Contents

Chapter 1
Objective-First Nanophotonic Design

Abstract The abstract for the book.

In this chapter, we introduce an "objective-first" strategy for designing nanophotonic devices.

General description:

- Only uses desired fields.
- Can attempt to design any linear device.
- Allows for non-physical fields.

This strategy is unique in that it asks the user only for what electromagnetic field the device should produce, and then attempts to generate the design without further user intervention.

1.1 The electromagnetic wave equation

1.1.1 Physics formulation

First, let's derive our wave equation, starting with the differential form of Maxwell's equations,

$$\nabla \times E = -\mu_0 \frac{\partial H}{\partial t} \tag{1.1}$$

$$\nabla \times H = J + \varepsilon \frac{\partial E}{\partial t}, \tag{1.2}$$

where E, H, and J are the electric, magnetic and electric current vector fields, respectively, ε is the permittivity and μ_0 is the permeability, which we assume to be that of vacuum everywhere.

For time dependence $\exp(-i\omega t)$, where ω is the angular frequency, these become

$$\nabla \times E = -i\mu_0 \omega H \tag{1.3}$$
$$\nabla \times H = J + i\varepsilon \omega E, \tag{1.4}$$

which we can combine to form our wave equation,

$$\nabla \times \varepsilon^{-1} \nabla \times H - \mu_0 \omega^2 H = \nabla \times \varepsilon^{-1} J. \tag{1.5}$$

For further information, as well as simplifications to the wave equation in reduced dimensions, please see the Appendix.

1.1.2 Numerical formulation

Now, on top of the analytical formulation of the wave equation (1.5) we will now add a numerical, or discretized, formulation. This will be needed in order to solve for arbitrary structures.

First, we discretize our computational space according to the Yee grid, which allows us to easily define the curl $(\nabla\times)$ operators in (1.5) as described in the Appendix. This allows us, with a change of variables to formulate (1.5) as

$$A(p)x = b(p), \tag{1.6}$$

where $H \rightarrow x$, $\varepsilon^{-1} \rightarrow p$; and where

$$A(p) = \nabla \times \varepsilon^{-1} \nabla \times -\mu_0 \omega^2 \tag{1.7}$$

and

$$b(p) = \nabla \times \varepsilon^{-1} J. \tag{1.8}$$

Note that our use of $A(p)$ and $b(p)$ instead of A and b simply serves to clarify the dependence of both A and b to p.

Additionally, we use periodic boundary conditions with stretched-coordinate perfectly matched layers where necessary for our examples.

1.1.3 Solving for H

With our numerical formulation, we can now solve for the H-field (the E-field can be computed from the H-field using (1.4)) by using general linear algebra solvers. Doing so is also simply known as a time-harmonic or a finite-difference frequency-domain (FDFD) simulation.

Now, while a full three-dimensional problem is computationally quite taxing; in one- and two-dimensions, (1.6) is easily solved using the standard sparse solver

included in Matlab, and this technique is regularly employed in the examples which follow.

1.1.4 Solving for ε^{-1}

The next step, after having built a field-solver or simulator (finds x given p) for our wave equation, is to build a structure-solver for it. In other words, we need to be able to solve for p given x.

To do so, we return to (1.5) and remark that $\varepsilon^{-1}(\nabla \times H) = (\nabla \times H)\varepsilon^{-1}$ and $\varepsilon^{-1}J = J\varepsilon^{-1}$ since scalar multiplication is communicative. This allows us to rearrange (1.5) as

$$\nabla \times (\nabla \times H)\varepsilon^{-1} - \nabla \times J\varepsilon^{-1} = \mu_0 \omega^2 H \tag{1.9}$$

which we now write as

$$B(x)p = d(x), \tag{1.10}$$

where

$$B(x) = \nabla \times (\nabla \times H) - \nabla \times J \tag{1.11}$$

and

$$d(x) = \mu_0 \omega^2 H. \tag{1.12}$$

Solving this system would now seem to allow us to choose an electromagnetic field and then find the structure to produce it; which strongly suggests that it will be useful in the design of nanophotonic devices.

In terms of computational complexity, as with (1.6), (1.10) in its current form can be solved using standard tools.

1.1.5 Bi-linearity of the wave equation

Although additional mathematical machinery must still be added in order to get a useful design tool, we have really shown so far is that the wave equation is separately linear or *bi-linear* in x and p. Namely that,

$$A(p)x - b(p) = B(x)p - d(x). \tag{1.13}$$

In other words, fixing p makes solving the wave equation for x a linear problem, and vice versa. Note that the joint problem, where both x and p are allowed to vary, is not linear.

The bi-linearity of the wave equation is fundamental in our objective-first strategy which relies on the fact that we already know how to solve linear systems well, and is the reason why we chose $\varepsilon^{-1} \to p$ instead of the more natual $\varepsilon \to p$. Indeed,

this property forms a natural division of labor in the objective-first scheme, which we outline below.

1.2 The objective-first design problem

We now build off of the field-solver and the structure-solver, as previously outlined, by formulating the design problem and outlining the objective-first strategy.

1.2.1 Design objectives

A design objective, $f(x)$, is simply defined as a function we wish to be minimal for the design to be produced.

For instance, in the design of a device which must transmit efficiently into a particular mode, we could choose $f(x)$ to be the negative power transmitted into that mode. Or, if the device was to be a low-loss resonator, we could choose $f(x)$ to be the amount of power leaking out of the device.

In general, there are multiple choices of $f(x)$ which can be used to describe the same objective. For example, $f(x)$ for a transmissive device may not only be the negative power transmitted into the desired output mode, but it could also be the amount of power lost to other modes, or even the error in the field values at the output port relative to the field values needed for perfect transmission. These design objectives are equivalent in the sense that, if minimized, all would produce structures with good performance. At the same time, we must consider that the computational cost and complexity of using one $f(x)$ over another may indeed vary greatly.

1.2.2 Convexity

Before formulating the design problem, we would like to inject a note regarding the complexity of various optimization problems.

Specifically, we want to introduce the notion of *convexity* and to simply note the difference between problems that are convex and those which are not. The difference is simply this: convex problems have a single optimum point (only one local optimum, which is therefore the global optimum) which we can reliably find using existing numerical software, whereas non-convex problems typically have multiple optima and are thus much more difficult to reliably solve.

That a convex problem can be reliably solved, in this case, means that regardless of the starting guess, convex optimization software will always arrive at the globally optimal solution and will be able to numerically prove global optimality as

well. Thus, formulating a design problem in terms of convex optimization problems virtually eliminates any ideas of chance or randomness.

For the examples presented in this chapter, we use CVX, a convex optimization software written for Matlab.

1.2.3 Typical design formulation

The typical, and most straightforward formulation of the design problem is

$$\underset{x,p}{\text{minimize}} \quad f(x) \tag{1.14}$$

$$\text{subject to} \quad A(p)x - b(p) = 0, \tag{1.15}$$

which states that we would like to vary x and p simultaneously in order to decrease $f(x)$ while always satisfying physics (the electromagnetic wave equation). Such a formulation, if solved using a steepest-descent method, is the well-known adjoint optimization method.

1.2.4 Objective-first design formulation

In contrast, the objective-first formulation switches the roles of the wave equation and the design objective with one another,

$$\underset{x,p}{\text{minimize}} \quad \|A(p)x - b(p)\|^2 \tag{1.16}$$

$$\text{subject to} \quad f(x) = f_{\text{ideal}}. \tag{1.17}$$

This first means that, as seen from (1.16), we allow for non-zero residual in the electromagnetic wave equation. This literally means that we allow for *non-physical* x and p, since $A(p)x - b(p) \neq 0$ is now allowed. For this reason, we denote $A(p)x - b(p)$ the *physics residual*.

Secondly, we see from (1.17) that we always force the device to exhibit ideal performance, even if doing so means breaking the laws of physics. As such, our strategy will be to vary x and p in order to decrease the physics residual (1.16) to zero, while always maintaining ideal performance.

Our initial motivation for doing this was that it allowed us to solve for x and p separately, as will be outlined below, and that always forcing ideal performance might provide a mechanism to "override" local optima in the optimization process.

To this end we have found that such a strategy actually allows us to design very unintuitive devices which exhibit very good performance, even when starting from completely non-functional initial guesses. Furthermore, we have found this to be true even true when the physics residual is never brought to exactly zero.

From a numerical standpoint, although the objective-first formulation is still non-convex in its original form, the bi-linearity of the physics residual term allows us to naturally break the original problem into two sub-problems which we outline below.

Lastly, we add an additional constraint to the original formulation, which is to set hard-limits on the allowable values of p, namely $p_0 \leq p \leq p_1$. This is actually a relaxation of the ideal constraint, which would be to allow p to only have discrete values, $p \in p_0, p_1$, but such a constraint would be essentially force us to only be able to perform brute force trial-and-error.

Our objective-first formulation is thus,

$$\begin{aligned} \underset{x,p}{\text{minimize}} \quad & \|A(p)x - b(p)\|^2 \\ \text{subject to} \quad & f(x) = f_{\text{ideal}} \\ & p_0 \leq p \leq p_1. \end{aligned} \tag{1.18}$$

1.2.5 Field sub-problem

Since the objective-first problem in its original form is still non-convex, we break it down into two convex sub-problems. The first of these is the field sub-problem, which simply involves fixing p and independently optimizing x,

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|A(p)x - b(p)\|^2 \\ \text{subject to} \quad & f(x) = f_{\text{ideal}}. \end{aligned} \tag{1.19}$$

This problem is convex, and actually quadratic, which means that it can even be solved in the same way as a simple least-squares problem.

1.2.6 Structure sub-problem

The second sub-problem is formulated by fixing x and independently optimizing p. At the same time, we use the bi-linearity property of the physics residual from (1.13) to rewrite the problem in a way that makes its convexity explicit,

$$\begin{aligned} \underset{p}{\text{minimize}} \quad & \|B(x)p - d(x)\|^2 \\ \text{subject to} \quad & p_0 \leq p \leq p_1. \end{aligned} \tag{1.20}$$

The structure sub-problem is also convex, but not quadratic. However, use of the CVX package still allows us to obtain the result quickly and reliably.

1.2.7 Alternating directions algorithm

We use a simple alternating directions scheme to peice together (1.19) and (1.20), which is to say that we simply alternately solve each and continue until we reach some stopping point, normally measured by how much the physics residual has decreased.

The advantage of such the alternating directions method is that the physics residual is guaranteed to monotonically decrease with every iteration, which is useful in that no safeguards are needed to guard against “rogue” steps in the optimization procedure. Note that this robustness stems from the fact that, among other things, each sub-problem does not rely on previous values of the variable which is being optimized, but only on the variable which is held constant.

The disadvantage of such a simple scheme is that the convergence is quite slow, although we have found it to be sufficient in our cases. See ref for related methods that exhibit far better convergence.

1.3 Metamaterials design

1.3.1 Modification of the design objective

1.3.2 Cloak devices

1.3.3 Mimic devices

1.4 Extending the method

1.4.1 3D

1.4.2 Multi-mode

1.4.3 Robustness

1.4.4 Binary structure

1.5 Appendix

1.5.1 Full 3D curl

1.5.2 1D

1.5.3 2D

1.5.4 2.5D