International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcon20

A monotonic method for nonlinear optimal control problems with concave dependence on the state

Julien Salomon [a] & Gabriel Turinici [a]

[a] CEREMADE, UMR CNRS 7534, Université Paris IX, Place du Marechal de Lattre de Tassigny, 75775 Paris, Cedex 16, France

Published online: 09 May 2011.

International Journal of Control
Vol. 84, No. 3, March 2011, 551–562

A monotonic method for nonlinear optimal control problems with concave dependence on the state

Julien Salomon* and Gabriel Turinici

CEREMADE, UMR CNRS 7534, Université Paris IX, Place du Marechal de Lattre de Tassigny, 75775 Paris, Cedex 16, France

(*Received 26 January 2010; final version received 9 February 2011*)

Initially introduced in the framework of quantum control, the so-called *monotonic algorithms* have demonstrated very good numerical performance when dealing with bilinear optimal control problems. This article presents a unified formulation that can be applied to more general nonlinear settings compatible with the hypothesis detailed below. In this framework, we show that the well-posedness of the general algorithm is related to a nonlinear evolution equation. We prove the existence of the solution to the evolution equation and give important properties of the optimal control functional. Finally we show how the algorithm works for selected models from the literature. We also compare the algorithm with the gradient algorithm.

Keywords: monotonic algorithms; nonlinear control; optimal control; non-convex optimisation

1. Introduction

This article presents a unified formulation of several algorithms that were proposed in different areas of nonlinear control (see works in Tannor, Kazakov, and Orlov 1992; Zhu and Rabitz 1998b; Lachapelle, Salomon, and Turinici 2010; Salomon and Carlier 2008). Given a cost functional $J(v)$ to be minimised by control input v, and a system $X(t)$ represented by evolution equation (4), these algorithms iteratively construct a sequence of solution v^k with the monotonic behaviour $J(v^{k+1}) \leq J(v^k)$. Thus, the algorithms were named 'monotonic' algorithms. A computational advantage of these procedures is that the monotonicity does not require any additional computational effort, but results from the definition of the procedure itself.

The monotonic algorithms have first been used in the field of quantum control where the dynamics is controlled by a laser field. In this framework the function that maps a control v to the corresponding state X is highly nonlinear. This results in poor performance of standard, gradient-based algorithms. The 'monotonic schemes' introduced in Tannor et al. (1992), Bartana and Kosloff (1997) and Zhu and Rabitz (1998b) were found to perform very well in this setting. These schemes were used in bilinear situations i.e. when the operator $A(t, v(t))$ is linear in $v(t)$ and for a cost functional $J(v) = G(X(T)) + C(v)$, that is, sum of a part $G(X(T))$ is quadratic in the final state $X(T)$ and a part $C(v)$ quadratic in the control v. Zhu et al. present variants of the monotonic scheme in Sugawara and Fujimura (1993), Zhu, Botina, and Rabitz (1998), Zhu and Rabitz (1998a), Schirmer, Girardeau, and Leahy (2000), Sugawara (2003) that included situations where $G(X)$ has negative semi-definite Hessian but $C(v)$ was still quadratic in the control and, most importantly, $A(t, v)$ was linear in v. In the works cited up to now, the function $G(X(T))$ depends only on the final state $X(T)$ but adaptations were proposed in Ohtsuki, Turinici, and Rabitz (2004) and Ohtsuki, Teranishi, Turinici, and Rabitz (2007) to deal with the case where G depends on the whole dynamics of the control process $X(t)$ at intermediary times or when the dynamics involve bilinear integro-differential equations.

Some similar procedures were also proposed in different control applications where the evolution equation is of parabolic type (Salomon and Carlier 2008; Lachapelle et al. 2010) or mixed hyperbolic–parabolic (Ohtsuki, Zhu, and Rabitz 1999; Ohtsuki 2003).

All works presented above considered bilinear situations, i.e. the evolution equation is linear in the state $X(t)$ and $A(t, v)$ is linear in the control v. Recently, different cases are documented in the literature where $A(t, v)$ is a polynomial in the control v up to power three; in Salomon, Dion, and Turinici (2005) and Lapert, Tehini, Turinici, and Sugny (2008) specific monotonic procedures were proposed that were shown to work in this setting. Ohtsuki and Nakagami (2008)

*Corresponding author. Email: julien.salomon@dauphine.fr

ISSN 0020–7179 print/ISSN 1366–5820 online

DOI: 10.1080/00207179.2011.562548
http://www.informaworld.com

presented an adaptation when $A(t, v)$ depends polynomially on a one-dimensional control $v(t) \in \mathbb{R}$. A model where the system is a nonlinear Bose–Einstein condensate was given in Sklarz and Tannor (2002).

In this article, we deal with an arbitrary nonlinear $A(t, v)$ under the requirement that the explicit dependence of J on X be concave (see the hypothesis detailed in this article). In all situations where monotonic algorithms were introduced, the well-posedness of the algorithms (i.e. the existence of v^{k+1} given v^k) was proved by *ad-hoc* techniques although the algebraic computations share similar points. The contribution of this article is to identify and exploit the similarities present in all these situations, and present a general setting of the 'monotonic' algorithms. This allows to tackle a large class of nonlinear situations that cannot be solved by the previous work in the literature. We prove the existence and convergence of a procedure that from a control v^k constructs a control v^{k+1} such that the cost functional is monotonic. The question of whether such a procedure exists has never been addressed before in the literature because until now the authors considered only particular cost functionals J and particular evolution equations. For each case, they proposed explicit analytic formulae for v^{k+1} applicable to their situation. On the contrary, we show in this article that for all problems that satisfy the hypothesis a control v^{k+1} can always be found to ensure $J(v^{k+1}) \leq J(v^k)$. In addition, we give a constructive procedure to compute it.

This article is structured as follows. Section 2 defines the general framework in which our procedure applies. The algorithm is presented in Section 3. We then show that the well-posedness of the algorithm is related to a nonlinear evolution equation. We prove the existence of the solution to the equation. We also give important properties of the optimal control functional. Numerical examples are presented in Section 4 to illustrate the performance of the algorithm.

2. Problem formulation

Let E, $\mathcal{H}$ and $\mathbf{V}$ be Hilbert spaces with $\mathbf{V}$ densely included in $\mathcal{H}$. We denote by $\cdot_E$ and $\langle \cdot, \cdot \rangle_{\mathbf{V}}$ the scalar product associated with E and $\mathbf{V}$. For any vector spaces $\mathcal{A}$ and $\mathcal{B}$, we denote by $\mathcal{L}(\mathcal{A}, \mathcal{B})$ the space of linear continuous operators between $\mathcal{A}$ and $\mathcal{B}$.

Given a real-valued function φ, we denote by $\nabla_x \varphi$ its gradient with respect to the variable x. We also denote by D_x and $D_{x,x}$ the first and the second derivative of vectorial functionals in the Fréchet sense.[1]

We consider the following optimal control problem:

$$\min_v J(v), \tag{1}$$

where

$$J(v) \triangleq \int_0^T F\big(t, v(t), X(t)\big)\mathrm{d}t + G\big(X(T)\big). \tag{2}$$

The functions $F: \mathbb{R} \times E \times \mathbf{V} \to \mathbb{R}$ and $G: \mathbf{V} \to \mathbb{R}$ are assumed to be differentiable and integral are assumed to exist. The state function $X(t) \in \mathbf{V}$ satisfies the following evolution equation:

$$\partial_t X + A(t, v(t))X = B(t, v(t)) \tag{3}$$

$$X(0) = X_0, \tag{4}$$

where $v: [0, T] \to E$ is the control. The operator $A(t, v): \mathbb{R} \times E \times \mathcal{H} \to \mathcal{H}$ is not required to be bounded and is such that for almost all $t \in [0, T]$ the domain of $A(t, v)^{1/2}$ includes $\mathbf{V}$; furthermore, we take $B(t, v)$ such that for almost all $t \in [0, T]$ and all $v \in E$ we have $B(t, v) \in \mathcal{L}(\mathcal{H}, \mathcal{H}) \cap \mathcal{L}(\mathbf{V}, \mathbf{V}^*)$.[2] Please refer to Section 3 for the precise (cf. Lemma 3.9, Theorem 3.11) formulation of additional regularity assumptions to be imposed on A, B, F and G.

Remark 2.1: E is not required to be 1-D. Furthermore, E may not be finite-dimensional (cf. Section 4.2). This implies that the control can not only be a set of several time-dependent functions but also a distributed control depending on a spatial variable.

We want to emphasise that although the equation is linear in X (for v fixed), the mapping $v \mapsto X$ is not linear; the term $A(t, v(t))$ multiplies the state X and as such the mapping is highly nonlinear (of non-commuting exponential type).

Remark 2.2: Most of the previous works considered a bilinear operator $A(t, v)$ i.e. $A(t, v)X = vX$; the only exceptions (cf. discussion in Section 1) were of the polynomial type (at most of order 3 in Salomon et al. (2005) and Lapert et al. (2008) and polynomial with $E = \mathbb{R}^1$ in Ohtsuki and Nakagami (2008). The techniques presented in the above papers cannot be used for general operators $A(t, v)$ and control sets E. On the contrary, the results in this work can not only handle all the situations considered in the bibliography but also apply to nonlinearities in v compatible with the hypothesis of Lemma 3.9 and Theorem 3.11.

The following concavity with respect to X will be assumed throughout this article:

$$\forall X, X' \in \mathbf{V},\ G(X') - G(X) \leq \langle \nabla_X G(X), X' - X \rangle_{\mathbf{V}}, \tag{5}$$

$$\forall t \in \mathbb{R}, \quad \forall v \in E, \quad \forall X, X' \in \mathbf{V},$$
$$F(t, v, X') - F(t, v, X) \leq \langle \nabla_X F(t, v, X), X' - X \rangle_{\mathbf{V}}. \tag{6}$$

Remark 2.3: Unlike the more technical hypothesis that will be assumed later, the properties (5), (6) and the linearity of (3) are crucial to the existence of the monotonic algorithms. Nevertheless, some comments on ways to avoid these assumptions are given in Appendix 1.

3. Monotonic algorithms

We now present the structure of our optimisation procedure together with the general algorithm.

3.1 *Tools for monotonic algorithms*

The general idea of the monotonic algorithms is to exploit a specific factorisation, that is the consequence of the results described in this section. To simplify notations, we will make explicit the dependence of X on v, i.e. we will write X_v instead of X in Equations (3)–(4).

We define the adjoint state Y_v (Lions 1971; Gabasov, Kirillova, and Prischepova 1995) which is the solution of the following evolution equation:

$$\partial_t Y_v - A^*\big(t, v(t)\big) Y_v + \nabla_X F\big(t, v(t), X_v(t)\big) = 0 \tag{7}$$

$$Y_v(T) = \nabla_X G\big(X_v(T)\big). \tag{8}$$

A first estimate about the variations in J can be obtained by the following lemma.

Lemma 3.1: *For any $v', v : [0, T] \to E$ denote*

$$\begin{aligned}\Upsilon\Big(t, X_v(t), v(t), v'(t), Y_v(t), X_{v'}(t)\Big) &\triangleq -\langle Y_v(t), \\ &\Big(A\big(t, v'(t)\big) - A\big(t, v(t)\big)\Big) X_{v'}(t)\rangle \\ &\langle Y_v(t), B\big(t, v'(t)\big) - B\big(t, v(t)\big)\rangle + F\big(t, v'(t), X_{v'}(t)\big) \\ &- F\big(t, v(t), X_{v'}(t)\big).\end{aligned} \tag{9}$$

Then

$$J(v') - J(v) \leq \int_0^T \Upsilon\Big(t, X_v(t), v(t), v'(t), Y_v(t), X_{v'}(t)\Big) \mathrm{d}t. \tag{10}$$

Proof: Using successively (5), (6), (3) and finally (8), we find that

$$\begin{aligned}J(v') - J(v) &= \int_0^T F\big(t, v(t), X_{v'}(t)\big) - F\big(t, v(t), X_v(t)\big) \\ &\quad + F\big(t, v'(t), X_{v'}(t)\big) - F\big(t, v(t), X_{v'}(t)\big) \mathrm{d}t \\ &\quad + G\big(X_{v'}(T)\big) - G\big(X_v(T)\big) \\ &\leq \int_0^T \langle \nabla_X F\big(t, v(t), X_v(t)\big), X_{v'}(t) - X_v(t)\rangle_{\mathbf{V}} \\ &\quad + F\big(t, v'(t), X_{v'}(t)\big) - F\big(t, v(t), X_{v'}(t)\big) \mathrm{d}t \\ &\quad + \langle Y_v(T), X_{v'}(T) - X_v(T)\rangle_{\mathbf{V}} \\ &\leq \int_0^T \Big\langle \frac{\partial}{\partial t} Y_v(t) - A\big(t, v(t)\big)^* Y_v(t) \\ &\quad + \nabla_X F\big(t, v(t), X_v(t)\big), X_{v'}(t) - X_v(t) \Big\rangle_{\mathbf{V}} \\ &\quad - \Big\langle Y_v(t), \Big(A\big(t, v'(t)\big) - A\big(t, v(t)\big)\Big) X_{v'}(t) \Big\rangle_{\mathbf{V}} \\ &\quad + \langle Y_v(t), B(t, v'(t)) - B(t, v(t))\rangle_{\mathbf{V}} \\ &\quad + F\big(t, v'(t), X_{v'}(t)\big) - F\big(t, v(t), X_{v'}(t)\big) \mathrm{d}t.\end{aligned}$$

By (7), the first term of the right-hand side of this last inequality cancels and the result follows. □

Remark 3.2: The purpose of Lemma 3.1 is not to obtain an estimation of the increment $J(v') - J(v)$ via the adjoint (which is well documented in optimal control theory, cf. Lions (1971) and Gabasov et al. (1995)); we rather emphasise that the evaluation of the integrand Υ at time t requires information on the control $v(s)$ for all $s \in [0, T]$ (in order to compute $X_v(T)$ then $Y_v(t)$) but on the second control $v'(s)$ only for $s \in [0, t]$ (because this is enough to compute $X_{v'}(t)$). This estimate can be useful to decide at time t if the current value of the control $v'(t)$ will imply an increase or decrease of $J(v')$. This localisation property is a consequence of the concavity of F and G (in X) and bilinearity induced by A. The purpose of this article is to construct and theoretically support a general numerical algorithm that exploits this remark.

Remark 3.3: We can intuitively note that Υ has the factorised form

$$\begin{aligned}&\Upsilon\Big(t, X_v(t), v(t), v'(t), Y_v(t), X_{v'}(t)\Big) \\ &\quad = \Delta(v, v')(t) \cdot_E \big(v'(t) - v(t)\big),\end{aligned} \tag{11}$$

with $\cdot_E$ the E scalar product. Thus v' can always be chosen so as to make it negative (in the worse case set it null by the choice $v' = v$). We will come back with a formal definition of $\Delta(v, v')(t)$ and a proof of the previous relation in Section 3.3.

A more general formulation can be obtained if we assume that the backward propagation of the adjoint state is performed with the intermediate field $\widetilde{v}$ (cf. also Maday and Turinici 2003), i.e. according to the equation:

$$\frac{\partial}{\partial t} Y_{\tilde{v}} - A^*\big(t, \widetilde{v}(t)\big) Y_{\tilde{v}} + \nabla_X F\big(t, v(t), X_v(t)\big) = 0$$
$$Y_{\tilde{v}}(T) = \nabla_X G\big(X_v(T)\big).$$

Note that because of its final condition, $Y_{\tilde{v}}$ also depends on v. Nevertheless, for the sake of simplicity, we keep the previous notation. We then obtain the following lemma whose proof we leave as exercise for the reader.

Lemma 3.4: *For any* $v', \widetilde{v}, v : [0, T] \to E$,

$$\begin{aligned}
J(v') - J(v) \\
\leq \int_0^T &-\Big\langle Y_{\tilde{v}}(t), \Big(A\big(t, v'(t)\big) - A\big(t, \widetilde{v}(t)\big)\Big)X_{v'}(t)\Big\rangle_{\mathbf{V}} \\
&+ \big\langle Y_{\tilde{v}}(t), B\big(t, v'(t)\big) - B\big(t, \widetilde{v}(t)\big)\big\rangle_{\mathbf{V}} \\
&+ F\big(t, v'(t), X_{v'}(t)\big) - F\big(t, \widetilde{v}(t), X_{v'}(t)\big)\mathrm{d}t \\
&+ \int_0^T -\Big\langle Y_{\tilde{v}}(t), \Big(A\big(t, \widetilde{v}(t)\big) - A\big(t, v(t)\big)\Big)X_v(t)\Big\rangle_{\mathbf{V}} \\
&+ \big\langle Y_{\tilde{v}(t)}(t), B\big(t, \widetilde{v}(t)\big) - B\big(t, v(t)\big)\big\rangle_{\mathbf{V}} \\
&+ F\big(t, \widetilde{v}(t), X_{\tilde{v}}(t)\big) - F\big(t, v(t), X_{\tilde{v}}(t)\big)\mathrm{d}t.
\end{aligned}$$

In this lemma, the variation in the cost functional J is expressed as the sum of two terms, and can be considered as factorised with respect to $v' - \widetilde{v}$ and $\widetilde{v} - v$.

Remark 3.5: Lemmas 3.1 and 3.4 are generalisations of previous results that were proved in the bilinear case. To the best of our knowledge, only specific corollaries requiring additional assumptions have appeared in the literature up to now.

3.2 *The algorithms*

The factorisation (11) that will be proved in Lemma 3.8 enables to design various ways to ensure that $J(v') \leq J(v)$, i.e. guarantee the monotonicity resulting from the update $v' \leftarrow v$. This allows to present a general structure for our class of optimisation algorithms. We focus on the one that results from Lemma 3.1.

Algorithm 3.6 (Monotonic algorithm)**:** Given an initial control v^0, the sequence $(v^k)_{k \in \mathbb{N}}$ is computed iteratively by:

(i) Compute the solution X_{v^k} of (4–5) with $v = v^k$.
(ii) Compute the solution Y_{v^k} of (8–9) with $v = v^k$ backward in time from

$$Y_{v^k}(T) \triangleq \nabla_X G\big(X_{v^k}(T)\big).$$

(iii) Define (as explained later) v^{k+1} together with $X_{v^{k+1}}$ such that for all $t \leq T$ the following monotonicity condition be satisfied:

$$\Delta(v^{k+1}, v^k)(t) \cdot_E \Big(v^{k+1}(t) - v^k(t)\Big) \leq 0. \tag{12}$$

Lemma 3.1 then guarantees that $J(v^{k+1}) \leq J(v^k)$. Several strategies can be used to ensure (12); we will present one below. Its importance stems from the fact that no further optimisation is necessary once this condition is fulfilled. In order to guarantee (12), many authors (Tannor et al. 1992; Zhu and Rabitz 1998b; Maday and Turinici 2003) consider an update formula of the form:

$$v^{k+1}(t) - v^k(t) = -\frac{1}{\theta}\Delta(v^{k+1}, v^k)(t), \tag{13}$$

where θ is a positive number that can also depend on k and t. In what follows, we focus on the existence of solution of (13), and on practical methods to compute it. If v^{k+1} satisfies (13), the variations in J satisfy

$$J(v^{k+1}) - J(v^k) \leq -\theta \int_0^T (v^{k+1}(t) - v^k(t))^2 \,\mathrm{d}t.$$

Note that (13) reads as an update formula combining on the one hand a gradient method:

$$v^{k+1}(t) - v^k(t) = -\frac{1}{\theta}\Delta(v^k, v^k)(t),$$

and on the other hand the so-called proximal algorithm (as described in Attouch and Bolte (2009)) which prescribes

$$v^{k+1}(t) - v^k(t) = -\frac{1}{\theta}\Delta(v^{k+1}, v^{k+1})(t).$$

Remark 3.7: When $F = 0$ and A is independent of v, i.e. linear control with final objective, (13) coincides with a gradient method.

3.3 *Well-posedness of the algorithm*

In this section, we focus on the procedure obtained when using Algorithm 3.6 with the update formula (13). To the best of our knowledge, no theoretical result exists in the literature to prove the existence of a solution to Equation (13) for general choices of $A(t, v)$ and general space of controls E because previous works only dealt with particular choices of functionals F, G, operators A, B and managed to obtain in each case an analytic solution; we provide here such a proof together with a convergent procedure to compute it. Since this procedure involves the resolution of an implicit equation, the proof is non-trivial and has been split in three parts: two preparatory lemmas (Lemmas 3.8 and 3.9) and the final result in Theorem 3.11. As a by-product, we obtain a proof of the monotonicity of the algorithm.

Lemma 3.8: *Suppose that for any* $t \in [0, T]$ *such that*

- $\mathcal{A} : \mathbb{R} \times \mathbf{V} \times \mathbf{V} \times E \to \mathbb{R}$ *defined by* $\mathcal{A}(t, X, Y, v) = \langle Y, A(t, v)X\rangle_{\mathbf{V}}$ *is of* C^1 *class with respect to* v *for any* X, Y, v;

- $\mathcal{B}: \mathbb{R} \times \mathbf{V} \times E \to \mathbb{R}$ *with* $\mathcal{B}(t, Y, v) = \langle Y, B(t, v) \rangle_{\mathbf{V}}$ *is of* C^1 *class with respect to* v *for any* Y, v;
- F *is of* C^1 *class with respect to* $v \in E$ *for any* X, Y, v.

Then there exists $\Delta(\cdot, \cdot; t, X, Y) \in C^0(E^2, E)$ *such that, for all* $v, v' \in E$,

$$\Delta(v', v; t, X, Y) \cdot_E \left(v' - v\right) = -\left\langle Y, \left(A(t, v') - A(t, v)\right)X + B(t, v') - B(t, v)\right\rangle_{\mathbf{V}} + F(t, v', X) - F(t, v, X). \tag{14}$$

Moreover, $\Delta(\cdot, \cdot; t, X, Y)$ *can be defined through the explicit formula*

$$\Delta(v', v; t, X, Y) = \int_0^1 -\nabla_w\left(\left\langle Y, A(t, w)X - B(t, w)\right\rangle_{\mathbf{V}}\right)\Big|_{w = v + \lambda(v' - v)} + \nabla_v F(t, v + \lambda(v' - v), X)\mathrm{d}\lambda. \tag{15}$$

Proof: We denote by $\|\cdot\|$ the norm associated with E. Since $\mathcal{A}, \mathcal{B}, F$ are Fréchet differentiable with respect to v, the full expression in Equation (15) is of the form $\Xi(v') - \Xi(v)$ with $\Xi(v) = -\mathcal{A}(t, X, Y, v) + \mathcal{B}(t, Y, v) - F(t, v, X)$ differentiable with respect to v; Equation (16) is an application of the identity

$$\Xi(v') - \Xi(v) = \int_0^1 \nabla_v \Xi(v + \lambda(v' - v))\mathrm{d}\lambda \cdot_E (v' - v).$$

The continuity is obtained from $\nabla_v \Xi$. □

Lemma 3.9: *Suppose that*

- $\mathcal{A}, \mathcal{B}, F$ *are of (Fréchet)* C^2 *class with respect to* v *with* $D_{vv}A, D_{vv}B$ *uniformly bounded as soon as* X, Y *are in a bounded set*;
- $\nabla_v F$ *is of* C^1 *class with respect to* X;
- $D_{vv}F(t, \cdot, X)$ *is bounded by a positive, continuous, increasing function* $X \mapsto k(\|X\|)$.

Then given that $\varepsilon > 0$, $(t, v, X, Y) \in \mathbb{R} \times E \times \mathbf{V} \times \mathbf{V}$ *and a bounded neighbourhood* W *of* (t, v, X, Y), *there exists* $\theta^\star > 0$ *depending only on* ε, W, $\|v\|$, $\|X\|$ *and* $\|Y\|$ *such that, for any* $\theta > \theta^\star$,

(i) $\Delta(v', v; t, X, Y) = -\theta(v' - v)$ *has a unique solution* $v' = \mathcal{V}_\theta(t, v, X, Y) \in E$;
(ii) $\mathcal{V}_\theta(t, v, X, Y) = v$ *implies that*

$$-\nabla_v\left(\left\langle Y, A(t, v)X\right\rangle_{\mathbf{V}}\right)(v) + \nabla_v\left(\left\langle Y, B(t, v)\right\rangle_{\mathbf{V}}\right)(v) + \nabla_v F(t, v, X) = 0; \tag{16}$$

(iii) $\|\mathcal{V}_\theta(t, v, X, Y) - v\| \le \frac{\|X\|\|Y\| + \|Y\| + k(\|X\|)}{\theta}\{M_0(t) + M_1\|v\|\}$ *with* $M_0(t)$ *and* M_1 *independent of* v, X, Y. *If the dependence of* A, B, F *with respect to* t *is smooth then* $M_0(t)$ *is bounded on* $[0, T]$;
(iv) $\mathcal{V}_\theta(t, v, X, Y)$ *is continuous on* W;
(v) *Let* X *belong to a bounded set, then* $X \mapsto \mathcal{V}_\theta(t, v, X, Y)$ *is Lipschitz with the Lipschitz constant smaller than* ε.

Proof:

(i) Denote $h = v' - v$ and define $\mathcal{G}_{t,v,X,Y}(h) = \frac{-\Delta(v+h, v; t, X, Y)}{\theta}$. When the dependence is clear, we will write simply $\mathcal{G}(h)$ instead of $\mathcal{G}_{t,v,X,Y}(h)$. We thus look for a solution to the following fixed point problem: $\mathcal{G}(h) = h$. For θ large enough, the mapping $\mathcal{G}$ is a (strict) contraction and we obtain the conclusion by a Picard iteration. The uniqueness is a consequence of the contractivity of $\mathcal{G}$.
(ii) If $v' = v$ then $h = 0$, thus $\mathcal{G}(h) = 0$ which gives (16) after using (15).
(iii) For θ large enough, the mapping $\mathcal{G}$ is not only a contraction but also has its Lipschitz constant less than, say, 1/2. Because of the contractivity of $\mathcal{G}$, we have $\|h\| - \|\mathcal{G}(0)\| \le \|h - \mathcal{G}(0)\| = \|\mathcal{G}(h) - \mathcal{G}(0)\| \le \frac{1}{2}\|h\|$, which amounts to $\|h\| \le 2\|\mathcal{G}(0)\|$. Next, we note that

$$\|\mathcal{G}(0)\| \le \frac{\left\{\begin{array}{c}\|\Delta(v, v, t, X, Y) - \Delta(0, 0, t, X, Y)\| \\ + \|\Delta(0, 0, t, X, Y)\|\end{array}\right\}}{\theta} \le M_2\|v\| + M_3(t)$$

and the estimate follows.
(iv) Formula (16) shows that Δ depends continuously on t, v, v', X, Y. Consider converging sequences $t_n \to t$, $v_n \to v$, $X_n \to X$, $Y_n \to Y$ and define $h_n \triangleq \mathcal{V}_\theta(t_n, v_n, X_n, Y_n)$ and $h \triangleq \mathcal{V}_\theta(t, v, X, Y)$.

Given W and $\eta > 0$, consider a large enough value of θ such that:

- for any $(t', v', X', Y') \in W$, $\mathcal{G}_{t', v', X', Y'}$ is a contraction with Lipschitz constant less than 1/2.
- for any (t', v', X', Y'), $(t'', v'', X'', Y'') \in W$,

$$\|\Delta(v'' + h, v'', t'', X'', Y'') - \Delta(v' + h, v', t', X', Y')\| \le \eta.$$

This last property implies $\|\mathcal{G}_{t_n, v_n, X_n, Y_n} \times (h) - \mathcal{G}_{t,v,X,Y}(h)\| \le \frac{\eta}{\theta}$ for n large enough.

On the other hand,

$$\begin{aligned}\|h_n - h\| &= \|\mathcal{G}_{t_n,v_n,X_n,Y_n}(h_n) - \mathcal{G}_{t,v,X,Y}(h)\| \\ &\le \|\mathcal{G}_{t_n,v_n,X_n,Y_n}(h_n) - \mathcal{G}_{t_n,v_n,X_n,Y_n}(h)\| \\ &\quad + \|\mathcal{G}_{t_n,v_n,X_n,Y_n}(h) - \mathcal{G}_{t,v,X,Y}(h)\| \\ &\le \frac{1}{2}\|h_n - h\| + \frac{\eta}{\theta}.\end{aligned}$$

We have thus obtained that for n large enough: $\frac{1}{2}\|h_n - h\| \le \frac{\eta}{\theta}$ and the continuity follows.

(v) Subtracting the two equalities

$$\Delta(V_1, v; t, X_1, Y) = -\theta(V_1 - v),$$
$$\Delta(V_2, v; t, X_2, Y) = -\theta(V_2 - v)$$

and using that $\Delta(V, v; t, X, Y)$ is C^1 in X and v gives to first-order

$$\begin{aligned}&\Delta_V(\ldots)(V_1 - V_2) + \Delta_X(\ldots)(X_1 - X_2) \\ &\quad = -\theta(V_1 - V_2).\end{aligned}$$

For θ large enough the operator $\Delta_V(\ldots) + \theta \cdot Id$ is invertible and the conclusion follows. □

Remark 3.10: Note that $\theta^\star$ is proportional to $(\|X\|_{\mathbf{V}}\|Y\|_{\mathbf{V}} + \|Y\|_{\mathbf{V}} + k(\|X\|_{\mathbf{V}}))$.

We are now able to construct a procedure such that the existence of $v^{k+1}(t)$ satisfying (12) is guaranteed.

Theorem 3.11: *Suppose that A, B and F satisfy the hypothesis of Lemma 3.9. Also suppose that the operators A, B are such that Equations (3)–(4) and (7)–(8) have solutions for any $v \in L^\infty(0, T; E)$ with $v \mapsto X$, $v \mapsto Y$ locally Lipschitz. Then:*

(i) *For any $v \in L^\infty(0, T; E)$, there exists $\theta^\star > 0$ such that for any $\theta > \theta^\star$, the (nonlinear) evolution system*

$$\partial_t X_{v'}(t) + A(t, v')X_{v'}(t) = B(t, v') \tag{17}$$

$$v'(t) = \mathcal{V}_\theta(t, v(t), X_{v'}(t), Y_v(t)) \tag{18}$$

$$X_{v'}(0) = X_0 \tag{19}$$

has a solution. Here Y_v is the adjoint state defined by (7) and (8) and corresponding to control v.

(ii) *There exists a sequence $(\theta_k)_{k\in\mathbb{N}}$ such that the Algorithm 3.6 (cf. Section 3.2)*

a/ initialization $v^0 \in L^\infty(0, T; E)$,
b/ $v^{k+1}(t) = \mathcal{V}_{\theta_k}(t, v^k(t), X_{v^{k+1}}(t), Y_{v^k}(t))$ is monotonic and satisfies

$$J(v^{k+1}) - J(v^k) \le -\theta_k\|v^{k+1} - v^k\|^2_{L^2([0,T])}.$$

(iii) *With the notations above, if for all $t \in [0, T]$ $v^{k+1}(t) = v^k(t)$ (i.e. algorithm stops) then v^k is a critical point of J: $\nabla_v J(v^k) = 0$.*

Proof: Some part of the proof is contained in the previous lemmas. The part that still has to be proved is the existence of a solution to (17)–(19).

Given that $v \in L^\infty(0, T; E)$, consider the following iterative procedure:

$$v_0 = v, \ v_{l+1}(t) = \mathcal{V}_\theta(t, v(t), X_{v_l}(t), Y_v(t)).$$

We take a spherical neighbourhood $B_v(R)$ of v of radius R and suppose that $\forall k \le l$, $v_k \in B_v(R)$. Since the correspondence $v \mapsto X_v$ is continuous, it follows that the set of solutions $S_{v,R} \triangleq \{X_w;\ w \in B_v(R)\}$ of (4) is bounded. In particular, for $w = v_l$ by the item (iii) of Lemma 3.9 the quantity $\|\mathcal{V}_\theta(t, v(t), X_{v_l}(t), Y_v(t)) - v\|$ will be bounded by R for θ large enough (depending on R, independent of l), i.e. $v_{l+1} \in B_v(R)$. Thus $v_l \in B_v(R)$ for all $l \ge 1$.

Since $S_{v,R}$ is bounded, recall that by item (v) of Lemma 3.9 the mapping $X \mapsto \mathcal{V}_\theta(t, v(t), X, Y_v(t))$ has on $S_{v,R}$ a Lipschitz constant as small as desired. Since the mapping $w \mapsto X_w$ is Lipschitz, for θ large enough, $w \in B_v(R) \mapsto \mathcal{V}_\theta(t, v(t), X_w, Y_v(t))$ is a contraction. By a Picard argument, the sequence v_l is converging. The limit will be the solution of (17) and (18). □

4. Examples

We now present three examples that fit into the setting of Theorem 3.11. Due to space limit, we do not treat different variants (cf. references in Section 1).

Within the framework of control theory, nonlinear formulations prove useful nowadays in domains as diverse as the laser control of quantum phenomena (Warren, Rabitz, and Dahleh 1993; Weinacht, Ahn, and Bucksbaum 1999; Rabitz, de Vivie-Riedle, Motzkus, and Kompa 2000; Rice and Zhao 2000; Levis, Menkir, and Rabitz 2001; Rabitz, Hsieh, and Rosenthal 2004) or the modelling of an equilibrium (or again social beliefs, product prices, etc.) of a game with an infinite numbers of agents (Lasry and Lions 2006a, b, 2007). Yet, other applications arise from modern formulations of the Monge–Kantorovich mass transfer problem (Benamou and Brenier 2001, 2000; Salomon and Carlier 2008).

In the following, we present some examples from these fields of application. We also present the corresponding monotonic algorithm resulting from Theorem 3.11. Concerning the time-discretisation, we use the approach described in Appendix 2.

4.1 *(I): Quantum control*

4.1.1 *Setting*

The evolution of a quantum system is described by the Schrödinger equation

$$\partial_t X + iH(t)X = 0$$
$$X(0, z) = X_0(z),$$

where $i = \sqrt{-1}$, $H(t)$ is the Hamiltonian of the system and $z \in \mathbb{R}^\gamma$ the set of internal degrees of freedom. We assume that the Hamiltonian is a self-adjoint operator over $L^2(\mathbb{R}^\gamma; \mathbb{C})$, i.e. $H(t)^* = H(t)$.[3] Note that this implies the following norm conservation property:

$$\|X(t, \cdot)\|_{L^2(\mathbb{R}^\gamma;\mathbb{C})} = \|X_0\|_{L^2(\mathbb{R}^\gamma;\mathbb{C})} \quad \forall t > 0, \tag{20}$$

so that the state (also called wave-) function $X(t, z)$, evolves on the (complex) unit sphere $S \triangleq \{X \in L^2(\mathbb{R}^\gamma; \mathbb{C}) : \|X\|_{L^2(\mathbb{R}^\gamma;\mathbb{C})} = 1\}$.

The Hamiltonian is composed of two parts: a free evolution Hamiltonian H_0 and a part that describes the coupling of the system with an external laser source of intensity $v(t) \in \mathbb{R}$, $t \geq 0$; a first-order approximation leads to adding a time-independent dipole moment operator $\mu(z)$ resulting in the formula $H(t) = H_0 - v(t)\mu$ and the dynamics:

$$\partial_t X + i(H_0 - v(t)\mu)X = 0$$
$$X(0) = X_0.$$

The purpose of control may be formulated as to drive the system from its initial state X_0 to a final state X_{target} compatible with predefined requirements. Here, the control is the laser intensity $v(t)$. Because the control is multiplying the state, this formulation is called 'bilinear' control. The dependence $v \mapsto X(T)$ is of course not linear.

The optimal control approach can be implemented by introducing a cost functional. The following functionals are often considered:

$$J(v) \triangleq \|X(T) - X_{\text{target}}\|^2_{L^2(\mathbb{R}^\gamma;\mathbb{C})} + \int_0^T \alpha(t)v^2(t)\mathrm{d}t, \tag{21}$$

$$\widetilde{J}(v) \triangleq -\langle X(T), OX(T)\rangle_{L^2(\mathbb{R}^\gamma;\mathbb{C})} + \int_0^T \alpha(t)v^2(t)\mathrm{d}t, \tag{22}$$

where O is a positive linear operator defined on $\mathcal{H}$, characterising an observable quantity and $\alpha(t) > 0$ is a parameter that penalises large (in the L^2 sense) controls. The goal is to minimise these functionals with respect to v. According to (20), the cost functional J is equal to

$$J(v) \triangleq 2 - 2Re\langle X(T), X_{\text{target}}\rangle_{L^2(\mathbb{R}^\gamma;\mathbb{C})} + \int_0^T \alpha(t)v^2(t)\mathrm{d}t, \tag{23}$$

so that the functionals J and $\widetilde{J}$ satisfy assumptions (5) and (6).

4.1.2 *Mathematical formulation*

We have

- $A(t, v) = H_0 + v(t)\mu$ with (possibly) unbounded v-independent operator H_0 (but which generates a C^0 semi group) and bounded operator μ. The dependence of A on v is smooth (linear) and therefore all hypotheses on A are satisfied.
- $E = \mathbb{R}$, $\mathcal{H} = L^2(\mathbb{R}^d; \mathbb{C})$, $\mathbf{V} = \mathrm{dom}(H_0^{1/2})$ (over $\mathbb{C}$), or their realifications $\mathcal{H} = L^2 \times L^2$, $\mathbf{V} = \mathrm{dom}(H_0^{1/2}) \times \mathrm{dom}(H_0^{1/2})$ (over $\mathbb{R}$) as explained in Ito and Kunisch (2007).
- $B(t, v) = 0$.
- $F(t, v, X) = \alpha(t)v(t)^2$ with $\alpha(t) \in L^\infty(\mathbb{R})$; here the second derivative $D_{vv}F$ is obviously bounded. Since it is independent of X, it will be trivially concave.
- G is either (see e.g. Maday and Turinici 2003; Maday, Salomon, and Turinici 2007) $2 - 2Re\langle X_{\text{target}}, X(T)\rangle_{\mathbf{V}}$ or $-\langle X(T), OX(T)\rangle_{\mathbf{V}}$ where O is a positive semi-definite operator; both are concave in X.
- Here

  $$\Delta(v', v; t, X, Y) = -Re\langle Y, i\mu X\rangle_{\mathbf{V}} + \alpha(t)(v' + v) \tag{24}$$

  and the equation in v' is $\Delta(v', v;\ t, X, Y) = -\theta(v' - v)$ and has a unique solution $v' = \mathcal{V}_\theta(t, v, X, Y) \triangleq \frac{(\theta-\alpha(t))v + Re\langle Y, i\mu X\rangle_{\mathbf{V}}}{\theta+\alpha(t)}$ for θ large enough.
- At the $k+1$-th iteration, Theorem 3.11 guarantees the existence of the solution X^{k+1} of the following nonlinear evolution equation:

$$i\partial_t X^{k+1}(t) = \left(H_0 + \frac{(\theta - \alpha(t))v^k + Re\langle Y_{v^k}, i\mu X^{k+1}\rangle_{\mathbf{V}}}{\theta + \alpha(t)}\mu\right)X^{k+1}(t). \tag{25}$$

Then

$$v^{k+1} = \frac{(\theta - \alpha(t))v^k + Re\langle Y_{v^k}, i\mu X^{k+1}\rangle_{\mathbf{V}}}{\theta + \alpha(t)}, \quad X_{v^{k+1}} = X^{k+1}. \tag{26}$$

4.1.3 *Numerical test*

In order to test the performance of the algorithm, we have chosen a case already treated in the literature (Zhu and Rabitz 1998b). The system under consideration is the $O-H$ bond that vibrates in a Morse-type potential $V(z)=D_0(\exp(-\beta(z-z'))-1)^2-D_0$ and $H_0=-m\frac{\partial^2}{\partial z^2}+V(z)$. The dipole moment operator of this system is modelled by $\mu(z)=\mu_0\cdot ze^{-\frac{z}{z^\star}}$. The objective is to localise the wavefunction at time $T=1{,}31{,}000$ at a given location z_0; this is expressed through the requirement that the functional $\widetilde{J}$ is maximised, with the observable O defined by $O(z)=\frac{\gamma_0}{\sqrt{\pi}}e^{-\gamma_0^2(z-z_0)^2}$. The numerical values we use are given below:

D_0	β	z'	$z^\star$	z_0	γ_0	μ_0	m
0.1994	1.189	1.821	0.6	2.5	25	3.088	$2.8694.10^{-4}$

We consider a constant penalisation parameter $\alpha=1$ and optimisation parameter $\theta=10^{-2}$. To compare this procedure with a standard algorithm, we have also minimised $J(v)$ with an optimal step gradient method. The line search is achieved through a golden section search (cf. Press, Teukolsky, Vetterling, and Flannery 2002). Results are presented in Figure 1. This test shows that the gradient method fails in efficiently solving the problem, whereas the monotonic procedure ensures that the cost functional values rapidly decrease. Note that the non-convexity of the problem renders difficulty in the convergence of the gradient method. On the other hand, the monotonic scheme fully exploits the concavity of the cost functional with respect to the state variable. In our implementation the time of computation is about two times larger for the gradient method as the line search requires about three evaluations of the cost functional per iteration.

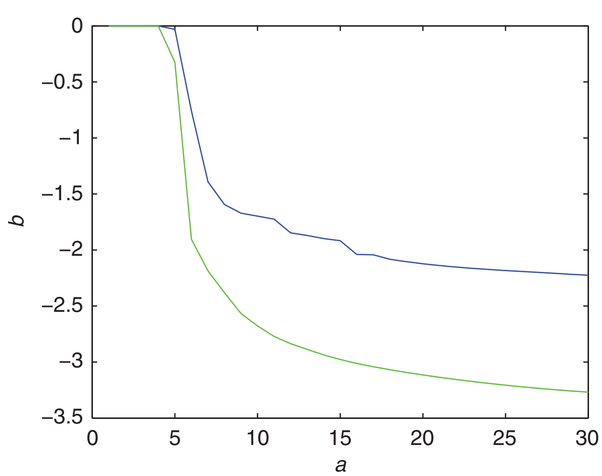

Figure 1. Numerical resolution of the example of Section 4.1. The cost functional $J(v)$ is optimised using the monotonic algorithm (26) (green line) and the optimal step gradient algorithm (blue line). Available in colour online.

4.2 *(II): Mean field games*

4.2.1 *Setting*

Although the Nash equilibrium in game theory has been initially formulated for a finite number of players, modern results (Lasry and Lions 2006a, b, 2007) indicate that it is possible to extend it to an infinite number of players and obtain the equations that describe this equilibrium; applications have already been proposed in the economic theory and other are expected in the behaviour of multi-agents ensembles and decision theory.

The equations describe evolution of the density $X(t,z)$ of players at time t and position $z\in Q=[0,1]$ in terms of a control $v(t,z)$ and a fixed parameter $\nu>0$:

$$\partial_t X-\nu\Delta X+\mathrm{div}(v(t,z)X)=0,$$
$$X(0)=X_0.$$

The control v is chosen to minimise the cost criterion (2). For reasons related to economic modelling, interesting examples include situations where F, G are concave in X, e.g. as in Lachapelle et al. (2010)

$$G=0,\quad F(t,z,X)=\int_Q p(t)(1-\beta z)X(t,z)+\frac{c_0\cdot z\cdot X(t,z)}{c_1+c_2X(t,z)}+\frac{v^2(t)}{2}X(t,z)\mathrm{d}z, \tag{27}$$

with positive constants β, c_0, c_1, c_2 and $p(t)$ a positive function. Another example is given in Salomon and Carlier (2008):

$$G(X(T))=\int_Q V(z)X(T,z)\mathrm{d}z,$$
$$F(t,z,X)=\int_Q X(t,z)v^2(t,z)\mathrm{d}z, \tag{28}$$

where V encodes a potential. The interpretation of this terminal cost is that the crowd aims at reaching zones of low potential V at the terminal time T while minimising the cost of changing state.

The numerical relevance of the monotonic algorithms to this setting has been established in several works (Salomon and Carlier 2008; Lachapelle et al. 2010).

4.2.2 *Mathematical formulation*

We have

- $E=W^{1,\infty}(0,1)$, $\mathcal{H}=L^2(0,1)$, $\mathbf{V}=H^1(0,1)$, see Lachapelle et al. (2010) and Dautray and Lions (1992, Chapter XVIII, Section 4.4).
- $A(t,v)=-\nu\Delta\cdot+\mathrm{div}(v\cdot)$. The dependence of A on v is smooth (linear) and therefore all hypotheses on A are satisfied ($D_{vv}A=0,\ldots$).

- $B(t, v) = 0$.
- With definitions in (27) $F(t, v, X) = \int_Q p(t)(1 - \beta z)X(t, z) + \frac{c_0 \cdot z \cdot X(t,z)}{c_1 + c_2 X(t,z)} + \frac{v(t,z)^2}{2} X(t, z)\mathrm{d}z$; F is concave in X; the second differential $D_{vv}F$ has all required properties.
- $G = 0$ (algorithm will apply in general when G is concave with respect to X).
- Here

$$\Delta(v', v; t, X, Y) = \nabla Y + \frac{v' + v}{2} \tag{29}$$

and the equation in v' is $\Delta(v', v; t, X, Y) = -\theta(v' - v)$ and has a unique solution $v' = \mathcal{V}_\theta(t, v, X, Y) \triangleq \frac{(\theta - 1/2)v - \nabla Y}{\theta + 1/2}$ for all $\theta > 0$.
- At the $k+1$-th iteration, Theorem 3.11 guarantees the existence of the solution X^{k+1} of the following nonlinear evolution equation:

$$\partial_t X^{k+1}(t) - \nu \Delta X^{k+1} + \mathrm{div}\left(\frac{(\theta - 1/2)v^k - \nabla Y_{v^k}}{\theta + 1/2} X^{k+1}\right) = 0. \tag{30}$$

Then

$$v^{k+1} = \frac{(\theta - 1/2)v^k - \nabla Y_{v^k}}{\theta + 1/2}, \quad X_{v^{k+1}} = X^{k+1}. \tag{31}$$

4.2.3 *Numerical test*

The algorithm is tested on the time interval [0, 1] with $p(t) = 1$ and the numerical values $\beta = 0.8$, $c_0 = c_2 = 1$ and $c_1 = 0.1$. The same gradient method as in Section 4.1.3 is also tested. Results are presented in Figure 2. In this example, the gradient method gives better results in the first iterations.

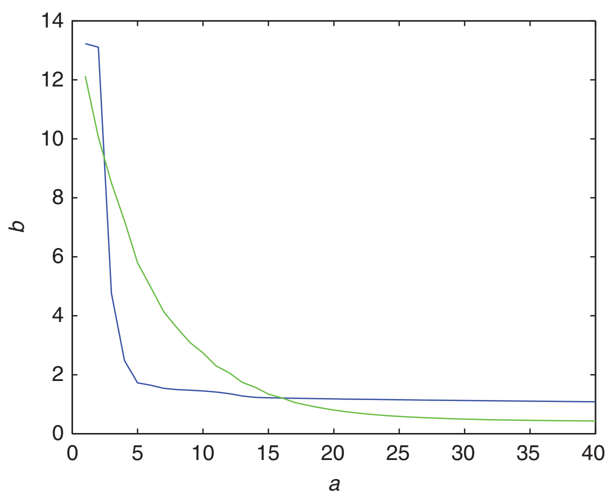

Figure 2. Numerical resolution of the example of Section 4.2. The cost functional $J(v)$ is optimised using the monotonic algorithm (31) (green line) and the optimal step gradient algorithm (blue line). Available in colour online.

However, the monotonic algorithm converges asymptotically faster.

4.3 *Additional application*

As a third example, we consider a nonlinear vectorial case from Friedrich and Herschbach (1995) and Tehini and Sugny (2008) which differs from that of Section 4.1 in that $v(t) = \binom{v_1}{v_2} \in E = \mathbb{R}^2$ and $A(t, v) = i[H_0 + (v_1(t)^2 + v_2(t)^2)\mu_1 + v_1(t)^2 v_2(t)\mu_2]$. Here, denoting $\xi_1 = -Re\langle Y, i\mu_1 X\rangle_{\mathbf{V}} + \alpha(t)$, $\xi_2 = -Re\langle Y, i\mu_2 X\rangle_{\mathbf{V}}$, we obtain

$$\Delta(v', v; t, X, Y) = \xi_1 \begin{pmatrix} v_1 + v_1' \\ v_2 + v_2' \end{pmatrix} + \xi_2 \begin{pmatrix} (v_1 + v_1')v_2' \\ (v_1)^2 \end{pmatrix} \tag{32}$$

and the equation in v' is $\Delta(v', v;\ t, X, Y) = -\theta(v' - v)$ and has a unique solution

$$v' = \mathcal{V}_\theta(t, v, X, Y) = \begin{pmatrix} \dfrac{(\theta - \xi_1)v_2 - \xi_2 v_1^2}{\theta + \xi_1} \\ -\dfrac{\theta - \xi_1 + \xi_2 \frac{(\theta - \xi_1)v_2 - \xi_2 v_1^2}{\theta + \xi_1}}{\theta + \xi_1 + \xi_2 \frac{(\theta - \xi_1)v_2 - \xi_2 v_1^2}{\theta + \xi_1}} v_1 \end{pmatrix}$$

for θ large enough. We leave the writing of the equation for X^{k+1} and the formula for v^{k+1} as an exercise to the reader.

This model corresponds to the problem of controlling the orientation γ of a molecule, considered as rigid rotator.

4.3.1 *Numerical test*

To test our approach, we have used the parameters of the molecule *CO* (Friedrich and Herschbach 1995; Tehini and Sugny 2008), namely $H_0 = BJ^2$, where B is the rotational constant and J is the angular momentum. We consider the basis given by the spherical harmonics; the corresponding matrix is diagonal with diagonal coefficients given by $(H_0)_{k,k} = k(k+1)$. The controlled is performed over an interval of length $T = 20T_{\mathrm{per}} = 20\frac{\pi}{B}$. We consider constant penalisation factor $\alpha = 10^{-1}$ and optimisation parameter $\theta = 10^3$.

The other parameters correspond to the polarisability and the hyperpolarisability components of the molecule. We have $\mu_1 = -\frac{1}{2}\lambda$, and $\mu_2 = -\frac{3}{4}\beta$, with $\lambda = \frac{1}{2}(\lambda_\parallel \cos^2\gamma + \lambda_\perp \sin^2\gamma)$, $\beta = \frac{1}{6}((\beta_\parallel - 3\beta_\perp)\cos^3\gamma + 3\beta_\perp \cos\gamma)$. The matrix $\cos\gamma$ is tridiagonal, with

$$(\cos\gamma)_{k,k} = 0,$$

$$(\cos\gamma)_{k,k+1} = (\cos\gamma)_{k+1,k} = \frac{k+1}{\sqrt{(2k+1)(2k+3)}}.$$

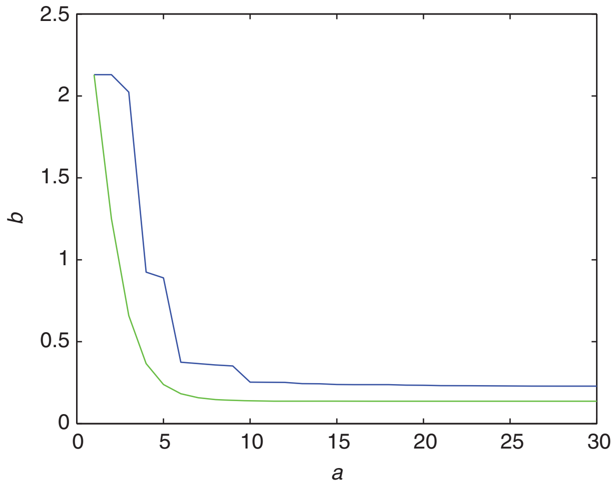

Figure 3. Numerical resolution of the example of Section 4.3. The cost functional $J(v)$ is optimised using the monotonic algorithm (32) (green line) and a gradient method (blue line). Available in colour online.

We use the numerical values given in Friedrich and Herschbach (1995) and Tehini and Sugny (2008):

B	$\lambda_\perp$	$\lambda_\parallel$	$\beta_\parallel$	$\beta_\perp$
1.93	11.73	15.65	28.35	6.64

A gradient method similar to the one that is used in Section 4.1.3 is also performed. The results are presented in Figure 3. The monotonic algorithm shows a fast convergence whereas the gradient method does not optimises efficiently the cost functional values.

5. Conclusion

Motivated by a set of control algorithms that were initially introduced in the specific context of quantum control, we have presented an abstract formulation that includes them all. We identified the theoretical assumptions to ensure that the evolution equation is well posed and has a solution. We provide an original and constructive proof to serve as basis for numerical approximations of the solution. We also proved several properties concerning the algorithms. Examples are provided to show how the proposed procedure solves cases from not only the previous literature but also new situation that were not previously considered. Numerical simulations indicate that the procedures have indeed the expected behaviour.

Acknowledgements

We thank the anonymous referees for their helpful comments. This work is partially supported by the French ANR programs OTARIE (grant ANR-07-BLAN-0235) C-QUID (grant ANR-06-BLAN-0052-04) and by a CNRS-NFS PICS grant. G. Turinici acknowledges partial support by INRIA Rocquencourt (MicMac and OMQP).

Notes

1. Recall that, given H_1 and H_2 two Banach spaces and $U \subset H_1$ an open subset of H_1, a function $f: U \to H_2$ is said to be Fréchet differentiable at $x \in U$ if there exists a continuous linear operator $A_x \in \mathcal{L}(H_1, H_2)$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - A_x(h)\|_{H_2}}{\|h\|_{H_1}} = 0.$$

The operator A_x is then called the Fréchet differential (or Fréchet derivative) of f at x and is denoted by $D_x f \triangleq A_x$. Let us also recall that given an open set $\Omega \subset \mathbb{R}^\gamma$ and a Hilbert space H_1, the set $L^\infty(\Omega; H_1)$ is the space of functions f from Ω with values in the Hilbert space H_1 such that for almost all $t \in \Omega$ the norm $\|f(t)\|_{H_1}$ is bounded by the same constant (the lowest of which is the $L^\infty(\Omega; H_1)$ norm of f). One can likewise define $L^2(\Omega; H_1)$:

$$L^2(\Omega; H_1) = \left\{ f : \Omega \to H_1 \text{ such that } \int_\Omega \|f(t)\|^2_{H_1} \mathrm{d}t < \infty \right\}.$$

When the derivatives of f are considered, the Sobolev spaces $W^{1,\infty}$ have to be introduced. We refer to Yosida (1995) and Adams and Fournier (2003) for further details.
2. For a space $\mathbf{V}$ we denote by $\mathbf{V}^*$ its dual space.
3. For any operator M, we denote by M^* its adjoint.

References

Adams, R., and Fournier, J. (2003), 'Sobolev Spaces', *Pure and Applied Mathematics (Amsterdam)* (Vol. 140, 2nd ed.), Amsterdam: Elsevier/Academic Press.

Attouch, H., and Bolte, J. (2009), 'On the Convergence of the Proximal Algorithm for Nonsmooth Functions Involving Analytic Features', *Mathematical Programming*, 116, 5–16.

Bartana, A., and Kosloff, R. (1997), 'Laser Cooling and Internal Degrees of Freedom. II', *Journal of Chemical Physics*, 106, 1435–1448.

Benamou, J.-D., and Brenier, Y. (2000), 'A Computational Fluid Mechanics Solution to the Monge–Kantorovich Mass Transfer Problem', *Numerical Mathematik*, 84, 375–393.

Benamou, J.D., and Brenier, Y. (2001), 'Mixed L^2-wasserstein Optimal Mapping Between Prescribed Density Functions', *Journal of Optimization Theory and Applications*, 111, 255–271.

Dautray, R., and Lions, J.-L. (1992), *Mathematical Analysis and Numerical Methods for Science and Technology* (Vol. 5), Berlin: Springer-Verlag. (Evolution problems. I).

Friedrich, B., and Herschbach, D. (1995), 'Alignment and Trapping of Molecules in Intense Laser Fields', *Physical Review Letters*, 74, 4623.

Gabasov, R., Kirillova, F.M., and Prischepova, S.V. (1995), *Optimal Feedback Control*, Lecture Notes in Control and Information Sciences (Vol. 207), London: Springer-Verlag London Ltd.

Ito, K., and Kunisch, K. (2007), 'Optimal Bilinear Control of an Abstract SchröDinger Equation', *SIAM Journal of Control Optimization*, 46, 274–287 (electronic).

Lachapelle, A., Salomon, J., and Turinici, G. (2010), 'Computation of Mean Field Equilibria in Economics', *Mathematical Models and Methods in Applied Science*, 20, 567–588.

Lapert, M., Tehini, R., Turinici, G., and Sugny, D. (2008), 'Monotonically Convergent Optimal Control Theory of Quantum Systems Under a Nonlinear Interaction with the Control Field', *Physical Review A*, 78(2), 023408.

Lasry, J.-M., and Lions, P.-L. (2006a), 'Jeux à Champ Moyen. I. Le Cas Stationnaire', *Comptes Rendus Mathematique Academic des Sciiences Paris*, 343, 619–625.

Lasry, J.-M., and Lions, P.-L. (2006b), 'Jeux à Champ Moyen. II. Horizon Fini et Contrôle Optimal', *Comptes Rendus Mathematique Academic des Sciiences Paris*, 343, 679–684.

Lasry, J.-M., and Lions, P.-L. (2007), 'Mean Field Games', *Japanese Journal of Mathematics*, 2, 229–260.

Levis, R.J., Menkir, G., and Rabitz, H. (2001), 'Selective Bond Dissociation and Rearrangement with Optimally Tailored, Strong-field Laser Pulses', *Science*, 292, 709–713.

Lions, J.-L. (1971), *Optimal Control of Systems Governed by Partial Differential Equations*, New York: Springer-Verlag.

Maday, Y., Salomon, J., and Turinici, G. (2007), 'Monotonic Parareal Control for Quantum Systems', *SIAM Journal on Numerical Analysis*, 45, 2468–2482.

Maday, Y., and Turinici, G. (2003), 'New Formulations of Monotonically Convergent Quantum Control Algorithms', *Journal of Chemical Physics*, 18, 8191–8196.

Ohtsuki, Y. (2003), 'Non-markovian Effects on Quantum Optimal Control of Dissipative Wave Packet Dynamics', *The Journal of Chemical Physics*, 119, 661–671.

Ohtsuki, Y., and Nakagami, K. (2008), 'Monotonically Convergent Algorithms for Solving Quantum Optimal Control Problems of a Dynamical System Nonlinearly Interacting with a Control', *Physical Review A*, 77, 033–414.

Ohtsuki, Y., Teranishi, Y., Turinici, G., and Rabitz, H. (2007), 'Monotonically Convergent Algorithms for Solving Quantum Optimal Control Problems Described by an Integro-differential Equation of Motion', *Physical Review A*, 75, 033407.

Ohtsuki, Y., Turinici, G., and Rabitz, H. (2004), 'Generalized Monotonically Convergent Algorithms for Solving Quantum Optimal Control Problems', *Journal of Chemical Physics*, 120, 5509–5517 (also published in the Virtual Journal of Ultrafast Science).

Ohtsuki, Y., Zhu, W., and Rabitz, H. (1999), 'Monotonically Convergent Algorithm for Quantum Optimal Control with Dissipation', *Journal of Chemical Physics*, 110, 9825–9832.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2002), *Numerical Recipes in C++* (2nd ed.), Cambridge: Cambridge University Press. (The art of scientific computing, updated for C++).

Rabitz, H., de Vivie-Riedle, R., Motzkus, M., and Kompa, K. (2000), 'Wither the Future of Controlling Quantum Phenomena?', *Science*, 288, 824–828.

Rabitz, H., Hsieh, R., and Rosenthal, C. (2004), 'Quantum Optimally Controlled Transition Landscapes', *Science*, 303, 1998.

Rice, S., and Zhao, M. (2000), *Optical Control of Quantum Dynamics*, New York: Wiley.

Salomon, J., and Carlier, G. (2008), 'A Monotonic Algorithm for the Optimal Control of the Fokker-planck Equation', in *Proceedings of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico.

Salomon, J., Dion, C., and Turinici, G. (2005), 'Optimal Molecular Alignment and Orientation Through Rotational Ladder Climbing', *Journal of Chemical Physics*, 123, 144310.

Schirmer, S., Girardeau, M., and Leahy, J. (2000), 'Efficient Algorithm for Optimal Control of Mixed-state Quantum Systems', *Physical Review A*, 61, 012101.

Sklarz, S.E., and Tannor, D.J. (2002), 'Loading a Bose–Einstein Condensate Onto an Optical Lattice: an Application of Optimal Control Theory to the Nonlinear SchröDinger Equation', *Physical Review A*, 66, 053619.

Sugawara, M. (2003), 'General Formulation of Locally Designed Coherent Control Theory for Quantum Systems', *Journal of Chemical Physics*, 118, 6784–6800.

Sugawara, M., and Fujimura, Y. (1993), 'Control of Quantum Dynamics by a Locally Optimised Laser Field. Application to Ring Puckering Isomerization', *Journal of Chemical Physics*, 100, 5646–5655.

Tannor, D.J., Kazakov, V., and Orlov, V. (1992), 'Control of Photochemical Branching: Novel Procedures for Finding Optimal Pulses and Global Upper Bounds', in NATO ASIB Proc. 299: *Time Dependent Quantum Molecular Dynamics*, eds. J. Broeckhove and L. Lathouwers, New York: Plenum, pp. 347–360.

Tehini, R., and Sugny, D. (2008), *Physical Review A*, 77, 023407.

Warren, W., Rabitz, H., and Dahleh, M. (1993), 'Coherent Control of Quantum Dynamics: The Dream is Alive', *Science*, 259, 1581–1589.

Weinacht, T., Ahn, J., and Bucksbaum, P. (1999), 'Controlling the Shape of a Quantum Wavefunction', *Nature*, 397, 233–235.

Yosida, K. (1995), *Functional Analysis*, Berlin: Springer-Verlag.

Zhu, W., Botina, J., and Rabitz, H. (1998), 'Rapidly Convergent Iteration Methods for Quantum Optimal Control of Population', *Journal of Chemical Physics*, 108, 1953–1963.

Zhu, W., and Rabitz, H. (1998a), 'Uniform Rapidly Convergent Algorithm for Quantum Optimal Control of Objectives with a Positive Semi-definite Hessian Matrix', *Physical Review A*, 58, 4741–4748.

Zhu, W., and Rabitz, H. (1998b), 'A Rapid Monotonically Convergent Iteration Algorithm for Quantum Optimal Control Over the Expectation Value of a Positive Definite Operator', *Journal of Chemical Physics*, 109, 385–391.

Appendix 1: Extension of the monotonic algorithms

In this section, we discuss the relaxation of some of the assumptions concerning either the concavity of (parts of) J or the linearity in Equation (3) (cf. Remark 2.3).

Relaxation of concavity assumptions for norm preserving evolution

In some cases, Equation (3) is endowed with additional properties that enable to relax the hypothesis of concavity of the cost functional J. For instance, in Section 4.1 the L^2 norm of X is preserved. Thus, for any G whose second differential with respect to $X \in L^2$ is bounded (e.g. by M), our algorithm applies: in this case use $G - M \cdot Id$ instead of G (see, e.g. Salomon et al. 2005). The same conclusions also hold for F.

General evolution equation

We consider a general form of the semi-group generator

$$\partial_t X_v + L(t, v(t), X_v(t)) = 0$$
$$X_v(0) = X_0.$$

For a given v, the corresponding adjoint state Y_v is

$$\partial_t Y_v - D_X L^*(t, v(t), X_v(t)) Y_v + \nabla_X F(t, v(t), X_v) = 0$$
$$Y_v(T) = \nabla_X G\big(X_v(T)\big).$$

In the case of the cost functional J defined in (3), the arguments of the proof of Lemma 3.1 apply, and we obtain the following result.

Lemma A.1: *For any $v', v : [0, T] \to E$,*

$$J(v') - J(v) \leq \int_0^T \mathcal{D}(v, v', t) \mathrm{d}t,$$

where

$$\begin{aligned}\mathcal{D}(v, v', t) = {} & F(t, v'(t), X_{v'}) - F(t, v(t), X_{v'}) \\ & + \langle Y_v(t), L(t, v(t), X_v(t)) - L(t, v'(t), X_{v'}(t)) \rangle_{\mathbf{V}} \\ & + \langle Y_v(t), D_X L(t, v(t), X_v(t))(X_{v'}(t) - X_v(t)) \rangle_{\mathbf{V}}.\end{aligned}$$

We note, however, that choosing at time t, $v'(t) = v(t)$ does not ensure in general that $\mathcal{D}(v, v', t)$ is zero; thus the factorisation of the form $\mathcal{D}(v, v', t) = \Delta^{NL}(v, v') \cdot_E (v' - v)$ is not true any more. In particular, we are not sure to be able to find a $v'(t)$ which sets this term negative. Manifestly the reason is that the adjoint is not adapted; we do not want to develop here on how to change the adjoint but we are lead to propose the following procedure: advance in time $v'(t)$ by solving for $v'(t)$ in the relation $\mathcal{D}(v, v', t) = -\theta(v'(t) - v(t))^2$ for as long as possible, say from $t_1 = 0$ to $t_2 \leq T$. Then one sets $v \leftarrow 1_{[0,t_2[} v'(t) + 1_{[t_2,T]} v(t)$, compute a new adjoint Y_v and advance again in time from t_2 to t_3, etc.

Appendix 2: Time discretised case

This section is devoted to the time-discretisation.

Setting

In order to reproduce at the discrete level the computation involved in the monotonic algorithms, one has to define a time discretised version of J and a scheme devoted to numerical resolution of (3)–(4).

Note that our optimisation method does not impose any restrictions thus any scheme with standard numerical properties (consistency, stability, convergence) is compatible with our procedure.

Since we only deal with optimisations problems, we consider arbitrary time-discretisations of the functional (2):

$$J_{\Delta t}(v) = \Delta t \sum_{n=0}^{N-1} F(v_n, x_n) + G(x_N),$$

together with the general numerical scheme

$$x_{n+1} = A_{\Delta t}(v_n) x_n + B_{\Delta t}(v_n), \tag{33}$$

where N is a positive integer, $\Delta t = T/N$ and $v = (v_n)_{n=0...N-1}$. We assume that the functions F and G have the same properties as in Section 2.

Discrete adjoint and factorisation

As in the continuous case, the adjoint operator definition directly follows from the state equation evolution. Given a numerical solver (33), the discrete adjoint operator is defined by

$$y_n = A^*_{\Delta t}(v_n) y_{n+1} + \Delta t \nabla_x F(v_n, x_n)$$
$$y_N = \nabla_x G\big(x_N\big).$$

With this definition, a factorisation similar to Lemma 3.1 can be obtained.

Lemma A.2: *For any $v' = (v'_n)_{n=0...N-1}$, $v = (v_n)_{n=0...N-1}$,*

$$\begin{aligned}J_{\Delta t}(v') - J_{\Delta t}(v) \leq {} & \sum_{n=0}^{N-1} \big\langle y_n, (A_{\Delta t}(v'_{n-1}) - A_{\Delta t}(v_{n-1})) x'_{n-1} \big\rangle_{\mathbf{V}} \\ & + \big\langle y_{n+1}, B_{\Delta t}(v'_n) - B_{\Delta t}(v_n) \big\rangle_{\mathbf{V}} \\ & + \Delta t \big(F(v'_n, x'_n) - F(v_n, x'_n)\big).\end{aligned}$$

By means of this lemma, we obtain a discrete version of monotonicity condition (12).

Depending on the way the functions A, B and F depend on v, the computation of a v'_n satisfying the discrete monotonic condition may requires an inner iterative solver.

In many cases this computation can anyway be parallelised. During an optimisation step, at a given time step n, the terms of the previous sum can be factorised with respect to each component of the vector $v'_n - v_n$ and made negative independently.

The fact that the computation of v'_n requires x'_n makes anyway the time resolution sequential. To solve this problem, some time parallelisations have been designed in the case of quantum control (Maday et al. 2007).