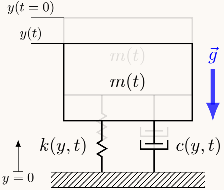
Adjoint intro and demo

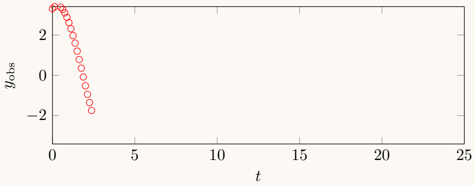
(Hear the word *adjoint*. Think *gradient*.)

Classical spring-mass-damper model

Given

Discrete observation data $(y_{obs}(t_i))$ and model configuration





Known

Conservation law (e.g. momentum $my'' = \sum_i F_i$):

$$m(t)\frac{d^2}{dt^2}y + c(t)\frac{d}{dt}y + k(t)y = 0$$

Find

$\vec{p} = [m(t_i), c(t_i), k(t_i)]^{\mathrm{T}}$ which minimizes $\mathcal{J} = \int [y(t) - y_{obs}(t)]^2 dt$

Variation of governing equations and cost

Noting the governing equation and cost function,

$$\mathcal{M}(y,\vec{p}) \equiv m(t)\frac{d^2}{dt^2}y + c(t)\frac{d}{dt}y + k(t) = 0$$

$$\mathcal{J} = \int_{t=0}^{T} \left[y - y_{obs.}\right]^2 dt,$$

take their variation (chain rule differention)

$$\delta\mathcal{M}(y,\vec{p}) = \frac{\partial\mathcal{M}}{\partial y}\delta y + \frac{\partial\mathcal{M}}{\partial\vec{p}}\delta\vec{p} = 0 \qquad (1)$$

$$\delta\mathcal{J} = \frac{\partial\mathcal{J}}{\partial y}\delta y + \frac{\partial\mathcal{J}}{\partial\vec{p}}\delta\vec{p} \qquad (2)$$

Multiply $\delta\mathcal{M}$ by Lagrange multiplier $y^\dagger$ and subtract from $\delta\mathcal{J}$

$$\delta\mathcal{J} = \left[\frac{\partial\mathcal{J}}{\partial y} - y^\dagger\frac{\partial\mathcal{M}}{\partial y}\right]\delta y + \left[\frac{\partial\mathcal{J}}{\partial\vec{p}} - y^\dagger\frac{\partial\mathcal{M}}{\partial\vec{p}}\right]\delta\vec{p}$$

Variation of governing equations and cost

Appears to have made things more complicated

$$\delta \mathcal{J} = \left[\frac{\partial \mathcal{J}}{\partial y} - y^\dagger \frac{\partial \mathcal{M}}{\partial y} \right] \delta y + \left[\frac{\partial \mathcal{J}}{\partial \vec{p}} - y^\dagger \frac{\partial \mathcal{M}}{\partial \vec{p}} \right] \delta \vec{p};$$

however, we can choose $y^\dagger$ to eliminate $[\ldots] \, \delta y$ by setting

$$\left[\frac{\partial \mathcal{J}}{\partial y} - y^\dagger \frac{\partial \mathcal{M}}{\partial y} \right] \delta y = 0$$

Leaving the gradient we seek

$$\frac{\delta \mathcal{J}}{\delta \vec{p}} = \left[\frac{\partial \mathcal{J}}{\partial \vec{p}} - y^\dagger \frac{\partial \mathcal{M}}{\partial \vec{p}} \right] \delta \vec{p};$$

Find the adjoint system via integration by parts of the original ode . . .

Adjoint derivation: Integration by parts

The mass term

$$\int_{t=0}^{T} y^{\dagger} m(t) \frac{d^2}{dt^2} \delta y \, dt = y^{\dagger} m(t) \frac{d\delta y}{dt}\bigg|_{t=0}^{T} - \int_{t=0}^{T} \frac{d\delta y}{dt} \frac{d}{dt} \left[y^{\dagger} m(t) \right] dt \tag{3}$$

$$\begin{aligned} \int_{t=0}^{T} y^{\dagger} m(t) \frac{d^2}{dt^2} \delta y \, dt = {} & y^{\dagger} m(t) \frac{d\delta y}{dt}\bigg|_{t=0}^{T} \\ & - \delta y \frac{d}{dt} \left[y^{\dagger} m(t) \right] \bigg|_{t=0}^{T} \\ & + \int_{t=0}^{T} \delta y \frac{d^2}{dt^2} \left[y^{\dagger} m(t) \right] dt \end{aligned} \tag{4}$$

The dissipation term

$$\int_{t=0}^{T} y^{\dagger} c(t) \frac{d}{dt} \delta y \, dt = y^{\dagger} c(t) \delta y \bigg|_{t=0}^{T} - \int_{t=0}^{T} \delta y \frac{d}{dt} \left[c(t) y^{\dagger} \right] dt \tag{5}$$

Adjoint derivation: Integration by parts

The spring term

$$\int_{t=0}^{T} y^\dagger k(t) \delta y \, dt. \tag{6}$$

Since we do not consider δy at $t = 0$ and we choose $y^\dagger = 0$ and $\frac{d}{dt} y^\dagger = 0$ at $t = T$, boundary terms are zero, leaving

$$\begin{aligned} \mathcal{L}^\dagger y^\dagger \equiv m(t) \frac{d^2}{dt^2} y^\dagger + \left\{ 2 \frac{dm(t)}{dt} - c(t) \right\} \frac{d}{dt} y^\dagger \\ + \left\{ \frac{d^2 m(t)}{dt^2} - \frac{dc(t)}{dt} + k(t) \right\} y^\dagger = 2[y(t) - y_{\mathrm{obs}}]. \end{aligned}$$

which is solved with same methods as the forward using final conditions $y^\dagger(T) = \frac{d}{dt} y^\dagger = 0$.

Summary of forward and adjoint systems

Forward system

$$\mathcal{M}(y,\vec{p}) \equiv m(t)\frac{d^2}{dt^2} + c(t)\frac{d}{dt} + k(t) = 0$$
$$y(0) = 3.3, v(0) = 1.0$$

Adjoint system

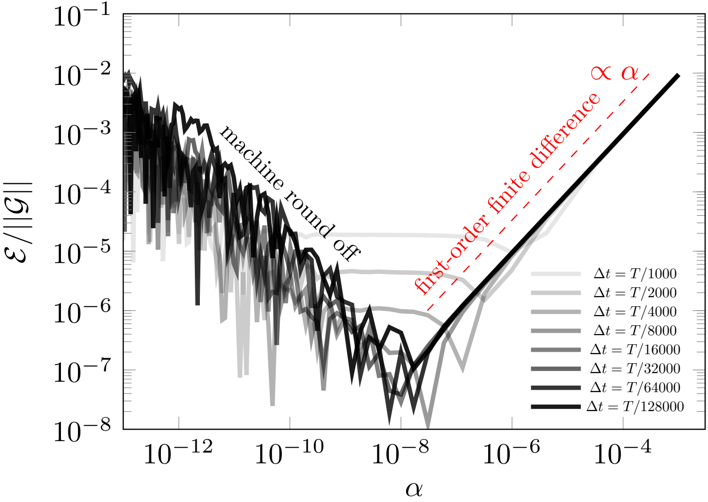
$$\mathcal{L}^\dagger y^\dagger = m(t)\frac{d^2}{dt^2}y^\dagger + \left\{2\frac{dm(t)}{dt} - c(t)\right\}\frac{d}{dt}y^\dagger + \left\{\frac{d^2 m(t)}{dt^2} - \frac{dc(t)}{dt} + k(t)\right\}y^\dagger$$
$$y^\dagger(t=T) = 0, v^\dagger(t=T) = 0$$

Iteratively reduce $\mathcal{J}$

$$m_i = m_{i-1} - \lambda\left(-y^\dagger \frac{d^2}{dt^2}y\right)$$
$$k_i = k_{i-1} - \lambda\left(-y^\dagger y\right)$$
$$c_i = c_{i-1} - \lambda\left(-y^\dagger \frac{d}{dt}y\right)$$

Adjoint derivation for 2$^{\text{nd}}$ order non-constant coefficient ODE in [Greenberg. *Applications of Green's Functions in Science and Engineering*. 1971. pg. 6-7.]

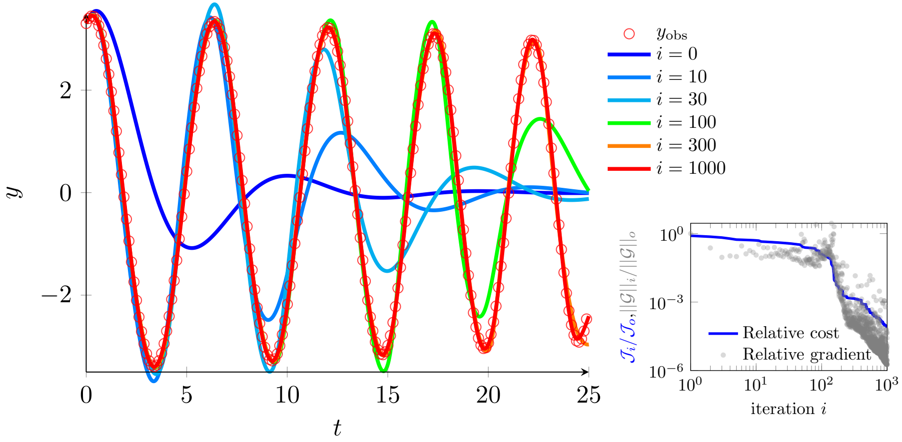
Is the derivation correct?



- Gradient converges with α in parameter space
- Not yet convinced? Let us optimize ...

Automatically reduce data mismatch

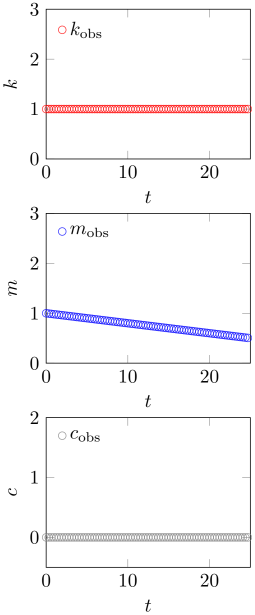
- Start $(i = 0)$ with guess $\vec{p}_o = [1, 1, 1]$; let computer do the work



- With a naive optimizer, relative error in observation falls below $\mathcal{J}_i/\mathcal{J}_o < 10^{-4}$

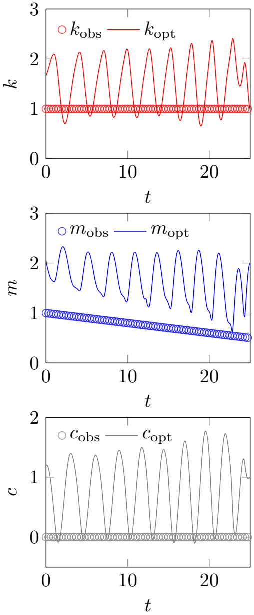
What do the optimal $[k(t), m(t), c(t)]$ look like?

- In reality, we do not know what created the observation data
- However, we can reveal what was used to actually generate it
- Results suggest a **different** local minimum attained!



What do the optimal $[k(t), m(t), c(t)]$ look like?

- In reality, we do not know what created the observation data
- However, we can reveal what was used to actually generate it
- Results suggest a **different** local minimum attained!



Summary of adjoint introduction

- Hear the word *adjoint*. Think *gradient*
- Cost is independent of $\mathrm{size}(\vec{p})$
- Gradients give a route to optimize, assimilate data, and inform parameter sensitivity
- Matlab code and slides for this portion available at: `https://github.com/buchta1/adjointExample-kmc.git` [$O(300)$ lines of code (includes plotting)]
- 3D, high-order finite-difference compressible Navier–Stokes solver with discrete-exact adjoint capability [$O(30k)$ lines of code]

This material is based in part upon work supported by the Department of Energy, National Nuclear Security Administration, under Award Number DE-NA0002374.

This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE–AC05–00OR22725.

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE–AC05–00OR22725.