

3.2 The direct model

Let Ω be a bounded domain (Ω Lipschitz). Let U , a Hilbert space, be the controls space (U Banach space only is potentially enough, but we consider it here Hilbert). Let V , a Hilbert space, be the states space.

We consider the following *state equation* (called *direct model* too):

$$\begin{cases} \text{Given } u \in U, \text{ find } y \in V \text{ such that:} \\ A(u; y) = L(u) \text{ in } \Omega \\ \text{with boundary conditions on } \partial\Omega \end{cases} \quad (3.3)$$

where A is an elliptic operator (with respect to the unknown y), defined from $U \times V$ into V' (dual of V). It is non-linear a-priori (both with respect to u and y). L is defined from U into V' .

3.2.1 Distributed control, boundary control

If the control u appears in the "bulk" (i.e. in the equation posed in Ω) then we say that it is a *distributed control*.

If u appears on the boundary conditions only, then we say it is a *boundary control*.

We assume

Assumption 3.2.1. *The state equation (3.3) is well posed in the sense that it has an unique solution $y \in V$, and this solution is continuous with respect to parameters (thus in particular with respect to u).*

3.2.2 Examples

Two linear basic examples of second order elliptic operators are as follows.

Example 1). We set: $U = L^2(\Omega)$, $V = H^1(\Omega)$, $A(u; y) = -u\Delta y$ with mixed boundary conditions: $y = y_0$ on Γ_0 ; $-u\partial_n y = \varphi$ on $\partial\Omega/\Gamma_0$, with y_0 and φ given.

The state equation models a diffusion phenomena (e.g. heat diffusion in a structure, concentration in a fluid, the elastic deformation of a membrane under external force f , the electrostatic field in a conducting media, etc). Here, u is a distributed control, it is the diffusivity coefficient of the material.

Exercise 3.2.1.

- Write the corresponding state equation, and prove that it has one and only (weak) solution in V .
- Prove that the unique solution y is continuous with respect to u i.e. the operator $\pi : u \in U \mapsto y^u \in V$ is continuous.

Solution.

- a) In vertu of Lax-Milgram theorem.
 b) The inequalities of continuity and coercivity give the result. $\square$

Example 2). As in Example 1), we set: $U = L^2(\Omega)$, $V = H^1(\Omega)$. But: $A(u; y) = -\text{div}(\lambda \nabla y)$ with mixed boundary conditions: $y = y_0$ on Γ_0 ; $-\lambda \partial_n y = u$ on $\partial\Omega/\Gamma_0$, with y_0 and λ given. Now, u is a boundary control, it represents the flux at boundary while λ , the diffusive coefficient, is given fixed.

As previously, the corresponding state equation has one and only solution in V (in vertu of Lax-Milgram theorem), the inequalities of continuity and coercivity give the continuity of y with respect to u .

3.3 The objective and cost functions

Let $J(u; y)$ be the objective function, $J : U \times V \rightarrow \mathbb{R}$.

The objective function is supposed to be of class C^1 . Then the cost function is defined by:

$$\begin{cases} j : U \rightarrow \mathbb{R}; j(u) = J(u, y^u) \\ \text{with } y^u \text{ be the unique solution of (3.3)} \end{cases} \quad (3.4)$$

3.3.1 A standard case: quadratic objective function

A typical objective function J is as follows:

$$J(u; y) = \|C(y) - z_d\|^2 + \|u\|_N^2 \quad (3.5)$$

where C is the *observation operator* defined from V into Z , z_d a given "target" in Z (Hilbert space).

N is an operator symmetrical positive definite. It defines the norm $\|\cdot\|_N^2 = (N\cdot, \cdot)_U$. We have: $\forall v \in U$, $(Nv, v)_U \geq c\|v\|_U^2$, $c > 0$.

The first term measures the discrepancy between a target and the computed solution in the *observation space* Z .

The second term measures the cost of the control in norm N .

Exercice 3.3.1. Let j be the cost function defined by $j(u) = J(u; y^u)$ with J defined by (3.5).

- a) Give a sufficient condition to have j of class C^1 ; and give an expression of $\frac{dj}{du}(u) \cdot \delta u$
 b) Is the cost function $j(u)$ a quadratic functional ? Precise your answer.

3.3.2 Differentiability of the cost function

In the following, we will need to differentiate the cost function j (with respect to its unique variable u). Thus, the following question is of main interest in order to well define the optimal control problem:

Is the cost function (continuously) differentiable ?

This question of differentiability is potentially difficult to answer for non-linear systems. For non-linear hyperbolic system in particular, the solution may be even not continuous with respect to the control variable u ...

An useful result to address this question of differentiability is the *implicit function theorem*.

Theorem 3.3.1. (*Implicit function theorem*) *Let us assume that:*

- i) the operator A and L in the state equation (3.3) are C^1 (i.e. $A \in C^1(U \times V)$ and $L \in C^1(U)$),*
- ii) the linearized problem is well-posed (i.e. given (u_0, y_0) the linearized operator $\partial_u A(u_0; y_0)$ is an isomorphism from V into V').*

Then, the implicit function $\mathcal{M} : u \mapsto y^u$, with y^u is the (unique) solution of the state equation, is locally C^1 (locally means it exists a neighborhood of u_0 such that). $\square$

In short, in view to apply the implicit function theorem one need to verify that the linearized problem is C^1 regular and well-posed.

The implicit function is in fact the "operator model" $\mathcal{M} : u \in U \mapsto y^u \in V$, y^u solution of the state equation.

3.4 The optimal control problem

Let U_{ad} , subset of U , be the admissible control set. The optimal control problem writes:

$$\begin{cases} \text{Find } u^* \in U_{ad} \text{ such that:} \\ j(u^*) = \min_{U_{ad}} j(u) \end{cases} \quad (3.6)$$

with the cost function j defined by (3.4).

Let us point out again that j depends on u through the "operator model" $\mathcal{M} : u \in U \mapsto y^u \in V$, where y^u is the (unique) solution of the state equation (3.3).

The optimal control problem (3.6) can be read as follows:

$$\begin{cases} \text{Minimize } J(u; y^u) \text{ in } U_{ad} \\ \text{under the "constraint model" (3.3)} \end{cases} \quad (3.7)$$

In other words, the problem is an optimization problem under the constraint "model must be satisfied". This point of view is the starting point of the Lagrangian approach presented later.

Exercise 3.4.1. *Write the first step of your practical work (step 1)a).*

3.5 A result of existence and uniqueness

We refer to [18, 19] for a full course on the subject.

In the "academic" *linear-quadratic* functional framework, we have existence and uniqueness of the optimal control solution.

Theorem 3.5.1. *Let us consider the state equation linear and coercive, the quadratic cost function defined by (3.5), and U_{ad} closed convex subset of U . Then, there exists a unique solution at the optimal control problem (3.6).*

The state equation is said to be coercive in V if for all $y \in V$, for all $u \in U$, there exists $\alpha > 0$ such that:

$$a(u; y, y) = \langle A(u; y), y \rangle_{V' \times V} \geq \alpha \|y\|^2$$

Proof. The proof is similar to those of Theorem 2.3. First let us reformulate the cost function.

Step 0) We consider $j(u)$ defined by (3.5). We rewrite j as follows:

$$j(u) = \|C(y^u - y^0) + Cy^0 - z_d\|^2 + \|u\|_N^2$$

We set:

$$\pi(u, v) = (C(y^u - y^0), C(y^v - y^0))_Z + (Nu, v)_N \text{ and } L(v) = (z_d - Cy^0, C(y^v - y^0))$$

Since the mapping model $\mathcal{M}$ is affine and continuous, the form π is bilinear symmetric in U and the form L is linear continuous in U .

Furthermore, the form π is coercive in U :

$$\forall u \in U, \quad \pi(u, u) \geq c_0 \|u\|^2, \quad c_0 > 0$$

The cost function writes:

$$j(u) = \pi(u, u) - 2L(u) + \|z_d - y^0\|^2$$

Thus j is continuous and satisfies:

$$j(u) \geq c_0 \|u\|^2 - c_1 \|u\| \tag{3.8}$$

A) Proof of existence. It is based on the convergence of minimizing sequence (calculus of variations, D. Hilbert, 1900 a.c. approx.).

Step 1). Let (u_n) be a minimizing sequence:

$$j(u_n) \rightarrow_n \inf\{j(u), u \in U_{ad}\} \quad (3.9)$$

From (3.8)(??), we obtain:

$$\|u_n\| \leq \text{constant}$$

Hence there exists a sub-sequence (u_{n_k}) which converges weakly to a control u in U_{ad} :

$$u_{n_k} \rightharpoonup u \text{ in } U_{ad}$$

Step 2). U_{ad} is a closed convex subset hence *weakly* closed. Hence $u \in U_{ad}$.

Step 3). Since the cost function $j(u)$ is continuous (lower semi-continuous would be enough), we have: $j(u) = \min_{v \in U_{ad}} j(v)$. In other words, u is solution of the optimal control problem.

B) Uniqueness. The bilinear form $v \mapsto \pi(v, v)$ is coercitive hence the cost function is strictly convex.

Then, the uniqueness is a straightforward consequence of the strict convexity of $j(u)$. $\square$

Exercise 3.5.1. *Detail the proof of uniqueness. (It is similar to the previous proof for an ODE).* $\square$

Let us point out that most of the control problems are not linear-quadratic since the model is not linear. Then, one can prove at best the existence of solutions u . Since generally the cost function is not convex (in the whole set U_{ad}), one do not have uniqueness of the solution.

In short, the cost function to be minimized would be in 2d probably more like a small part of the Pyrénées mountains than a TV parabola !

3.6 Computing a gradient

The final goal of the present course is the numerical resolution of optimal control problems for PDEs systems. To compute numerically the solution of (3.7), few approaches are a-priori possibles, *depending either if the CPU time T_{cpu}^j required to evaluate the cost function j is small or not.*

If T_{cpu}^j is small (let say in fractions of seconds using your laptop or a super-computer whatever), then one can adopt a global optimization approach based on stochastic algorithms (e.g. Monte-Carlo), or heuristic methods (e.g. genetic algorithms) or surface response approaches. Unfortunately, if the state equation is a PDE, generally T_{cpu}^j is not small enough. Thus we do not consider this case in the present course.

If T_{cpu}^j is not small (let say in minutes and more), then global optimization is a-priori not reachable. In such a case, we adopt a *local minimization* approach based on *algorithms of descent* (see the section of recalls). Thus, one needs to compute the gradient of the cost function.

To do so, first j must be differentiable of course (see above), second one must evaluate efficiently the gradient value at a given point u_0 .

Hence, one must address the question: given u_0 , *how to evaluate efficiently the differential function* $\frac{dj}{du}(u_0)$?

Recall that $\frac{dj}{du}(u_0) \in \mathcal{L}(U; \mathbb{R})$. Thus, the question may be:

Given $u_0 \in U_{ad}$ and $\delta u \in U$, how to evaluate the differential value $\frac{dj}{du}(u_0) \cdot \delta u$?

In a computational context, we have discretized the state equation with our favorite numerical methods (and adequate one upon the equation !) e.g. finite differences, finite elements, finite volumes etc. Then, the question reads as follows:

How to compute the gradient $\nabla j(u_0) \in \mathbb{R}^m$?
or $\nabla j(u_0) \cdot \delta u \in \mathbb{R}$ for a potentially large number of directions δu ?

m being the number of discrete control variable ($\dim(U_h) = m$).

The most simple approach: finite differences

The historical method (and the most simple one) is to approximate the gradient values using finite differences. Let U_h be the discrete control space, $\dim(U_h) = m$. Then, *given a direction* $\delta u \in \mathbb{R}^m$, one evaluate as follows:

$$\nabla j(u) \cdot \delta u \approx \frac{j(u + \varepsilon \delta u) - j(u)}{\varepsilon} \text{ at order 1 in } \varepsilon$$

Thus, the evaluation of the vector $\nabla j(u)$ requires $(m + 1)$ evaluation of j , hence $(m + 1)$ resolution of the direct model... If T_{cpu}^j is not small and if m not small (typically in geophysical flows, $m \approx 10^6$ or more, see next chapter), then the present finite difference method is not possible.

The finite difference approach is simple to implement, it is a non-intrusive method. Indeed, it is enough to run the direct model as much as required. It is possible if m tiny; at contrary if m large it is not possible.

Another drawback of this approach is its weak accuracy, depending on the empirical choice of ε .

An other approach would be to solve the linear tangent model but it is not more efficient in a computational time point of view (see next section). The right solution, if possible, is to solve the adjoint model (see later).

3.7 The linear tangent model

First we write:

Lemma 3.7.1. *Let u_0 in U , for all $\delta u \in U$, we have:*

$$\frac{dj}{du}(u_0) \cdot \delta u = \frac{\partial J}{\partial u}(u_0; y) \cdot \delta u + \frac{\partial J}{\partial y}(u_0; y) \cdot w^u \quad (3.10)$$

with: $w^u = \frac{dy}{du} \cdot \delta u$.

Exercise 3.7.1. *Prove this result (straightforward).* □

The function $w^u = \frac{dy}{du} \cdot \delta u$ can be obtained as the solution of the following *linear tangent model*:

$$\left\{ \begin{array}{l} \text{Given } u \in U \text{ and } y^u \text{ the corresponding solution of the state equation (3.3),} \\ \text{given a direction } \delta u \in U, \text{ find } w \in V \text{ such that:} \\ \frac{\partial A}{\partial y}(u; y^u) \cdot w = \left[\frac{\partial L}{\partial u}(u) - \frac{\partial A}{\partial u}(u; y^u) \right] \cdot \delta u \text{ in } \Omega \\ \text{with corresponding linearized boundary conditions on } \partial\Omega \end{array} \right. \quad (3.11)$$

Exercise 3.7.2. *Prove this result (straightforward).* □

We assume that the linearized tangent model is well-posed. Let us remark that if we have proved existence of solutions to the non-linear model, it is probable that we had to prove that the linearized model is well-posed.

In other respect, if the implicit function theorem holds (and has been applied to prove the differentiability of the state with respect to the control), then the linearized problem must be well-posed. Nevertheless, for many problems one do not know to prove any pertinent result concerning the linearized tangent model...

Let us point out that the linearized tangent model give the local *sensitivity of the state y^u* with respect to the control u , at the point u in the direction δu .

In a modeling point of view, this is an very interesting information (see examples later).

Nevertheless if $\dim(U_h) = m$, with the present approach one must solve m times the linear tangent model to obtain w^u in each direction ! Again, if T_{cpu}^j is not small and if m not small, then the present approach is prohibitive.

3.8 The adjoint equations and optimality system

We present below a fundamental result which introduces an auxiliary model: *the adjoint model*. The latter is a powerful mathematical trick which allows to compute the gradient by solving

one (1) extra system only, to be compared to $(m + 1)$ resolutions if using the finite difference approach or the linear tangent model.

Let us point out that excepted if the operator A of the state equation is self-adjoint (i.e. $A^* = A$, in other words A linear, symmetric), the adjoint model has a-priori no physical meaning.

Before deriving the adjoint equations, first we write the state equation and the linear tangent equation in weak forms.

3.8.1 Weak forms of the equations

Let us recall the *direct model*:

$$\begin{cases} \text{Given } u \in U, \text{ find } y \in V \text{ such that:} \\ A(u; y) = L(u) \text{ in } \Omega \\ \text{with boundary conditions on } \partial\Omega \end{cases}$$

We seek to solve this system in the *weak sense* (i.e. in $D'(\Omega)$). To do so, we write its weak (or variational) formulation in the adequate functional space (e.g. V is the Sobolev space $H^1(\Omega)$ for the Laplace equation). We define the following forms:

$$a(u; y, z) : U \times V \times V \rightarrow \mathbb{R} ; \quad a(u; y, z) = \langle A(u; y), z \rangle_{V' \times V}$$

$$b(u; z) : U \times V \rightarrow \mathbb{R} ; \quad b(u; z) = \langle L(u), z \rangle_{V' \times V}$$

Let us point out that both $z \mapsto a(., ., z)$ and $z \mapsto b(., z)$ are linear (i.e. linear with respect to the test function z).

Furthermore, if the direct model is linear (i.e. with respect to its unknown y) then the form $a(., y, z)$ is bilinear; and is not if not.

The weak formulation of the *direct model* (called again the state equation) is as follows:

$$\begin{cases} \text{Given } u \in U, \text{ find } y \in V \text{ such that:} \\ a(u; y, z) = b(u; z) \text{ for all } z \in V \end{cases} \quad (3.12)$$

One of the advantage to consider the direct model in its weak form is to take into account naturally the boundary conditions in the "single" equation (3.11). Then, it will be more convenient to derive the "right" adjoint equations.

Then the weak formulation of the *linear tangent model* is as follows:

$$\begin{cases} \text{Given } u \in U \text{ and } y^u \text{ solution of (3.11),} \\ \text{given } \delta u \in U, \text{ find } w \in V \text{ such that:} \\ \frac{\partial a}{\partial y}(u; y^u, z).w = [\frac{\partial b}{\partial u}(u; z) - \frac{\partial a}{\partial u}(u; y^u, z)].\delta u \text{ for all } z \in V \end{cases} \quad (3.13)$$

Thus, the resolution of the linear tangent model gives $w^u = \frac{dy}{du}(u).\delta u$, the derivative of the state with respect to the control, in the direction δu .

Let us remark that we have: $\mathcal{M} : u \in U \mapsto y^u \in V$, hence $\frac{dy}{du}(u) \in \mathcal{L}(U; V)$ hence $w^u \in V$.

3.8.2 Fundamental theorem: the adjoint equations

We have:

Theorem 3.8.1. *Let us assume that:*

- i) *the state equation (3.11) is well-posed,*
- ii) *its unique solution y^u is C^1 with respect to u ,*
- iii) *the objective function $J(u; y)$ is C^1 ,*
- iv) *the linearized tangent model (3.12) is well-posed.*

Then, the cost function $j(u)$ is C^1 , and: $\forall \delta u \in U$,

$$j'(u) \cdot \delta u \equiv \frac{dj}{du}(u) \cdot \delta u = \frac{\partial J}{\partial u}(u; y^u) \cdot \delta u - \left[\frac{\partial a}{\partial u}(u; y^u, p^u) - \frac{\partial b}{\partial u}(u; p^u) \right] \cdot \delta u \quad (3.14)$$

where y^u is the unique solution of the state equation (3.11), and p^u is solution of the adjoint state equation:

$$\begin{cases} \text{Given } u \text{ and } y^u \text{ the unique solution of (3.11),} \\ \text{find } p \in V \text{ satisfying:} \\ \frac{\partial a}{\partial y}(u; y^u, p) \cdot z = \frac{\partial J}{\partial y}(u, y^u) \cdot z \quad \forall z \in V \end{cases} \quad (3.15)$$

p^u exists and is unique.

Proof.

First, let us notice under the present assumptions, the implicit function theorem applies and it gives the differentiability of the state with respect to u ; i.e. the "operator model" $\mathcal{M} : u \in U \mapsto y^u \in V$ is C^1 . Then, we can differentiate j and we have:

$$\langle j'(u), \delta u \rangle_{U' \times U} = \langle \frac{\partial J}{\partial u}(u; y^u), \delta u \rangle_{U' \times U} + \langle \frac{\partial J}{\partial y}(u; y^u), w^u \rangle_{V' \times V} \quad \forall \delta u \in U$$

(see Lemma (3.9)) with $\langle \cdot, \cdot \rangle_{U' \times U}$, $\langle \cdot, \cdot \rangle_{V' \times V}$ the corresponding duality products.

The linear tangent model writes:

$$\langle \frac{\partial A}{\partial y}(u; y^u) \cdot w^u, z \rangle_{V' \times V} = \langle \frac{\partial L}{\partial u}(u) \cdot \delta u, z \rangle_{V' \times V} - \langle \frac{\partial A}{\partial u}(u; y^u) \cdot \delta u, z \rangle_{V' \times V} \quad \forall z \in V$$

By adding the two equalities above, we obtain:

$$\begin{aligned} \langle j'(u), \delta u \rangle_{U' \times U} &= \langle \frac{\partial J}{\partial u}(u; y^u), \delta u \rangle_{U' \times U} \\ &\quad - \langle [\frac{\partial A}{\partial u}(u; y^u) - \frac{\partial L}{\partial u}(u)] \cdot \delta u, z \rangle_{V' \times V} \\ &\quad + \langle [\frac{\partial J}{\partial y}(u; y^u) - (\frac{\partial A}{\partial y})^*(u; y^u) \cdot z], w^u \rangle_{V' \times V} \quad \forall \delta u \in U \quad \forall z \in V \end{aligned}$$

where $(\frac{\partial A}{\partial y})^*$ is the adjoint operator of the (linearized direct) operator $\frac{\partial A}{\partial y}$.

In the expression of $j'(u).\delta u$ above, we want to make vanish the term w^u . To do so, we introduce the following equations. It is the so-called adjoint equations.

Let $p^u \in V$ be solution of the problem:

$$\langle (\frac{\partial A}{\partial y})^*(u; y^u).p^u, z \rangle_{V' \times V} = \langle \frac{\partial J}{\partial y}(u, y^u), z \rangle_{V' \times V} \quad \forall z \in V$$

The linearized problem is well-posed, hence the operator $\partial_y A(u; y^u)$ is an isomorphism from V into V' , and its adjoint operator $(\partial_y A)^*(u; y^u)$ is an isomorphism from V into V' too, see e.g. [7].

Thus, we obtain the expression:

$$\langle j'(u), \delta u \rangle_{U' \times U} = \langle \frac{\partial J}{\partial u}(u; y^u), \delta u \rangle_{U' \times U} - \langle [\frac{\partial A}{\partial u}(u; y^u) - \frac{\partial L}{\partial u}(u)].\delta u, p^u \rangle_{V' \times V} \quad \forall \delta u \in U$$

hence the result. $\square$

Some fundamental remarks

Let us remark the following properties.

- The expression of $j'(u)$ in the direction δu obtained in Theorem 3.8.1 does not depend anymore on w^u . Its expression is explicit with respect to the direction δu . Thus, to compute the gradient $\nabla j(u)$, we have to solve numerically the direct model plus the adjoint model and *all components* of the gradient follow (i.e. the gradient in all directions). Let us recall that after discretization (in the finite dimension space U_h), we have:

$$\nabla j(u) \in \mathbb{R}^m, \quad \langle \nabla j(u), \delta u \rangle_{\mathbb{R}^m} = \langle j'(u), \delta u \rangle_{\mathbb{R}^m} \quad \text{for all } \delta u \in U_h$$

- By definition, the adjoint model is linear whatever if the direct model is linear or not. (Recall the adjoint is the adjoint operator of the linearized direct operator).

- If the direct operator is self-adjoint, in other words if $a(u, v)$ is bilinear symmetric, then the adjoint operator is the same as the direct operator (but not the right-hand side). Indeed, in such a case, we have:

$$\partial_y a(u; y^u, p).z = a(u; z, p) = a(u; p, z)$$

Only the source term and the boundary conditions differ from the state equation. The differential operator (hence the numerical method and solver) is the same.

3.8.3 The optimality system

We call the optimality system the set of equations characterizing the optimal control solution: the state equation, the adjoint state equation and the first order necessary optimality condition.

In the case, $K = U_{ad} = V$, the optimality system reads: the optimal control solution u^* of Problem (3.6) satisfies:

$$\begin{cases} a(u; y^u, z) = l(u; z) & \forall z \in V \\ \partial_y a(u; y^u, p).z = \partial_y J(u, y^u).z & \forall z \in V \\ j'(u).\delta u = 0 = \partial_u J(u; y^u).\delta u - [\partial_u a(u; y^u, p^u) - \partial_u b(u; p^u)].\delta u & \forall \delta u \in U \end{cases} \quad (3.16)$$

3.9 Complement: adjoint equations for a coupled system

Roughly, if the direct model is composed by two PDEs equations *weakly* coupled, then the adjoint system is composed by the corresponding adjoint equations weakly coupled too by in the reverse way.

If the direct model is composed by two PDEs equations coupled (fully), then the adjoint system is composed by the corresponding adjoint equations (fully) coupled too.

We do not develop more in detail this section since no time enough.

3.10 Example

Control of the elastic deformation of a membrane.

Let us consider the deformation of an elastic membrane, fixed at its boundaries, and distorted by an external force f . We denote by d the displacement of the membrane, $d : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$. It satisfies the following linear second order elliptic system:

$$\begin{cases} -\Delta d = f + u & \text{in } \Omega \\ d = 0 & \text{on } \partial\Omega \end{cases} \quad (3.17)$$

where u is the control of the system, applied on ω a subset of Ω .

The goal is to characterize u such that the displacement d is as close as possible to a target displacement d_0 given. Thus, we define the cost function:

$$j(u) = \int_{\Omega} |d - d_0|^2 dx + \alpha \int_{\omega} |u|^2 dx$$

where d is the solution of (3.16) given u . The second term of the cost has two consequences. In a mechanical point of view, it implies that one seek a control as small as possible (in L^2 norm). In a mathematical point of view, it regularizes the cost function: it is the quadratic term essential in the previous existence-uniqueness theorem. In a numerical point of view it "convexifies" the functional; it is what we call a Tikhonov's regularization (see later).

The weight coefficient α , $\alpha > 0$ makes a balance between the two terms, it has to be tuned by hand empirically (multi-objective function to be minimized...).

We set $U = (L^2(\Omega))^N$. The set of admissible control K is as follows:

$$U = \{u \in U, u = 0 \text{ in } \Omega/\omega\}$$

an the optimal control reads:

$$\min_{u \in K} j(u) \quad (3.18)$$

Existence and uniqueness of solution.

i) Let us point out that in vertu of Lax-Milgram's theorem, given u , the direct model has one and one weak solution d in $V = (H_0^1(\Omega))^N$.

ii) Furthermore, since the mapping $\mathcal{M} : u \in U \mapsto d \in V$ is affine, the cost function j is strongly convex ($j(u) \geq \alpha \|u\|_{L^2}^2$).

In other respect, K is a closed convex set of U . Therefore, in vertu of Theorem 3.5.1, there exists an unique minimum u to the optimal control problem (3.17).

Exercise 3.10.1. Write the optimality system which characterizes the solution u of the optimal control problem (3.17).

3.11 The Lagrangian point of view

An other way to introduce the adjoint equation is to formulate the optimization problem like (3.7). In other words, the model is viewed as a constraint equality of the minimization problem. Then, we write the corresponding lagrangian $\mathcal{L}$:

$$\mathcal{L}(u; y, p) = J(u; y) + \langle A(u; y) - L(u), p \rangle_{V' \times V}$$

The necessary condition of minimum writes:

$$\begin{cases} \partial_u \mathcal{L}(u; y, p) \cdot \delta u = 0 & \forall \delta u \in U \\ \partial_y \mathcal{L}(u; y, p) \cdot \delta y = 0 & \forall \delta y \in V \\ \partial_p \mathcal{L}(u; y, p) \cdot \delta p = 0 & \forall \delta p \in V \end{cases} \quad (3.19)$$

The 1st equation of (3.18) gives the necessary condition "the gradient must vanish":

$$\partial_u J(u; y) \cdot \delta u - [\partial_u a(u; y, p) - \partial_u b(u; p)] \cdot \delta u = j'(u) \cdot \delta u = 0 \quad \forall \delta u \in U$$

The 2nd equation of (3.18) gives the adjoint equation:

$$\partial_y a(u; y, p) \cdot \delta y = \partial_y J(u, y) \cdot \delta y \quad \forall \delta y \in V$$

The 3rd equation of (3.18) gives the state equation:

$$a(u; y, \delta p) = l(u; \delta p) \quad \forall \delta p \in V$$

In short, the optimality system (3.15) is equivalent to the conditions (3.18) of stationary point of the Lagrangian; and, the adjoint state p is the lagrangian multiplier associated to the "constraint model".

Exercise 3.11.1. Write by yourself the equations above. □

3.12 The computational control algorithm

In practice, it is classical to solve the optimization problem (3.6) and the optimality system (3.15) by an iterative descent algorithm.

Given a *first guess* u_0 , we seek iterates $(u^m)_m$ which make decrease the cost function using a first order *descent algorithm* (eg the Quasi-Newton method BFGS). To do so, at each iterate, see Fig. (3.3):

- 1) compute the cost function using the direct model,
- 2) compute the gradient using the adjoint model,
- 3) given the current iterate u^n , the current values $j(u^n)$ and $\nabla j(u^n)$, compute a new iterate u^{n+1} such that:

$$j(u^{n+1}) < j(u^n)$$

*) Iterate till convergence

Remark 3.12.1. *In order to solve the optimality system (3.15), the 'natural' method would be use the Newton algorithm. Such a choice is a-priori efficient (quadratic order of convergence) but it requires the computation of the Hessian of j . Generally, its computation is either very complex and/or CPU time consuming, that is why in many cases, we solve the optimality system (3.15) using a first-order method only.*

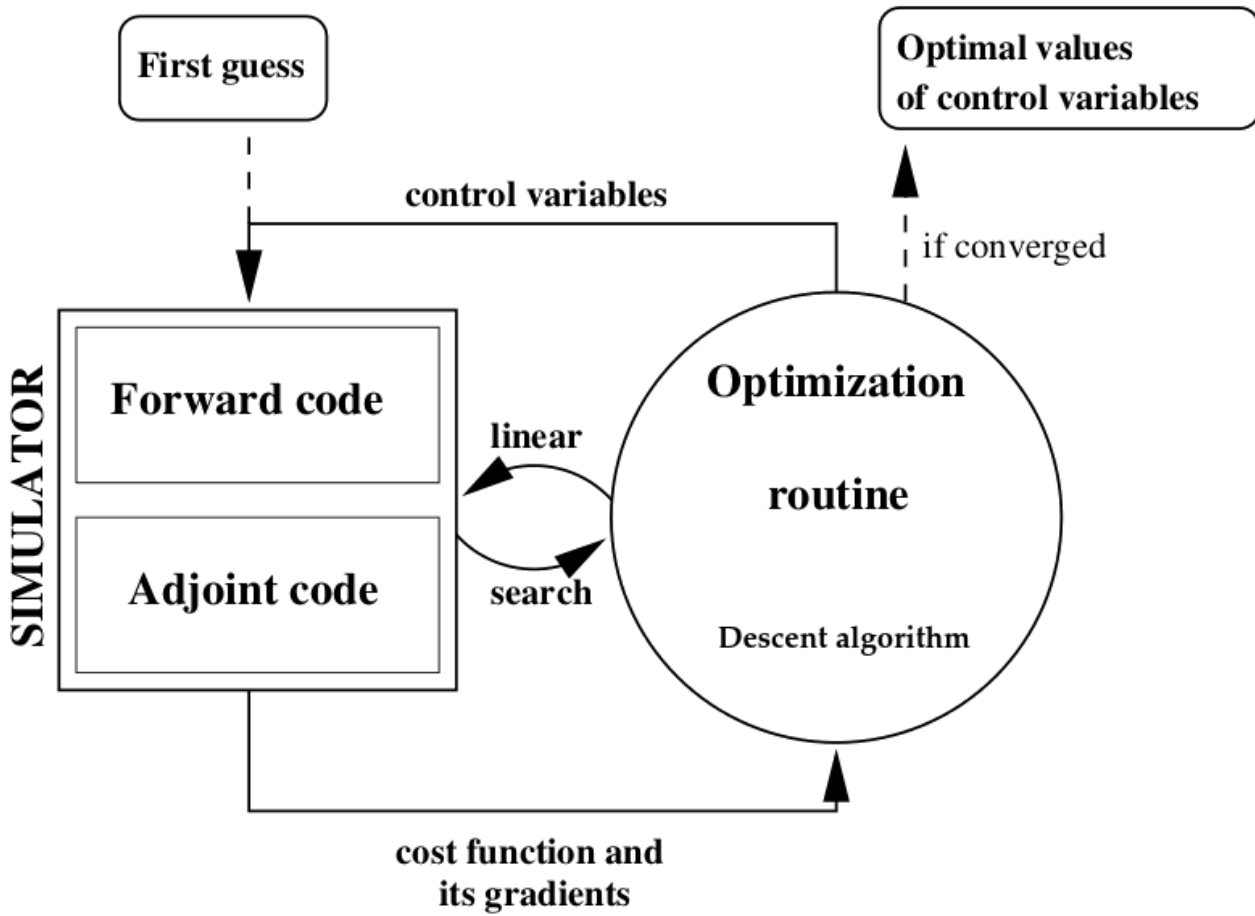


Figure 3.3: Algorithm of optimal control of the PDE system (identification process)

3.13 Practical work: Identification of a pollutant source

In a famous touristic lake, few measurements showed an abnormal concentration of a pollutant. Then, the authorities want to know: 1) a precise distribution of the pollutant in the lake, 2) an estimate of the pollutant volume entering in the lake a day.

Your job, as an "expert", is to give scientific data to the authorities. Thus you will try to answer the request using numerical computations, given the data available. Data are either sparse or uncertain. For few reasons, one does not have accurate measurements of the incoming fluxes of pollutant in the lake....

The pollutant concentration c is supposed to be at equilibrium (steady-state phenomena) and it propagates by diffusion and advection. The lake is supposed to be mono-layer 2d; its currents (fluid velocity v) are known and given. Then, the (simplified) model writes as follows:

$$\begin{cases} -\operatorname{div}(\lambda \nabla c) + v \cdot \nabla c & = 0 \text{ in } \Omega \\ \text{with boundary conditions} \end{cases} \quad (3.20)$$

The diffusivity coefficient λ is a-priori non constant in space, and it is estimated with quite large uncertainties.