Some exercices on the adjoint method

Exercise 1: a simple example of the adjoint method
We consider the ordinary differential equation

$$\begin{cases} -bu"(x) + cu'(x) = f(x) \qquad x \in]0,1[\\ u(0) = u(1) = 0 \end{cases} \tag{1}$$

We want to identify the two parameters b and c using an observation $u_{\text{obs}}(x)$.
Give a mathematical formulation for this inverse problem, and a method for solving it using an adjoint approach.

Exercise 2: control of the initial condition of an ODE with an adjoint method
Let the ordinary differential equation

$$\begin{cases} u'(t) = f(u(t)) \qquad t \in]0,T[\\ u(0) = a \end{cases} \tag{2}$$

We want to identify the initial condition a using an observation d of $u(T)$.
Give a mathematical formulation for this inverse problem, and a method for solving it using an adjoint approach.

Exercise 3: control of the Burgers' equation
Let the Burgers' equation

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\,\dfrac{\partial u}{\partial x} - \nu\,\dfrac{\partial^2 u}{\partial x^2} = f \qquad x \in]0,L[, t \in [0,T] \\ u(0,t) = \psi_1(t) \qquad t \in [0,T] \\ u(L,t) = \psi_2(t) \qquad t \in [0,T] \\ u(x,0) = u_0(x) \qquad x \in [0,L] \end{cases} \tag{3}$$

This equation can be interpreted as a 1-D version of the Navier-Stokes equations. Note the non-linear advection term.

1. We want to identify simultaneously the initial condition u_0 and the boundary conditions ψ_1, ψ_2. We have observations $u_{\text{obs}}(x,t)$ in $L^2(]0,L[\times]0,T[)$, and we want to minimize

$$J(u_0,\psi_1,\psi_2) = \frac{1}{2}\int_0^T \int_0^L (u - u_{\text{obs}})^2\, dxdt$$

Write the optimality system leading to the computation of $u_0^*, \psi_1^*, \psi_2^*$.

2. We now assume that the diffusion coefficient ν is unknown. How is the optimality system modified ?

Exercise 4: control of a diffusion coefficient
We consider the diffusion equation

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial}{\partial x}\left(K(x)\dfrac{\partial u}{\partial x}\right) \qquad x \in]0,1[, t \in [0,T] \\ u(0,t) = u(1,t) = 0 \quad \forall t \in [0,T] \\ u(x,0) = u_0(x) \text{ given on } [0,1] \end{cases} \tag{4}$$

The coefficient K is a function of $\mathcal{C}^1(]0,1[)$, and is time independent.
Let $u_{\text{obs}}(x,t)$ an observation in $L^2(]0,1[\times]0,T[)$. K is supposed to be unknown, and we want to estimate it given the observation u_{obs}. Let:

$$J(K) = \frac{1}{2}\int_0^T \int_0^1 (u_K - u_{\text{obs}})^2 \, dxdt$$

where u_K is the solution (assumed to be unique) of Eq.(4). The optimal value K^* is determined by $J(K^*) = \min J(K)$.

1. Determine the optimality system leading to the estimation of K^*.

2. Let $\Lambda = (\lambda_1, \dots, \lambda_n) \in \mathbf{R}^n$. We seek now for $K(x)$ under the form: $K(x) = \sum_{i=1}^{n} \lambda_i \, e_i(x)$ where the e_i are given functions in $\mathcal{C}^1(]0,1[)$. J becomes a function of Λ. Give the expression of the gradient of J w.r.t. Λ.

Exercise 5: control with constraint
We consider the problem of the control of the initial condition, given a constraint. We want this initial condition to satisfy an additional condition, corresponding for instance to some physical equilibrium.

1. Let the Burgers' equation

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial u^2}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f \qquad x \in]0, L[, t \in [0, T]$$

with the initial condition $u(x,0) = u_0(x)$ and the boundary conditions $u(0,t) = u(L,t) = 0$. We want to control u_0 by minimizing the cost function

$$J(u_0) = \frac{1}{2}\int_0^T \int_0^L (u - u_{\text{obs}})^2$$

with the constraint of a null averaged value : $\int_0^L u_0(x)\, dx = 0$.
How can this problem be solved using a Lagrange multiplier ?

2. The previous method is generalized to the following model (discretized in space):

$$\begin{cases} \dfrac{dX}{dt} = M(X) \\ X(t=0) = U \end{cases}$$

where the state variable X, of dimension n, represents the values of the variables at the model gridpoints. Thus X is a function from $[0,T]$ to $\mathbf{R}^n$, which is assumed to be $\mathcal{C}^1$. The observation is a given function Y_{obs}, from $[0,T]$ to $\mathbf{R}^m$ (X is not fully observed). The observation operator $H \in \mathcal{L}(\mathbf{R}^n, \mathbf{R}^m)$ (i.e. linear application from $\mathbf{R}^n$ to $\mathbf{R}^m$) is assumed to be time independent. We define the cost function:

$$J(U) = \frac{1}{2}\int_0^T \|HX - Y_{\text{obs}}\|^2$$

and we want to minimize $J(U)$ with the linear constraint $GU = 0$, where $G \in \mathcal{L}(\mathbf{R}^p, \mathbf{R}^n)$.

Find the optimality condition minimizing J and satisfying the constraint. (indication: use a Lagrange multiplier $\lambda \in \mathbf{R}^p$).

3. Another method, called penalty method, consists in defining the function

$$J_\varepsilon(U) = J(U) + \frac{1}{\varepsilon}\|GU\|^2$$

and minimizing $J_\varepsilon(U)$ without contrainst.

a. Write the optimality conditions.

b. What happens when ε tends to 0 ?

c. How could this penalty method be applied to question 1 ?

Exercise 6: equivalence between Kalman filter and 4D-Var

We consider a dynamical system $\mathbf{x}$, in finite dimension, which is modelled by

$$\mathbf{x}^t(t_{k+1}) = \mathbf{M}(t_k, t_{k+1})\mathbf{x}^t(t_k) + \mathbf{e}(t_k)$$

where $\mathbf{x}^t$ is the state vector (exponent ^{t} stands for *true*), and where the t_k are successive instants. **The model M is linear**. $\mathbf{e}(t_k)$ is the forecast error (or *model error*) at time t_k. It is assumed to be unbiased, with a known covariance matrix $\mathbf{Q}_k$, and uncorrelated in time ($\mathbf{e}(t_k)$ and $\mathbf{e}(t_l)$ are independent for $k \neq l$).
Observations $\mathbf{y}_k$ are available at time t_k, with an error ε_k, unbiased, and which covariance matrix is $\mathbf{R}_k$. These observations are linked to $\mathbf{x}(t_k)$ by the linear observation operator $\mathbf{H}_k$: $\mathbf{y}_k = \mathbf{H}_k\mathbf{x}^t(t_k) + \varepsilon_k$ with $E(\varepsilon_k\varepsilon_k^T) = \mathbf{R}_k$

With the usual hypotheses related to the independence of the errors, we can define the **Kalman filter** by the algorithm:

Initialization :

$$\begin{aligned}
\mathbf{x}^a(t_0) &= \mathbf{x}_b \\
\mathbf{P}^a(t_0) &= \mathbf{P}_b
\end{aligned}$$

where $\mathbf{x}_b$ is an approximation of the initial state, with $\mathbf{P}_b$ the corresponding error covariance matrix.

Step k (forecast - correction) :

$$\begin{aligned}
\mathbf{x}^f(t_{k+1}) &= \mathbf{M}(t_k, t_{k+1})\mathbf{x}^a(t_k) && \text{(KF1)} \\
\mathbf{P}^f(t_{k+1}) &= \mathbf{M}(t_k, t_{k+1})\mathbf{P}^a(t_k)\mathbf{M}^T(t_k, t_{k+1}) + \mathbf{Q}_k && \text{(KF2)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{x}^a(t_{k+1}) &= \mathbf{x}^f(t_{k+1}) + \mathbf{K}_{k+1}\left[\mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{x}^f(t_{k+1})\right] && \text{(KF3)} \\
\mathbf{K}_{k+1} &= \mathbf{P}^f(t_{k+1})\mathbf{H}_{k+1}^T\left[\mathbf{H}_{k+1}\mathbf{P}^f(t_{k+1})\mathbf{H}_{k+1}^T + \mathbf{R}_{k+1}\right]^{-1} && \text{(KF4)} \\
\mathbf{P}^a(t_{k+1}) &= \mathbf{P}^f(t_{k+1}) - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}^f(t_{k+1}) && \text{(KF5)}
\end{aligned}$$

where exponents ^{f} et ^{a} respectively stand for *forecast* and *analysis*.

A variational approach to this inverse problem is the **4D-Var**, which consists in minimizing the cost function:

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T P_b^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)^T \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)$$

with $\mathbf{x}_k = \mathbf{x}(t_k) = \mathbf{M}(t_0, t_k)\mathbf{x}_0$.

The aim of this exercise is to demonstrate that, when the model is perfect ($\mathbf{Q}_k = 0, \quad \forall k$), the 4D-Var and the Kalman filter lead to the same estimation $\mathbf{x}_N$ at the final time (and only at the final time !).

a. We consider a single step of the Kalman filter, with observations available at time t_1. Demonstrate that $\mathbf{x}_1^a$ is the value obtained by the minimization of

$$J_1(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{P}_b^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y}_1 - \mathbf{H}_1\mathbf{x}_1)^T \mathbf{R}_1^{-1}(\mathbf{y}_1 - \mathbf{H}_1\mathbf{x}_1)$$

b. We now consider two successive steps, with observations available at times t_1 and t_2. We start by minimizing the function $J_1(\mathbf{x}_0)$ associated with the problem defined on $[t_0, t_1]$. Let $\mathbf{x}_0^a$ the value corresponding to its minimum, and $\mathbf{P}_0^a$ its error covariance matrix (inverse of the Hessian matrix at the optimum). We perform then a second minimization, using the results of the first minimization in a new background term:

$$J_2(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^a)^T (\mathbf{P}_0^a)^{-1}(\mathbf{x}_0 - \mathbf{x}_0^a) + \frac{1}{2}(\mathbf{y}_2 - \mathbf{H}_2\mathbf{x}_2)^T \mathbf{R}_2^{-1}(\mathbf{y}_2 - \mathbf{H}_2\mathbf{x}_2)$$

Show that: $J_1(\mathbf{x}_0) = J_1(\mathbf{x}_0^a) + \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^a)^T (\mathbf{P}_0^a)^{-1}(\mathbf{x}_0 - \mathbf{x}_0^a)$.
Then show that: $J(\mathbf{x}_0) = J_1(\mathbf{x}_0^a) + J_2(\mathbf{x}_0)$.

c. Give some form of equivalence between the Kalman filter and the 4D-Var.

Exercise 7: Control of the Lorenz model

The Lorenz model (1963) is a nonlinear system of three ordinary differential equations. It is very famous in meteorological and climate sciences as a simple example of a chaotic system having an attractor. Its expression is:

$$\begin{cases} \dfrac{\mathrm{d}x}{\mathrm{d}t} = \alpha(y - x) \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = \beta x - y - xz \qquad\qquad t \in [0, T] \\ \dfrac{\mathrm{d}z}{\mathrm{d}t} = -\gamma z + xy \end{cases}$$

where x, y et z are functions of t, and where α, β and γ are fixed parameters. The initial condition is $(x(0), y(0), z(0)) = (x_0, y_0, z_0)$.

1. Let suppose than we observe $z(t)$ continuously during a time period $[0, T]$. We denote this (noisy) observation by $z_{\text{obs}}(t)$, and we define a cost function to be minimized

$$J_o(x_0, y_0, z_0) = \frac{1}{2}\int_0^T (z(t) - z_{\text{obs}}(t))^2 \, dt$$

We want to apply the adjoint method to get the expression of the gradient of J_o.

1.a. What is the expression for the Gâteaux derivative of J_o ?
1.b. What is the tangent linear model corresponding to the Lorenz model ?
1.c. Writing this tangent linear model in matrix form

$$\frac{\mathrm{d}\hat{X}}{\mathrm{d}t} = A\hat{X} \qquad \text{with } \hat{X}(t) = \begin{pmatrix} \hat{x}(t) \\ \hat{y}(t) \\ \hat{z}(t) \end{pmatrix} \text{ and } A \text{ a matrix,}$$

find the associated adjoint system (be careful in the definition of the scalar product).
1.d. What is the gradient of J_o ?

2. How are the preceding results modified if one controls the coefficients α, β, γ instead of the initial condition ?

3. Another approach to determine the adjoint system in question 1 consists in considering the inverse problem as the minimization of $J_o(x, y, z, x_0, y_0, z_0)$ under the constraint that the Lorenz model equations be satisfied.
3.a. Write the Lagrangian function associated to this problem.
3.b. Modify its expression by integration by part, and compute its gradient w.r.t. the different variables.
3.c. Show that, at the optimum, we get the adjoint model derived in question 1.

Exercise 8: control of the initial condition in a 1D shallow water model

We consider the linear shallow water equations:

$$\left\{ \begin{array}{l} \dfrac{\partial u}{\partial t} + V\dfrac{\partial u}{\partial x} + g\dfrac{\partial h}{\partial x} + ru = f \\[2ex] \dfrac{\partial h}{\partial t} + V\dfrac{\partial h}{\partial x} + D\dfrac{\partial u}{\partial x} + Gu = 0 \end{array} \right. \qquad x \in]0, L[, t \in]0, T[$$

with periodic boundary conditions, and with initial conditions $(u_0(x), h_0(x))$.
u is the velocity, h is the height of the free surface with regard to a reference level, V and D are constant values of velocity and surface height (around which the nonlinear shallow water equations were linearized), g is the gravity parameter, $r \geq 0$ is a friction coefficient, G is the topographic gradient, and f is an external forcing field.
We assume that h is fully observed, and that u is not observed.
Let $J_o(u_0, h_0) = \dfrac{1}{2}\displaystyle\int_0^T \int_0^L (h - h^{\mathrm{obs}})^2$

Use the adjoint method to estimate the gradient of J.