anisotropic media. Maxwell's equations are

$$\begin{aligned}\nabla \times \mathbf{H} &= \frac{\partial D}{\partial t} + \mathbf{J}_\sigma + \mathbf{J} \\ -\nabla \times \mathbf{E} &= \frac{\partial \mathbf{B}}{\partial t},\end{aligned}\quad (9)$$

where  $\mathbf{J}$  denotes the density of current sources. The constitutive relations are

$$\begin{aligned}\mathbf{B} &= \mu \mathbf{H} \\ \mathbf{D} &= \varepsilon \mathbf{E} \\ \mathbf{J}_\sigma &= \sigma \mathbf{E},\end{aligned}\quad (10)$$

where  $\mu$  denotes the permeability tensor,  $\varepsilon$  the permittivity tensor, and  $\sigma$  the specific conductivity tensor. These equations can be combined into

$$\nabla \times \mu^{-1} \nabla \times \mathbf{E} + \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma \frac{\partial \mathbf{E}}{\partial t} = -\frac{\partial \mathbf{J}}{\partial t}.\quad (11)$$

In our optical gratings, the specific conductivity tensor  $\sigma$  vanishes. Furthermore, we write our electric field (and source) as  $\mathbf{E}(t) = \text{Re}(\bar{\mathbf{E}}e^{-i\omega t})$ , where  $\bar{\mathbf{E}} \in \mathbb{C}$ . Note that the permittivity  $\varepsilon$  describes the layout of our optical grating and therefore directly depends on our design parameters. We rewrite Eq (11) as

$$\mathbf{R}(\bar{\mathbf{E}}, \mathbf{p}) = \mathcal{M}\bar{\mathbf{E}} - i\omega\bar{\mathbf{J}} = 0,\quad (12)$$

where we introduced

$$\mathcal{M}\bar{\mathbf{E}} = \nabla \times \mu^{-1} \nabla \times \bar{\mathbf{E}} - \omega^2 \varepsilon \bar{\mathbf{E}}.\quad (13)$$

To compute an electric field  $\bar{\mathbf{E}}$  that satisfies Eq. (11), we can use any convenient Maxwell solver, such as any FDTD [19, 20] solver. The simulation returns the complex electric field amplitude at each point of the computational grid.

Our target function  $T$  is the electric field strength in a specific region  $\Omega$

$$T(\bar{\mathbf{E}}, \mathbf{p}) = \frac{1}{N} \int_{\Omega} d\Omega |\bar{\mathbf{E}}|^2 = \bar{\mathbf{E}}_0^\dagger \bar{\mathbf{E}},\quad (14)$$

where  $\bar{\mathbf{E}}_0 = \bar{\mathbf{E}}$  within the region  $\Omega$  and zero outside, and  $N$  is some (irrelevant) normalization constant introduced for mathematical rigor, only. With this notation, we also introduce the additional hypothesis that  $\bar{\mathbf{E}}$  varies predominantly in norm (rather than in phase), as we vary  $\mathbf{p}$ . This will allow us to approximate the derivative of our target function with respect to the electric field as  $\bar{\mathbf{E}}_0^\dagger$ .

At this point, we are able to derive the adjoint equation for this problem. First, note that for isotropic dielectric media where  $\varepsilon^\dagger = \varepsilon$  and  $\mu^\dagger = \mu$  we have  $\mathcal{M}^\dagger = \mathcal{M}$ . Second, we use  $\bar{\mathbf{E}}$  computed as the numerical solution of Eq. (12) in Eq. (7) and find  $\bar{\mathbf{E}}_0^\dagger = \bar{\mathbf{v}}^\dagger \mathcal{M}$ . This equation will allow us to compute the adjoint field  $\bar{\mathbf{v}}$ . Third, we rewrite this equation in a way that shows its correspondence to the direct problem Eq. (12) and conclude that

$$\mathcal{M}\bar{\mathbf{v}} - i\omega \left( \frac{1}{i\omega} \bar{\mathbf{E}}_0 \right) = 0.\quad (15)$$

Both the direct problem Eq. (12) and the adjoint problem Eq. (15) can be simulated using the same Maxwell solver. In fact, we see that the adjoint field  $\bar{\mathbf{v}}$  can be interpreted as an electric field that was created by a particular set of sources. The adjoint sources term,  $\frac{1}{i\omega}\bar{\mathbf{E}}_0$ , is proportional to the field of the direct solution, wherever we are optimizing the field, and zero everywhere else.

Computing the relevant gradient Eq. (8) is computationally trivial once we have the solutions of the direct problem,  $\bar{\mathbf{E}}$ , and of the adjoint problem,  $\bar{\mathbf{v}}$ . Our target function (14) does not depend on the design parameters  $\mathbf{p}$ . Hence,  $\frac{\partial T}{\partial p_k} = 0$ . Also, the only element of the governing vector equation (12) that depends on the design parameters is the spatially varying dielectric constant  $\varepsilon$ . Therefore,  $\frac{\partial \mathbf{R}}{\partial p_k} = \frac{\partial \mathcal{M}}{\partial p_k} \bar{\mathbf{E}} = \frac{\partial \varepsilon}{\partial p_k} \bar{\mathbf{E}}$  and the gradient Eq. (8) is simply

$$\frac{dT}{dp_k} = -\bar{\mathbf{v}}^\dagger \frac{\partial \varepsilon}{\partial p_k} \bar{\mathbf{E}}. \quad (16)$$

Note that this expression includes simple vector-matrix multiplication and finding the derivative of our spatially varying dielectric constant  $\varepsilon$  with respect to each of the design parameters  $p_k$ . However, we certainly know the current dielectric layout of the device we are trying to optimize. Hence, we can compute  $\frac{\partial \varepsilon}{\partial p_k}$  either analytically, or by using computationally cheap numerical methods. This shows that we can ultimately compute the derivative of our target function with respect to our design parameters,  $\frac{dT}{dp_k}$ , with only two non-trivial computations: to find the direct solution  $\bar{\mathbf{E}}$ , and the adjoint solution  $\bar{\mathbf{v}}$ .

## 5. Example 1: waveguide slab

Consider the text-book example of a waveguide slab of thickness  $h$ , described by the cover material index  $n_c$ , the guiding film index  $n_f$ , and the substrate index  $n_s$ . We would like to know what modes are supported by this waveguide slab for a given wavelength  $\lambda$ . The vacuum wave vector is  $k_0 = \frac{2\pi}{\lambda}$ . The characteristic equation for transverse electric (TE) modes is

$$\tan(h\kappa_f) = \frac{\gamma_c + \gamma_s}{\kappa_f \left(1 - \frac{\gamma_c \gamma_s}{\kappa_f^2}\right)}, \quad (17)$$

where  $\kappa_f = \sqrt{k_0^2 n_f^2 - \beta^2}$  is the transverse wave vector,  $\gamma_c = \sqrt{\beta^2 - k_0^2 n_c^2}$  and  $\gamma_s = \sqrt{\beta^2 - k_0^2 n_s^2}$  are the attenuation coefficients of cover and substrate, respectively, which all depend on the longitudinal wave vector  $\beta$ . Therefore, solving Eq. (17) means finding one or all  $\beta$  for which this equation holds.

If we were to ask how a specific value of  $\beta$  changes as we modify, for example, the height of the slab, we would usually have to solve Eq. (17) for different values of  $h$ . Worse, even, if we were to modify the indices  $n_c$ ,  $n_f$ , and  $n_s$ , too, we would have to solve that same equation over and over again to derive how  $\beta$  depends on them. Nowadays computers can solve this transcendental equation really easily, but for the sake of explaining the adjoint method, pretend that this is an arduous task.

To use the adjoint method, we rewrite Eq. (17) as

$$R = \tan(h\kappa_f) - \frac{\gamma_c + \gamma_s}{\kappa_f \left(1 - \frac{\gamma_c \gamma_s}{\kappa_f^2}\right)}, \quad (18)$$

knowing that the value of  $R(\beta, \mathbf{p})$  depends on the longitudinal wave vector, as well as a number of design parameters  $\mathbf{p} = (h, n_c, n_f, n_s)^T$ , and vanishes for all solutions  $\beta(\mathbf{p})$ . For the sake of

simplicity, we shall define our target function  $T$  to be simply the longitudinal wave vector,

$$T = \beta, \quad (19)$$

We know from the above Eq. (6) that we can avoid these repetitive calculations if we find a  $v$  such that

$$\frac{\partial T}{\partial \beta} = 1 = v \frac{\partial R}{\partial \beta}. \quad (20)$$

In this case,

$$\frac{dT}{dp_k} = \frac{d\beta}{dp_k} = -v \frac{\partial R}{\partial p_k}. \quad (21)$$

The derivative of our characteristic equation with respect to the longitudinal wave vector is

$$\frac{\partial R}{\partial \beta} = \frac{\partial}{\partial \beta} (\tan(h\kappa_f)) - \frac{\partial}{\partial \beta} \left( \frac{\gamma_c + \gamma_s}{\kappa_f \left( 1 - \frac{\gamma_c \gamma_s}{\kappa_f^2} \right)} \right), \quad (22)$$

where

$$\frac{\partial}{\partial \beta} (\tan(h\kappa_f)) = \frac{-h\beta}{\kappa_f \cos^2(h\kappa_f)}, \quad (23)$$

and, introducing  $q = \kappa_f \left( 1 - \frac{\gamma_c \gamma_s}{\kappa_f^2} \right)$ ,

$$\frac{\partial}{\partial \beta} \left( \frac{\gamma_c + \gamma_s}{q} \right) = \frac{\frac{\beta}{\gamma_c} + \frac{\beta}{\gamma_s}}{q} - \frac{\gamma_c + \gamma_s}{q^2} \frac{\partial q}{\partial \beta}, \quad (24)$$

where

$$\frac{\partial q}{\partial \beta} = -\frac{\beta}{\kappa_f} \left[ 1 + \frac{\gamma_s}{\gamma_c} + \frac{\gamma_c}{\gamma_s} + \frac{\gamma_c \gamma_s}{\kappa_f^2} \right]. \quad (25)$$

Collecting all these terms, we find the solution to our adjoint Eq. (20),

$$\frac{1}{v} = -\beta \left[ \frac{h}{\kappa_f \cos^2(h\kappa_f)} + \frac{(\gamma_c + \gamma_s)\kappa_f}{\kappa_f^2 - \gamma_c \gamma_s} \left( \frac{1}{\gamma_s \gamma_c} + \frac{1 + \frac{\gamma_s}{\gamma_c} + \frac{\gamma_c}{\gamma_s} + \frac{\gamma_c \gamma_s}{\kappa_f^2}}{\kappa_f^2 - \gamma_c \gamma_s} \right) \right], \quad (26)$$

and we know that by defining

$$v = \frac{1}{\frac{\partial R}{\partial \beta}} \quad (27)$$

we can compute the relevant gradient as

$$\frac{d\beta}{dp_k} = -v \frac{\partial R}{\partial p_k}. \quad (28)$$

For example, if we are interested in computing how  $\beta$  changes as we vary the height of the guiding film,  $p_k = h$ , and we therefore need to compute

$$\frac{dR}{dh} = \frac{\kappa_f}{\cos^2(h\kappa_f)} \quad (29)$$

and finally, following Eq. (28),

$$\frac{d\beta}{dh} = -v \frac{\kappa_f}{\cos^2(h\kappa_f)}. \quad (30)$$

As we can see in Fig. 1, the adjoint method leads to the correct gradient. However, is far more efficient for computing multi-variable optimizations than finite-differences gradients.

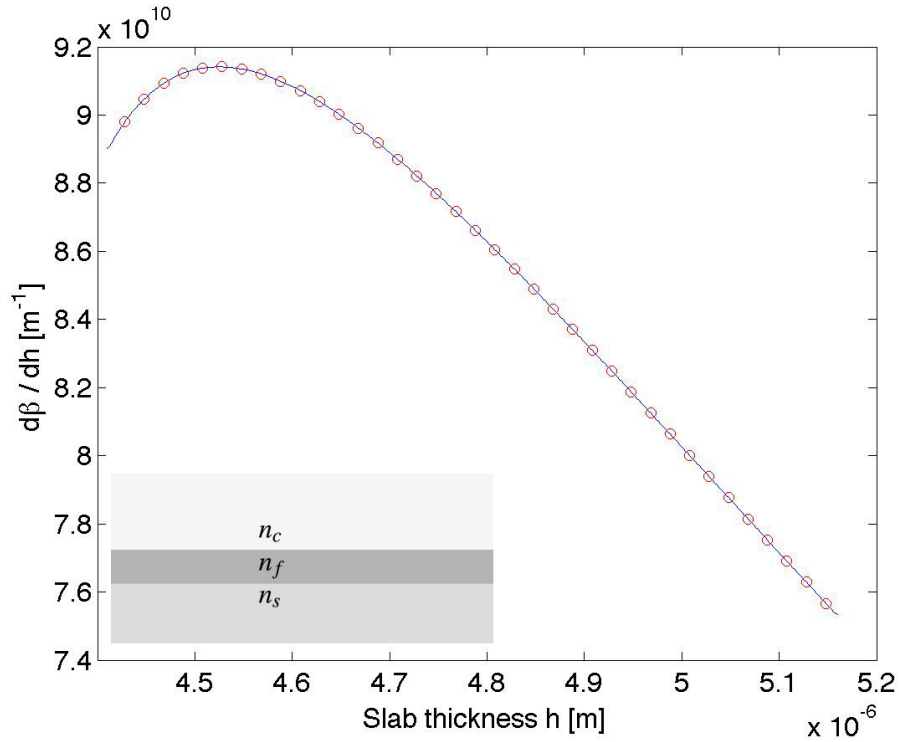


Fig. 1. The derivative of the longitudinal wave vector  $\beta$  with respect to the thickness of a waveguide slab can faithfully be reproduced using adjoints: the comparison between the direct calculation (blue line) and adjoint calculation (red dots) of  $\frac{d}{dh}\beta$  shows that the two methods deliver exactly the same result. However, the adjoint method requires no further expensive solutions of the eigenvalue system in order to determine the derivatives of  $\beta$  with respect to other design variables. Thus, the adjoint method delivers the relevant gradients of a system depending on multiple design parameters faster than the finite-differences approach. The inset shows a schematic of the waveguide slab structure comprised of the cover material  $n_c$ , the guiding film  $n_f$ , and the substrate  $n_s$ .

## 6. Example 2: grating coupler

Silicon grating couplers are used in photonic circuits to couple light between an on-chip waveguide and an off-chip optical fiber [21]. The exact groove pattern (width, position, depth of each groove) can be optimized to provide the best possible coupling efficiency [22]. For the sake of illustrating the adjoint method, we will restrict the design space to two variables, the uniform pitch  $p$  and duty cycle  $d$  of the grating grooves. Our simulations are assuming a waveguide thickness of  $0.22\mu\text{m}$ , a groove-depth of  $0.08\mu\text{m}$ , a wavelength of  $1.55\mu\text{m}$ .

The vector of our design parameters is

$$\mathbf{p} = \begin{pmatrix} p \\ d \end{pmatrix} \quad (31)$$

Figure 2 shows the effect of these design parameters on the design of our grating coupler.

To illustrate the effect of adjoint-based optimization, we computed the full topology of our optimization problem for this example (colored background in Fig. 3). We see the areas of best coupling in dark red, and those of worst coupling in dark blue. Of course, computing the



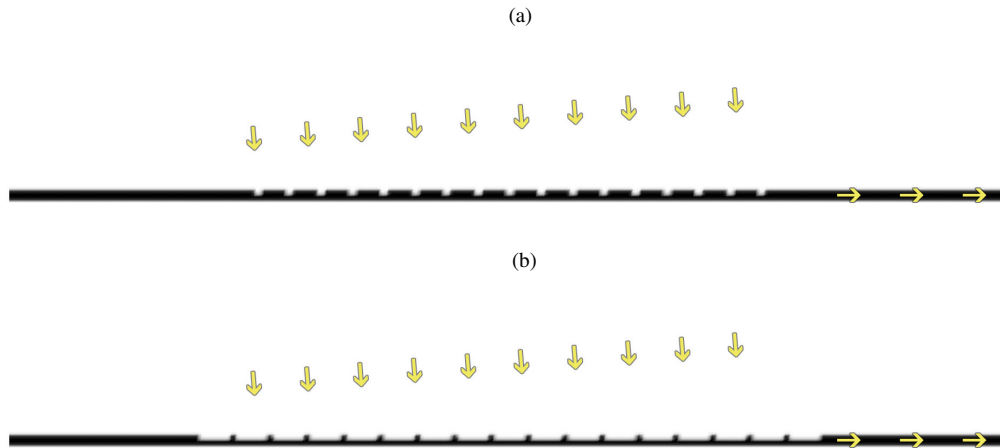


Fig. 2. Our SiO<sub>2</sub> grating coupler redirects free-space plane-waves (vertical arrows) to a guided mode inside the waveguide (horizontal arrows). We show the device at (a) a pitch of 0.58 and a duty cycle of 0.2, and (b) at a pitch of 0.68 and a duty cycle of 0.8. Both the pitch and the duty cycle are optimized to maximize the coupling between the incident light and the light inside the waveguide.

full topology is computationally impossible for most real-world applications. Here, it serves intuition and helps to verify that our algorithm works, indeed, as steepest ascent. Furthermore, it helped us select rather poor starting points for our optimization.

To perform the actual adjoint-based optimization, we select a number of starting points that serve as initial parameters. For each point, we

1. compute the direct electromagnetic field for specific design parameters (12),
2. compute the adjoint field for the same design parameters (15),
3. combine the direct and adjoint solutions to (trivially) compute the gradient (16),
4. update the design parameters for steepest ascent,
5. repeat steps 1-4 until convergence.

Note that sensitivity analysis can be performed by executing steps 1-3 for any desired set of design parameters.

Figure 3 shows the result of the full optimization (steps 1-5), and the (usually unknown) topology of our problem. In order to compute these optimization runs, finite-differences gradient optimization would have required three non-trivial field computations per gradient. Adjoint-based optimization, in contrast, only required two non-trivial field computations to find the same gradient. Thus, adjoint based optimization is already slightly faster than finite-differences based optimization for two design parameters.

Consider a grating optimization starting at a pitch of 0.62 and a duty cycle of 0.73 (black star). Adjoint-based optimization evolves these initial parameters to a pitch of 0.60 and a duty cycle of 0.275. Figure 4 shows the result of Meep [23] simulations of  $\text{real}(\vec{E})$  before and after optimization. The coupling of the incident light from the top to inside the waveguide (inside the black square) has improved from 0.39 to 0.88.

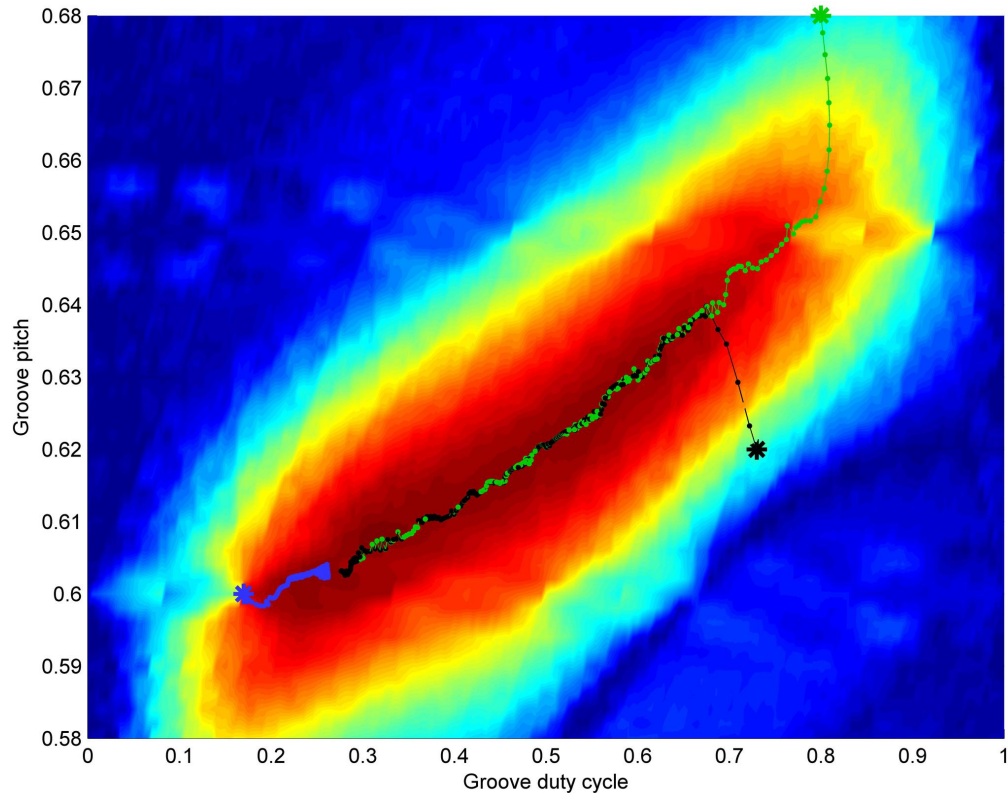


Fig. 3. Adjoint-based optimization convergence: from three initial parameters (stars) in parameter regions of poor optical coupling (blue, green, yellow), adjoint-based optimization evolves the design parameters to the parameter regions of optimal coupling (dark red). The computation of the relevant gradients requires only two Maxwell solutions, independently of the number of design parameters. The finite differences approach involves roughly one Maxwell solution per design variable. Thus, this figure shows that adjoints help optimize the system faster than the finite-differences approach.

For  $n$  design parameters, the advantage of adjoint-based optimization would be even more striking: direct computation of the finite-differences gradient would require at least  $n + 1$  field computations per gradient, whereas the adjoint method still only requires two.

## 7. Conclusion

We have shown how to use the adjoint method to efficiently design of optical structures. The key of this method is to compute the gradient of the target function with respect to the design parameters using only two non-trivial computations, independently of the number of design parameters. We illustrated this method in two cases: the optical waveguide slab, and a grating coupler simulated using a standard Maxwell solver. For the latter, the converged electric field  $\vec{E}$  and the corresponding adjoint field  $\vec{v}$  are the only non-trivial computations required to calculate the gradient, needed to perform gradient ascent (or gradient descent) on the design parameters. We believe this method will prove itself immensely valuable in the design of various grating structures, including a host of non-periodic diffractive structures.

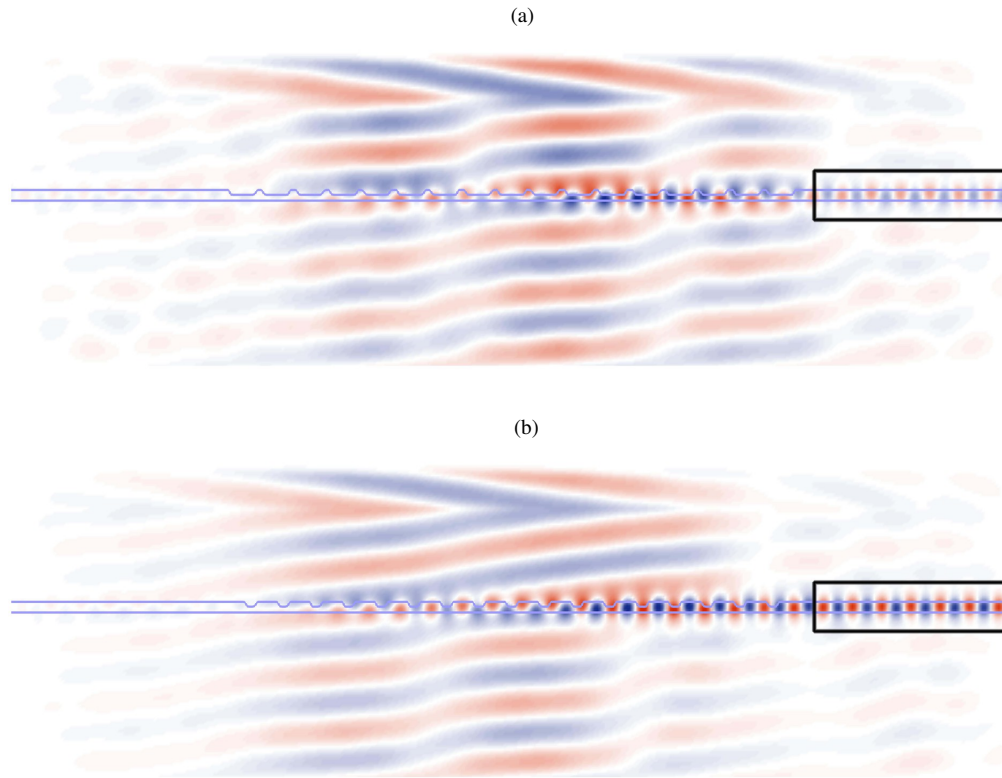


Fig. 4. Electric field in and around the waveguide (outlined in light blue) before and after adjoint-based optimization: (a) the grating with the initial design parameters couples the incident light (top) poorly to the waveguide (inside black box), with a complex amplitude of 0.39 (b) After adjoint-based optimization, the grating couples the light very well to the waveguide (inside black box), with a complex amplitude of 0.88.

### Acknowledgments

A.N. thanks Hideo Mabuchi, Erling Riis, and Thomas Baer for their mentorship during his time at Stanford University. Furthermore, he thanks Charles Santori, Marco Fiorentino, Sonny Vo, and Wayne Sorin for interesting discussions on how to design optical devices. This work was supported by the SU2P Program of Stanford University and Research Councils UK.