Adjoint methods

Enrique Zuazua

AU - AvH
enrique.zuazua@fau.de

April 25, 2020

In this lecture we introduce the adjoint method technique that is of great important when characterizing optimal control and developing gradient descent algorithms.
To fix ideas and simplify the presentation we consider the linear finite-dimensional system

$$Ax = b, \quad \text{in } \mathbb{R}^N. \tag{1}$$

But the same methods can be extended to the nonlinear context, time evolving ODEs and PDEs, etc.
Given a target x^* we aim at computing the control b such that the solution x of the system gets as close as possible to x.
Of course in such a simple case it would suffice to take $b = Ax^*$.
But our goal here is to understand how the optimal control strategy could be implemented.

For, we define the functional

$$J(b) = \frac{1}{2}|x - x^*|^2 + \frac{1}{2}|b|^2 \tag{2}$$

that we minimize with respect to b in the search of a compromise between the goal that the state x gets as close as possible to the target x^* but without using a control b that is too large.
By the Direct Method of the Calculus of Variations it is easy to see that there is an unique minimizer $\bar{b}$ that is characterized by the Euler-Lagrange equations

$$J'(\bar{b}) = (\bar{x} - x^*, y) + (\bar{b}, c) = 0, \quad \forall c \in R^N,$$

where $\bar{x}$ is the solution of the system associated to the minimizer $\bar{b}$ and y is the solution of

$$Ay = c.$$

Note however that this Optimality System (OS) is constituted by N equations and that it requires solving the N systems corresponding to a basis of all the possible right hand side terms c.
This is in practice expensive when the dimension N is large.
The same occurs when, at each step of a gradient descent iteration, we need to compute the gradient of J.
The adjoint method provide a way of significantly reduced the computational cost.
Consider the adjoint system

$$A^* p = \bar{x} - x^*,$$

where A^* is the adjoint matrix of A.

The key observation is as simple as useful.

Once the adjoint system is solved and p is computed, then, the fist term in J', namely,

$$(\bar{x} - x^*, y)$$

can simply be rewritten as follows

$$(\bar{x} - x^*, y) = (A^* p, y) = (p, Ay) = (p, c).$$

Then the gradient of J can simply be rewritten as follows

$$J'(\bar{b}) = (p + \bar{b}, c), \quad \forall c \in R^N,$$

This new expression does not require solving the system $Ay = c$ for each c but just once the adjoint equation.

And this reduced dramatically the computational cost.