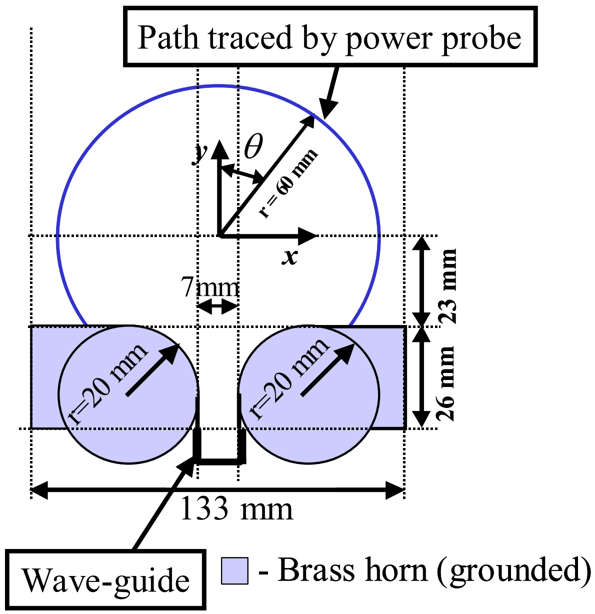
A practical introduction to adjoint methods

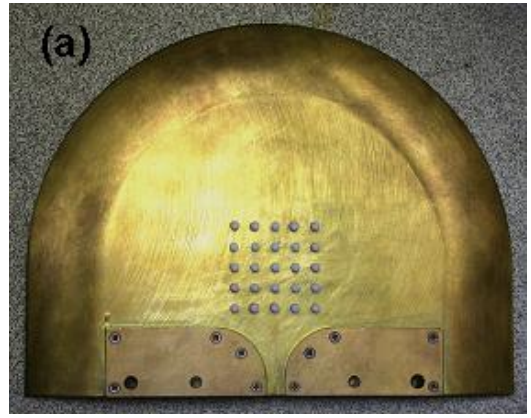
Leslie Greengard
CCM

NYU COURANT INSTITUTE OF MATHEMATICAL SCIENCES

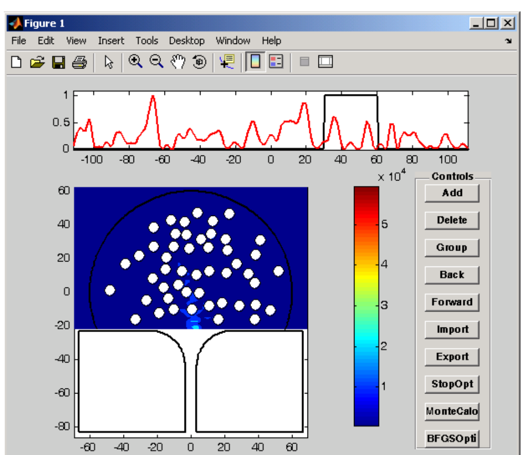
FLATIRON INSTITUTE
Center for Computational Mathematics

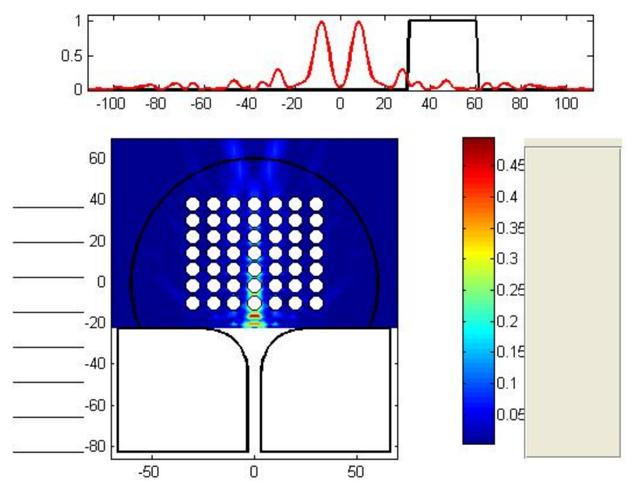
Model Problem: Electromagnetic Design



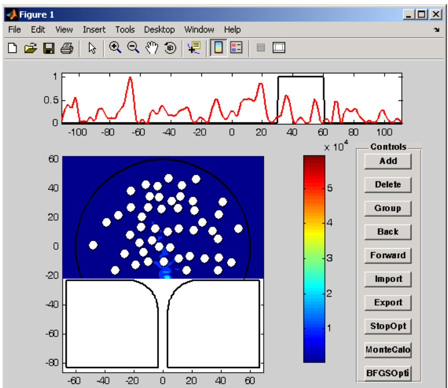


Gheorma, Haas, Levi (2004)
Zhao, Cheng, Gimbutas, G- (2007)

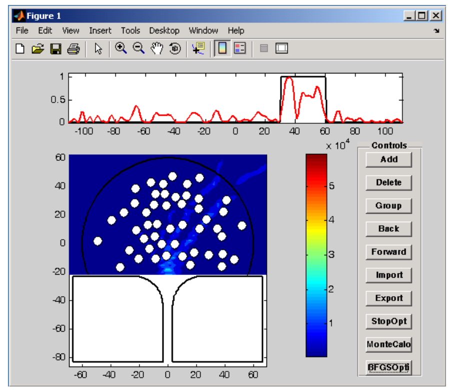




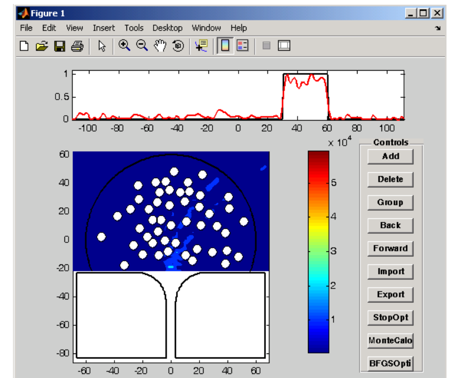
Model Problem: Electromagnetic Design



Snapshot of the GUI for the initial configuration (randomly place rods).



Snapshot of the GUI after 14 BFGS iterations.

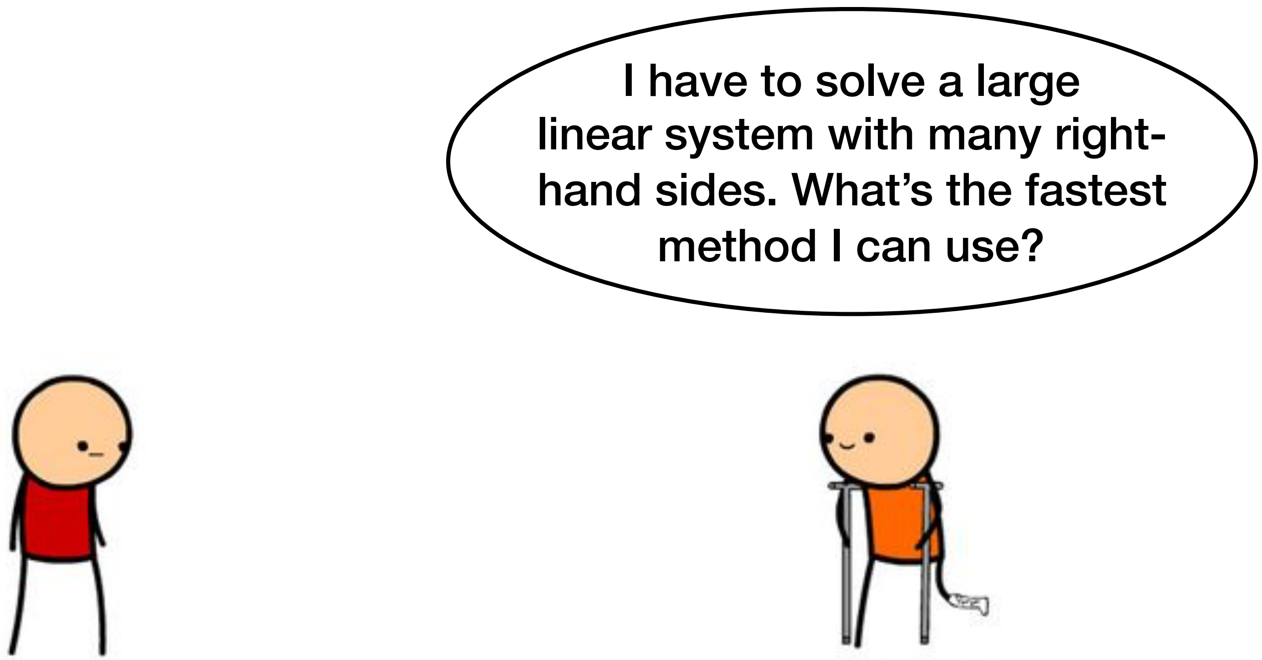


Snapshot of the GUI after 60 BFGS iterations.

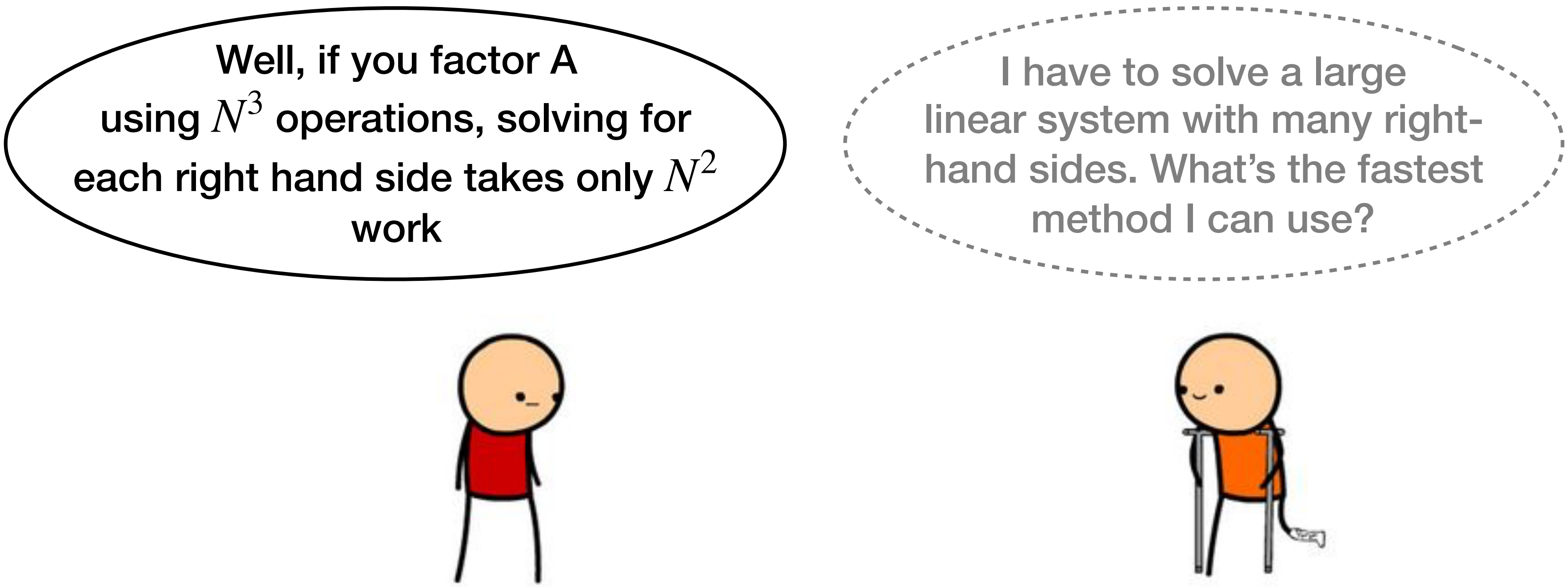
Adjoint methods allow for the calculation of the gradient of the objective function with respect to all cylinder center coordinates using only **4** solutions of the adjoint of the scattering problem. (Finite difference approximation of the gradient would require **2N** scattering calculations, where **N** is the number of rods.)

FLATIRON INSTITUTE
Center for Computational Mathematics

Scenario 1


I have to solve a large linear system with many right-hand sides. What's the fastest method I can use?

Scenario 1



$$A = LU \quad \Rightarrow \quad A^{-1} = U^{-1}L^{-1}$$

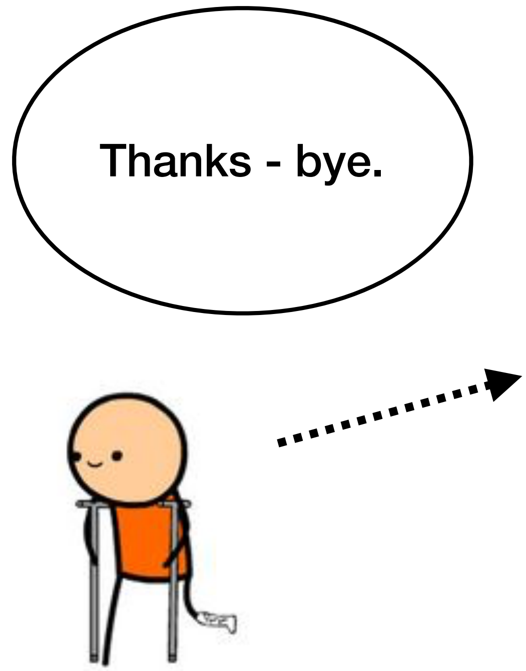
$$A \in R^{N \times N}$$

Algorithm 1

1) **Compute** $A = LU$ $\leftarrow O(N^3)$ work

2) For $i = 1, \ldots, M$

$$x_i = U^{-1}L^{-1}b_i \qquad \leftarrow O(MN^2) \text{ work}$$

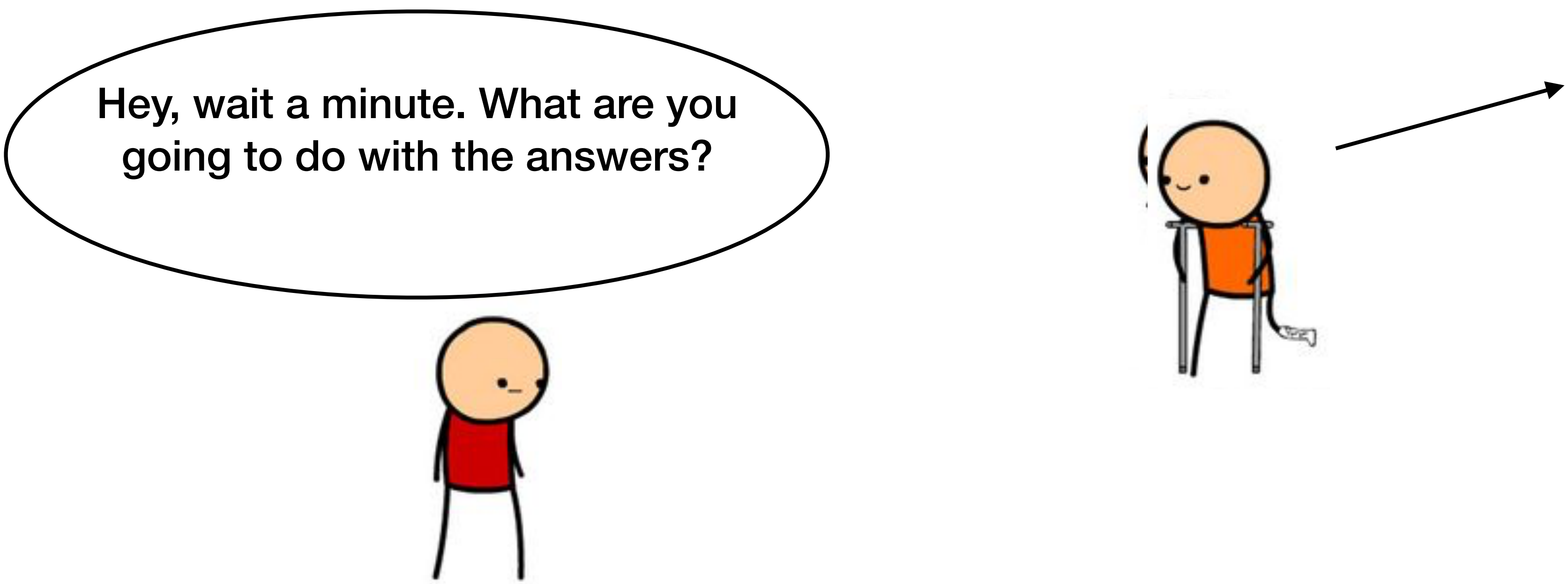


Algorithm 1

1) **Compute** $A = LU$

2) For $i = 1, \ldots, M$

$$x_i = U^{-1} L^{-1} b_i$$



Algorithm 1

1) **Compute** $A = LU$

2) For $i = 1, \dots, M$

$$x_i = U^{-1}L^{-1}b_i$$

Compute their average
$a_i = \langle c, x_i \rangle = c^T x_i$

Hey, wait a minute. What are you going to do with the answers?

$$c = \frac{1}{N}(1, 1, \dots, 1)$$

Algorithm 2 (Adjoint State Method)

Compute their average

$a_i = \langle c, x_i \rangle = c^T x_i$

But then I have a *much better idea*

$a_i = c^T x_i = c^T A^{-1} b_i = b_i^T A^{-T} c$

1) Compute $A = LU$ $\leftarrow O(N^3)$ work

3) $q \equiv A^{-T} c = L^{-T} U^{-T} c$ $\leftarrow O(N^2)$ work

4) For $i = 1, \ldots, M$

$a_i = b_i^T q$ $\leftarrow O(MN)$ work!

Gradient Evaluation

- Suppose $\mathbf{x}$ is the solution of a linear system (linear PDE, etc.) that involves $\mathbf{p} = (p_1, p_2, \ldots, p_M)$ design parameters or control variables
- We would like to compute the gradient of some objective function $g(\mathbf{p}, \mathbf{x})$ with respect to parameters $p_1, p_2, \ldots, p_M$:

$$\frac{dg}{dp_j} = \frac{\partial g}{\partial p_j} + \sum_{i=1}^{N} \frac{\partial g}{\partial x_i} \frac{\partial x_i}{\partial p_j}$$

$$\frac{dg}{d\mathbf{p}} = g_{\mathbf{p}} + g_{\mathbf{x}} \mathbf{x}_{\mathbf{p}}$$

$1 \times M$ $\quad$ $1 \times N$ $\quad$ $N \times M$

Gradient Evaluation

$$\frac{dg}{dp_j} = \frac{\partial g}{\partial p_j} + \sum_{i=1}^{N} \frac{\partial g}{\partial x_i}\frac{\partial x_i}{\partial p_j}$$

$$\frac{dg}{d\mathbf{p}} = g_{\mathbf{p}} + g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}}$$

- Reasonable to assume g is known analytically so that $\frac{\partial g}{\partial p_j}, \frac{\partial g}{\partial x_i}$ are easily computed.
- $\mathbf{x}$ on the other hand is the solution to $A\mathbf{x} = \mathbf{b}$ so that

$$A_{p_j}\mathbf{x} + A\mathbf{x}_{p_j} = b_{p_j} \quad \Rightarrow \quad \mathbf{x}_{p_j} = A^{-1}[b_{p_j} - A_{p_j}\mathbf{x}]$$

M applications of A^{-1}

Gradient Evaluation

$$\frac{dg}{dp_j} = \frac{\partial g}{\partial p_j} + \sum_{i=1}^{N} \frac{\partial g}{\partial x_i}\frac{\partial x_i}{\partial p_j} \qquad \frac{dg}{d\mathbf{p}} = g_{\mathbf{p}} + g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}}$$

The most expensive term (from previous slide) can be written:

$$g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}} = g_{\mathbf{x}}A^{-1}[b_{\mathbf{p}} - A_{\mathbf{p}}\mathbf{x}]$$

$1 \times N$ ($g_{\mathbf{x}}$), $N \times M$ ($\mathbf{x}_{\mathbf{p}}$), $N \times N$ (A^{-1}), $N \times M$ ($[b_{\mathbf{p}} - A_{\mathbf{p}}\mathbf{x}]$)

We have $g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}} = [g_{\mathbf{x}}A^{-1}][b_{\mathbf{p}} - A_{\mathbf{p}}\mathbf{x}] = \mathbf{q}[b_{\mathbf{p}} - A_{\mathbf{p}}\mathbf{x}]$

$1 \times N$ $\quad$ $N \times M$

where. $\mathbf{q}^T = A^{-T}g_{\mathbf{x}}^T$

Nonlinear Problems

$$\frac{dg}{dp_j} = \frac{\partial g}{\partial p_j} + \sum_{i=1}^{N} \frac{\partial g}{\partial x_i}\frac{\partial x_i}{\partial p_j} \qquad \frac{dg}{d\mathbf{p}} = g_{\mathbf{p}} + g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}}$$

(still)

Suppose

$$\mathbf{f}(\mathbf{x}, \mathbf{p}) = 0 \quad \Rightarrow \mathbf{f}_{\mathbf{x}}\mathbf{x}_{\mathbf{p}} + \mathbf{f}_{\mathbf{p}} = 0 \Rightarrow \mathbf{x}_{\mathbf{p}} = -\,\mathbf{f}_{\mathbf{x}}^{-1}\mathbf{f}_{\mathbf{p}} = -\,\mathbf{J}^{-1}\,\mathbf{f}_{\mathbf{p}}$$

Then. $g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}} = [g_{\mathbf{x}}\mathbf{J}^{-1}]\,\mathbf{f}_{\mathbf{p}} = \mathbf{q}\,\mathbf{f}_{\mathbf{p}}$

$N \times N$ **Jacobian**

where $\mathbf{q}^T = \mathbf{J}^{-T} g_{\mathbf{x}}^T$

(Recall this is just a linearization about current guess)

Initial Value Problems

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\,\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{b} \quad \Rightarrow \mathbf{x}(t) = e^{\mathbf{B}t}\,\mathbf{b}$$

Looks like $A\mathbf{x} = \mathbf{b}$ with $A = e^{-\mathbf{B}t}$. Suppose now that we want to optimize some function $g(\mathbf{p}, \mathbf{x}(t))$ based on the solution $\mathbf{x}$ at time t.

Then $\quad g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}} = g_{\mathbf{x}}A^{-1}[b_{\mathbf{p}} - A_{\mathbf{p}}\mathbf{x}] = \mathbf{q}[b_{\mathbf{p}} - A_{\mathbf{p}}\mathbf{x}]$

where $\mathbf{q}^T = A^{-T}g_{\mathbf{x}}^T = e^{B^T t}g_{\mathbf{x}}^T$

That is, $\dfrac{d\mathbf{q}^T}{dt} = \mathbf{B}^T\mathbf{q}^T, \mathbf{q}^T(0) = g_{\mathbf{x}}^T$

$$\frac{dg}{d\mathbf{p}} = g_{\mathbf{p}} + g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}} = g_{\mathbf{p}} + \mathbf{q}b_{\mathbf{p}} + \int_0^t \mathbf{q}(t - t')B_{\mathbf{p}}\mathbf{x}(t')dt'$$

Summary

- **Start with the big picture**
- **Always consider adjoint ideas when computing gradients of functionals (objective functions)**

References

- **Steven Johnson, Notes on Adjoint Methods for 18.335** **https://math.mit.edu/~stevenj/notes.html**
- Grégoire Allaire, A review of adjoint methods for sensitivity analysis, uncertainty quantification and optimization in numerical codes, Ingénieurs de l'Automobile, SIA, 2015, 836, pp.33-36. hal-01242950
- O.Pironneau, Optimal Shape Design for Elliptic Systems. Springer-Verlag, NewYork, 1984.
- Y. Cao, S. Li, L. Petzold, and R. Serban, "Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution," *SIAM J. Sci. Com- put.*, vol. 24, no. 3, pp. 1076–1089, 2003.
- G. Strang, *Computational Science and Engineering*. Wellesley, MA: Wellesley-Cambridge Press, 2007.