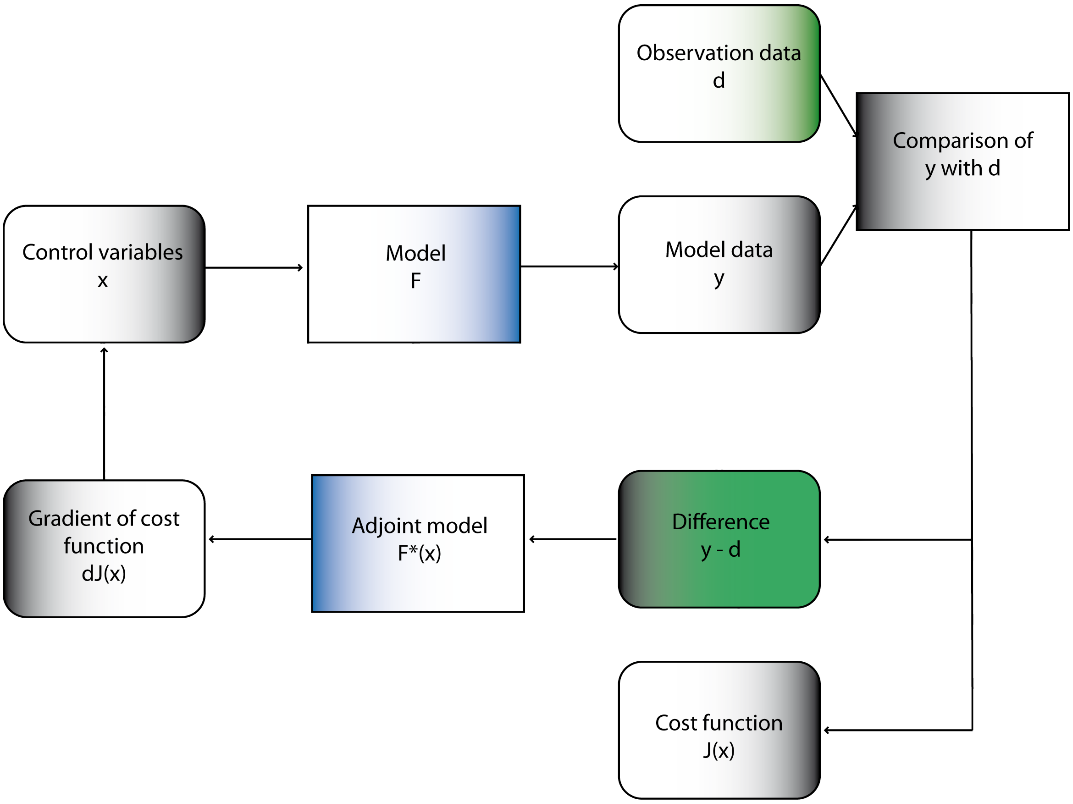
ADJOINT METHOD FOR INVERSE PROBLEMS

Adjoint Methods (I)

- A very general approach for solving inverse problems... including Machine Learning!
- Variational DA is based on an adjoint approach.



Adjoint Methods (II) - definition

Definition 1. *An* **adjoint method** *is a general mathematical technique, based on variational calculus, that enables the computation of the gradient of an objective, or cost functional with respect to the model parameters in a very efficient manner.*

Adjoint Methods (III) - continuous formulation

- Let $\mathbf{u}(\mathbf{x},t)$ be the state of a *dynamical system* whose behavior depends on model parameters $\mathbf{m}(\mathbf{x},t)$ and is described by a differential operator equation
$$\mathbf{L}(\mathbf{u},\mathbf{m}) = \mathbf{f},$$
where $\mathbf{f}(\mathbf{x},t)$ represents external forces.

- Define a *cost function* $J(\mathbf{m})$ as an energy functional[1] or, more commonly, as a misfit functional that quantifies the error (L^2-distance[2]) between the observation and the model prediction $\mathbf{u}(\mathbf{x},t;\mathbf{m})$.

[1] A functional is a generalization of a function. The functional depends on functions, whereas a function depends on variables. We then say that a functional is mapping from a space of functions into the real numbers.

[2] The L^2-space is a Hilbert space of (measurable) functions that are square-integrable (in Lebesgue sense).

For example,

$$J(\mathbf{m}) = \int_0^T \int_\Omega \left(\mathbf{u}(\mathbf{x}, t; \mathbf{m}) - \mathbf{u}^{\mathrm{obs}}(\mathbf{x}, t)\right)^2 \,\mathrm{d}\mathbf{x}\mathrm{d}t,$$

where $\mathbf{x} \in \Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$, and $0 \leq t \leq T$.

- Our objective is to choose the model parameters $\mathbf{m}$ as a function of the observed output $\mathbf{u}^{\mathrm{obs}}$, such that the cost function $J(\mathbf{m})$ is minimized.

- The minimization is most frequently performed by a gradient-based method, the simplest of which is steepest gradient, though usually some variant of a quasi-Newton approach is used—see [Asch2022, Nocedal2016].

- If we can obtain an expression for the gradient, then the minimization will be considerably facilitated.

- This is the objective of the adjoint method that provides an **explicit formula for the gradient of $J(\mathbf{m})$.**

Adjoint Methods (IV) - optimization formulation

- Suppose we are given a (P)DE,

$$F(\mathbf{u}; \mathbf{m}) = 0, \tag{1}$$

  where

  $\Rightarrow$ $\mathbf{u}$ is the state vector,
  $\Rightarrow$ $\mathbf{m}$ is the parameter vector, and
  $\Rightarrow$ F includes the partial differential operator $\mathbf{L}$, the right-hand side (source) $\mathbf{f}$, boundary and initial conditions.

- Note that the components of $\mathbf{m}$ can appear as any combination of

  $\Rightarrow$ coefficients in the equation,
  $\Rightarrow$ the source,

- $\Rightarrow$ or as components of the boundary/initial conditions.

- To solve this very general parameter estimation problem, we are given a cost function $J(\mathbf{m};\mathbf{u})$.
  - $\Rightarrow$ The constrained optimization problem is then

$$\begin{cases} \text{minimize}_{\mathbf{m}} & J\left(\mathbf{u}(\mathbf{m}),\mathbf{m}\right) \\ \text{subject to} & F(\mathbf{u};\mathbf{m}) = 0, \end{cases}$$

    where J can depend on both $\mathbf{u}$ and on $\mathbf{m}$ explicitly in the presence of eventual regularization terms.
  - $\Rightarrow$ Note:
    - $\rightarrow$ the constraint is a (partial) differential equation and
    - $\rightarrow$ the minimization is with respect to a (vector) function.
  - $\Rightarrow$ This type of optimization is the subject of variational calculus, a generalization of classical calculus where differentiation is performed with respect to a variable, not a function.

- The gradient of J with respect to $\mathbf{m}$ (also known as the *sensitivity*) is then obtained by the chain-rule,

$$\nabla_{\mathbf{m}} J = \frac{\partial J}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{m}} + \frac{\partial J}{\partial \mathbf{m}}.$$

  ⇒ The partial derivatives of J with respect to $\mathbf{u}$ and $\mathbf{m}$ are readily computed from the expression for J,

  ⇒ but the derivative of $\mathbf{u}$ with respect to $\mathbf{m}$ requires a potentially very large number of evaluations, corresponding to the product of the dimensions of $\mathbf{u}$ and $\mathbf{m}$ that can both be very large.

- The adjoint method is a way to avoid calculating all of these derivatives.

- We use the fact that if $F(\mathbf{u}; \mathbf{m}) = 0$ everywhere, then this implies that the total derivative of F with respect to $\mathbf{m}$ is equal to zero everywhere too.

- Differentiating the PDE (1), we can thus write

$$\frac{\partial F}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{m}} + \nabla_{\mathbf{m}} F = 0.$$

- This can be solved for the untractable derivative of $\mathbf{u}$ with respect to $\mathbf{m}$, to give

$$\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = -\left(\frac{\partial F}{\partial \mathbf{u}}\right)^{-1} \nabla_{\mathbf{m}} F$$

assuming that the inverse of $\partial F/\partial \mathbf{u}$ exists.

- Substituting in the expression for the gradient of J, we obtain

$$\begin{aligned}\nabla_{\mathbf{m}} J &= -\frac{\partial J}{\partial \mathbf{u}}\left(\frac{\partial F}{\partial \mathbf{u}}\right)^{-1} \nabla_{\mathbf{m}} F + \frac{\partial J}{\partial \mathbf{m}}, \\ &= \mathbf{p}\nabla_{\mathbf{m}} F + \frac{\partial J}{\partial \mathbf{m}}\end{aligned}$$

where

$\Rightarrow$ we have conveniently defined $\mathbf{p}$ as the solution of the *adjoint equation*

$$\left(\frac{\partial F}{\partial \mathbf{u}}\right)^{\mathrm{T}} \mathbf{p} = -\frac{\partial J}{\partial \mathbf{u}}.$$

Adjoint Methods (V) - summing up

In summary, we have a three-step procedure that combines a model-based approach (through the PDE) with a data-driven approach (through the cost function):

1. Solve the *adjoint equation* for the adjoint state, $\mathbf{p}$.

2. Using the adjoint state, compute the *gradient* of the cost function J.

3. Using the gradient, solve the *optimization* problem to estimate the parameters $\mathbf{m}$ that *minimize the mismatch* between model and observations.

This key result enables us to compute the desired gradient, $\nabla_{\mathbf{m}} J$, without the explicit knowledge of the variation of $\mathbf{u}$.

- A number of important remarks can be made.

  1. We obtain *explicit* formulas—in terms of the adjoint state—for the gradient with respect to each/any model parameter. Note that this has been done in a completely general setting, without any restrictions on the operator F or on the model parameters $\mathbf{m}.$
  2. The *computational cost* is one solution of the adjoint equation which is usually of the same order as (if not identical to) the direct equation, but with a reversal of time. Note that for nonlinear equations this may not be the case and one may require four or five times the computational effort.
  3. The *variation* (Gâteaux derivative) of F with respect to the model parameters $\mathbf{m}$ is, in general, straightforward to compute.
  4. We have not explicitly considered boundary (or initial) conditions in the above, general approach. In real cases, these are potential sources of difficulties for the use of the adjoint approach—the *discrete adjoint* can provide a way

to overcome this hurdle.

5. For complete mathematical rigor, the above development should be performed in an appropriate *Hilbert space* setting that guarantees the existence of all the inner products and adjoint operators. The interested reader could consul [Troltzsch2010].
6. In many real problems, the optimization of the misfit functional leads to *multiple local minima* and often to very "flat" cost functions. These are hard problems for gradient-based optimization methods. These difficulties can be (partially) overcome by a panoply of tools:
   (a) *Regularization* terms can alleviate the non-uniqueness problem.
   (b) *Rescaling* the parameters and/or variables in the equations can help with the "flatness." This technique is often employed in numerical optimization.
   (c) *Hybrid* algorithms, that combine stochastic and deterministic optimization (e.g., Simulated Annealing), can be used to avoid local

minima.

(d) Judicious use of machine learning techniques and methods.

7. When measurement and modeling errors can be modeled by Gaussian distributions and a background (prior) solution exists, the objective function may be generalized by including suitable *covariance matrices*. This is the approach employed systematically in data assimilation—see below.

Adjoint Methods (VI) - Examples

- We will now present a series of examples where we apply the adjoint approach to increasingly complex cases of inverse problems, for which we have

  $\Rightarrow$ a cost/mismatch function to minimze
  $\Rightarrow$ subject to the constraint of a (P)DE.

- There are two alternative methods for the derivation of the adjoint equation:

  $\Rightarrow$ the Lagrange multiplier approach
  $\Rightarrow$ and the tangent linear model (TLM) approach.

- After seeing the two in action, the reader can adopt the one that suits her/him best.

- Note that the Lagrangian approach supposes that we perturb the sought for parameters and is thus not applicable to inverting for constant-valued parameters, in which case we must resort to the tangent linear model approach.