

## CONVERGENCE OF DISCONTINUOUS GALERKIN APPROXIMATIONS OF AN OPTIMAL CONTROL PROBLEM ASSOCIATED TO SEMILINEAR PARABOLIC PDE'S

KONSTANTINOS CHRYSAFINOS<sup>1</sup>

**Abstract.** A discontinuous Galerkin finite element method for an optimal control problem related to semilinear parabolic PDE's is examined. The schemes under consideration are discontinuous in time but conforming in space. Convergence of discrete schemes of arbitrary order is proven. In addition, the convergence of discontinuous Galerkin approximations of the associated optimality system to the solutions of the continuous optimality system is shown. The proof is based on stability estimates at arbitrary time points under minimal regularity assumptions, and a discrete compactness argument for discontinuous Galerkin schemes (see Walkington [*SINUM* (June 2008) (submitted), preprint available at <http://www.math.cmu.edu/~noelw>], Sects. 3, 4).

**Mathematics Subject Classification.** 65M60, 49J20.

Received February 12, 2009. Revised July 15, 2009.

Published online December 16, 2009.

### 1. INTRODUCTION

The optimal control problem considered here is associated to the minimization of the tracking functional

$$J(y, g) = \frac{1}{2} \int_0^T \|y - U\|_{L^2(\Omega)}^2 dt + \frac{\alpha}{2} \int_0^T \|g\|_{L^2(\Omega)}^2 dt \quad (1.1)$$

subject to the constraints,

$$\begin{cases} y_t - \operatorname{div}[A(x)\nabla y] + \phi(y) = f + g & \text{in } (0, T) \times \Omega \\ y = 0 & \text{on } (0, T) \times \Gamma \\ y(0, x) = y_0 & \text{in } \Omega. \end{cases} \quad (1.2)$$

Here,  $\Omega$  denotes a bounded domain in  $\mathbb{R}^2$ , with Lipschitz boundary  $\Gamma$ ,  $y_0$ ,  $f$  denote the initial data and the forcing term respectively,  $g$  denotes the control variable of distributed type,  $U$  is the target function, and  $\alpha$  is a penalty parameter. The nonlinear mapping  $\phi$  satisfies certain continuity and monotonicity properties,

---

*Keywords and phrases.* Discontinuous Galerkin approximations, distributed controls, stability estimates, semi-linear parabolic PDE's.

<sup>1</sup> National Technical University of Athens, Department of Mathematics, Zografou Campus, Athens 15780, Greece.  
chrysafinos@math.ntua.gr

and  $A(x) \in C^1(\bar{\Omega})$  is a symmetric matrix valued function that is uniformly positive definite. The physical meaning of the optimization problem is to seek states  $y$  and controls  $g$  such that  $y$  is as close as possible to the given target  $U$ .

It is worth noting that several problems arise in the analysis of numerical algorithms of optimal control problems constrained to evolutionary PDE's. Solutions of such optimal control problems as well as of their corresponding optimality systems (first order necessary conditions), satisfy low regularity properties. Furthermore, the associated optimality system consists of a state (forward in time) equation and an adjoint (backwards in time) equation which are coupled through an optimality condition, and nonlinear terms (see, *e.g.* [18,21,29,36]). Hence, techniques developed for uncontrolled parabolic problems are not easily applicable. The size of the parameter  $\alpha$  also plays an important role in many interesting applications, since it effectively determines the size of the control  $g$ , and hence the speed of convergence (see also [21] for relevant discussions).

The scope of this work is the analysis of classical discontinuous Galerkin (DG) schemes which are discontinuous in time and conforming in space. First, it is shown that DG schemes of arbitrary order converge to the optimal solution for all  $\alpha > 0$ , for any data  $f \in L^2[0, T; H^{-1}(\Omega)]$ ,  $y_0 \in L^2(\Omega)$  satisfying minimal regularity assumptions, and for any target  $U \in L^2[0, T; L^2(\Omega)]$ . The key ingredient of the proof is the application of a recently developed discrete compactness property (see Walkington [42], Thm. 3.1) for DG schemes of arbitrary order, combined with stability estimates at arbitrary time-points. The dependence upon  $\alpha$  of various constants appearing in these estimates is quantified, and the technique presented here, allows us to avoid any exponential dependence. In addition, it is shown that the DG approximations of the corresponding optimality system converge to the solution of the continuous optimality system under minimal regularity assumptions on data,  $f$ ,  $y_0$ , target  $U$ , and for any choice of  $\alpha > 0$ .

The motivation for using a DG approach stems from its performance in a vast area of problems where the given data satisfy low regularity properties, such as optimal control problems. The main difficulty in handling high-order discrete schemes within the framework of the DG methodology for nonlinear evolutionary PDE's, stems from the lack of control on the discrete time-derivative. Recall that in the continuous case, standard regularity theory implies that under certain assumptions on  $\phi$ , the weak solution  $y$  of (1.2) belongs to  $W(0, T) \equiv L^2[0, T; H_0^1(\Omega)] \cap H^1[0, T; H^{-1}(\Omega)]$  when  $y_0 \in L^2(\Omega)$  and  $f \in L^2[0, T; H^{-1}(\Omega)]$ . Therefore, the nonlinear terms can be treated by using the classical Aubin-Lions compactness Lemma (see [38,45]) which allows to establish strong convergence in an appropriate norm. In the discrete case, the presence of discontinuities imply that discrete time derivative is not integrable and hence this line of argument fails. For the uncontrolled case and for low order DG schemes, *i.e.* for piecewise constants or piecewise linear approximations in time, one may circumvent this difficulty by deriving estimates at arbitrary times *via* estimates at the partition points and at the energy norm, provided that the solution is sufficiently smooth. However this technique is not easily applicable in the optimal control setting due to the lack of regularity, and the nonlinear coupling of the forward/backward in time optimality system.

In this work, we present an analysis of DG schemes of arbitrary order which is suitable for optimal control problems. A synopsis of our work and related results follows.

## 1.1. Synopsis

After introducing the necessary notation in Section 2, we define the continuous optimal control problem and its corresponding optimality system. In Section 3, a key stability estimate at arbitrary time points for the solution of the discrete optimal control problem is obtained. The proof is based on the construction of a suitable polynomial approximation of discrete characteristic functions (developed in [5]) combined with a “boot-strap” argument. A key feature of our stability estimates is that the time-step  $\tau$  can be chosen independent of the size of the spatial parameter  $h$ . These estimates together with the discrete compactness argument of [42] are used to show the existence of the corresponding discrete optimal solution and to prove convergence of discrete schemes of arbitrary order. In Section 4, using a “boot-strap” argument combined with approximation properties of a suitable polynomial interpolant, we establish stability estimates on arbitrary time-points for the adjoint variable. Then, using once more the discrete compactness theorem of [42] we show convergence of the DG approximations

of the associated discrete optimality system to the continuous optimality system. To our best knowledge the proposed technique and results presented here are new.

### 1.2. Related results

Several problems with distributed controls have been studied before analytically in [18,21,29,30,36] (see also references within). Issues related to the analysis of numerical algorithms for optimal control problems constrained to time-dependent problems were studied in [2,8,10,17,19,22,24–28,33,37,40,41,43,44]. In the recent works of [4,31,32,34,35] discontinuous Galerkin schemes were analyzed for distributed optimal control problems constrained to linear parabolic PDE's. In particular, *a posteriori* estimates for DG schemes were studied in [31,32] for distributed control problems related to linear parabolic PDE's, while in [34] an adaptive space-time finite element algorithm is analyzed. *A priori* error estimates for an optimal control problem of distributed type, having states constrained to the heat equation are presented at the energy norm in the recent work of [35]. Finally, in [4] *a priori* error estimates for DG schemes for the tracking problem related to linear parabolic PDE's with non-selfadjoint elliptic part with time dependent coefficients are established.

The literature related to DG schemes for the solution of parabolic equations (without applying controls) is quite extensive (see *e.g.* [39] and references therein). The relation of the DG method to adaptive techniques was studied in [11,12,39]. Results related to finite element approximation of semi-linear and general nonlinear parabolic problems are presented in [1,13–15].

## 2. PRELIMINARIES

### 2.1. Notation

We use standard notation for Hilbert spaces  $L^2(\Omega)$ ,  $H^s(\Omega)$ ,  $0 < s \in \mathbb{R}$ ,  $H_0^1(\Omega) \equiv \{v \in H^1(\Omega) : v|_\Gamma = 0\}$ , related norms and inner products (see *e.g.* [16], Chap. 5). We denote by  $H^{-1}(\Omega)$  the dual of  $H_0^1(\Omega)$  and the corresponding duality pairing by  $\langle \cdot, \cdot \rangle$ . For any Banach space  $X$ , we denote by  $L^p[0, T; X]$ ,  $L^\infty[0, T; X]$  the time-space spaces, endowed with norms,

$$\|v\|_{L^p[0,T;X]} = \left( \int_0^T \|v\|_X^p dt \right)^{\frac{1}{p}}, \quad \|v\|_{L^\infty[0,T;X]} = \text{esssup}_{t \in [0,T]} \|v\|_X.$$

The set of all continuous functions  $v : [0, T] \rightarrow X$ , is denoted by  $C[0, T; X]$ , with norm defined by  $\|v\|_{C[0,T;X]} = \max_{t \in [0,T]} \|v(t)\|_X$ . Finally, we denote by  $H^1[0, T; X]$ ,

$$\|v\|_{H^1[0,T;X]} = \left( \int_0^T \|v\|_X^2 dt \right)^{\frac{1}{2}} + \left( \int_0^T \|v_t\|_X^2 dt \right)^{\frac{1}{2}},$$

and the solution space by  $W(0, T) = L^2[0, T; H_0^1(\Omega)] \cap H^1[0, T; H^{-1}(\Omega)]$  with norm

$$\|v\|_{W(0,T)}^2 = \|v\|_{L^2[0,T;H^1(\Omega)]}^2 + \|v_t\|_{L^2[0,T;H^{-1}(\Omega)]}^2.$$

The bilinear form associated to our operator, is defined by

$$a(y, v) = \int_\Omega A(x) \nabla y \nabla v dx \quad \forall y, v \in H^1(\Omega),$$

and satisfies the standard coercivity and continuity conditions

$$a(y, y) \geq \eta \|y\|_{H^1(\Omega)}^2, \quad a(y, v) \leq C_c \|y\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \quad \forall y, v \in H_0^1(\Omega).$$

A weak formulation of (1.2) is then defined as follows: we seek  $y \in W(0, T)$  such that for a.e.  $t \in (0, T]$ ,

$$\begin{cases} \langle y_t, v \rangle + a(y, v) + \langle \phi(y), v \rangle = \langle f, v \rangle + (g, v) \\ (y(0), v) = (y_0, v), \end{cases} \quad (2.1)$$

for all  $v \in H_0^1(\Omega)$ . A weak formulation suitable for the DG schemes considered here, is to seek  $y \in W(0, T)$  such that

$$(y(T), v(T)) + \int_0^T \left( -\langle y, v_t \rangle + a(y, v) + \langle \phi(y), v \rangle \right) dt = (y_0, v(0)) + \int_0^T \left( \langle f, v \rangle + (g, v) \right) dt, \quad (2.2)$$

for all  $v \in W(0, T)$ . The data satisfy the minimal regularity assumptions which guarantee the existence of a weak solution  $y \in W(0, T)$ , i.e.,

$$f \in L^2[0, T; H^{-1}(\Omega)], \quad y_0 \in L^2(\Omega)$$

while the distributed control will be sought in the space

$$g \in L^2[0, T; L^2(\Omega)].$$

The above choice of the control space significantly simplifies the implementation of the finite element algorithm, since it leads to an algebraic optimality condition. Hence, it avoids the use of spaces of fractional order, or the solution of an extra PDE which typically occur when other norms of  $g$  are included in the functional (see e.g. [21,23]).

For the subsequent analysis and it suffices that the target  $U \in L^2[0, T; L^2(\Omega)]$ . However in most cases  $U$  is actually smoother, since the target typically corresponds to the solution a parabolic PDE, and hence it can be assumed that  $U \in W(0, T)$ . The semi-linear term is required to fulfill the following structural assumptions.

**Assumption 2.1.** *The semi-linear term  $\phi \in C^1(\mathbb{R}; \mathbb{R})$  satisfy the following monotonicity and growth properties. There exists  $C > 0$ , such that*

$$\phi'(s) \geq 0, \quad |\phi(s)| \leq C|s|^p, \quad |\phi'(s)| \leq C|s|^{p-1}, \quad s\phi(s) \geq C|s|^{p+1}, \quad \text{for } 1 < p < 3.$$

We close this preliminary section, by recalling generalized Hölder's and Young's inequalities and the Gagliardo-Nirenberg interpolation inequality (see e.g. [3,16,20,45]) for two dimensional domains, which will be used subsequently.

**Generalized Hölder's inequality.** For any measurable set  $E$ , of any dimension and for  $(1/s_1) + (1/s_2) + (1/s_3) = 1$ ,  $s_i \geq 1$ ,

$$\int_E f_1 f_2 f_3 dE \leq \|f_1\|_{L^{s_1}(E)} \|f_2\|_{L^{s_2}(E)} \|f_3\|_{L^{s_3}(E)}.$$

**Young's inequality.** For any  $a, b \geq 0$ ,  $\delta > 0$ , and  $s_1, s_2 > 1$

$$ab \leq \delta a^{s_1} + C(\delta) b^{s_2}, \quad \text{with } (1/s_1) + (1/s_2) = 1.$$

**Gagliardo-Nirenberg inequality.** Let  $1 \leq q \leq p < \infty$ . Then, for  $s = 1 - (q/p)$ ,

$$\|u\|_{L^p(\Omega)} \leq C \|u\|_{L^q(\Omega)}^{1-s} \|u\|_{H^1(\Omega)}^s, \quad \forall u \in H^1(\Omega).$$

Next, we formulate the optimal control problem and state results regarding the existence of optimal solution(s) and its corresponding optimality system.

## 2.2. The continuous optimal control problem

First, we quote a result regarding the solvability of weak problem (2.2) on the natural energy space under minimal regularity assumptions.

**Theorem 2.2.** *Let  $f \in L^2[0, T; H^{-1}(\Omega)]$ ,  $y_0 \in L^2(\Omega)$ ,  $g \in L^2[0, T; L^2(\Omega)]$ . Then, there exists a unique solution  $y \in W(0, T)$  which satisfies the following energy estimate*

$$\|y\|_{W(0, T)} \leq C \left( \|f\|_{L^2[0, T; H^{-1}(\Omega)]} + \|y_0\|_{L^2(\Omega)} + \|g\|_{L^2[0, T; L^2(\Omega)]} \right).$$

Here  $C > 0$  depends on the continuity and coercivity constants  $C_c, \eta$  and  $\Omega$ .

*Proof.* The proof is standard (see e.g. [8, 16, 45]). □

Next, we state the definition of the set of admissible solutions  $\mathcal{A}_{ad}$  and of the optimal control problem respectively.

**Definition 2.3.** Let  $f \in L^2[0, T; H^{-1}(\Omega)]$ ,  $y_0 \in L^2(\Omega)$ , and  $U \in L^2[0, T; L^2(\Omega)]$ .

- (1) The pair  $(y, g)$  is said to be an admissible element (pair) if  $y \in W(0, T)$ ,  $g \in L^2[0, T; L^2(\Omega)]$  satisfy (2.2). (Note that  $J(y, g) < \infty$ , due to Thm. 2.2).
- (2) The pair  $(y, g) \in \mathcal{A}_{ad}$  is said to be an optimal solution if  $J(y, g) \leq J(w, h) \forall (w, h) \in \mathcal{A}_{ad}$ , when  $\|y - w\|_{W(0, T)} + \|g - h\|_{L^2[0, T; L^2(\Omega)]} \leq \delta$  for  $\delta > 0$  appropriately chosen.

Below, we state the main result concerning the existence of an optimal solution for the minimization of the functional (1.1).

**Theorem 2.4.** *Suppose  $y_0 \in L^2(\Omega)$ ,  $f \in L^2[0, T; H^{-1}(\Omega)]$ ,  $U \in L^2[0, T; L^2(\Omega)]$ . Then, the optimal control problem has solution  $(y, g) \in W(0, T) \times L^2[0, T; L^2(\Omega)]$ .*

*Proof.* Similar to [8, 18, 29]. □

**Remark 2.5.** The solution of optimal control problems having states constrained to nonlinear parabolic PDE's is in general not unique. However, note that if  $f, g, U \in L^2[0, T; L^2(\Omega)]$ ,  $y_0 \in H_0^1(\Omega)$ , and  $\phi$  is a continuous concave increasing function, with  $s\phi(s) \geq 0$  then it is proved that there exists a unique optimal control  $g$  (see e.g. [30], Chap. 3, p. 43). In addition, if  $\phi'$  is continuous, then the corresponding optimality system admits a unique solution. For more results regarding existence and uniqueness we refer the reader to [18].

## 2.3. The continuous optimality system

Suppose now that  $(y, g) \in \mathcal{A}_{ad}$  is an optimal solution in the sense of Definition 2.3. Then, an optimality system corresponding to the optimal control problem of Definition 2.3 can be easily derived based on well known Lagrange multiplier techniques (see e.g. [8, 18, 29, 36]). In particular, given  $f, y_0, U$  satisfying the assumptions of Definition 2.3, we seek a state (primal) variable  $y \in W(0, T)$  and an adjoint (dual) variable  $\mu \in W(0, T)$  such that for a.e.  $t \in (0, T]$ ,

$$\begin{cases} \langle y_t, v \rangle + a(y, v) + \langle \phi(y), v \rangle = \langle f, v \rangle + (g, v) & \text{in } (0, T] \times \Omega \\ (y(0), v) = (y_0, v) & \text{in } \Omega \\ -\langle \mu_t, v \rangle + a(\mu, v) + \langle \phi'(y)\mu, v \rangle = (y - U, v) & \text{in } [0, T) \times \Omega \\ \mu(T) = 0 & \text{in } \Omega \\ \alpha g + \mu = 0 & \text{in } [0, T) \times \Omega, \end{cases}$$

for all  $v \in H_0^1(\Omega)$ . Using the optimality condition, we may replace  $g = -\frac{1}{\alpha}\mu$  from the forward in time equation which leads to the following weak formulation which is suitable for DG approximations. Given  $f, y_0, U$

satisfying the assumptions of Definition 2.3, we seek  $y, \mu \in W(0, T)$  such that

$$\begin{cases} (y(T), v(T)) + \int_0^T \left( -\langle y, v_t \rangle + a(y, v) + \langle \phi(y), v \rangle \right) dt \\ = (y_0, v(0)) + \int_0^T \left( \langle f, v \rangle - (1/\alpha)(\mu, v) \right) dt \\ y(0, x) = y_0, \end{cases} \quad (2.3)$$

$$\begin{cases} \int_0^T \left( \langle \mu, v_t \rangle + a(\mu, v) + \langle \phi'(y)\mu, v \rangle \right) dt = -(\mu(0), v(0)) + \int_0^T (y - U, v) dt \\ \mu(T, x) = 0, \end{cases} \quad (2.4)$$

for all  $v \in W(0, T)$ .

**Remark 2.6.** Note that due to optimality condition we obtain that the control  $g$  is actually smoother, *i.e.*,  $g = -(1/\alpha)\mu \in W(0, T)$ . The later can be used to obtain improved regularity results for the state and adjoint variables *via* a “boot-strap” argument, when additional regularity on  $U, f, y_0$  is available.

### 3. THE DISCRETE OPTIMAL CONTROL PROBLEM

#### 3.1. The semi-discrete (in time) optimal control problem

We first state the definition of the semi-discrete (in time) optimal control problem. We will use the DG method for the discretization of the state equation (2.2) in time. Approximations will be constructed on a partition  $0 = t^0 < t^1 < \dots < t^N = T$  of  $[0, T]$ . On each interval of the form  $(t^{n-1}, t^n]$ , we impose that the semi-discrete (in time) associated functions are polynomials of degree  $k$ , *i.e.*, they belong to the space

$$\mathcal{U} = \{y \in L^2[0, T; H_0^1(\Omega)] : y|_{(t^{n-1}, t^n]} \in \mathcal{P}_k[t^{n-1}, t^n; H_0^1(\Omega)]\}.$$

Here  $\mathcal{P}_k[t^{n-1}, t^n; H_0^1(\Omega)]$  denotes the space of polynomials of degree  $k$  or less having values in  $H_0^1(\Omega)$ . The admissible pairs of the semi-discrete (in time) approximate problem can be defined analogously to the continuous case. Therefore, the semi-discrete (in time) optimal control problem is to seek state  $y \in \mathcal{U}$ , and control  $g \in L^2[0, T; L^2(\Omega)]$  such that the functional  $J(y, g)$  is minimized subject to the constraints,

$$(y^n, v^n) + \int_{t^{n-1}}^{t^n} \left( -\langle y, v_t \rangle + a(y, v) + \langle \phi(y), v \rangle \right) dt = (y^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} \left( \langle f, v \rangle + (g, v) \right) dt \\ \forall v \in \mathcal{P}_k[t^{n-1}, t^n; H_0^1(\Omega)], \quad (3.1)$$

for  $n = 1, \dots, N$ , and  $y^0 \equiv y(0)$ . Here we assume that the functions of  $\mathcal{U}$  are left continuous with right limits and we write  $y^n$  for  $y(t^n) = y(t^n_-)$ ,  $y_+^n$  for  $y(t^n_+)$  while the jump term is denoted by  $[y^n] = y_+^n - y^n$ . A few comments regarding DG approximations follow.

**Remark 3.1.** Note that the continuous weak solution  $y$  satisfies an analogous to (3.1) weak form, when  $v \in \mathcal{P}_k[t^{n-1}, t^n; H_0^1(\Omega)]$  are being used as test functions into (2.2). The control  $g$  needs only to satisfy  $g \in L^2[0, T; L^2(\Omega)]$ , *i.e.* it will be sought in the continuous space while the semi-discrete (in time) state variable satisfies (3.1). However, due to the algebraic structure of optimality condition  $\alpha g + \mu = 0$ , the semi-discrete (in time) approximation of  $g$ , can be implicitly defined and computed as the DG approximation of the adjoint variable (see also [35]).

**Remark 3.2.** Recall, that for the uncontrolled problem the existence of DG approximations can be easily proved when  $k = 0, 1$  (see *e.g.* [15,39]), for linear and semi-linear problems. For  $k > 1$ , existence and uniqueness can be proved under local Lipschitz continuity properties for more general nonlinear problems by using fixed point arguments (see, *e.g.* [1] and references within).

The analysis of the semi-discrete (in time) optimal control problem is similar to the fully-discrete case, if we restrict ourselves to conforming finite element subspaces, *i.e.*, for  $U_h \subset H_0^1(\Omega)$ . The analysis will be presented for the fully-discrete case.

### 3.2. The fully-discrete optimal control problem

The fully-discrete approximations are constructed on a partition  $0 = t^0 < t^1 < \dots < t^N = T$  of  $[0, T]$ . On each interval of the form  $(t^{n-1}, t^n]$  a subspace  $U_h$  of  $H_0^1(\Omega)$  is specified, and it is assumed that  $U_h$  satisfies the classical approximation theory results (see *e.g.* [9]). We seek approximate solutions who belong to the space

$$\mathcal{U}_h = \{y_h \in L^2[0, T; H_0^1(\Omega)] : y_h|_{(t^{n-1}, t^n]} \in \mathcal{P}_k[t^{n-1}, t^n; U_h]\}.$$

Here  $\mathcal{P}_k[t^{n-1}, t^n; U_h]$  denotes the space of polynomials of degree  $k$  or less having values in  $U_h$ . The discretization of the control can be effectively achieved through the discretization of the adjoint variable.

Similar to the semi-discrete (in time) case, by convention, the functions of  $\mathcal{U}_h$  are left continuous with right limits and hence we will subsequently write (abusing the notation)  $y^n$  for  $y_h(t^n) = y_h(t^n_-)$ , and  $y_+^n$  for  $y_h(t^n_+)$ . The jump at  $t^n$  will be occasionally denoted by  $[y^n] = y_+^n - y^n$ .

The discrete optimal control problem is now defined as follows. Under the assumptions of Definition 2.3, we seek state  $y_h \in \mathcal{U}_h$ , and control  $g_h \in L^2[0, T; U_h]$  such that the functional  $J(y_h, g_h)$  is minimized subject to the constraints:

$$(y^n, v^n) + \int_{t^{n-1}}^{t^n} \left( -\langle y_h, v_{ht} \rangle + a(y_h, v_h) + (\phi(y_h), v_h) \right) dt = (y^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} \left( \langle f, v_h \rangle + (g_h, v_h) \right) dt$$

$$\forall v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h], \quad (3.2)$$

for  $n = 1, \dots, N$ . Here  $y^0$  denotes the given initial approximation of  $y(0)$ . Similar, to the semi-discrete case, we note that  $g_h$  needs only to satisfy  $L^2[0, T; L^2(\Omega)]$  regularity. However, motivated by the optimality condition, we discretize the control by using the same discrete space  $\mathcal{U}_h$  with the discrete state variable  $y_h$ .

The proof of existence of optimal solution of the discrete problem and its corresponding discrete optimality system of equations (first order necessary conditions) require stability estimates for the solution of (3.2).

The key ingredient is a stability result at interior time points when  $k$  is arbitrary. For the later, we'll use a suitable polynomial approximation of discrete characteristic functions (see *e.g.* [5]). The main advantage of this approach, within the context of optimal control problems, is that the proof does not need any additional regularity, apart from the one needed to guarantee the existence of a weak solution. In particular, we do not assume that  $u_t \in L^2[0, T; L^2(\Omega)]$  which is frequently used in the literature for DG approximations of parabolic PDE's (even without controls), and it is not suitable in the current optimal control setting.

### 3.3. Quotation of results related to the discrete characteristic function

Note that the derivation of stability estimates at arbitrary times  $t \in [t^{n-1}, t^n]$  can be facilitated by substituting  $v_h = \chi_{[t^{n-1}, t)} y_h$  into the discrete equations. However, this choice is not available since  $\chi_{[t^{n-1}, t)} y_h$  is not a member of  $\mathcal{U}_h$ , unless  $t$  coincides with a partition point. Therefore approximations of such functions need to be constructed. This is done in [5], Section 2.3. For completeness we state the main results. The approximations are constructed on the interval  $[0, \tau)$ , where  $\tau = t^n - t^{n-1}$  and they are invariant under translations.

Let  $t \in (0, \tau)$ . We consider polynomials  $s \in \mathcal{P}_k[0, \tau]$ , and we denote the discrete approximation of  $\chi_{[0, t)} s$  by the polynomial  $\hat{s} \in \{\hat{s} \in \mathcal{P}_k(0, \tau), \hat{s}(0) = s(0)\}$  which satisfies

$$\int_0^\tau \hat{s}q = \int_0^t sq \quad \forall q \in \mathcal{P}_{k-1}[0, \tau].$$

The motivation for the above construction stems from the elementary observation that for  $q = s'$  we obtain  $\int_0^\tau s' \hat{s} = \int_0^t s s' = \frac{1}{2}(s^2(t) - s^2(0))$ .

The construction can be extended to approximations of  $\chi_{[0,t]}v$  for  $v \in \mathcal{P}_k[0, \tau; V]$  where  $V$  is a linear space. The discrete approximation of  $\chi_{[0,t]}v$  in  $\mathcal{P}_k[0, \tau; V]$  is defined by  $\hat{v} = \sum_{i=0}^k \hat{s}_i(t)v_i$  and if  $V$  is a semi-inner product space then,

$$\hat{v}(0) = v(0), \quad \text{and} \quad \int_0^\tau (\hat{v}, w)_V = \int_0^t (v, w)_V \quad \forall w \in \mathcal{P}_{k-1}[0, \tau; V].$$

Finally, we quote the main result from [5]. In the rest of this paper, we denote by  $C_k$ , constants depending only on  $k$ .

**Proposition 3.3.** *Suppose that  $V$  is a (semi-) inner product space. Then the mapping  $\sum_{i=0}^k s_i(t)v_i \rightarrow \sum_{i=0}^k \hat{s}_i(t)v_i$  on  $\mathcal{P}_k[0, \tau; V]$  is continuous in  $\|\cdot\|_{L^2[0, \tau; V]}$ . In particular,*

$$\|\hat{v}\|_{L^2[0, \tau; V]} \leq C_k \|v\|_{L^2[0, \tau; V]}, \quad \|\hat{v} - \chi_{[0,t]}v\|_{L^2[0, \tau; V]} \leq C_k \|v\|_{L^2[0, \tau; V]}$$

where  $C_k$  is a constant depending on  $k$ .

*Proof.* See [5], Lemma 2.4. □

**Remark 3.4.** Combining the above estimate with standard scaling arguments and the finite dimensionality of  $\mathcal{P}_k[0, \tau]$  we also obtain an estimate of the form

$$\|\hat{v}\|_{L^\infty[0, \tau; L^2(\Omega)]} \leq C_k \|v\|_{L^\infty[0, \tau; L^2(\Omega)]}.$$

For various extensions of these results we refer the reader to [6].

**Remark 3.5.** The estimates of Proposition 3.3 hold when  $V = H_0^1(\Omega)$  as well as when  $V$  is replaced by conforming finite element subspaces  $U_h \subset H_0^1(\Omega)$ .

### 3.4. Stability estimates

Now we are ready to prove stability estimates for the discrete optimal control problem under minimal regularity assumptions, which are needed in order to obtain the existence of a discrete optimal solution and its convergence to the optimal solution.

**Lemma 3.6.** *Suppose that  $y_0 \in L^2(\Omega)$ ,  $U \in L^2[0, T; L^2(\Omega)]$ ,  $f \in L^2[0, T; H^{-1}(\Omega)]$  are given functions, and let  $\phi$  satisfy Assumption 2.1. If  $(y_h, g_h) \in \mathcal{U}_h \times L^2[0, T; U_h]$  denotes a solution of the discrete optimal control problem, then*

$$\int_0^T \|y_h - U\|_{L^2(\Omega)}^2 dt + (\alpha/2) \int_0^T \|g_h\|_{L^2(\Omega)}^2 dt \leq C \left( \|y^0\|_{L^2(\Omega)}^2 + (1/\eta) \int_0^T \|f\|_{H^{-1}(\Omega)}^2 dt + \int_0^T \|U\|_{L^2(\Omega)}^2 dt \right) \equiv C_{st}$$

where  $C$  is a constant depending only on  $\Omega$ . In addition, for all  $n = 1, \dots, N$

$$\|y^n\|_{L^2(\Omega)}^2 + \sum_{i=0}^{n-1} \|[y^i]\|_{L^2(\Omega)}^2 + \int_0^{t^n} \left( \eta \|y_h\|_{H^1(\Omega)}^2 + \|y_h\|_{L^{p+1}(\Omega)}^{p+1} \right) dt \leq D_{yst},$$

with  $D_{yst} \equiv C_{st} \max\{1, 1/\alpha\}$ . Let  $\tau \equiv \max_{i=1, \dots, n} \tau_i$ , with  $\tau_i = t^i - t^{i-1}$ . If  $\tau \leq \min\{(1/8D_{yst}^{(p-1)/2} C_k)^{2/(3-p)}, (\alpha/8)\}$ , then

$$\|y_h\|_{L^\infty[0, T; L^2(\Omega)]}^2 \leq CD_{yst}$$

where  $C$  depends on  $(C_c/\eta), C_k$  and  $\Omega$  but not on  $\alpha, \tau, h$ . Here  $C_c, \eta$  denote the continuity and coercivity constants of the bilinear form  $a(\cdot, \cdot)$ .



*Proof.* For the first estimate note that  $(\tilde{y}_h, 0)$  is an admissible pair for the discrete problem, and hence  $J(y_h, g_h) \leq J(\tilde{y}_h, 0) \leq (1/2) \int_0^T \|\tilde{y}_h - U\|_{L^2(\Omega)}^2 \leq C_{st}$ , where  $C_{st}$  is a constant independent of  $\alpha$ . The estimate on  $\|\tilde{y}_h\|_{L^2[0,T;L^2(\Omega)]}$  can be easily derived (see e.g. [5], Sect. 2) since it corresponds to the stability estimate without control.

Setting  $v_h = y_h$  into (3.2) and using the monotonicity of  $\phi$  and Young's inequalities, we easily derive

$$\begin{aligned} (1/2)\|y^n\|_{L^2(\Omega)}^2 + (1/2)\|[y^{n-1}]\|_{L^2(\Omega)}^2 - (1/2)\|y^{n-1}\|_{L^2(\Omega)}^2 + \int_{t^{n-1}}^{t^n} \left( \eta \|y_h\|_{H^1(\Omega)}^2 + \|y_h\|_{L^{p+1}(\Omega)}^{p+1} \right) dt \\ \leq (C/\eta) \int_{t^{n-1}}^{t^n} \|f\|_{H^{-1}(\Omega)}^2 dt + \alpha \int_{t^{n-1}}^{t^n} \|g_h\|_{L^2(\Omega)}^2 dt + (C/\alpha) \int_{t^{n-1}}^{t^n} \|y_h\|_{L^2(\Omega)}^2 dt. \end{aligned} \quad (3.3)$$

Summing the resulting inequalities from  $i = 1$  to  $n$ , and dropping positive terms on the left we obtain the estimate at partition points by using the previous bounds on  $\alpha \int_0^T \|g_h\|_{L^2(\Omega)}^2 dt$ ,  $\int_0^T \|y_h\|_{L^2(\Omega)}^2 dt$ . The estimate at the energy norm follows upon summation from 1 to  $N$ . It remains to obtain a bound at arbitrary time-points. To achieve this, we will use the approximation of the discrete characteristic. Fix  $t \in [t^{n-1}, t^n]$  and set  $v_h = \hat{y}_h$  into (3.2), where  $\hat{y}_h$  denotes the approximation of  $\chi_{[t^{n-1}, t]} y_h$  defined as in Proposition 3.3. Then, using the definition of  $\hat{y}_h$ , we obtain,

$$\begin{aligned} (1/2)\|y_h(t)\|_{L^2(\Omega)}^2 + (1/2)\|[y^{n-1}]\|_{L^2(\Omega)}^2 - (1/2)\|y^{n-1}\|_{L^2(\Omega)}^2 + \int_{t^{n-1}}^{t^n} \langle \phi(y_h), \hat{y}_h \rangle dt \\ \leq \int_{t^{n-1}}^{t^n} |a(y_h, \hat{y}_h)| dt + \int_{t^{n-1}}^{t^n} |\langle f, \hat{y}_h \rangle| dt + \int_{t^{n-1}}^{t^n} |(g_h, \hat{y}_h)| dt. \end{aligned}$$

Using Proposition 3.3, we may bound  $\hat{y}_h$  in terms of  $y_h$  in various norms. In particular, using Young's inequalities with appropriately chosen  $\delta > 0$ ,

$$\begin{aligned} \int_{t^{n-1}}^{t^n} |a(y_h, \hat{y}_h)| dt &\leq C_k C_c \int_{t^{n-1}}^{t^n} \|y_h\|_{H^1(\Omega)}^2 dt \\ \int_{t^{n-1}}^{t^n} |\langle f, \hat{y}_h \rangle| dt &\leq (C_k/\eta) \int_{t^{n-1}}^{t^n} \|f\|_{H^{-1}(\Omega)}^2 dt + \eta \int_{t^{n-1}}^{t^n} \|y_h\|_{H^1(\Omega)}^2 dt \\ \int_{t^{n-1}}^{t^n} |(g_h, \hat{y}_h)| dt &\leq \alpha \int_{t^{n-1}}^{t^n} \|g_h\|_{L^2(\Omega)}^2 dt + (C_k/\alpha) \int_{t^{n-1}}^{t^n} \|y_h\|_{L^2(\Omega)}^2 dt. \end{aligned}$$

Therefore, collecting the above inequalities and using standard algebra, we obtain

$$\begin{aligned} (1/2)\|y_h(t)\|_{L^2(\Omega)}^2 + (1/2)\|[y^{n-1}]\|_{L^2(\Omega)}^2 - (1/2)\|y^{n-1}\|_{L^2(\Omega)}^2 + \int_{t^{n-1}}^{t^n} \langle \phi(y_h), \hat{y}_h \rangle dt \\ \leq C_k \int_{t^{n-1}}^{t^n} \left( \|f\|_{H^{-1}(\Omega)}^2 + (C_c + \eta) \|y_h\|_{H^1(\Omega)}^2 + \alpha \|g_h\|_{L^2(\Omega)}^2 + (1/\alpha) \|y_h\|_{L^2(\Omega)}^2 \right) dt. \end{aligned}$$

For the last term note that  $(1/\alpha) \int_{t^{n-1}}^{t^n} \|y_h\|_{L^2(\Omega)}^2 dt \leq (\tau_n/\alpha) \|y_h\|_{L^\infty[t^{n-1}, t^n; L^2(\Omega)]}^2$ . It remains to bound the semi-linear term. For this purpose, note the growth condition and Young's inequality with  $s_1 = (p+1)/p$ ,

$s_2 = p + 1$ , imply,

$$\begin{aligned} \int_{t^{n-1}}^{t^n} \langle \phi(y_h), \hat{y}_h \rangle dt &\leq C \int_{t^{n-1}}^{t^n} \int_{\Omega} |y_h|^p |\hat{y}_h| dt \\ &\leq \int_{t^{n-1}}^{t^n} \|y_h\|_{L^{p+1}(\Omega)}^{p+1} dt + C \int_{t^{n-1}}^{t^n} \|\hat{y}_h\|_{L^{p+1}(\Omega)}^{p+1} dt. \end{aligned}$$

For the last term on the right hand side, the Gagliardo-Nirenberg interpolation inequality states that

$$\|\hat{y}_h\|_{L^{p+1}(\Omega)} \leq C \|\hat{y}_h\|_{L^2(\Omega)}^{1-s} \|\hat{y}_h\|_{H^1(\Omega)}^s,$$

with  $s = 1 - (2/p + 1) = (p - 1)/(p + 1)$ . Hence,  $1 - s = 2/(p + 1)$  and

$$\begin{aligned} \int_{t^{n-1}}^{t^n} \|\hat{y}_h\|_{L^{p+1}(\Omega)}^{p+1} dt &\leq C \int_{t^{n-1}}^{t^n} \|\hat{y}_h\|_{L^2(\Omega)}^2 \|\hat{y}_h\|_{H^1(\Omega)}^{p-1} dt \\ &\leq C \|\hat{y}_h\|_{L^\infty[t^{n-1}, t^n; L^2(\Omega)]}^2 \int_{t^{n-1}}^{t^n} \|\hat{y}_h\|_{H^1(\Omega)}^{p-1} dt \\ &\leq C \|\hat{y}_h\|_{L^\infty[t^{n-1}, t^n; L^2(\Omega)]}^2 \left( \int_{t^{n-1}}^{t^n} 1 dt \right)^{(3-p)/2} \left( \int_{t^{n-1}}^{t^n} \|\hat{y}_h\|_{H^1(\Omega)}^2 dt \right)^{(p-1)/2} \\ &\leq C_k \|y_h\|_{L^\infty[t^{n-1}, t^n; L^2(\Omega)]}^2 \tau_n^{(3-p)/2} D_{yst}^{(p-1)/2}. \end{aligned}$$

Here we have used the generalized Hölder inequality with  $s_1 = 2/(p - 1) > 1$ ,  $s_2 = 2/(3 - p) > 1$  (recall  $1 < p < 3$ ), Proposition 3.3 to bound  $\hat{y}_h$  in terms of  $y_h$ , and the stability estimates at the energy norm. Hence, selecting  $t$  such that  $\|y_h(t)\|_{L^2(\Omega)}^2 = \sup_{s \in (t^{n-1}, t^n]} \|y_h(s)\|_{L^2(\Omega)}^2$  and choosing  $\tau_n > 0$  in way to satisfy  $C_k \tau_n^{(3-p)/2} D_{yst}^{(p-1)/2} \leq (1/8)$  and  $(\tau_n/\alpha) \leq (1/8)$ , *i.e.*, for  $\tau_n \leq \min\{(1/8 D_{yst}^{(p-1)/2} C_k)^{2/3-p}, (\alpha/8)\}$  we obtain,

$$\begin{aligned} (1/4) \|y_h\|_{L^\infty[t^{n-1}, t^n; L^2(\Omega)]}^2 &\leq \|y^{n-1}\|_{L^2(\Omega)}^2 \\ &\quad + C_k \int_{t^{n-1}}^{t^n} \left( \|f\|_{H^{-1}(\Omega)}^2 + (C_c + \eta) \|y_h\|_{H^1(\Omega)}^2 + \|y_h\|_{L^{p+1}(\Omega)}^{p+1} + \alpha \|g_h\|_{L^2(\Omega)}^2 \right) dt. \end{aligned}$$

The estimate now follows by using the previously derived estimates at the energy norm and at partition points.  $\square$

In order to handle the nonlinear terms within the DG setting some form of strong convergence needs to be established. For the later we will employ the following compactness argument of Walkington (see [42], Thm. 3.1).

### 3.5. The discrete compactness theorem

The problems considered in [42], involve the numerical approximations of solutions  $u : [0, T] \rightarrow U$  of general evolution equations of the form

$$u_t + A(u) = f(u) \quad u(0) = u_0, \quad (3.4)$$

where  $U$  is a Banach space and each term of the equation takes values in  $U^*$ . Here, both  $A(u) = A(t, u)$  and  $f(u) = f(t, u)$  may depend upon  $t$  and are allowed to be nonlinear. We assume that  $U \subset H \subset U^*$  (with continuous embeddings) form the standard evolution triple, *i.e.*, the pivot space  $H$  is a Hilbert space. The numerical schemes approximate the weak form of (3.4), *i.e.*,

$$\langle u_t, v \rangle + a(u, v) = \langle f(u), v \rangle \quad \forall v \in U, \quad (3.5)$$

where  $a : U \times U \rightarrow \mathbb{R}$  is defined by  $a(u, v) = (A(u), v)$ . Recall, that for each subspace  $U_h \subset U$  and partition  $0 = t^0 < t^1 < \dots < t^N = T$  of  $[0, T]$  the DG scheme constructs a function in  $\mathcal{P}_k[t^{n-1}, t^n; U_h]$  on each  $(t^{n-1}, t^n)$ , which satisfies for  $n = 1, \dots, N$  and for all  $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$ ,

$$\int_{t^{n-1}}^{t^n} \left( (u_{ht}, v_h) + a(u_h, v_h) \right) dt + (u_+^{n-1} - u^{n-1}, v_+^{n-1}) = \int_{t^{n-1}}^{t^n} (f(u_h), v_h) dt. \tag{3.6}$$

Here,  $u^0$  is a given approximation of  $u_0$ . Set  $F(u) \equiv f(u) - A(u)$ . Then the following theorem [42], Theorem 3.1, establishes the compactness property of the discrete approximation.

**Theorem 3.7.** *Let  $H$  be a Hilbert space,  $U$  be a Banach space and  $U \subset H \subset U^*$  be dense and compact embeddings. Fix integer  $k \geq 0$  and  $1 \leq p, q < \infty$ . Let  $h > 0$  be the mesh parameter, and let  $\{t^i\}_{i=0}^N$  denote a uniform partition of  $[0, T]$ . Assume that*

- (1)  $u_h \in \{u_h \in L^p[0, T; U] \mid u_h|_{(t^{n-1}, t^n)} \in \mathcal{P}_k[t^{n-1}, t^n; U_h]\}$  and on each interval,

$$\int_{t^{n-1}}^{t^n} (u_{ht}, v_h) dt + (u_+^{n-1} - u^{n-1}, v_+^{n-1}) = \int_{t^{n-1}}^{t^n} (F(u_h), v_h) dt$$

holds for every  $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$ .

- (2)  $\{u_h\}_{h>0}$  is bounded in  $L^p[0, T; U]$  and  $\{\|F(u_h)\|_{L^q[0, T; U^*]}\}_{h>0}$  is also bounded.

Then,

- (1) If  $p > 1$  then  $\{u_h\}_{h>0}$  is compact in  $L^r[0, T; H]$  for  $1 \leq r < 2p$ .
- (2) If  $1 \leq (1/p) + (1/q) < 2$ , and  $\sum_{i=1}^N \|u_h\|_H^2 < C$  is bounded independent of  $h$ , then  $\{u_h\}_{h>0}$  is compact in  $L^r[0, T; H]$  for  $1 \leq r < 2/((1/p) + (1/q) - 1)$ .

*Proof.* See [42]. □

### 3.6. Convergence of the discrete optimal control problem

Once, we have shown the stability estimates in  $L^\infty[0, T; L^2(\Omega)] \cap L^2[0, T; H_0^1(\Omega)]$ , we may apply the discrete compactness Theorem 3.7, to obtain the existence of an optimal discrete solution, and its convergence to the continuous optimal solution.

**Theorem 3.8.** *Suppose that  $f \in L^2[0, T; H^{-1}(\Omega)]$ ,  $y_0 \in L^2(\Omega)$ ,  $U \in L^2[0, T; L^2(\Omega)]$ . Let  $h > 0$  and a uniform partition  $0 = t^0 < t^1 < \dots < t^N = T$  of  $[0, T]$  fixed, with  $\tau = \max_{i=1, \dots, N} \tau_i$ ,  $\tau_i = t^i - t^{i-1}$ , satisfying the assumptions of Lemma 3.6. Then,*

- (1) For  $\alpha > 0$ , there exist  $y_h \in \mathcal{U}_h$  and  $g_h \in L^2[0, T; L^2(\Omega)]$  such that the pair  $(y_h, g_h)$  satisfies the discrete equation (3.2) and the functional  $J(y_h, g_h)$  is minimized.
- (2) For  $\alpha > 0$ ,  $(y_h, g_h)$  converges to the solution  $(y, g)$  of the continuous optimal control problem as  $h, \tau \rightarrow 0$ .

*Proof.* We present the proof for  $3/2 \leq p < 3$ . The case  $1 < p \leq 3/2$ , can be treated similarly.

(1). Let  $h > 0$  and  $0 = t^0 < t^1, \dots, t^N = T$  be a fixed uniform partition of  $[0, T]$  with  $\tau, h$  satisfying the assumptions of Lemma 3.6. The discrete admissible set

$$\mathcal{A}_{ad}^d \equiv \{(y_h, g_h) \in \mathcal{U}_h \times L^2[0, T; U_h] \text{ such that (3.2) is satisfied}\}$$

is not empty since  $(\tilde{y}_h, 0)$  belongs in it. Now let  $(y_{hm}, g_{hm}) \in \mathcal{A}_{ad}^d$  be minimizing sequence where  $y_{hm}$  denotes the corresponding solution of (3.2) with right hand side  $g_{hm}$ . In fact, we may choose the minimizing sequence such that  $J(y_{hm}, g_{hm}) \leq M$ , with  $M$  be the value of the functional for an admissible element, say  $(\tilde{y}_h, 0)$ . Hence, the stability estimates (independent of  $h, \tau$ ) imply that (passing to a subsequence, if necessary), as  $m \rightarrow \infty$ ,

$$\begin{aligned} y_{hm} &\rightarrow y_h && \text{weakly in } L^2[0, T; H^1(\Omega)], && y_{hm} &\rightarrow y_h && \text{weakly-* in } L^\infty[0, T; L^2(\Omega)], \\ g_{hm} &\rightarrow g_h && \text{weakly in } L^2[0, T; L^2(\Omega)]. \end{aligned}$$

The proof now follows by using standard arguments and the finite dimensionality of the subspaces. We may pass to the limit to show that  $(y_h, g_h) \in \mathcal{A}_{ad}^d$  satisfy the discrete equation (3.2) (see also part (2)). The weak lower semi-continuity of the functional finishes the proof.

(2). Recall that  $y_h$  is bounded in  $L^\infty[0, T; L^2(\Omega)] \cap L^2[0, T; H^1(\Omega)]$  by constants independent of  $\tau, h$ , and similarly  $g_h$  is bounded in  $L^2[0, T; L^2(\Omega)]$ . Hence, we may extract subsequences (still denoted by  $y_h, g_h$ ), converging weakly to  $y, g$  respectively in the following sense,

$$y_h \rightharpoonup y \quad \text{weakly in } L^2[0, T; H^1(\Omega)], \quad y_h \rightharpoonup^* y \quad \text{weakly-}^* \text{ in } L^\infty[0, T; L^2(\Omega)],$$

$$g_h \rightharpoonup g \quad \text{weakly in } L^2[0, T; L^2(\Omega)].$$

Using the discrete compactness Theorem 3.7 we will prove the strong convergence of  $y_h$  to  $y$  in  $L^2[0, T; L^2(\Omega)]$ . To verify the assumptions of Theorem 3.7 set  $U = H_0^1(\Omega)$ ,  $H = L^2(\Omega)$  and define,  $\langle F(y), v \rangle = -a(y, v) - \langle \phi(y), v \rangle + (g, v) + \langle f, v \rangle$ ,  $\forall y, v \in H_0^1(\Omega)$ . It is evident by the stability Lemma 3.6 and in particular by the estimates on  $y_h$  in  $L^2[0, T; H^1(\Omega)]$ ,  $L^\infty[0, T; L^2(\Omega)]$ , on  $g_h$  in  $L^2[0, T; L^2(\Omega)]$ , and the assumptions on the semi-linear term, that  $\{\|y_h\|_{L^2[0, T; H^1(\Omega)]}\}_{h>0}$ ,  $\{\|F(y_h)\|_{L^{4/3}[0, T; H^{-1}(\Omega)]}\}_{h>0}$  remain bounded independent of  $h, \tau$ . Indeed, to bound the later term, we only need to consider the semi-linear term. Let  $v \in L^4[0, T; H_0^1(\Omega)]$ . Using Hölder's inequalities, the embedding  $H^1(\Omega) \subset L^4(\Omega)$  and the Gagliardo-Nirenberg interpolation inequality (note that  $2 \leq \frac{4p}{3}$  when  $3/2 \leq p$ )

$$\|y_h\|_{L^{4p/3}(\Omega)} \leq C \|y_h\|_{L^2(\Omega)}^{1-s} \|y_h\|_{H^1(\Omega)}^s,$$

with  $s = 1 - \frac{2}{(4p/3)} \equiv \frac{2p-3}{2p}$ ,  $1-s = 3/2p$ , we obtain

$$\begin{aligned} \int_0^T \int_\Omega |y_h|^p v_h \, dx dt &\leq C \int_0^T \|y_h\|_{L^{4p/3}(\Omega)}^p \|v\|_{L^4(\Omega)} \, dt \\ &\leq C \left( \int_0^T \|y_h\|_{L^{4p/3}(\Omega)}^{4p/3} \, dt \right)^{3/4} \left( \int_0^T \|v\|_{H^1(\Omega)}^4 \, dt \right)^{1/4} \\ &\leq C \left( \int_0^T \|y_h\|_{L^2(\Omega)}^2 \|y_h\|_{H^1(\Omega)}^{\frac{2}{3} \times (2p-3)} \, dt \right)^{3/4} \|v\|_{L^4[0, T; H_0^1(\Omega)]} \leq C, \end{aligned}$$

where at the last step we have used the stability bounds for  $y_h$  and the fact that  $\frac{2}{3} \times (2p-3) < 2$ , for  $1 < p < 3$ . Hence, Theorem 3.7 is applicable, with  $p = 2$ ,  $q = 4/3$ ,  $r = 2$  and the strong convergence of  $y_h$  to  $y$  is proven in  $L^2[0, T; L^2(\Omega)]$  norm. It remains to show that the solution  $(y_h, g_h) \in \mathcal{A}_{ad}^d$  of (3.2) converges to the solution  $(y, g)$  of (2.2). Then, the weak lower semi-continuity of the functional finishes the proof. Suppose now that we choose  $v_h \in C[0, T; U_h] \cap \mathcal{U}_h$  with  $v_h(T) = 0$ . Then, equation (3.2) takes the form,

$$\int_0^T \left( -\langle y_h, v_{ht} \rangle + a(y_h, v_h) + \langle \phi(y_h), v_h \rangle \right) dt = \int_0^T \left( \langle f, v_h \rangle + (g_h, v_h) \right) dt + (y^0, v_h(0)).$$

Recall, that there exists a subsequence (still denoted by  $y_h, g_h$ ) such that  $y_h \rightharpoonup y$  weakly in  $L^2[0, T; H^1(\Omega)]$ ,  $g_h \rightharpoonup g$  weakly in  $L^2[0, T; L^2(\Omega)]$  and  $y_h \rightarrow y$  strongly in  $L^2[0, T; L^2(\Omega)]$ . Hence, we may pass the limit term by term into the above equation to get equation (2.2). A standard density argument completes the proof.  $\square$

#### 4. THE DISCRETE OPTIMALITY SYSTEM

In the last section, we proved convergence of the solutions of the discrete optimality system to the solutions of the continuous optimality system. The fully-discrete optimality system is defined as follows: we seek  $y_h, \mu_h \in \mathcal{U}_h$

such that for  $n = 1, \dots, N$  and for every  $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$ ,

$$(y^n, v^n) + \int_{t^{n-1}}^{t^n} \left( -\langle y_h, v_{ht} \rangle + a(y_h, v_h) + (\phi(y_h), v_h) \right) dt = (y^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} \left( \langle f, v_h \rangle + (g_h, v_h) \right) dt, \tag{4.1}$$

$$-(\mu_+^n, v^n) + \int_{t^{n-1}}^{t^n} \left( \langle \mu_h, v_{ht} \rangle + a(v_h, \mu_h) + (\phi'(y_h)\mu_h, v_h) \right) dt = -(\mu_+^{n-1}, v_+^{n-1}) + \int_{t^{n-1}}^{t^n} (y_h - U, v_h) dt, \tag{4.2}$$

and

$$\int_0^T (\alpha g_h + \mu_h, v_h) dt = 0 \quad \forall v_h \in L^2[0, T; U_h]. \tag{4.3}$$

Here,  $y^0, \mu_+^N = 0, f, U$  are given data, and  $y^0$  denotes an approximation of  $y(0)$ .

**Remark 4.1.** The existence of the discrete optimality system can be proved similar to the linear case (see e.g. [35]), by using the previously developed stability estimates. Note that the optimality condition (4.3) is equivalent to  $\int_{t^{n-1}}^{t^n} (\alpha g_h + \mu_h, v_h) = 0$  for all  $v_h \in \mathcal{P}_k[t^{n-1}, t^n; U_h]$ , and  $n = 1, \dots, N$  and hence we may replace the control function from equation (4.1), similar to the continuous case.

A “boot-strap” argument will be applied in order to derive estimates on the adjoint variable at arbitrary time points. For this purpose, an exponential interpolant of  $e^{-\lambda(t^n-t)}\mu_h, \lambda > 0$ , needs to be constructed.

#### 4.1. An exponential interpolant

An  $L^\infty[0, T; L^2(\Omega)]$  bound for the adjoint variable will be obtained, by using the following polynomial interpolant.

**Definition 4.1.** Let  $V$  be a linear space, and  $\lambda > 0$  be given. If  $v = \sum_{i=0}^k r_i(t)v_i \in \mathcal{P}_k[t^{n-1}, t^n; V]$ , with  $r_i \in \mathcal{P}_k[t^{n-1}, t^n], v_i \in V$ , we define the exponential interpolant of  $v$  by

$$\bar{v} = \sum_{i=0}^k \bar{r}_i v_i$$

where  $\bar{r}_i \in \mathcal{P}_k[t^{n-1}, t^n]$  is the approximation of  $r_i(t)e^{-\lambda(t^n-t)}$  satisfying  $\bar{r}_i(t^n) = r_i(t^n)$  and

$$\int_{t^{n-1}}^{t^n} \bar{r}_i(t)q(t)dt = \int_{t^{n-1}}^{t^n} r_i(t)q(t)e^{-\lambda(t^n-t)}dt, \quad q \in \mathcal{P}_{k-1}[t^{n-1}, t^n].$$

The above construction is the analogue of [7], Definition 3.3, suitably modified for the adjoint equation (4.2) which is posed backwards in time. In particular, we use the extra degree of freedom to match the interpolant at the end point of each time interval  $(t^{n-1}, t^n]$  instead of the initial point which is used when dealing with the forward in time problem. An analogue of [7], Lemma 3.4, can be proved using exactly the same arguments (see also [5], Lems. 2.3 and 2.4).

**Lemma 4.2.** Let  $V$  and  $Q$  be linear spaces and  $v \rightarrow \bar{v}$  be the map constructed in Definition 4.1, with parameter  $\lambda > 0$ . If  $L(\cdot, \cdot) : V \times Q \rightarrow \mathbb{R}$  is a bilinear mapping and  $v \in \mathcal{P}_k[t^{n-1}, t^n; V]$ , then

$$\int_{t^{n-1}}^{t^n} L(v(t), q(t))dt = \int_{t^{n-1}}^{t^n} L(v(t), q(t))e^{-\lambda(t^n-t)}dt, \quad q \in \mathcal{P}_{k-1}[t^{n-1}, t^n; Q].$$

If  $(\cdot, \cdot)_V$  is a (semi-) inner-product on  $V$ , then there exists a constant  $C_k$  depending on  $k$  (and not on  $\lambda$ ), such that

$$\|v - \bar{v}\|_{L^2[t^{n-1}, t^n; V]} \leq C_k \lambda (t^n - t^{n-1}) \|v\|_{L^2[t^{n-1}, t^n; V]},$$

and in particular,

$$\|\bar{v}\|_{L^2[t^{n-1}, t^n; V]} \leq C_k \|v\|_{L^2[t^{n-1}, t^n; V]}.$$

*Proof (sketch).* The key step is to show that

$$\int_{t^{n-1}}^{t^n} (r_i - \bar{r}_i)^2 dt \leq C_k (1 - e^{-\lambda(t^n - t^{n-1})})^2 \int_{t^{n-1}}^{t^n} r_i^2 dt.$$

The rest of the proof follows by standard calculations as in [5], Lemma 2.4. For this purpose, note that since  $r_i(t^n) = \bar{r}_i(t^n)$ , there exists  $p_i \in \mathcal{P}_{k-1}[t^{n-1}, t^n]$  such that  $r_i - \bar{r}_i = (t^n - t)p_i$ . The last relation, and the definition of the interpolant imply that for all  $q \in \mathcal{P}_{k-1}[t^{n-1}, t^n]$ ,

$$\int_{t^{n-1}}^{t^n} (t^n - t)p_i q dt = \int_{t^{n-1}}^{t^n} (r_i - \bar{r}_i)q dt = \int_{t^{n-1}}^{t^n} (1 - e^{-\lambda(t^n - t)})r_i q dt.$$

Setting  $q = p_i$ , and using Hölder's inequality, we obtain

$$\int_{t^{n-1}}^{t^n} (t^n - t)p_i^2 dt \leq (1 - e^{-\lambda(t^n - t^{n-1})}) \|r_i\|_{L^2[t^{n-1}, t^n]} \|p_i\|_{L^2[t^{n-1}, t^n]}.$$

The equivalence of norms in  $\mathcal{P}_k[t^{n-1}, t^n]$ , and the last inequality show that

$$C_k (t^n - t^{n-1}) \int_{t^{n-1}}^{t^n} p_i^2 dt \leq (1 - e^{-\lambda(t^n - t^{n-1})}) \|r_i\|_{L^2[t^{n-1}, t^n]} \|p_i\|_{L^2[t^{n-1}, t^n]},$$

or equivalently,

$$C_k (t^n - t^{n-1}) \|p_i\|_{L^2[t^{n-1}, t^n]} dt \leq (1 - e^{-\lambda(t^n - t^{n-1})}) \|r_i\|_{L^2[t^{n-1}, t^n]}.$$

The desired bound follows by the norm equivalence in  $\mathcal{P}_k[t^{n-1}, t^n]$ .  $\square$

## 4.2. Convergence of the discrete optimality system

Finally, we establish stability estimates under minimal regularity assumptions for the adjoint variable. These estimates combined with the discrete compactness Theorem 3.7 will be used to establish convergence of the discrete optimality system to continuous one.

**Lemma 4.3.** *Suppose that  $y_0 \in L^2(\Omega)$ ,  $U \in L^2[0, T; L^2(\Omega)]$ ,  $f \in L^2[0, T; H^{-1}(\Omega)]$  are given functions, let  $\phi$  satisfy the growth condition Assumption 2.1. If  $(y_h, g_h)$  denote the discrete optimal solution and  $(y_h, \mu_h, g_h)$  satisfy (4.1)–(4.3) then*

$$\|\mu_+^0\|_{L^2(\Omega)}^2 + \sum_{i=1}^N \|\mu^i\|_{L^2(\Omega)}^2 + \eta \int_0^T \|\mu_h\|_{H^1(\Omega)}^2 dt \leq C_{st} \alpha^{1/2}$$

and for  $n = 1, \dots, N$

$$\|\mu_+^{n-1}\|_{L^2(\Omega)}^2 \leq C_{st} \alpha^{1/2},$$

where  $C_{st}$  is defined in Lemma 3.6. Let  $\tau \equiv \max_{i=1,\dots,n} \tau_i$ , with  $\tau_i = t^i - t^{i-1}$  satisfy the assumption of Lemma 3.6. Then

$$\|\mu_h\|_{L^\infty[0,T;L^2(\Omega)]}^2 \leq CC_{st}\alpha^{1/2} \left(1 + D_{yst}^{(p-1)/2}\right) \equiv D_{\mu st},$$

where  $C$  does not depend on  $\alpha$ ,  $\tau$ ,  $h$ , but only on  $C_c/\eta$ ,  $C_k$ ,  $\Omega$  and  $D_{yst}$  denotes the constant of Lemma 3.6. Here  $C_c$ ,  $\eta$  denote the continuity and coercivity constants of bilinear form  $a(\cdot, \cdot)$ .

*Proof.* First, note that the optimality condition and the estimate of Lemma 3.6 on  $g_h$  clearly imply that  $\int_0^T \|\mu_h\|_{L^2(\Omega)}^2 dt \leq C_{st}\alpha$ . Now, setting  $v_h = \mu_h$  to the adjoint equation (4.2), using the monotonicity of  $\phi$  and Young's inequality we obtain

$$\begin{aligned} & - (1/2)\|\mu_+^n\|_{L^2(\Omega)}^2 + (1/2)\|[\mu^n]\|_{L^2(\Omega)}^2 + (1/2)\|\mu_+^{n-1}\|_{L^2(\Omega)}^2 + \eta \int_{t^{n-1}}^{t^n} \|\mu_h\|_{H^1(\Omega)}^2 dt \\ & \leq \alpha^{1/2} \int_{t^{n-1}}^{t^n} \|y_h - U\|_{L^2(\Omega)}^2 dt + (1/4\alpha^{1/2}) \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)}^2 dt. \end{aligned}$$

Summing the above inequalities from  $N$  to 1 and using the bounds on  $y_h - U$ ,  $\mu_h$  and  $\mu_+^N = 0$ , we obtain the first estimate. The second estimate follows upon summing the inequalities from  $N$  to  $n$ . It remains to obtain a bound at arbitrary time-points. For this purpose let  $\bar{\mu}_h$  denote the exponential interpolant of  $\mu_h$  as constructed in Definition 4.1. Integrating by parts (in time) (4.2) and setting  $v_h = \bar{\mu}_h$  we obtain

$$\int_{t^{n-1}}^{t^n} \left( -\langle \mu_{ht}, \bar{\mu}_h \rangle + a(\mu_h, \bar{\mu}_h) + (\phi'(y_h)\mu_h, \bar{\mu}_h) \right) dt - ([\mu^n], \mu^n) = \int_{t^{n-1}}^{t^n} (y_h - U, \bar{\mu}_h) dt.$$

Using the Definition 4.1, and integration by parts (in time) the first term of the above equality can be written as,

$$\begin{aligned} - \int_{t^{n-1}}^{t^n} \langle \mu_{ht}, \bar{\mu}_h \rangle dt &= - \int_{t^{n-1}}^{t^n} (\mu_{ht}, \mu_h) e^{-\lambda(t^n-t)} dt = -(1/2) \int_{t^{n-1}}^{t^n} \frac{d}{dt} \|\mu_h\|_{L^2(\Omega)}^2 e^{-\lambda(t^n-t)} dt \\ &= -(1/2)\|\mu^n\|_{L^2(\Omega)}^2 + (1/2)\|\mu_+^{n-1}\|_{L^2(\Omega)}^2 e^{-\lambda(t^n-t^{n-1})} + (\lambda/2) \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)}^2 e^{-\lambda(t^n-t)} dt. \end{aligned}$$

Hence, combining the last two relations, and standard algebra, we obtain

$$\begin{aligned} & (1/2)\|\mu_+^{n-1}\|_{L^2(\Omega)}^2 e^{-\lambda(t^n-t^{n-1})} - (1/2)\|\mu_+^n\|_{L^2(\Omega)}^2 + (1/2)\|[\mu^n]\|_{L^2(\Omega)}^2 \\ & + (\lambda/2) \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)}^2 e^{-\lambda(t^n-t)} dt + \int_{t^{n-1}}^{t^n} (\phi'(y_h)\mu_h, \bar{\mu}_h) dt = - \int_{t^{n-1}}^{t^n} \left( a(\mu_h, \bar{\mu}_h) + (y_h - U, \bar{\mu}_h) \right) dt. \quad (4.4) \end{aligned}$$

Recall that Lemma 4.2 implies that

$$\int_{t^{n-1}}^{t^n} a(\mu_h, \bar{\mu}_h) dt \leq C_k C_c \int_{t^{n-1}}^{t^n} \|\mu_h\|_{H^1(\Omega)}^2 dt,$$

and

$$\int_{t^{n-1}}^{t^n} (y_h - U, \bar{\mu}_h) dt \leq (C_k/\alpha^{1/2}) \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)}^2 dt + \alpha^{1/2} \int_{t^{n-1}}^{t^n} \|y_h - U\|_{L^2(\Omega)}^2 dt.$$

It remains to treat the semi-linear term. Adding and subtracting  $\mu_h$ , and using the monotonicity of  $\phi$ , we obtain

$$\int_{t^{n-1}}^{t^n} (\phi'(y_h)\mu_h, \bar{\mu}_h) dt \geq \int_{t^{n-1}}^{t^n} (\phi'(y_h)\mu_h, \bar{\mu}_h - \mu_h) dt.$$

For the later term, the growth condition on  $\phi'$ , the Hölder's inequality (with  $s_1 = 2/(p-1)$ ,  $s_2 = s_3 = 4/(3-p)$ ), and the continuous embedding  $H^1(\Omega) \subset L^{4/(3-p)}(\Omega)$ , imply

$$\begin{aligned} \left| \int_{t^{n-1}}^{t^n} (\phi'(y_h)\mu_h, \bar{\mu}_h - \mu_h) dt \right| &\leq C \int_{t^{n-1}}^{t^n} \|y_h\|_{L^2(\Omega)}^{p-1} \|\mu_h\|_{L^{4/(3-p)}(\Omega)} \|\bar{\mu}_h - \mu_h\|_{L^{4/(3-p)}(\Omega)} dt \\ &\leq C \|y_h\|_{L^\infty[t^{n-1}, t^n; L^2(\Omega)]}^{p-1} \int_{t^{n-1}}^{t^n} \|\mu_h\|_{H^1(\Omega)} \|\bar{\mu}_h - \mu_h\|_{H^1(\Omega)} dt \\ &\leq CD_{yst}^{(p-1)/2} \|\mu_h\|_{L^2[t^{n-1}, t^n; H^1(\Omega)]} \|\bar{\mu}_h - \mu_h\|_{L^2[t^{n-1}, t^n; H^1(\Omega)]} \\ &\leq CD_{yst}^{(p-1)/2} \lambda \tau_n \|\mu_h\|_{L^2[t^{n-1}, t^n; H^1(\Omega)]}^2, \end{aligned}$$

where we have used the stability estimate of Lemma 3.6 on  $\|y_h\|_{L^\infty[0, T; L^2(\Omega)]}^2 \leq D_{yst}$ , and Lemma 4.2 to bound  $\bar{\mu}_h - \mu_h$  in terms of  $\mu_h$ . Therefore substituting the last inequality into (4.4), we obtain

$$\begin{aligned} (1/2) \|\mu_+^{n-1}\|_{L^2(\Omega)}^2 e^{-\lambda(t^n - t^{n-1})} - (1/2) \|\mu_+^n\|_{L^2(\Omega)}^2 + (1/2) \|[\mu^n]\|_{L^2(\Omega)}^2 + (\lambda/2) \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)}^2 e^{-\lambda(t^n - t)} dt \\ \leq \int_{t^{n-1}}^{t^n} \left( C_c C_k \|\mu_h\|_{H^1(\Omega)}^2 + (1/\alpha^{1/2}) \|\mu_h\|_{L^2(\Omega)}^2 + \alpha^{1/2} \|y_h - U\|_{L^2(\Omega)}^2 \right) dt + CD_{yst}^{(p-1)/2} \lambda \tau_n \int_{t^{n-1}}^{t^n} \|\mu_h\|_{H^1(\Omega)}^2 dt. \end{aligned}$$

Setting now  $\lambda = 1/\tau_n$ , we obtain

$$\begin{aligned} (1/2) \|\mu_+^{n-1}\|_{L^2(\Omega)}^2 e^{-1} - (1/2) \|\mu_+^n\|_{L^2(\Omega)}^2 + (1/2) \|[\mu^n]\|_{L^2(\Omega)}^2 + (e^{-1}/2\tau_n) \int_{t^{n-1}}^{t^n} \|\mu_h\|_{L^2(\Omega)}^2 dt \\ \leq \int_{t^{n-1}}^{t^n} \left( C_k C_c \|\mu_h\|_{H^1(\Omega)}^2 + (1/\alpha^{1/2}) \|\mu_h\|_{L^2(\Omega)}^2 + \alpha^{1/2} \|y_h - U\|_{L^2(\Omega)}^2 \right) dt + CD_{yst}^{(p-1)/2} C_k \int_{t^{n-1}}^{t^n} \|\mu_h\|_{H^1(\Omega)}^2 dt. \end{aligned}$$

The proof now follows, after using the inverse inequality  $\|\mu_h\|_{L^\infty[t^{n-1}, t^n; L^2(\Omega)]}^2 \leq (C_k/\tau_n) \|\mu_h\|_{L^2[t^{n-1}, t^n; L^2(\Omega)]}^2$  and the previously developed estimates of  $y_h$ ,  $\mu_h$  in various norms.  $\square$

Note that stability estimates on  $\mu_h$  scale better in terms of  $\alpha$  compared to  $y_h$ , as expected. Using the above estimates, which are independent of  $\tau, h$ , we may pass the limit into the discrete optimality system to prove convergence.

**Theorem 4.4.** *Suppose that  $f \in L^2[0, T; H^{-1}(\Omega)]$ ,  $y_0 \in L^2(\Omega)$ ,  $U \in L^2[0, T; L^2(\Omega)]$  and that  $\phi$  satisfies Assumption 2.1. In addition, let  $\phi \in C^2(\mathbb{R}; \mathbb{R})$  with  $|\phi''(s)| \leq C|s|^{p-2}$  for  $2 < p < 3$  or  $\phi'$  be uniformly continuous. Let  $h > 0$  and a uniform partition  $\{t^i\}_{i=0}^N$  of  $[0, T]$  with  $\tau_i = t^i - t^{i-1}$  and  $\tau = \max_{i=1, \dots, N} \tau_i$ , satisfying the assumptions of Lemmas 3.6 and 4.3. Then, for  $\alpha > 0$ ,  $(y_h, \mu_h)$  converges to the solution  $(y, \mu)$  of the continuous optimality system as  $\tau, h \rightarrow 0$ .*

*Proof (sketch).* The proof follows similarly to Theorem 3.8. We treat the case  $3/2 < p < 3$  (the case  $1 < p \leq 3/2$  can be treated similarly). Recall that  $y_h, g_h$  are bounded in  $L^\infty[0, T; L^2(\Omega)] \cap L^2[0, T; H^1(\Omega)]$  and  $L^2[0, T; L^2(\Omega)]$  by constants independent of  $\tau, h$ . Similarly  $\mu_h$  is bounded in  $L^\infty[0, T; L^2(\Omega)] \cap L^2[0, T; L^2(\Omega)]$  by a constant independent of  $\tau, h$ . Hence, we may extract subsequences, converging weakly to  $y, g, \mu$  respectively in the above norms, while an application of the discrete compactness Theorem 3.7 guarantees the strong convergence



of  $y_h$  to  $y$  in  $L^2[0, T; L^2(\Omega)]$  and of  $\mu_h$  to  $\mu$  in  $L^2[0, T; L^2(\Omega)]$ . Note that in order to apply Theorem 3.7, we now define  $\langle F(\mu), v \rangle \equiv -a(\mu, v) - \langle \phi'(y)\mu, v \rangle + (y - U, v)$ . It is evident by the stability estimates that  $\{\|\mu_h\|_{L^2[0, T; H^1(\Omega)]}\}_{h>0}$ ,  $\{\|F(\mu_h)\|_{L^{4/3}[0, T; H^{-1}(\Omega)]}\}_{h>0}$  remain bounded independent of  $h, \tau$ . For the later term, we only need to estimate the term,

$$\begin{aligned} \int_0^T \langle \phi'(y_h)\mu_h, v_h \rangle dt &\leq C\|\mu_h\|_{L^4[0, T; L^4(\Omega)]} \| |y_h|^{p-1} \|_{L^2[0, T; L^2(\Omega)]} \|v\|_{L^4[0, T; L^4(\Omega)]} \\ &\leq C\|\mu_h\|_{L^\infty[0, T; L^2(\Omega)]}^{1/2} \|\mu_h\|_{L^2[0, T; H^1(\Omega)]}^{1/2} \| |y_h|^{p-1} \|_{L^2[0, T; L^2(\Omega)]} \|v\|_{L^4[0, T; H^1(\Omega)]}. \end{aligned}$$

Note that using the embedding  $L^4(\Omega) \subset L^{2(p-1)}(\Omega)$  (recall that  $3/2 \leq p < 3$ ), and the interpolation inequality  $\|\cdot\|_{L^4(\Omega)}^2 \leq C\|\cdot\|_{L^2(\Omega)}\|\cdot\|_{H^1(\Omega)}$ , we may show that  $\| |y_h|^{p-1} \|_{L^2[0, T; L^2(\Omega)]} < \infty$  since  $y_h$  remains bounded (with constant independent of  $h, \tau$ ) in  $L^\infty[0, T; L^2(\Omega)] \cap L^2[0, T; H^1(\Omega)]$ . All other terms are easy to handle. The proof is completed after noting that we may pass the limit into equations (4.1)–(4.3), and into the functional (1.1). Indeed, for the adjoint equation, choosing  $v_h \in C[0, T; U_h] \cap \mathcal{U}_h$  with  $v_h(0) = 0$ , equation (4.2) takes the form,

$$\int_0^T (\langle \mu_h, v_{ht} \rangle + a(\mu_h, v_h) + \langle \phi'(y_h)\mu_h, v_h \rangle) dt = \int_0^T (y_h - U, v_h) dt.$$

Using the strong convergence of  $y_h, \mu_h$  in  $L^2[0, T; L^2(\Omega)]$  we may pass the limit term by term. For the semilinear term note that

$$\int_0^T \langle \phi'(y)\mu - \phi'(y_h)\mu_h, v_h \rangle dt = \int_0^T \langle (\phi'(y) - \phi'(y_h))\mu, v_h \rangle dt + \int_0^T \langle \phi'(y_h)(\mu - \mu_h), v_h \rangle dt.$$

The first term can be treated using the growth condition on  $\phi''$  (or by the uniform continuity of  $\phi'$ ), the regularity of  $\mu, v_h$  and the strong convergence of  $y_h$  in  $L^2[0, T; L^2(\Omega)]$ . The growth condition on  $\phi'$ , the stability estimates on  $y_h$ , and the strong convergence of  $\mu_h$  allow to pass the limit through the second term.  $\square$

*Acknowledgements.* The author would like to thank Prof. Noel J. Walkington, for providing the compactness result [42], Theorem 3.1, which led to significant improvements of the results.

## REFERENCES

- [1] G. Akrivis and C. Makridakis, Galerkin time-stepping methods for nonlinear parabolic equations. *ESAIM: M2AN* **38** (2004) 261–289.
- [2] A. Borzi and R. Griesse, Distributed optimal control for lambda-omega systems. *J. Numer. Math.* **14** (2006) 17–40.
- [3] H. Brezis, *Analyse fonctionnelle – Theorie et applications*. Masson, Paris, France (1983).
- [4] K. Chrysafinos, Discontinuous Galerkin approximations for distributed optimal control problems constrained to linear parabolic PDE’s. *Int. J. Numer. Anal. Mod.* **4** (2007) 690–712.
- [5] K. Chrysafinos and N.J. Walkington, Error estimates for the discontinuous Galerkin methods for parabolic equations. *SIAM J. Numer. Anal.* **44** (2006) 349–366.
- [6] K. Chrysafinos and N.J. Walkington, Lagrangian and moving mesh methods for the convection diffusion equation. *ESAIM: M2AN* **42** (2008) 25–55.
- [7] K. Chrysafinos and N.J. Walkington, Discontinuous Galerkin approximations of the Stokes and Navier-Stokes equations. *Math. Comp.* (to appear), available at <http://www.math.cmu.edu/~noelw>.
- [8] K. Chrysafinos, M.D. Gunzburger and L.S. Hou, Semidiscrete approximations of optimal Robin boundary control problems constrained by semilinear parabolic PDE. *J. Math. Anal. Appl.* **323** (2006) 891–912.
- [9] P.G. Ciarlet, *The finite element method for elliptic problems, Classics in Applied Mathematics* **40**. SIAM (2002).
- [10] K. Dechelnick and M. Hinze, Semidiscretization and error estimates for distributed control of the instationary Navier-Stokes equations. *Numer. Math.* **97** (2004) 297–320.
- [11] K. Eriksson and C. Johnson, Adaptive finite element methods for parabolic problems. I. A linear model problem. *SIAM J. Numer. Anal.* **28** (1991) 43–77.
- [12] K. Eriksson and C. Johnson, Adaptive finite element methods for parabolic problems. II. Optimal error estimates in  $L_\infty(L^2)$  and  $L_\infty(L_\infty)$ . *SIAM J. Numer. Anal.* **32** (1995) 706–740.
- [13] K. Eriksson and C. Johnson, Adaptive finite element methods for parabolic problems IV: Nonlinear problems. *SIAM J. Numer. Anal.* **32** (1995) 1729–1749.

- [14] K. Eriksson, C. Johnson and V. Thomée, Time discretization of parabolic problems by the discontinuous Galerkin method. *RAIRO Modél. Math. Anal. Numér.* **29** (1985) 611–643.
- [15] D. Estep and S. Larsson, The discontinuous Galerkin method for semilinear parabolic equations. *RAIRO Modél. Math. Anal. Numér.* **27** (1993) 35–54.
- [16] L. Evans, *Partial Differential Equations*. AMS, Providence, USA (1998).
- [17] R. Falk, Approximation of a class of optimal control problems with order of convergence estimates. *J. Math. Anal. Appl.* **44** (1973) 28–47.
- [18] A. Fursikov, *Optimal control of distributed systems – Theory and applications*. AMS, Providence, USA (2000).
- [19] M. Garvie and C. Trenchea, Optimal control of a nutrient-phytoplankton-zooplankton-fish system. *SIAM J. Control Optim.* **46** (2007) 775–791.
- [20] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes*. Springer-Verlag, New York, USA (1986).
- [21] M.D. Gunzburger, *Perspectives in flow control and optimization, Advances in Design and Control*. SIAM, Philadelphia, USA (2003).
- [22] M.D. Gunzburger and S. Manservigi, Analysis and approximation of the velocity tracking problem for Navier-Stokes flows with distributed control. *SIAM J. Numer. Anal.* **37** (2000) 1481–1512.
- [23] M.D. Gunzburger, L.S. Hou and T. Svobodny, Analysis and finite element approximation of optimal control problems for the stationary Navier-Stokes equations with Dirichlet controls. *RAIRO Modél. Math. Anal. Numer.* **25** (1991) 711–748.
- [24] M.D. Gunzburger, S.-D. Yang, and W. Zhu, Analysis and discretization of an optimal control problem for the forced Fisher equation. *Discrete Contin. Dyn. Syst. Ser. B* **8** (2007) 569–587.
- [25] M. Hinze and K. Kunisch, Second order methods for optimal control of time-dependent fluid flow. *SIAM J. Control Optim.* **40** (2001) 925–946.
- [26] L.S. Hou, and H.-D. Kwon, Analysis and approximations of a terminal-state optimal control problem constrained by semilinear parabolic PDEs. *Int. J. Numer. Anal. Model.* **4** (2007) 713–728.
- [27] G. Knowles, Finite element approximation of parabolic time optimal control problems. *SIAM J. Control Optim.* **20** (1982) 414–427.
- [28] I. Lasiecka, Rietz-Galerkin approximation of the time optimal boundary control problem for parabolic systems with Dirichlet boundary conditions. *SIAM J. Control Optim.* **22** (1984) 477–500.
- [29] I. Lasiecka and R. Triggiani, *Control theory for partial differential equations*. Cambridge University Press, Cambridge, UK (2000).
- [30] J.-L. Lions, *Some aspects of the control of distributed parameter systems*. Conference Board of the Mathematical Sciences, SIAM (1972).
- [31] W.-B. Liu and N. Yan, A posteriori error estimates for optimal control problems governed by parabolic equations. *Numer. Math.* **93** (2003) 497–521.
- [32] W.-B. Liu, H.-P. Ma, T. Tang and N. Yan, A posteriori error estimates for DG time-stepping method for optimal control problems governed by parabolic equations. *SIAM J. Numer. Anal.* **42** (2004) 1032–1061.
- [33] K. Malanowski, Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems. *Appl. Math. Optim.* **8** (1981) 69–95.
- [34] D. Meidner and B. Vexler, Adaptive space-time finite element methods for parabolic optimization problems. *SIAM J. Control Optim.* **46** (2007) 116–142.
- [35] D. Meidner and B. Vexler, A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part I: Problems without control constraints. *SIAM J. Control Optim.* **47** (2008) 1150–1177.
- [36] P. Neittaanmaki and D. Tiba, *Optimal control of nonlinear parabolic systems – Theory, algorithms and applications*. M. Dekker, New York, USA (1994).
- [37] A. Rösch, Error estimates for parabolic optimal control problems with control constraints. *Zeitschrift Anal. Anwendungen* **23** (2004) 353–376.
- [38] R. Temam, *Navier-Stokes equations*. North Holland (1977).
- [39] V. Thomée, *Galerkin finite element methods for parabolic problems*. Springer-Verlag, Berlin, Germany (1997).
- [40] F. Tröltzsch, Semidiscrete Ritz-Galerkin approximation of nonlinear parabolic boundary control problems. *International Series of Numerical Math.* **111** (1993) 57–68.
- [41] F. Tröltzsch, Semidiscrete Ritz-Galerkin approximation of nonlinear parabolic boundary control problems – Strong convergence of optimal controls. *Appl. Math. Optim.* **29** (1994) 309–329.
- [42] N.J. Walkington, Compactness properties of the DG and CG time stepping schemes for parabolic equations. *SINUM (June 2008)* (submitted), preprint available at <http://www.math.cmu.edu/~noelw>.
- [43] R. Winther, Error estimates for a Galerkin approximation of a parabolic control problem. *Ann. Math. Pura Appl.* **117** (1978) 173–206.
- [44] R. Winther, Initial value methods for parabolic control problems. *Math. Comp.* **34** (1980) 115–125.
- [45] E. Zeidler, *Nonlinear functional analysis and its applications, II/B Nonlinear monotone operators*. Springer-Verlag, New York, USA (1990).