

Method of Adjoint for Unsteady Problems

Vishal Srivastava

1 A Review of the Method of Adjoint

Let there be a steady-state physical system, mathematically modeled by a system of equations denoted by $\mathcal{R}(\mathbf{u}, \boldsymbol{\beta}) = \mathbf{0}$, with $\mathbf{u} \in \mathbb{R}^n$, $\mathcal{R} \in \mathbb{R}^n$, $\boldsymbol{\beta} \in \mathbb{R}^m$, so that \mathbf{u} are the states of the system and $\boldsymbol{\beta}$ are the inputs to the system. Now, say one needs to solve for the required inputs $\boldsymbol{\beta}$, given desired values of observables $\mathbf{d}(\mathbf{u}, \boldsymbol{\beta})$ as \mathbf{d}_0 . This is an inverse problem and is usually solved via optimization techniques to find an optimum $\boldsymbol{\beta}$. The objective function \mathcal{J} for this optimization consists of a cost function \mathcal{C} and a regularization \mathcal{R} to keep $\boldsymbol{\beta}$ from assuming extravagant values (since the problem might very well be an ill-posed one). Thus, the problem can then be written as follows.

$$\boldsymbol{\beta}^* = \arg \min_{\boldsymbol{\beta}} \mathcal{J}(\mathbf{u}, \boldsymbol{\beta})$$

$$s.t. \quad \mathcal{R}(\mathbf{u}, \boldsymbol{\beta}) = \mathbf{0}$$

$$\text{where } \mathcal{J} = \mathcal{C} + \lambda \mathcal{R}$$

The cost function \mathcal{C} represents an aggregate discrepancy measure between $\mathbf{d}(\mathbf{u}, \boldsymbol{\beta})$ and \mathbf{d}_0 . For example,

$$\mathcal{C} = \|\mathbf{d}(\mathbf{u}, \boldsymbol{\beta}) - \mathbf{d}_0\|_2^2$$

The regularization represents how much does $\boldsymbol{\beta}$ wanders away from some baseline value. For example,

$$\mathcal{R} = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2^2$$

Now, if the dimension of $\boldsymbol{\beta}$, viz. m , is too large, a gradient-based optimization technique would be a feasible choice to solve the inverse problem. This then requires evaluation of sensitivities of the objective function \mathcal{J} w.r.t. $\boldsymbol{\beta}$. Since, the states \mathbf{u} are determined from the inputs, $\boldsymbol{\beta}$, provided to the physical system, the sensitivity can be expressed as,

$$\frac{d\mathcal{J}}{d\boldsymbol{\beta}} = \frac{\partial \mathcal{J}}{\partial \boldsymbol{\beta}} + \frac{\partial \mathcal{J}}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \boldsymbol{\beta}}$$

One can evaluate the term $\frac{\partial \mathbf{u}}{\partial \boldsymbol{\beta}}$ from the mathematical model as,

$$\frac{\partial \mathcal{R}}{\partial \boldsymbol{\beta}} + \frac{\partial \mathcal{R}}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \boldsymbol{\beta}} = 0 \quad \Rightarrow \quad \frac{\partial \mathbf{u}}{\partial \boldsymbol{\beta}} = - \left[\frac{\partial \mathcal{R}}{\partial \mathbf{u}} \right]^{-1} \frac{\partial \mathcal{R}}{\partial \boldsymbol{\beta}}$$

Note that one would need to solve m linear systems in this case. Finally, one can write the sensitivity vector as,

$$\frac{d\mathcal{J}}{d\boldsymbol{\beta}} = \frac{\partial \mathcal{J}}{\partial \boldsymbol{\beta}} - \frac{\partial \mathcal{J}}{\partial \mathbf{u}} \left[\frac{\partial \mathcal{R}}{\partial \mathbf{u}} \right]^{-1} \frac{\partial \mathcal{R}}{\partial \boldsymbol{\beta}}$$

Now, however, if we solve $\frac{\partial \mathcal{J}}{\partial \mathbf{u}} \left[\frac{\partial \mathcal{R}}{\partial \mathbf{u}} \right]^{-1}$ (one system) instead of $\left[\frac{\partial \mathcal{R}}{\partial \mathbf{u}} \right]^{-1} \frac{\partial \mathcal{R}}{\partial \boldsymbol{\beta}}$ (m systems), we can cut down the computational cost of sensitivity evaluation by a considerable amount. The adjoint variable $\boldsymbol{\psi}$, is the solution of the former such that,

$$\begin{aligned} \frac{d\mathcal{J}}{d\boldsymbol{\beta}} &= \frac{\partial \mathcal{J}}{\partial \boldsymbol{\beta}} + \boldsymbol{\psi}^T \frac{\partial \mathcal{R}}{\partial \boldsymbol{\beta}} \\ s.t. \quad \frac{\partial \mathcal{J}}{\partial \mathbf{u}} + \boldsymbol{\psi}^T \frac{\partial \mathcal{R}}{\partial \mathbf{u}} &= \mathbf{0} \end{aligned}$$

And now, if noticed closely, one can see that the adjoint variable is nothing but a Lagrange multiplier to solve a constrained optimization problem for the Lagrange function $\mathcal{L} = \mathcal{J} + \boldsymbol{\psi}^T \mathcal{R}$.

2 Method of Adjoints for Unsteady Problems

Let there be a physical system with discretized time-varying states $\mathbf{u}(t) \in \mathbb{R}^n$ and time-varying inputs $\beta(t) \in \mathbb{R}^m$. Let us also have a discretized time-stepping scheme for any time step k (with stencil spanning from $k - s_k$ to $k + p_k$),

$$\mathcal{S}^k(\mathbf{u}^{k+p_k}, \mathbf{u}^{k+p_k-1}, \dots, \mathbf{u}^{k-s_k+1}, \mathbf{u}^{k-s_k}, \beta^k) = \mathbf{0}$$

Now let the observable at time step k , and spatial location i , be given as d_i^k and their desired values as $d_{0,i}^k$. The motive is to match these values at all spatial locations and all time steps. Thus, an example for the cost function becomes,

$$\mathcal{C} = \sum_k \sum_i (d_i^k - d_{0,i}^k)^2$$

Now, building up the Lagrange function \mathcal{L} for the resulting \mathcal{J} under the constraints that the physical laws should be followed at *every* time step, we have,

$$\mathcal{L} = \mathcal{J} + \sum_{k=0}^K \psi_k^T \mathcal{S}^k(\mathbf{u}^{k+p_k}, \mathbf{u}^{k+p_k-1}, \dots, \mathbf{u}^{k-s_k+1}, \mathbf{u}^{k-s_k}, \beta^k)$$

Now, taking the Jacobians of the Lagrange function w.r.t. states at all time steps and setting them to zero we have for every state variable instance in time (K being the final time step),

$$\frac{\partial \mathcal{J}}{\partial \mathbf{u}^\ell} + \sum_{j=\max(0, k-p)}^{\min(K, k+s)} \psi_j^T \frac{\partial \mathcal{S}^j}{\partial \mathbf{u}^\ell} = 0 \quad \forall \quad \ell \in [p_0, K + p_K]$$

Please note here that the states $\mathbf{u}^{-s_0}, \mathbf{u}^{-s_0+1}, \dots, \mathbf{u}^{p_0-1}$ constitute the initial conditions for the unsteady problem and thus have no dependence on β_0 or any input at any time step for that matter. Solving this system of equations across all time steps one obtains adjoint variables at all time steps. Finally, the sensitivities w.r.t. inputs at time step k , β^k , can be obtained as,

$$\frac{d\mathcal{J}}{d\beta^k} = \frac{\partial \mathcal{J}}{\partial \beta^k} + \psi_k^T \frac{\partial \mathcal{S}^k}{\partial \beta^k} \quad \forall \quad k \in [0, K]$$

3 Example: Unsteady Heat Conduction (Forward Euler)

Let the problem be defined as follows. The system being considered here lies in the 2-D domain $[-1, 1] \times [-1, 1]$, where the states consist of a purely diffusive field $u(x, y, t)$. The system follows the Laplace Equation

$$\begin{aligned} \frac{\partial u}{\partial t} &= \alpha \frac{\partial^2 u}{\partial x^2} + \alpha \frac{\partial^2 u}{\partial y^2} \\ \text{s.t.} \quad \frac{\partial u}{\partial x} &= 0 \quad x \in \{-1, 1\}, y \in [-1, 1] \\ \frac{\partial u}{\partial y} &= 0 \quad x \in [-1, 1], y = 1 \\ \frac{\partial u}{\partial y} &= \beta(x, t) \alpha_b (u_b - u) \quad x \in [-1, 1], y = -1 \end{aligned}$$

Let the observable be the time history of the state at one given point (here taken to be $x=0, y=1$), the target for which is taken to be

$$u_0(0, 1, t) = 2.2(\exp(-t^2/10) - 1)$$

and the objective function is evaluated as (no regularization used in this case)

$$J = \|u(0, 1, t) - u_0(0, 1, t)\|_2^2$$

Discretizing the field and defining the Fourier number as

$$F = \alpha \frac{\Delta t}{\Delta x^2}$$

we have for the internal nodes using forward Euler,

$$\mathbf{u}_{i,j}^k = F \left(\mathbf{u}_{i-1,j}^{k-1} + \mathbf{u}_{i+1,j}^{k-1} + \mathbf{u}_{i,j-1}^{k-1} + \mathbf{u}_{i,j+1}^{k-1} + \left(\frac{1}{F} - 4 \right) \mathbf{u}_{i,j}^{k-1} \right)$$

The boundary conditions can be set trivially using a second order finite difference approximation for the first derivatives. The update in states for every time step can then be expressed as $\mathbf{u}^k = \mathcal{G}^k(\mathbf{u}^{k-1}, \boldsymbol{\beta}^k)$ and so we can say,

$$\mathcal{S}^k = \mathcal{G}^k(\mathbf{u}^{k-1}, \boldsymbol{\beta}^k) - \mathbf{u}^k$$

Now from the previous section we can write the following:

$$\psi_{k-1}^T = \frac{\partial J}{\partial \mathbf{u}^{k-1}} + \psi_k^T \frac{\partial \mathcal{G}^k}{\partial \mathbf{u}^{k-1}}$$

$$\frac{dJ}{d\boldsymbol{\beta}^k} = \frac{\partial J}{\partial \boldsymbol{\beta}^k} + \psi_k^T \frac{\partial \mathcal{G}^k}{\partial \boldsymbol{\beta}^k}$$

Having evaluated the sensitivity, the time-varying 1-D field $\boldsymbol{\beta}$ is solved for using a steepest gradient method, the results for which have been shown as follows.

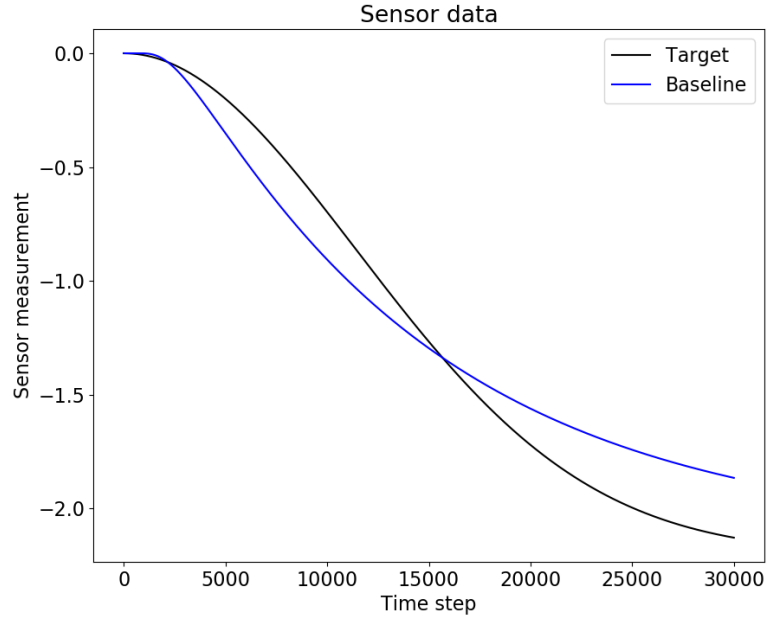


Figure 1: Difference in the baseline and target time histories at the sensor location

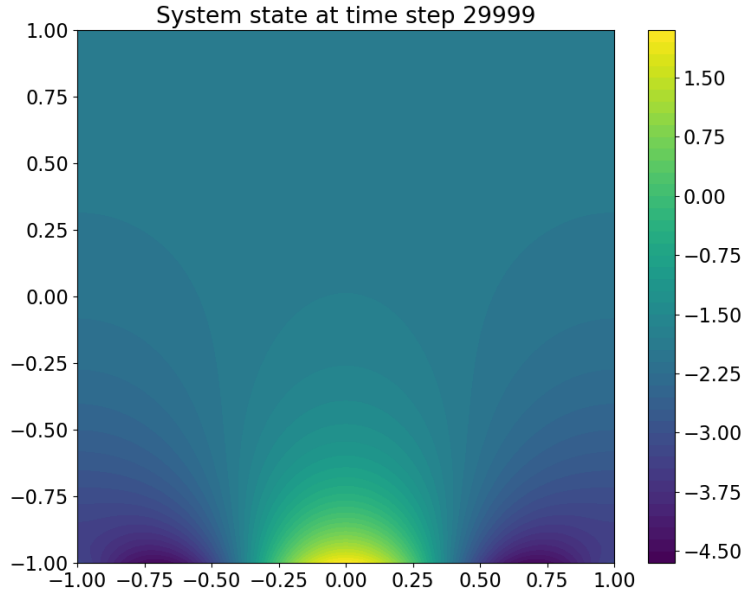


Figure 2: Baseline state field

The optimization reduces the objective function J by more than 2 orders of magnitude and the corresponding plots have been shown as follows

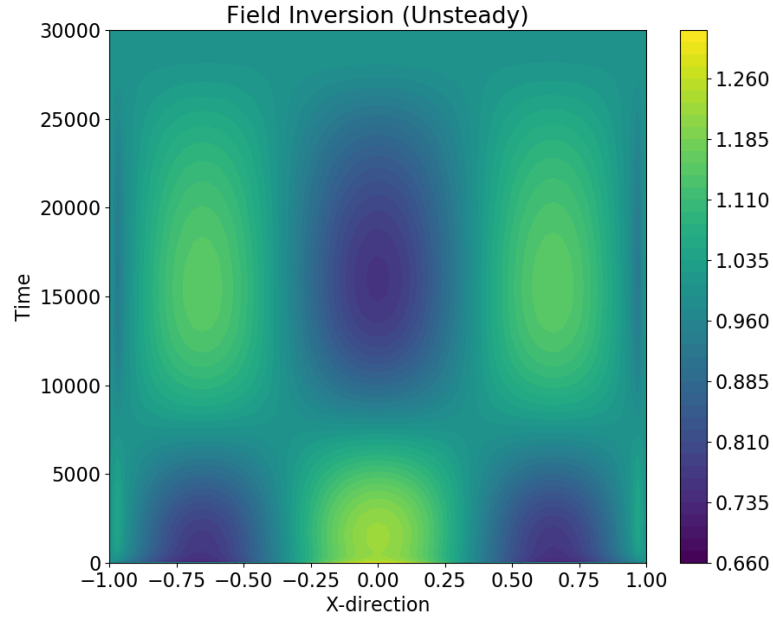


Figure 3: Optimal 1-D time varying β field

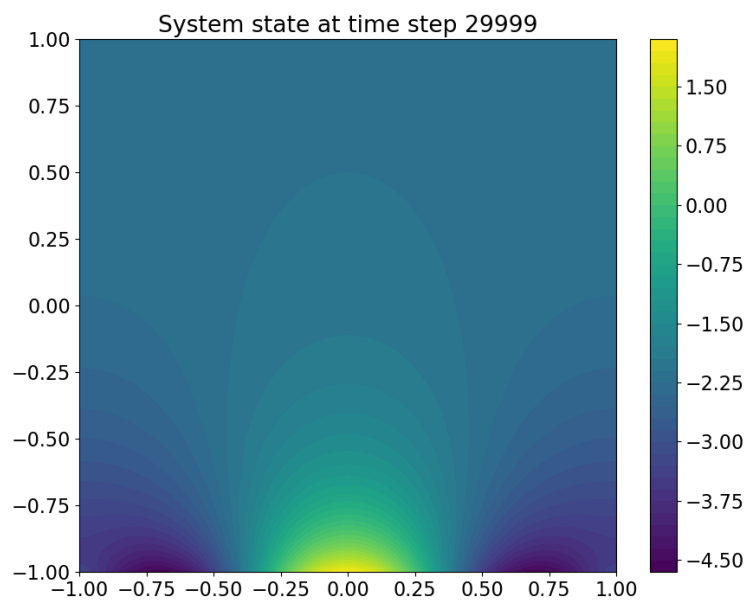


Figure 4: Corrected state field

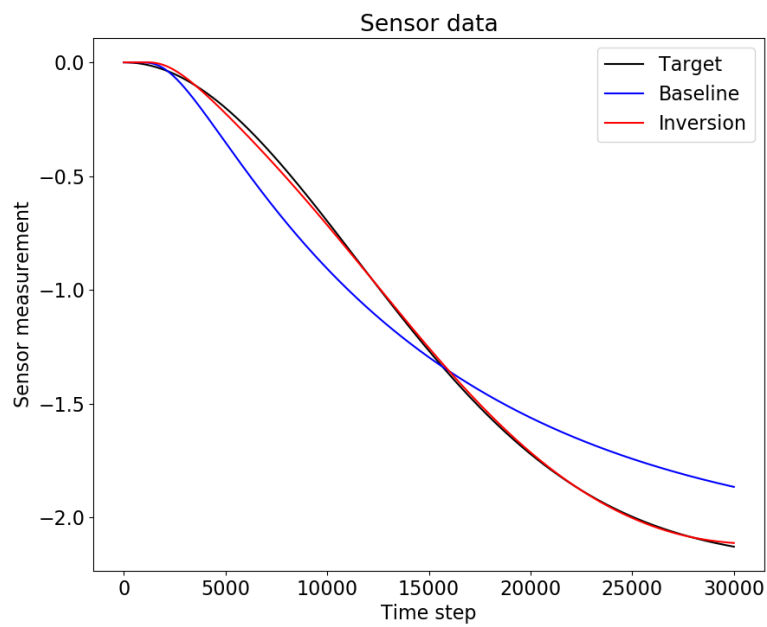


Figure 5: Correction in the time history recorded at the sensor location