Adjoint Design Method

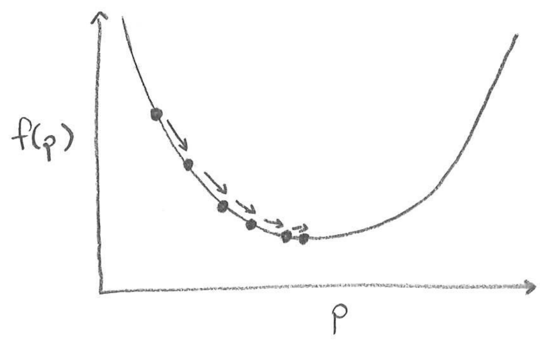
We typically cast design as a minimization problem:

$$\text{minimize } f(p)$$

$p = [p_1, p_2, \ldots, p_M]$ = design variables

$f(p)$ = objective function (power in desired mode, field at point, etc)

An efficient way to minimize $f(p)$ is to use gradient descent:



The hard part is finding the gradient $\frac{df}{dp}$.

We could use brute force:

$$\frac{\partial f}{\partial p_i} \approx \frac{f(p_1, \ldots, p_i + \Delta, \ldots, p_M) - f(p_1, \ldots, p_i, \ldots, p_M)}{\Delta}$$

This is very inefficient: we need to evaluate $f(p)$ once for every design variable, and every evaluation of $f(p)$ requires at least one EM simulation. This quickly becomes infeasible if we have $10^3 - 10^6$ design variables!

Thus, we need to calculate the gradient analytically.

Basic Idea

In an electromagnetic design problem, we have the following typical structure:

$$\underbrace{\underset{\substack{\text{design}\\\text{variables}}}{p} \longrightarrow \underset{\text{permittivity}}{\varepsilon} \longrightarrow \underset{\substack{\text{electric}\\\text{field}}}{E}}_{\substack{\text{multi-dimensional}\\\text{(possibly up to } 10^8 \text{ dims)}}} \longrightarrow \underbrace{\underset{\substack{\text{objective}\\\text{function}}}{f}}_{\text{scalar}}$$

We can write this explicitly as

$$f(p) = \underbrace{f}_{\substack{\text{Scalar}\\\text{Valued}}}\underbrace{(E(\mathcal{E}(p)))}_{\substack{\text{vector}\\\text{valued}}}$$

More generally, when optimizing physical systems, the objective f is a composition of functions of the form

$$f(p) = \underbrace{f_0}_{\substack{\text{Scalar}\\\text{Valued}}}\underbrace{(f_1(\cdots f_n(p)\cdots))}_{\substack{\text{Vector}\\\text{valued}}}$$

To find the gradient of f, use chain rule:

$$\frac{df}{dp} = \underbrace{\frac{df_0}{df_1}}_{\substack{\text{1D}\\\text{Vector}}}\underbrace{\frac{df_1}{df_2}\cdots\frac{df_n}{dp}}_{\substack{\text{2D}\\\text{matrices}}}$$

We can evaluate this from right $\rightarrow$ left or left $\rightarrow$ right:

right $\rightarrow$ left: propagate 2D matrix through chain

left $\rightarrow$ right: —— " —— 1D vector —— " —— : MUCH MORE EFFICIENT!!

We take transpose to make the desired left-to-right evaluation explicit:

$$\frac{df}{dp} = \left(\frac{df_n}{dp}^T \cdots \frac{df_1}{df_2}^T \frac{df_0}{df_1}^T\right)^T$$

$\longleftarrow$

evaluate in this direction

This is the basis of the adjoint design method: evaluating chain rule in the most sensible direction!

<u>Note</u>: this is essentially backpropagation for neural networks!

<u>Adjoint method for EM</u>

Assume all simulations are performed with FDFD. The problem we want to solve is

$$\begin{aligned} &\text{minimize} && f(e) \\ &\text{subject to} && A(p)e = b \end{aligned}$$

where

$p \in \mathbb{R}^P$ design variables $\qquad A(p) \in \mathbb{C}^{N\times N}$ FDFD matrix

$e \in \mathbb{C}^N$ electric field $\qquad b \in \mathbb{C}^N$ sources

The electric field e is implicitly a function of the design variables p through the FDFD equation $A(p)e = b$. Finding the gradient of f with the chain rule,

$$\underbrace{\frac{df}{dp}}_{1\times P} = \underbrace{\frac{df}{de}}_{1\times N} \underbrace{\frac{de}{dp}}_{N\times P}$$

We can usually find df/de analytically. The problematic term is de/dp, which is a large, dense matrix. We find this term using the FDFD equation:

$$Ae = b$$

$$\frac{\partial}{\partial p_i}(Ae) = \frac{\partial}{\partial p_i}(b)$$

$$\frac{\partial A}{\partial p_i} e + A \frac{\partial e}{\partial p_i} = 0$$

$$\therefore \underbrace{\frac{\partial e}{\partial p_i}}_{N\times 1} = - \underbrace{A^{-1}}_{N\times N} \underbrace{\frac{\partial A}{\partial p_i}}_{N\times N} \underbrace{e}_{N\times 1}$$

The full Jacobian is:

$$\frac{de}{dp} = \left[\frac{\partial e}{\partial p_1} \cdots \frac{\partial e}{\partial p_P}\right] = -A^{-1}\left[\frac{\partial A}{\partial p_1}e \cdots \frac{\partial A}{\partial p_P}e\right]$$

The gradient of the objective is then

$$\frac{df}{dp} = \frac{df}{de}\frac{de}{dp} = -\underbrace{\frac{df}{de}}_{1\times N} \underbrace{A^{-1}}_{N\times N} \underbrace{\left[\frac{\partial A}{\partial p_1}e \cdots \frac{\partial A}{\partial p_P}e\right]}_{N\times P}$$

We can evaluate:

right → left: P matrix solves (i.e. one simulation per design variable)

left → right: 1 matrix solve !!

Making left → right evaluation explicit with transpose:

$$\frac{df}{dp} = -\left(A^{-T}\frac{df}{de}^T\right)^T\left[\frac{\partial A}{\partial p_1}e \cdots \frac{\partial A}{\partial p_P}e\right]$$

We thus only need two FDFD simulations for arbitrarily many design variables !!

Forward problem: $Ae = b$ (our original system)

Adjoint problem: $A^T e_{adj} = \frac{df}{de}^T$ (using derivative of objective function as source)

Example: Optimizing power in waveguide mode

Suppose we want to optimize the power in a waveguide mode with modal fields $\mathcal{E}_m$ and $\mathcal{H}_m$:

$$f[E, H] = \left\| \int (E \times \mathcal{H}_m + \mathcal{E}_m \times H) \cdot \hat{z} \, dx \, dy \right\|^2$$

In terms of linear algebra, this is:

$$f(e) = \| m^+ e \|^2 = e^+ m m^+ e \qquad \text{where } m \in \mathbb{C}^N$$

$+$ = conjugate transpose

Using chain rule:

$$\frac{df}{dp} = \left(\frac{de^+}{dp}\right) m m^+ e + e^+ m m^+ \left(\frac{de}{dp}\right)$$

$$= 2 \operatorname{Re}\left\{ (e^+ m) \, m^+ \frac{de}{dp} \right\}$$

$$= -2 \operatorname{Re}\left\{ (e^+ m) \, m^+ A^{-1} \left[\frac{\partial A}{\partial p_1} e \cdots \frac{\partial A}{\partial p_P} e \right] \right\} \qquad \text{since } \frac{de}{dp} = -A^{-1} \left[\frac{\partial A}{\partial p_1} e \cdots \frac{\partial A}{\partial p_P} e \right]$$

$$= -2 \operatorname{Re}\left\{ (e^+ m) (A^{-+} m) \left[\frac{\partial A}{\partial p_1} e \cdots \frac{\partial A}{\partial p_P} e \right] \right\}$$

The two FDFD simulations are:

Forward problem: $Ae = b$ (same as always)

Adjoint problem: $A^+ e_{adj} = m$ (using desired output mode as source)