MATHEMATICS OF COMPUTATION
VOLUME 40, NUMBER 161
JANUARY 1983, PAGES 151-173

Sharp Error Estimates for a Finite Element-Penalty Approach to a Class of Regulator Problems*

By Goong Chen, Wendell H. Mills, Jr.,** Shunhua Sun and David A. Yost

Abstract. Quadratic cost optimal controls can be solved by penalizing the governing linear differential equation [2], [9]. In this paper, we study the numerical analysis of this approach using finite elements. We formulate the geometric *condition* (H) which requires that pairs of certain related finite-dimensional approximation spaces form "angles" which are bounded away from the "180° angle". Under condition (H), we prove that the penalty parameter ε and the discretization parameter h are independent in the error bounds, thereby giving sharp asymptotic error estimates. This condition (H) is shown to be also a necessary condition for such independence. Examples and numerical evidence are also provided.

0. Introduction. Consider the optimal control problem: Given the quadratic cost functional J,

$$J(x, u) \equiv \int_0^T \left[\langle \dot{x}, N_1 \dot{x} \rangle_{\mathbf{R}^n} + \langle x, N_2 x \rangle_{\mathbf{R}^n} + \langle u, Mu \rangle_{\mathbf{R}^m} \right] dt,$$

solve

$$\min_{(x, u) \in H^1_{0n} \times L^2_m} J(x, u) \tag{0.1}$$

subject to

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) + f(t), & t \in [0, T], \\ x(0) = 0, \end{cases} \tag{0.2}$$

where $x(t) \in \mathbf{R}^n$ is the state at time t, $u(t) \in \mathbf{R}^m$ is the control at t, $A(t)$ and $B(t)$ are, respectively, $n \times n$ and $n \times m$ time-varying matrices, and f is the inhomogeneous forcing term.

In the cost functional J, we assume

$$\begin{cases} N_1, N_2, M \text{ are constant } n \times n,\ n \times n, \text{ and } m \times m \text{ symmetric} \\ \qquad\qquad\qquad\qquad \text{positive semi-definite matrices,} \\ \langle x, N_1 x \rangle_{\mathbf{R}^n} \geq \nu \|x\|^2_{\mathbf{R}^n},\ \langle u, Mu \rangle_{\mathbf{R}^m} \geq \nu \|u\|^2_{\mathbf{R}^m} \quad \text{for all } x \in \mathbf{R}^n,\ u \in \mathbf{R}^m, \\ \qquad\qquad\qquad\qquad \text{where } \nu > 0, \text{ is independent of } x \text{ and } u. \end{cases} \tag{0.3}$$

The standard Sobolev norms and spaces used are as follows.

Received November 17, 1981; revised February 26, 1982 and May 3, 1982.

1980 *Mathematics Subject Classification.* Primary 34H05, 49D30, 65N30; Secondary 41A65.

* Supported in part by NSF Grant MCS 81-01892. Work completed while the third author was visiting Purdue University.

** *Current address*: Research and Development Department, Standard Oil of Ohio, Cleveland, Ohio 44128.

0025-5718/82/0000-0724/$07.50

$$\|y\|^2_{H_l^k} \equiv \sum_{j=0}^{k} \int_0^T \|y^{(j)}(t)\|^2_{\mathbf{R}^l}\, dt,$$

$$H_l^k \equiv H_l^k(0,T) \equiv \{ y : [0,T] \to \mathbf{R}^l \mid y^{(i)} \text{ is absolutely continuous,}$$
$$0 \leqslant i \leqslant k-1,\ \|y\|_{H_l^k} < \infty \},$$

$$H^1_{0l} \equiv \{ y \in H^1_l \mid y(0) = 0 \}, \qquad \|y\|_{H^1_{0l}} \equiv \|y\|_{H^1_l},$$
$$L^2_l \equiv L^2_l(0,T) \equiv H^0_l(0,T),$$

for $l \in \mathbf{Z}^+ \equiv \{1,2,3,\dots\}$ and $k \in \mathbf{N} \equiv \{0,1,2,\dots\}$.

Using calculus of variations, (0.1) and (0.2) are equivalent to the variational formulation

$$(0.4) \quad a\left(\begin{bmatrix} x \\ u \end{bmatrix}, \begin{bmatrix} y \\ v \end{bmatrix}\right) = 0 \quad \text{for all } \begin{bmatrix} y \\ v \end{bmatrix} \in H^1_{0n} \times L^2_m, \dot{y} = Ay + Bv, y(0) = 0,$$

where

$$a\left(\begin{bmatrix} x \\ u \end{bmatrix}, \begin{bmatrix} y \\ v \end{bmatrix}\right) \equiv \int_0^T [\langle \dot{x}, N_1 \dot{y} \rangle_{\mathbf{R}^n} + \langle x, N_2 y \rangle_{\mathbf{R}^n} + \langle u, Mv \rangle_{\mathbf{R}^m}]\, dt.$$

A feasible approach for computing the optimal control $\hat{u}$ and the corresponding optimal state $\hat{x}$ is by penalizing the governing equation (0.2): we solve the unconstrained problem

$$(0.5) \qquad \underset{(x,u)\in H^1_{0n}\times L^2_m}{\operatorname{Min}} \quad J_\varepsilon(x,u,f) \equiv J(x,u) + \frac{1}{\varepsilon}\|\dot{x} - Ax - Bu - f\|^2_{L^2_n}, \qquad \varepsilon > 0,$$

and let ε tend to zero to obtain convergence. This approach was first introduced by A. V. Balakrishnan [2] and J. L. Lions [9].

Note that the form of the cost functional J requires that the (weighted) rate of change of the state $\dot{x}$ be minimized, in addition to both the (weighted) state x and control u. This is an important technical assumption in our paper. We also note that an inhomogeneous initial condition $x(0) = x_0$ can be reduced to the zero initial condition (as in (0.2)) by the change of variable $y(t) = x(t) - x_0$.

From the Poincaré inequality, the expressions

$$\int_0^T \langle \dot{x}(t), N_1\dot{x}(t)\rangle\, dt, \quad \int_0^T [\langle \dot{x}(t), N_1\dot{x}(t)\rangle + \langle x(t), N_2 x(t)\rangle]\, dt$$

in J define equivalent norms in the Hilbert space H^1_{0n}. We assume $A(t)$ is sufficiently smooth such that, for $y \in H^1_{0n}$,

$$(0.6) \quad \|\dot{y} - Ay\|_{L^2_n} = \left(\int_0^T |\dot{y}(t) - A(t)y(t)|^2\, dt\right)^{1/2} \geqslant c\|y\|_{H^1_n} \quad \text{for some } c > 0$$

(i.e., the above defines an equivalent norm in H^1_{0n}). It is trivial to see that this holds when A is a constant matrix.

Let $S^1_{h_i} \subset H^1_{0n}$ and $S^2_{h_i} \subset L^2_m$ be two sequences of increasing finite-dimensional spaces such that

$$\lim_{h_i \downarrow 0} \inf_{y \in S^1_{h_i}} \|x - y\|_{H^1_{0n}} = 0, \quad \lim_{h_i \downarrow 0} \inf_{v \in S^2_{h_i}} \|u - v\|_{L^2_m} = 0 \quad \forall (x,u) \in H^1_{0n} \times L^2_m.$$

The finite element approximation is to minimize (0.5) over $S_h^1 \times S_h^2 \subset H_{0n}^1 \times L_m^2$ ($h = h_i$ for some i). Let $(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon)$ be the unique minimizer of (0.5) in $S_h^1 \times S_h^2$. Our goal is to analyze the error

$$\|\hat{x}_h^\varepsilon - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_h^\varepsilon - \hat{u}\|_{L_m^2}$$

with respect to the penalty parameter ε and the discretization parameter h.

Numerical analysis of penalty problems of similar nature has been studied in [6], [7], for example. In those works, the penalty parameter ε is often found to be coupled with (or dependent upon) h. In [4], Chen and Mills give some sharp estimates for a primal-penalty-finite element computation scheme which show that, in the error bounds for that problem and approach, ε and h are actually *independent* of each other. The problem in [4] is simpler than the one being studied here because the penalized constraint is finite-dimensional. As we will see later on, for the problem and approach considered here, the independence of ε and h will not hold in general.

The main result of our paper is as follows. We show that sharp estimates (cf. (3.4)) hold if and only if the approximating finite-dimensional spaces $\{(S_{h_i}^1, S_{h_i}^2)\}_{i=1}^\infty$ satisfy a certain special property, namely, Condition (H) in Section 3. This condition requires that pairs of certain related finite-dimensional subspaces form "angles" which are bounded away from the "180° angle".

In Section 1, we introduce some basic facts about penalty and study the well-posedness of exact solutions and penalized solutions with respect to the inhomogeneous data f. The relations between the solutions and the inhomogeneous data f are linear and expressed in terms of certain linear operators $\mathfrak{L}$, $\mathfrak{L}_\varepsilon$, $\mathfrak{L}_h$ and $\mathfrak{L}_h^\varepsilon$. Basic errors between the exact (or, the discretized) solution and the penalized solution are estimated.

It is found in this paper that the validity of sharp error bounds can be studied in terms of an *abstract approximation problem*. This problem has considerable theoretical interest in its own right and is thoroughly examined in Section 2. Necessary and sufficient conditions are formulated for the solvability of this problem

In Section 3, we give the main estimates in Theorem 12. Condition (H), which is obtained through the study of the abstract approximation problem in Section 2, is seen to be necessary and sufficient for Theorem 12 to hold. Error bounds in the case without (H) are given in Theorem 20.

In Section 4, we apply the theory to some examples. Numerical computations obtain suggest that the errors indicated in Theorem 12 are sharp.

As with the penalty method, the stiffness matrix (associated with the quadratic form J_ε in (0.5)) usually has a large condition number, thereby producing considerable numerical instability. This instability can be circumvented by using the standard matrix iterative refinement technique. Numerical results indicate that the penalty method is quite accurate and efficient when compared with other methods, e.g., the primal [4] or the dual [3] methods.

In the design of optimal regulators, the matrix N_1 appearing in J is usually 0 [11]. In this situation, computationally, the penalty method also produces accurate results. Nevertheless, at this stage, no satisfactory error estimates like (3.4) for such problems are available. It remains a challenging research work yet to be completed by control theorists and numerical analysts.

1. Finite Element Approximations of the Penalized Problem. We consider the unconstrained penalized problem

$$\min_{(x,u)\in H^1_{0n}\times L^2_m} J_\varepsilon(x, u, f), \tag{1.1}$$

where

$$J_\varepsilon(x, u, f) \equiv \langle \dot{x}, N_1\dot{x}\rangle_{L^2_n} + \langle x, N_2 x\rangle_{L^2_n} + \langle u, Mu\rangle_{L^2_m} + \frac{1}{\varepsilon}\|\dot{x} - Ax - Bu - f\|^2_{L^2_n}, \qquad \varepsilon > 0,$$

and (0.3), (0.6) hold.

Let $(\hat{x}_\varepsilon, \hat{u}_\varepsilon)$ be the unique minimizer of (1.1) and let $(\hat{x}, \hat{u})$ be the optimal state and control of problem (0.1) and (0.2). From the work of Polyak [10], we see that

$$\|\hat{x}_\varepsilon - \hat{x}\|_{H^1_n} = O(\varepsilon), \tag{1.2}$$

$$\|\hat{u}_\varepsilon - \hat{u}\|_{L^2_m} = O(\varepsilon), \tag{1.3}$$

as $\varepsilon \downarrow 0$. The analysis of these estimates is central to our development.

For each $f \in L^2_n$, let $(\hat{x}(f), \hat{u}(f)) \in H^1_{0n} \times L^2_m$ be the solution of (0.1) and (0.2) corresponding to this f. This induces a mapping $\mathfrak{L}: L^2_n \to H^1_{0n} \times L^2_m$ defined by

$$\mathfrak{L}(f) \equiv (\hat{x}(f), \hat{u}(f)). \tag{1.4}$$

The following lemma states that the optimal control problem (0.1) and (0.2) is well-posed with respect to the inhomogeneous data f.

LEMMA 1. *Let* (0.3) *and* (0.6) *hold. Then the mapping* $\mathfrak{L}$ *defined by* (1.4) *is a bounded linear transformation from* L^2_n *into* $H^1_{0n} \times L^2_m$.

Proof. This can be easily verified from the variational equation (0.4) satisfied by $\hat{u}$, using the primal theory. □

Let $f_\varepsilon \equiv \dot{\hat{x}}_\varepsilon - A\hat{x}_\varepsilon - B\hat{u}_\varepsilon$. It is obvious that $(\hat{x}_\varepsilon, \hat{u}_\varepsilon)$ is the unique solution of

$$\begin{aligned} &\text{Min } J(x, u) \\ &\text{Subject to} \\ &\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + f_\varepsilon(t), & t \in [0, T], \\ x(0) = 0. \end{cases} \end{aligned}$$

This implies that

$$\mathfrak{L}(f_\varepsilon) = (\hat{x}_\varepsilon, \hat{u}_\varepsilon). \tag{1.5}$$

Therefore, by Lemma 1,

$$\|\hat{x} - \hat{x}_\varepsilon\|_{H^1_{0n}} + \|\hat{u} - \hat{u}_\varepsilon\|_{L^2_m} = \|\mathfrak{L}(f) - \mathfrak{L}(f_\varepsilon)\|_{H^1_{0n}\times L^2_m} \leqslant \|\mathfrak{L}\|\,\|f - f_\varepsilon\|_{L^2_n}. \tag{1.6}$$

Since $(\hat{x}_\varepsilon, \hat{u}_\varepsilon)$ minimizes J_ε, we have

$$\begin{aligned} J(\hat{x}_\varepsilon, \hat{u}_\varepsilon) \leqslant J_\varepsilon(\hat{x}_\varepsilon, \hat{u}_\varepsilon, f) &= J(\hat{x}_\varepsilon, \hat{u}_\varepsilon) + \frac{1}{\varepsilon}\|\dot{\hat{x}}_\varepsilon - A\hat{x}_\varepsilon - B\hat{u}_\varepsilon - f\|^2_{L^2_n} \\ &\leqslant J_\varepsilon(\hat{x}, \hat{u}, f) = J(\hat{x}, \hat{u}). \end{aligned} \tag{1.7}$$

Thus

$$(1.8)\quad \frac{1}{\varepsilon}\|\dot{\hat{x}}_\varepsilon - A\hat{x}_\varepsilon - B\hat{u}_\varepsilon - f\|^2_{L^2_n} = \frac{1}{\varepsilon}\|f_\varepsilon - f\|^2_{L^2_n} \leq J(\hat{x},\hat{u}) - J(\hat{x}_\varepsilon,\hat{u}_\varepsilon)$$
$$= 2\cdot a\left(\begin{bmatrix}\hat{x}_\varepsilon\\ \hat{u}_\varepsilon\end{bmatrix},\begin{bmatrix}\hat{x}-\hat{x}_\varepsilon\\ \hat{u}-\hat{u}_\varepsilon\end{bmatrix}\right) + J(\hat{x}-\hat{x}_\varepsilon,\hat{u}-\hat{u}_\varepsilon).$$

Therefore

$$\frac{1}{\varepsilon}\|f_\varepsilon - f\|^2 \leq K\big(\|(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\|_{H^1_{0n}\times L^2_m}\|(\hat{x}-\hat{x}_\varepsilon,\hat{u}-\hat{u}_\varepsilon)\|_{H^1_{0n}\times L^2_m}\big)$$
$$+K\Big(\|(\hat{x}-\hat{x}_\varepsilon,\hat{u}-\hat{u}_\varepsilon)\|^2_{H^1_{0n}\times L^2_m}\Big),$$

where K is a positive constant satisfying

$$2\left|a\left(\begin{bmatrix}y_1\\ v_1\end{bmatrix},\begin{bmatrix}y_2\\ v_2\end{bmatrix}\right)\right| \leq K\|(y_1,v_1)\|_{H^1_{0n}\times L^2_m}\|(y_2,v_2)\|_{H^1_{0n}\times L^2_m},$$
$$J(y,v) \leq K\|(y,v)\|^2_{H^1_{0n}\times L^2_m}.$$

By (1.6) we obtain

$$\frac{1}{\varepsilon}\|f_\varepsilon - f\|^2_{L^2_n} \leq K\big(\|(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\| + \|(\hat{x}-\hat{x}_\varepsilon,\hat{u}-\hat{u}_\varepsilon)\|\big)\|(\hat{x}-\hat{x}_\varepsilon,\hat{u}-\hat{u}_\varepsilon)\|$$
$$\leq K\big(\|(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\| + \|(\hat{x}-\hat{x}_\varepsilon,\hat{u}-\hat{u}_\varepsilon)\|\big)\|\mathcal{L}\|\,\|f_\varepsilon - f\|_{L^2_n}.$$

So

$$(1.9)\qquad \|f_\varepsilon - f\|_{L^2_n} \leq K\big(\|(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\| + \|(\hat{x}-\hat{x}_\varepsilon,\hat{u}-\hat{u}_\varepsilon)\|\big)\|\mathcal{L}\|\cdot\varepsilon.$$

By (1.7), we have

$$(1.10)\qquad \|(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\| \leq \left[\frac{1}{\nu}J(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\right]^{1/2} \leq \left[\frac{1}{\nu}J(\hat{x},\hat{u})\right]^{1/2}$$
$$\leq \left[\frac{1}{\nu}K\|(\hat{x},\hat{u})\|^2\right]^{1/2} = \nu^{-1/2}K^{1/2}\|(\hat{x},\hat{u})\|,$$

where ν is the positive constant in (0.3). So, using (1.6), (1.9), and (1.10) we get

$$\|\hat{x}_\varepsilon - \hat{x}\|_{H^1_{0n}} + \|\hat{u}_\varepsilon - \hat{u}\|_{L^2_m} \leq \|\mathcal{L}\|\,\|f_\varepsilon - f\|_{L^2_n}$$
$$\leq \|\mathcal{L}\|\cdot K\big(\|(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\| + \|(\hat{x},\hat{u})\| + \|(\hat{x}_\varepsilon,\hat{u}_\varepsilon)\|\big)\|\mathcal{L}\|\cdot\varepsilon$$
$$\leq \|\mathcal{L}\|^2\cdot K\cdot\varepsilon\big[(1+2\nu^{-1/2}K^{1/2})\|(\hat{x},\hat{u})\|\big]$$
$$\equiv \bar{K}\cdot\|(\hat{x},\hat{u})\|\,\|\mathcal{L}\|^2\cdot\varepsilon, \qquad \bar{K} \equiv K(1+2\nu^{-1/2}K^{1/2}).$$

We summarize the above in

THEOREM 2. *Let* $(\hat{x}_\varepsilon,\hat{u}_\varepsilon)$ *solve* (1.1) *and* $(\hat{x},\hat{u})$ *solve* (0.1), (0.2). *If* (0.3) *and* (0.6) *hold, then*

$$\|\hat{x}_\varepsilon - \hat{x}\|_{H^1_{0n}} + \|\hat{u}_\varepsilon - \hat{u}\|_{L^2_m} \leq \bar{K}\cdot\|(\hat{x},\hat{u})\|\,\|\mathcal{L}\|^2\cdot\varepsilon$$

for all $\varepsilon > 0$, *where* $\bar{K} > 0$ *is independent of* $\hat{x}$, $\hat{u}$ *and* ε. □

For any $f \in L^2_n$, define $\mathcal{L}_\varepsilon\colon L^2_n \to H^1_{0n}\times L^2_m$ by $\mathcal{L}_\varepsilon(f) \equiv (\hat{x}_\varepsilon,\hat{u}_\varepsilon)$, where $(\hat{x}_\varepsilon,\hat{u}_\varepsilon)$ is the unique solution of (1.1) corresponding to this f. From (1.5) we have $\mathcal{L}_\varepsilon(f) = \mathcal{L}(f_\varepsilon)$.

LEMMA 3. *Let* (0.3) *and* (0.6) *hold. Then the mapping* $\mathfrak{L}_\varepsilon$ *is a bounded linear transformation from* L_n^2 *into* $H_{0n}^1 \times L_m^2$. *Furthermore*,

$$\overline{\lim_{\varepsilon\downarrow 0}} \left\{ \|\mathfrak{L}_\varepsilon\| \right\} < \infty. \tag{1.11}$$

Proof. The fact that $\mathfrak{L}_\varepsilon$ is bounded linear can be verified from the variational equation for (1.1), see (1.13) below. By Theorem 2, we have

$$\mathfrak{L}_\varepsilon f \to \mathfrak{L} f \quad \text{as } \varepsilon \downarrow 0, \quad \forall f \in L_n^2. \tag{1.12}$$

Applying the uniform boundedness principle gives (1.11). □

Remark 4. The penalized problem (1.1) is equivalent to the variational equation

$$a_\varepsilon\left(\begin{bmatrix} \hat{x}_\varepsilon \\ \hat{u}_\varepsilon \end{bmatrix}, \begin{bmatrix} y \\ v \end{bmatrix}\right) = \theta_\varepsilon\left(\begin{bmatrix} y \\ v \end{bmatrix}\right), \quad \forall \begin{bmatrix} y \\ v \end{bmatrix} \in H_{0n}^1 \times L_m^2, \tag{1.13}$$

where

$$a_\varepsilon\left(\begin{bmatrix} x_1 \\ u_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ u_2 \end{bmatrix}\right) \equiv a\left(\begin{bmatrix} x_1 \\ u_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ u_2 \end{bmatrix}\right) + \frac{1}{\varepsilon} \langle \dot{x}_1 - Ax_1 - Bu_1, \dot{x}_2 - Bx_2 - Bu_2 \rangle_{L_n^2}$$

and

$$\theta_\varepsilon\left(\begin{bmatrix} x \\ u \end{bmatrix}\right) \equiv \frac{1}{\varepsilon} \langle f, \dot{x} - Ax - Bu \rangle_{L_n^2}. \quad \square$$

From [10], we know that

$$s\text{-}\lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\left(\dot{\hat{x}}_\varepsilon - A\hat{x}_\varepsilon - B\hat{u}_\varepsilon - f\right) = \hat{p} \quad \text{in } L_n^2$$

for some $\hat{p} \in L_n^2$, which is the Lagrange multiplier. In the limit (1.13) becomes

$$a\left(\begin{bmatrix} \hat{x} \\ \hat{u} \end{bmatrix}, \begin{bmatrix} y \\ v \end{bmatrix}\right) + \langle \hat{p}, \dot{y} - Ay - Bv \rangle_{L_n^2} = 0 \quad \forall (y, v) \in H_{0n}^1 \times L_m^2, \tag{1.14}$$

the variational equation for $(\hat{x}, \hat{u})$.

We now approximate the penalized problem (1.1) by finite elements. Let $\{S_h^1 \times S_h^2 \mid 0 \leq h \leq h_0\}$ be a one-parameter family of products of finite-dimensional subspaces S_h^1 and S_h^2 satisfying

$$\begin{cases} S_h^1 \times S_h^2 \subseteq H_{0n}^1 \times L_m^2, \\ \lim\limits_{h\downarrow 0} \inf\limits_{y_h \in S_h^1} \|y - y_h\|_{H_{0n}^1} = 0 \quad \text{for any } y \in H_{0n}^1, \\ \lim\limits_{h\downarrow 0} \inf\limits_{v_h \in S_h^2} \|v - v_h\|_{L_m^2} = 0 \quad \text{for any } v \in L_m^2. \end{cases} \tag{1.15}$$

The approximation is to solve (1.1) over $S_h^1 \times S_h^2$. For each $\varepsilon > 0$, $h > 0$ let $(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon)$ be the unique solution of

$$\operatorname*{Min}_{(x,u)\in S_h^1 \times S_h^2} J_\varepsilon(x, u, f). \tag{1.16}$$

This is equivalent to the variational problem

$$a_\varepsilon\left(\begin{bmatrix} \hat{x}_h^\varepsilon \\ \hat{u}_h^\varepsilon \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right) = \theta_\varepsilon\left(\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}\right) \quad \forall (v_1, v_2) \in S_h^1 \times S_h^2. \tag{1.17}$$

If $\{\psi_i\}_{i=1}^{K_1}$, $\{\varphi_i\}_{i=1}^{K_2}$ are bases for S_h^1, S_h^2, respectively, (1.17) is a matrix equation $M_{\varepsilon,h}\bar{q}_{\varepsilon,h} = \bar{\theta}_{\varepsilon,h}$, where

$$[M_{\varepsilon,h}]_{ij} = a_\varepsilon(\Phi_j, \Phi_i), \qquad (\bar{\theta}_{\varepsilon,h})_j = \theta_\varepsilon(\Phi_j),$$

and $\{\Phi_i\}_{i=1}^{K_1+K_2}$ is the basis for $S_h^1 \times S_h^2$ induced by $\{\psi_i\}$ and $\{\varphi_i\}$. More specifically,

$$M_{\varepsilon,h} = \left[\begin{array}{c:c} \begin{array}{l} \dfrac{1}{\varepsilon}\displaystyle\int_0^T \left\langle \dot{\psi}_j - A\psi_j, \dot{\psi}_i - A\psi_i \right\rangle_{\mathbf{R}^n} dt \\ +\displaystyle\int_0^T \left[\left\langle \psi_j, N_2\psi_i \right\rangle_{\mathbf{R}^n} + \left\langle \dot{\psi}_j, N_1\dot{\psi}_i \right\rangle_{\mathbf{R}^n}\right] dt \end{array} & -\dfrac{1}{\varepsilon}\displaystyle\int_0^T \left\langle B\varphi_j, \dot{\psi}_i - A\psi_i \right\rangle_{\mathbf{R}^n} dt \\ \hdashline -\dfrac{1}{\varepsilon}\displaystyle\int_0^T \left\langle \dot{\psi}_j - A\psi_j, B\varphi_i \right\rangle_{\mathbf{R}^n} dt & \begin{array}{l} \dfrac{1}{\varepsilon}\displaystyle\int_0^T \left\langle B\varphi_j, B\varphi_i \right\rangle_{\mathbf{R}^n} dt \\ +\displaystyle\int_0^T \left\langle \varphi_j, M\varphi_i \right\rangle dt \end{array} \end{array}\right]$$

and

$$\bar{\theta}_{\varepsilon,h} = \left[\begin{array}{c} \dfrac{1}{\varepsilon}\displaystyle\int_0^T \left\langle f, \dot{\psi}_i - A\psi_i \right\rangle_{\mathbf{R}^n} dt \\ \hdashline -\dfrac{1}{\varepsilon}\displaystyle\int_0^T \left\langle f, B\phi_i \right\rangle_{\mathbf{R}^n} dt \end{array}\right].$$

Examples in Section 4 show that this matrix is of a block banded structure for certain choices of approximating spaces.

The analysis of the errors in this approximation is quite subtle. We begin by introducing certain subspaces of L_n^2 associated with S_h^1 and S_h^2. We define

$$V_h^1 \equiv \{\dot{y}_h - Ay_h \mid y_h \in S_h^1\}, \tag{1.18}$$

$$V_h^2 \equiv \{Bv_h \mid v_h \in S_h^2\}. \tag{1.19}$$

We denote by $V_h^1 + V_h^2$ the closed linear span of $V_h^1 \cup V_h^2$. Then $V_h^1 + V_h^2$ becomes a finite-dimensional subspace of L_n^2. It is easy to verify that

$$\lim_{h \downarrow 0} \inf_{w_h \in V_h^1 + V_h^2} \|w - w_h\|_{L_n^2} = 0$$

is satisfied, for all $w \in L_n^2$, provided (1.15) holds.

For a given Hilbert space H with some closed subspace H_1, we let $\mathbf{P}_{H_1}$ denote the *orthogonal projection* of H onto H_1. The error analysis hinges on the behavior of the operators $\mathfrak{L}_h^\varepsilon$: $L_n^2 \to H_{0n}^1 \times L_m^2$ defined by

$$\mathfrak{L}_h^\varepsilon(f) \equiv (\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon), \tag{1.20}$$

where $(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon)$ solves (1.16).

LEMMA 5. *Let* (0.3), (0.6), *and* (1.15) *hold. Then the mapping* $\mathfrak{L}_h^\varepsilon$ *defined above is a bounded linear transformation from* L^2 *into* $H_{0n}^1 \times L_m^2$ *with* Range$(\mathfrak{L}_h^\varepsilon) \subseteq S_h^1 \times S_h^2$. $(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon)$, *the image of* f *under* $\mathfrak{L}_h^\varepsilon$, *is characterized by the variational equation*

$$a_\varepsilon\left(\begin{bmatrix} \hat{x}_h^\varepsilon \\ \hat{u}_h^\varepsilon \end{bmatrix}, \begin{bmatrix} y_h \\ v_h \end{bmatrix}\right) = \theta_\varepsilon(f_h) \tag{1.21}$$

for all $(y_h, v_h) \in S_h^1 \times S_h^2$, *where* $f_h \equiv \mathbf{P}_{V_h^1 + V_h^2} f$.

Proof. Since $\langle f, \dot{y}_h - Ay_h - Bv_h\rangle = \langle f_h, \dot{y}_h - Ay_h - Bv_h\rangle$, we replace f by f_h in (1.17) and get (1.21). □

An immediate consequence of the above lemma is

$$\mathfrak{L}_h^\varepsilon(f) = \mathfrak{L}_h^\varepsilon(f_h). \tag{1.22}$$

In solving (1.16), unless $f \in V_h^1 + V_h^2$, it is in general true that

$$J_\varepsilon(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon, f) \to \infty, \quad \text{as } \varepsilon \downarrow 0, \tag{1.23}$$

because no (y_h, v_h) in $S_h^1 \times S_h^2$ can satisfy the constraint $\dot{y}_h - Ay_h - Bv_h - f = 0$. This makes one wonder whether the solution $(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon)$ in Lemma 5 will converge as $\varepsilon \downarrow 0$. However, solving (1.16) is equivalent to solving

$$\operatorname*{Min}_{(x,u)\in S_h^1\times S_h^2} J_\varepsilon(x, u, f_h). \tag{1.24}$$

Because solutions $(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon)$ of (1.24) do converge as $\varepsilon \downarrow 0$, we conclude that

$$s\text{-}\lim_{\varepsilon\downarrow 0} (\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon) = (\hat{x}_h, \hat{u}_h) \quad \text{in } H_{0n}^1 \times L_m^2$$

for some unique $(\hat{x}_h, \hat{u}_h)$. This defines a mapping $\mathfrak{L}_h$ by

$$\mathfrak{L}_h\colon L_n^2 \to H_{0n}^1 \times L_m^2, \qquad \mathfrak{L}_h(f) \equiv (\hat{x}_h, \hat{u}_h). \tag{1.25}$$

Then Range$(\mathfrak{L}_n) \subset S_h^1 \times S_h^2$. Arguing in the same manner as in the proof of Theorem 2, we have

COROLLARY 6. *Let* (0.3), (0.6), *and* (1.15) *hold. Then, for each* $h > 0$, *the mapping* $\mathfrak{L}_h$ *defined above is a bounded linear transformation from* L_n^2 *into* $H_{0n}^1 \times L_m^2$ *with* Range$(\mathfrak{L}_h) \subseteq S_h^1 \times S_h^2$. $(\hat{x}_h, \hat{u}_h)$, *the image of* f *under* $\mathfrak{L}_h$, *is the unique solution of*

$$\begin{cases} \text{Min } J(x,u) \\ \text{subject to} \\ (x,u) \in S_h^1 \times S_h^2 \\ \dot{x} = Ax + Bu + f_h, \qquad f_h = \mathbf{P}_{V_h^1+V_h^2} f \\ x(0) = 0. \end{cases} \tag{1.26}$$

Furthermore, for any $f \in L_n^2$, *we have*

$$\|\hat{x}_h^\varepsilon - \hat{x}_h\|_{H_{0n}^1} + \|\hat{u}_h^\varepsilon - \hat{u}_h\|_{L_m^2} \leqslant \overline{K} \cdot \|(\hat{x}_h, \hat{u}_h)\|_{H_{0n}^1 \times L_m^2} \cdot \|\mathfrak{L}_n\|^2 \cdot \varepsilon \tag{1.27}$$

for all $\varepsilon > 0$. □

It is obvious that the properties

$$\mathfrak{L}_h(f) = \mathfrak{L}_h(f_h), \qquad \dot{\hat{x}}_h - A\hat{x}_h - B\hat{u}_h = f_h = \mathbf{P}_{V_h^1+V_h^2} f \tag{1.28}$$

are satisfied for all $f \in L_n^2$. Also, the analogue of (1.14) for problem (1.16) is

$$a\left(\begin{bmatrix} \hat{x}_h \\ \hat{u}_h \end{bmatrix}, \begin{bmatrix} y_h \\ v_h \end{bmatrix}\right) + \langle \hat{p}_h, \dot{y}_h - Ay_h - Bv_h \rangle_{L_n^2} = 0 \quad \forall (y_h, v_h) \in S_h^1 \times S_h^2. \tag{1.29}$$

2. An Abstract Approximation Problem. In order to obtain optimal error estimates for solutions of (1.16), it is necessary to consider the following *abstract approximation problem*: "Let $\{G_i^1\}$, $\{G_i^2\}$ be two families of increasing finite-dimensional subspaces of a Hilbert space H such that

$$\lim_{i\to\infty} \inf_{g_i \in G_i^1} \|x - g_i\|_H = 0 \quad \forall x \in H. \tag{2.1}$$

$$\text{(2.2)}\qquad \begin{cases} G_i^1 \subseteq G_j^1,\ G_i^2 \subseteq G_j^2 & \text{for } i<j, \\ G_{i_{k_1}}^1 \subsetneqq G_{i_{k_2}}^1,\ G_{i_{k_1}}^2 \subsetneqq G_{i_{k_2}}^2 & \text{if } k_1 < k_2, \text{ for some increasing sequence } \{i_k\} \subseteq \mathbf{N}. \end{cases}$$

Then, for any $f \in H$, does there always exist a sequence

$$\{(g_i^1, g_i^2) \mid g_i^1 \in G_i^1,\ g_i^2 \in G_i^2\}_{i=1}^{\infty}$$

satisfying

$$\text{(2.3)}\qquad \mathbf{P}_{G_i^1+G_i^2} f = g_i^1 + g_i^2$$

such that

$$\text{(2.4)}\qquad \varlimsup_{i\to\infty} \left[\|g_i^1\| + \|g_i^2\|\right] < \infty?\text{''}.$$

Since this problem has quite independent interest, we study it separately in this section. In Section 3, we will apply these results using $G_i^1 = V_{h_i}^1$ and $G_i^2 = V_{h_i}^2$, the spaces defined in (1.18) and (1.19).

We let

$$\text{(2.5)}\qquad \tilde{G}_i^2 \equiv G_i^2 \ominus (G_i^1 \cap G_i^2)$$

be the orthogonal complement of $G_i^1 \cap G_i^2$ in G_i^2 and form the orthogonal decomposition

$$\text{(2.6)}\qquad G_i^2 = (G_i^1 \cap G_i^2) \oplus \tilde{G}_i^2.$$

Then

$$\text{(2.7)}\qquad G_i^1 + G_i^2 = G_i^1 \dotplus \tilde{G}_i^2$$

holds, where the right-hand side becomes a (in general, nonorthogonal) direct sum.

Definition 7. For any two closed subspaces H_1 and H_2 of a Hilbert space H, we define [8]

$$\cos(H_1, H_2) \equiv \begin{cases} \sup\limits_{\|f_1\|=\|f_2\|=1} |\langle f_1, f_2 \rangle| & \text{if } H_1 \neq \{0\},\ H_2 \neq \{0\}, \\ 0 & \text{if } H_1 = \{0\} \text{ or } H_2 = \{0\}. \end{cases} \quad \square$$

THEOREM 8. *Let $\{G_i^1\}_i$ and $\{G_i^2\}_i$ be two sequences of finite-dimensional subspaces in H satisfying* (2.1) *and* (2.2) *For any $f \in H$, there exists a sequence $\{(g_i^1, g_i^2) \in G_i^1 \times G_i^2 \mid$* (2.3) *holds$\}$ satisfying* (2.4) *if and only if there exists some $\mu \geqslant 0$ such that*

$$\text{(2.8)}\qquad \cos(G_i^1, \tilde{G}_i^2) \leqslant \mu < 1 \quad \forall i \in \mathbf{Z}^+.$$

Proof. (only if) Assume the contrary that (2.8) does not hold. Then there exists a sequence $\{i_j\}$ such that

$$\text{(2.9)}\qquad \cos\left(G_{i_j}^1, \tilde{G}_{i_j}^2\right) \equiv \mu_{i_j} \to 1 \quad \text{as } j \to \infty.$$

This is easily seen to be equivalent to

$$\text{(2.10)}\qquad \inf_{\substack{\|x\|=1 \\ x \in \tilde{G}_{i_j}^2}} \left\|x - \mathbf{P}_{G_{i_j}^1} x\right\| \equiv \alpha_{i_j} \to 0 \quad \text{as } j \to \infty.$$

Here we assume that $\tilde{G}^2_{i_j}$ is nontrivial ($\not\equiv \{0\}$), thus $\alpha_{i_j} > 0$. Because $\tilde{G}^2_{i_j}$ is finite-dimensional, (2.10) attains its minimum at some $x_{i_j} \in \tilde{G}^2_{i_j}$:

$$\|x_{i_j} - \mathbf{P}_{G^1_{i_j}} x_{i_j}\| = \inf_{\substack{\|x\|=1 \\ x \in \tilde{G}^2_{i_j}}} \|x - \mathbf{P}_{G^1_{i_j}} x\| = \alpha_{i_j}, \qquad \|x_{i_j}\| = 1. \tag{2.11}$$

$$\lim_{j\to\infty} \alpha_{i_j} = 0. \tag{2.12}$$

For any $x \in G^1_i + G^2_i$, by (2.7), we have a unique representation

$$x = g^1_i(x) + g^2_i(x), \quad g^1_i(x) \in G^1_i,\, g^2_i(x) \in \tilde{G}^2_i.$$

We define

$$P_i\colon G^1_i + G^2_i \to G^1_i, \qquad P_i x \equiv g^{(1)}_i(x), \tag{2.13}$$

and let $\|P_i\|_{\mathcal{L}(G^1_i+G^2_i, G^1_i)} \equiv \gamma_i$, $i \in \mathbf{Z}^+$. Note here that $\gamma_i \geqslant 1$.

Now we choose a subsequence $\{i_k\}_{k\in\mathbf{Z}^+}$ of positive integers and a sequence $\{\beta_k\}_{k\in\mathbf{Z}^+}$ of increasing positive real numbers with the following properties:

$$\begin{cases} \beta_k \equiv \left(\alpha_{i_k} 2^k\right)^{-1/3} \alpha_1^{1/3}, \quad k \in \mathbf{Z}^+, \alpha_{i_1} \equiv \alpha_1, \\ \beta_k \geqslant \gamma_{i_l} \qquad (1 \leqslant l < k), k \in \mathbf{Z}^+, \\ \lim\limits_{k\to\infty} \left[\beta_k - \left(1 + \gamma_{i_k}\right) \sum\limits_{j=1}^{k-1} \beta_j\right] = +\infty. \end{cases} \tag{2.14}$$

The constructability of such sequences $\{i_k\}$ and $\{\beta_k\}$ is guaranteed by (2.12).

Let

$$f_0 \equiv \sum_{k=1}^{\infty} \beta_k \left(x_{i_k} - \mathbf{P}_{G^1_{i_k}} x_{i_k}\right), \tag{2.15}$$

where $x_{i_k} \in \tilde{G}^2_{i_k}$ is defined through (2.11). Using (2.12)–(2.14), we easily verify that $f_0 \in H$.

Let

$$\xi_l \equiv \sum_{j=1}^{l} \beta_j \left(x_{i_j} - \mathbf{P}_{G^1_{i_j}} x_{i_j}\right), \qquad \eta_l \equiv f_0 - \xi_l. \tag{2.16}$$

Since $\xi_l \in G^1_{i_l} + G^2_{i_l}$, we have

$$\begin{aligned} \mathbf{P}_{G^1_{i_l}+G^2_{i_l}} f_0 &= \mathbf{P}_{G^1_{i_l}+G^2_{i_l}}[\xi_l + \eta_l] = \xi_l + \mathbf{P}_{G^1_{i_l}+G^2_{i_l}}[\eta_l] \\ &= \left(I - P_{i_l}\right)\left[\beta_l x_{i_l} + \sum_{j=1}^{l-1} \beta_j x_{i_j} + \mathbf{P}_{G^1_{i_l}+G^2_{i_l}} \eta_l\right] \\ &\quad + \left[P_{i_l}\left(\sum_{j=1}^{l} \beta_j x_{i_j} + \mathbf{P}_{G^1_{i_l}+G^2_{i_l}} \eta_l\right) - \sum_{j=1}^{l} \beta_j \mathbf{P}_{G^1_{i_j}} x_{i_j}\right]. \end{aligned} \tag{2.17}$$

In the above, the first term belongs to $\tilde{G}^2_{i_l}$ and the second belongs to $G^1_{i_l}$. Thus we have the unique representation

$$\mathbf{P}_{G^1_{i_l}+G^2_{i_l}} f_0 = (2.17) \equiv g^2_{i_l}(f_0) + g^1_{i_l}(f_0). \tag{2.18}$$

We wish to show that $\lim_{l\to\infty} \| g_{i_l}^2(f_0)\| = \infty$. From (2.16), (2.13), and (2.14),

$$\|\eta_l\| \leq \sum_{j=l+1}^{\infty} \beta_j \left\| x_{i_j} - \mathbf{P}_{G_{i_j}^1} x_{i_j} \right\| = \sum_{j=l+1}^{\infty} \beta_j \alpha_{i_j} = \sum_{j=l+1}^{\infty} \frac{\alpha_1}{\beta_j^2 2^j} \tag{2.19}$$
$$\leq \frac{\alpha_1}{\beta_{l+1}^2} \leq \frac{\alpha_1}{\beta_{l+1}} \frac{1}{\gamma_{i_l}} = \frac{\alpha_1}{\beta_{l+1}} \left\| P_{i_l} \right\|^{-1}.$$

Because $x_{i_l} \in \tilde{G}_{i_l}^2$, $(I - P_{i_l})x_{i_l} = x_{i_l}$, and because $\mathbf{P}_{G_{i_l}^1 + G_{i_l}^2}$ is an orthogonal projection, (2.13), (2.17)–(2.19) give

$$\left\| g_{i_l}^2(f_0) \right\| \geq \left\| \beta_l x_{i_l} \right\| - \left\| (I - P_{i_l}) \sum_{j=1}^{l-1} \beta_j x_{i_j} \right\| - \left\| (I - P_{i_l}) \mathbf{P}_{G_{i_l}^1 + G_{i_l}^2} \eta_l \right\| \tag{2.20}$$
$$\geq \left[\beta_l - \left(1 + \left\| P_{i_l} \right\|\right) \sum_{j=1}^{l-1} \beta_j \right] - \left(1 + \left\| P_{i_l} \right\|\right) \frac{\alpha_1}{\beta_{l+1}} \left\| P_{i_l} \right\|^{-1}.$$

The bracketed term in (2.20) tends to $+\infty$ as $l \to \infty$ by (2.14), and $\| P_{i_l} \| \geq 1$ gives

$$\left(1 + \left\| P_{i_l} \right\|\right) \frac{\alpha_1}{\beta_{l+1}} \left\| P_{i_l} \right\|^{-1} \to 0 \quad \text{as } l \to \infty.$$

Hence

$$\lim_{l\to\infty} \left\| g_{i_l}^2(f_0) \right\| = \infty. \tag{2.21}$$

For this f_0, if there still exist sequences $\{\bar{g}_i^1 \in G_i^1\}$, $\{\bar{g}_i^2 \in G_i^2\}$ satisfying (2.3) and (2.4), then

$$\mathbf{P}_{G_{i_l}^1 + G_{i_l}^2} f_0 = (2.17) = g_{i_l}^1(f_0) + g_{i_l}^2(f_0) = (2.3) = \bar{g}_{i_l}^1 + \bar{g}_{i_l}^2.$$

Therefore

$$g_{i_l}^1(f_0) - \bar{g}_{i_l}^1 = \bar{g}_{i_l}^2(f_0) \in G_{i_l}^1 \cap G_{i_l}^2,$$

since the left-hand side belongs to $G_{i_l}^1$ while the right-hand side belongs to $G_{i_l}^2$. So by (2.5), using $g_{i_l}^2(f_0) \in \tilde{G}_{i_l}^2$, we get

$$\left\langle \bar{g}_{i_l}^2 - g_{i_l}^2(f_0), g_{i_l}^2(f_0) \right\rangle = 0,$$
$$\left\| g_{i_l}^2(f_0) \right\|^2 = \left\langle \bar{g}_{i_l}^2, g_{i_l}^2(f_0) \right\rangle \leq \tfrac{1}{2} \left\| g_{i_l}^2(f_0) \right\|^2 + \tfrac{1}{2} \left\| \bar{g}_{i_l}^2 \right\|^2,$$

implying $\| g_{i_l}^2(f_0)\|^2 \leq \| \bar{g}_{i_l}^2 \|^2$. But the right-hand side is bounded by (2.4), contradicting (2.21). Hence the proof of the "only if" part is complete.

(if) For any $x \in G_i^1 + G_i^2$, we have a unique representation

$$x = g_i^1(x) + g_i^2(x), \qquad \left(g_i^1(x), g_i^2(x)\right) \in G_i^1 \times \tilde{G}_i^2.$$

We define P_i as in (2.13). Then for any $x \in G_i^1 + G_i^2$, we have

$$x = P_i x + (I - P_i)x,$$

So

$$\begin{aligned} \|x\|^2 = \langle x, x\rangle &\geqslant \|P_i x\|^2 + \|(I-P_i)x\|^2 - 2|\langle P_i x, (I-P_i)x\rangle| \\ &\geqslant \|P_i x\|^2 + \|(I-P_i)x\|^2 - 2\mu\|P_i x\|\,\|(I-P_i)x\| \\ &= (1-\mu^2)\|P_i x\|^2 + \left(\mu\|P_i x\| - \|(I-P_i)x\|\right)^2 \\ &\geqslant (1-\mu^2)\|P_i x\|^2. \end{aligned} \tag{2.22}$$

For any $f \in H$, we now let

$$g_i^1(f) \equiv P_i\left(\mathbf{P}_{G_i^1+G_i^2} f\right) \in G_i^1.$$

From (2.22), we get

$$\|g_i^1(f)\| \leqslant \frac{1}{(1-\mu^2)^{1/2}}\|\mathbf{P}_{G_i^1+G_i^2} f\| \leqslant \frac{1}{(1-\mu^2)^{1/2}}\|f\| \quad \forall f \in H, \forall i \in \mathbf{Z}^+,$$

and

$$\|g_i^2(f)\| \equiv \|\mathbf{P}_{G_i^1+G_i^2} f - g_i^1(f)\| \leqslant \|\mathbf{P}_{G_i^1+G_i^2} f\| + \|g_i^1(f)\| \leqslant \left[1 + \frac{1}{(1-\mu^2)^{1/2}}\right]\|f\|.$$

This proves (2.4) □.

COROLLARY 9. *Let $\{G_i^1\}_i$ and $\{G_i^2\}_i$ be two sequences of finite-dimensional subspaces in H satisfying* (2.1) *and* (2.2). *If they satisfy*

$$G_i^1 \cap G_i^2 = \{0\} \quad \forall i \in \mathbf{Z}^+,$$

then

$$\sup_i \cos\left(G_i^1, \tilde{G}_i^2\right) = 1.$$

Therefore there exist some $f \in H$ such that (2.3) *and* (2.4) *fail to hold.*

Proof. Obvious. □

Remark 10. A careful examination of the proof of Theorem 8 shows that (2.4) remains valid under the more general assumption that $\{G_i^1\}$, $\{G_i^2\}$ are sequences of *closed* subspaces only, i.e., $\{G_i^1\}$, $\{G_i^2\}$ need not be finite-dimensional. □

Remark 11. According to von Neumann's alternating projection theorem [5], for any two closed subspaces G^1, G^2 in a Hilbert space H, one has

$$s\text{-}\lim_{k\to\infty} (\mathbf{P}_{G^1}\mathbf{P}_{G^2})^k x = \mathbf{P}_{G^1\cap G^2} x \quad \forall x \in H.$$

Therefore, for any $f \in H$, its component in $G_i^1 \cap G_i^2$ can be obtained iteratively as above. □

3. Finite Element-Penalty Error Estimates. In Section 1, we have assumed that the family of products of finite-dimensional spaces $\{S_h^1 \times S_h^2 \subset H_{0n}^1 \times L_m^2 \mid 0 \leqslant h \leqslant h_0\}$ is a *continuous* one parameter-family. In this section, we consider instead the simpler case, namely, we assume that we have a discrete one-parameter (sub-)family of finite-dimensional product spaces $\{S_{h_i}^1 \times S_{h_i}^2 \mid 0 \leqslant h_i \leqslant h_0,\ i \in \mathbf{Z}^+\}$ with the following properties.

(3.1)
$$\begin{cases} \lim\limits_{h_i \downarrow 0} \inf\limits_{y_{h_i} \in S^1_{h_i}} \|y - y_{h_i}\|_{H^1_{0n}} = 0 \quad \forall y \in H^1_{0n}, \\ \lim\limits_{h_i \downarrow 0} \inf\limits_{v_{h_i} \in S^2_{h_i}} \|v - v_{h_i}\|_{L^2_m} = 0 \quad \forall v \in L^2_m, \end{cases}$$

(3.2)
$$S^1_{h_i} \subsetneq S^1_{h_j}, S^2_{h_i} \subsetneq S^2_{h_j} \quad \text{if } h_i > h_j.$$

From $S^1_{h_i}$ and $S^2_{h_i}$, we construct $V^1_{h_i}$ and $V^2_{h_i}$ as in (1.18) and (1.19). In the sequel, we will denote $S^1_{h_i}$, $S^2_{h_i}$, $V^1_{h_i}$, and $V^2_{h_i}$ simply as S^1_i, S^2_i, V^1_i, and V^2_i, respectively.

Condition (H). We say that the family $\{S^1_i \times S^2_i\}_{i \in \mathbf{Z}^+}$ satisfies *condition* (H) if the associated family $\{V^1_i, V^2_i\}_{i \in \mathbf{Z}^+}$ satisfies

$$\cos\left(V^1_i, \tilde{V}^2_i\right) \leq \mu < 1 \quad \forall i \in \mathbf{Z}^+,$$

where $\tilde{V}^2_i = V^2_i \ominus (V^1_i \cap V^2_i)$ (cf. (2.5)).

A further auxiliary condition on B will be needed. From now on we assume that B and $\{S^2_i\}$ satisfy the following condition:

(3.3) for any sequence $\{w_i \mid w_i \in V^2_i\}$ satisfying $\overline{\lim}_{i\to\infty} \|w_i\| < \infty$, there exists a sequence $\{v_i \mid v_i \in S^2_i\}$ such that $w_i = Bv_i$ and $\overline{\lim}_{i\to\infty} \|v_i\| < \infty$.

It is easy to see that if B is 1-1, then (3.3) is valid for any $\{S^2_i\}$.

We are now in a position to prove the main theorem in this paper.

THEOREM 12 (MAIN ESTIMATES). *Given a family of finite-dimensional subspaces* $\{S^1_i \times S^2_i \mid i \in \mathbf{Z}^+\}$ *satisfying* (3.1)–(3.3), *let* $\{(\hat{x}^\varepsilon_i, \hat{u}^\varepsilon_i) (\equiv (\hat{x}^\varepsilon_{h_i}, \hat{u}^\varepsilon_{h_i})) \mid i \in \mathbf{Z}^+\}$ *be the solutions of* (1.16). *Let* (0.3) *and* (0.6) *hold and let* $(\hat{x}, \hat{u})$ *be the solution of the optimal control problem* (0.1) *and* (0.2). *If condition* (H) *is satisfied, then, for every* $h_i > 0$, $\varepsilon > 0$, *and* $f \in L^2_n$, *we have*

(3.4)
$$\|\hat{x}^\varepsilon_i - \hat{x}\|_{H^1_{0n}} + \|\hat{u}^\varepsilon_i - \hat{u}\|_{L^2_m} \leq C_1(\hat{x}, \hat{u})\varepsilon + C_2\left[\inf_{y_i \in S^1_i} \|\hat{x} - y_i\|_{H^1_{0n}} + \inf_{v_i \in S^2_i} \|\hat{u} - v_i\|_{L^2_m}\right],$$

where $C_1(\hat{x}, \hat{u})$ *depends on* $\|(\hat{x}, \hat{u})\|$ (*or equivalently, on* $\|f\|$) *with a linear growth rate and* $C_2 > 0$ *is a constant independent of* $\hat{x}$, $\hat{u}$, f, ε, *and* h_i.

We first prove the following two lemmas.

LEMMA 13. *Let the hypothesis of Theorem* 12 *hold. Let* $\mathfrak{L}_i \equiv \mathfrak{L}_{h_i}$ *be the mapping defined in* (1.25), *and* $\mathfrak{L}$ *be as in* (1.4). *Then*

(3.5)
$$\mathfrak{L}_i f \to \mathfrak{L} f \qquad \text{in } H^1_{0n} \times L^2_m, \text{ as } h_i \downarrow 0 \ \forall f \in L^2_n$$

if and only if condition (H) *holds.*

Proof. (if) Assume (H). By Theorem 8 there exists a sequence $\{(w^1_i, w^2_i) \in V^1_i \times V^2_i \mid f_i \equiv \mathbf{P}_{V^1_i + V^2_i} f = w^1_i + w^2_i\}$ such that

(3.6)
$$\overline{\lim_{i\to\infty}} \left[\|w^1_i\| + \|w^2_i\|\right] < \infty.$$

Since $\mathfrak{L}_i f = (\hat{x}_i, \hat{u}_i)$, (1.28) gives

$$f_i = \mathbf{P}_{V^1_i + V^2_i} f = \left(\dot{\hat{x}}_i - A\hat{x}_i\right) + \left(-B\hat{u}_i\right) \equiv \bar{w}^1_i + \bar{w}^2_i,$$

where $(\hat{x}_i, \hat{u}_i)$ is uniquely characterized by

$$J(\hat{x}_i, \hat{u}_i) = \operatorname*{Min}_{(x,u)\in S_i^1\times S_i^2} J(x,u) \quad \text{subject to } \dot{x} - Ax - Bu = f_i,\ x(0) = 0.$$

Since $(w_i^1, w_i^2) \in V_i^1 \times V_i^2$, there exists $(y_i, v_i) \in S_i^1 \times S_i^2$ such that

$$w_i^1 = \dot{y}_i - Ay_i, \qquad w_i^2 = Bv_i.$$

Therefore

$$J(\hat{x}_i, \hat{u}_i) \leqslant J(y_i, v_i). \tag{3.7}$$

But, by (3.6) and (3.3), we can choose v_i such that $\overline{\lim}\|v_i\| < \infty$, so

$$\overline{\lim_{i\to\infty}}\, J(y_i, v_i) \leqslant C\, \overline{\lim_{i\to\infty}} \left\{\|w_i^1\|_{L_n^2} + \|v_i\|_{L_n^2}\right\} < \infty$$

for some constant C depending on N_1, N_2, and M only. Therefore, from (3.7), $\{(\hat{x}_i, \hat{u}_i)\}$ has a subsequence converging weakly in $H_{0n}^1 \times L_m^2$ to some $(\bar{x}, \bar{u})$. Because of the lower semicontinuity of J in $H_{0n}^1 \times L_m^2$, it is easy to see that this weak convergence is also strong, and the weak limit $(\bar{x}, \bar{u})$ satisfies

$$\dot{\bar{x}} - A\bar{x} - B\bar{u} = \lim_{i\to\infty} f_i = f.$$

Thus from uniqueness we see that $(\bar{x}, \bar{u}) = (\hat{x}, \hat{u})$, the unique solution to the optimal control problem (0.1) and (0.2). Since every subsequence of $(\hat{x}_i, \hat{u}_i)$ converges to $(\hat{x}, \hat{u})$ strongly, we conclude

$$\operatorname*{s-lim}_{i\to\infty} (\hat{x}_i, \hat{u}_i) = \operatorname*{s-lim}_{i\to\infty} \mathfrak{L}_i f = (\hat{x}, \hat{u}) = \mathfrak{L} f.$$

So (3.5) is proved.

(only if) If $\mathfrak{L}_i f$ converges to $\mathfrak{L} f$ for every f, we can choose

$$w_i^1 \equiv \dot{\hat{x}}_i - A\hat{x}_i, \qquad w_i^2 \equiv -B\hat{u}_i,$$

where $(\hat{x}_i, \hat{u}_i) = \mathfrak{L}_i f$. Because $(\hat{x}_i, \hat{u}_i) = \mathfrak{L}_i f$ converges to $(\hat{x}, \hat{u}) = \mathfrak{L} f$ strongly in $H_{0n}^1 \times L_m^2$, we have

$$\lim_{i\to\infty} \left[\|w_i^1\|_{L_n^2} + \|w_i^2\|_{L_n^2}\right] = \|\dot{\hat{x}} - A\hat{x}\|_{L_n^2} + \|B\hat{u}\|_{L_n^2} < \infty,$$

proving (H). $\square$

Remark 14. By the uniform boundedness principle and Lemma 13, we conclude

$$\sup_{i\in\mathbf{Z}^+} \{\|\mathfrak{L}_i\|\} < \infty \tag{3.8}$$

under condition (H). $\square$

LEMMA 15. *Assume the hypotheses of Theorem* 12. *Let* $(\tilde{x}_i, \tilde{u}_i) \in S_i^1 \times S_i^2$ *satisfy*

$$a\left(\begin{bmatrix}\hat{x} - \tilde{x}_i\\ \hat{u} - \tilde{u}_i\end{bmatrix}, \begin{bmatrix} y_i \\ v_i\end{bmatrix}\right) = 0 \tag{3.9}$$

for all $(y_i, v_i) \in S_i^1 \times S_i^2$. *Then we have*

$$\|\tilde{x}_i - \hat{x}_i\|_{H_{0n}^1} + \|\tilde{u}_i - \hat{u}_i\|_{L_m^2} \leqslant C\|\mathfrak{L}_i\| \left[\inf_{y_i\in S_i^1} \|\hat{x} - y_i\|_{H_{0n}^1} + \inf_{v_i\in S_i^2} \|\hat{u} - v_i\|_{L_m^2}\right] \tag{3.10}$$

for some constant $C > 0$ *independent of* $(\hat{x}, \hat{u})$.

Proof. Let $\tilde{f}_i$ be defined by

$$\tilde{f}_i \equiv \dot{\tilde{x}}_i - A\tilde{x}_i - B\tilde{u}_i.$$

From (3.9) and [1] we have

$$\|\hat{x} - \tilde{x}_i\|_{H^1_{0n}} + \|\hat{u} - \tilde{u}_i\|_{L^2_m} \le C_3\left[\inf_{y_i \in S_i^1} \|\hat{x} - y_i\|_{H^1_{0n}} + \inf_{v_i \in S_i^2} \|\hat{u} - v_i\|_{L^2_m}\right] \tag{3.11}$$

for some constant C_3 depending on N_1, N_2, and M only. Therefore, by (3.11)

$$\begin{aligned} \|f - \tilde{f}_i\|_{L^2_n} &= \|(\dot{\hat{x}} - A\hat{x} - B\hat{u}) - (\dot{\tilde{x}}_i - A\tilde{x}_i - B\tilde{u}_i)\|_{L^2_n} \\ &\le \|(\dot{\hat{x}} - \dot{\tilde{x}}_i) - A(\hat{x} - \tilde{x}_i)\|_{L^2_n} + \|B(\hat{u} - \tilde{u}_i)\|_{L^2_n} \\ &\le C_4\left[\inf_{y_i \in S_i^1} \|\hat{x} - y_i\|_{H^1_{0n}} + \inf_{v_i \in S_i^2} \|\hat{u} - v_i\|_{L^2_m}\right] \end{aligned} \tag{3.12}$$

for some constant C_4 independent of $\hat{x}$, $\hat{u}$, f, S_i^1 and S_i^2.

From (1.14) and (3.9), we see that $(\tilde{x}_i, \tilde{u}_i)$ is characterized by the variational equation

$$a\left(\begin{bmatrix} \tilde{x}_i \\ \tilde{u}_i \end{bmatrix}, \begin{bmatrix} y_i \\ v_i \end{bmatrix}\right) + \langle \hat{p}, \dot{y}_i - Ay_i - Bv_i \rangle_{L^2_n} = 0 \quad \forall (y_i, v_i) \in S_i^1 \times S_i^2.$$

Therefore $(\tilde{x}_i, \tilde{u}_i)$ is the unique solution in $S_i^1 \times S_i^2$ of

$$\begin{cases} \displaystyle\operatorname*{Min}_{(x,u) \in S_i \times S_i^2} \quad J(x, u) \\ \text{subject to} \\ \dot{x} - Ax - Bu = \tilde{f}_i, \qquad x(0) = 0. \end{cases} \tag{3.13}$$

Now, consider $(\tilde{x}_i, \tilde{u}_i)$. Let $f_i \equiv \mathbf{P}_{V_i^1 + V_i^2} f$. We have

$$\begin{aligned} \|f - f_i\|_{L^2_n} &= \inf_{g_i \in V_i^1 + V_i^2} \|f - g_i\|_{L^2_n} \\ &= \inf_{y_i, v_i} \|(\dot{\hat{x}} - A\hat{x} - B\hat{u}) - (\dot{y}_i - Ay_i - Bv_i)\|_{L^2_n} \\ &\le \inf_{y_i} \|[(\dot{\hat{x}} - \dot{y}_i) - A(\hat{x} - y_i)]\|_{L^2_n} + \inf_{v_i} \|B(\hat{u} - v_i)\|_{L^2_n} \\ &\le C_5\left[\inf_{y_i \in S_i^1} \|\hat{x} - y_i\|_{H^1_{0n}} + \inf_{v_i \in S_i^2} \|\hat{u} - v_i\|_{L^2_m}\right] \end{aligned} \tag{3.14}$$

for some constant C_5 independent of $\hat{x}$, $\hat{u}$, f, S_i^1 and S_i^2. Hence, (1.26), (3.13), (3.12), and (3.14) give

$$\begin{aligned} \|\tilde{x}_i - \hat{x}_i\|_{H^1_{0n}} + \|\tilde{u}_i - \hat{u}_i\|_{L^2_m} &= \|\mathfrak{L}_i(\tilde{f}_i) - \mathfrak{L}_i(f_i)\| \\ &\le \|\mathfrak{L}_i\| \, \|\tilde{f}_i - f_i\|_{L^2_n} \le \|\mathfrak{L}_i\| \left[\|\tilde{f}_i - f\|_{L^2_n} + \|f - f_i\|_{L^2_n}\right] \\ &\le C_6\|\mathfrak{L}_i\|\left[\inf_{y_i \in S_i^1} \|\hat{x} - y_i\|_{H^1_{0n}} + \inf_{v_i \in S_i^2} \|\hat{u} - v_i\|_{L^2_m}\right] \end{aligned}$$

with $C_6 \equiv \max(C_4, C_5)$. Thus (3.10) is proved. $\square$

We now give

Proof of Theorem 12. We use the triangle inequality

$$(3.15)\quad \begin{aligned} &\|\hat{x}_i^\varepsilon - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_i^\varepsilon - \hat{u}\|_{L_m^2} \\ &\leq \|\hat{x}_i^\varepsilon - \hat{x}_i\|_{H_{0n}^1} + \|\hat{x}_i - \tilde{x}_i\|_{H_{0n}^1} + \|\tilde{x}_i - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_i^\varepsilon - \hat{u}_i\|_{L_m^2} \\ &\quad + \|\hat{u}_i - \tilde{u}_i\|_{L_m^2} + \|\tilde{u}_i - \hat{u}\|_{L_m^2} \equiv \sum_{i=1}^{6} T_i. \end{aligned}$$

By Corollary 6, we have

$$(3.16)\qquad T_1 + T_4 \leq \overline{K} \cdot \|(\hat{x}_i, \hat{u}_i)\| \cdot \|\mathfrak{L}_i\|^2 \cdot \varepsilon \quad \forall \varepsilon > 0.$$

By Lemma 15, we have

$$(3.17)\qquad T_2 + T_5 \leq C_7 \|\mathfrak{L}_i\| \left[\inf_{y_i \in S_i^1} \|\hat{x} - y_i\|_{H_{0n}^1} + \inf_{v_i \in S_i^2} \|\hat{u} - v_i\|_{L_m^2} \right]$$

for some $C_7 > 0$.

By (3.11), we have

$$(3.18)\qquad T_3 + T_6 \leq C_8 \left[\inf_{y_i \in S_i^1} \|\hat{x} - y_i\|_{H_{0n}^1} + \inf_{v_i \in S_i^1} \|\hat{u} - v_i\|_{L_m^2} \right]$$

for some $C_8 > 0$.

We define

$$(3.19)\qquad C_1(\hat{x}, \hat{u}) \equiv \sup_i \overline{K} \cdot \|(\hat{x}_i, \hat{u}_i)\| \cdot \|\mathfrak{L}_i\|^2.$$

By (3.8) and the strong convergence of $(\hat{x}_i, \hat{u}_i)$ to $(\hat{x}, \hat{u})$, we see that $C_1(\hat{x}, \hat{u})$ is finite, and it depends on $\|(\hat{x}, \hat{u})\|$ with a linear growth rate. We also define

$$(3.20)\qquad C_2 \equiv C_8 + C_7 \cdot \sup_i \|\mathfrak{L}_i\|.$$

Then $C_2 < \infty$ by (3.8).

Combining (3.15)–(3.20), we conclude (3.4). □

Remark 16. One might try to prove Lemma 13 (and Theorem 12) from the boundedness of the multiplier $\hat{p}_h$ in (1.29). Actually, our argument above shows that if condition (H) is not satisfied, then not all multipliers $\hat{p}_h$ are bounded as $h \downarrow 0$. □

As in [1], we say that $S_h \subset H_l^s$ is an (r, s)-*system* if for all $v \in H_l^k(0, T)$, there exists $v_h \in S_h$ such that

$$\|v - v_h\|_{H_l^\eta} \leq C h^\mu \|v\|_{H_l^{\mu+\eta}} \quad \forall 0 \leq \eta \leq \min\{k, s\}, \eta \in \mathbf{N},$$

where $\mu = \min\{r - \eta, k - \eta\}$ and C is independent of h and v.

COROLLARY 17. *Let* $(\hat{x}_i^\varepsilon, \hat{u}_i^\varepsilon)$ *solve* (1.16). *Let* (0.3) *and* (0.6) *hold and* $(\hat{x}, \hat{u})$ *solve* (0.1) *and* (0.2) *with* $\hat{x} \in H_n^{l_1}$ *and* $\hat{u} \in H_m^{l_2}$. *Let* $S_i^1 \subseteq H_{0n}^1$ *be an* $(r_1, 1)$-*system and let* $S_i^2 \subseteq L_m^2$ *be an* $(r_2, 0)$-*system such that* (3.3) *and condition* (H) *are satisfied. Then for each* $h_i > 0$, $f \in L_n^2$, *and* $\varepsilon > 0$, *there exist constants* $C_1(\hat{x}, \hat{u})$ (*depending on* $\|(\hat{x}, \hat{u})\|$) *and* C_2 *such that*

$$(3.21)\quad \|\hat{x}_i^\varepsilon - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_i^\varepsilon - \hat{u}\|_{L_m^2} \leq C_1(\hat{x}, \hat{u})\varepsilon + C_2\left[h_i^{\mu_1} \|\hat{x}\|_{H_n^{\mu_1+1}} + h_i^{\mu_2} \|\hat{u}\|_{H_m^{\mu_2}}\right],$$

where $\mu_1 = \min(r_1 - 1, l_1 - 1)$ *and* $\mu_2 = \min(r_2, l_2)$. □

From (3.21) we see that optimal error bounds are obtained when $\mu_1 = \mu_2$ and $\varepsilon = O(h_i^{\mu_1})$. Therefore, if $(\hat{x}, \hat{u}) \in (H_n^{l+1} \cap H_{0n}^1) \oplus H_m^l$ for some $l \in \mathbf{Z}^+$, we usually choose

$$r_1 - 1 = r_2. \tag{3.22}$$

Condition (H) is a very strong assumption. If it is not satisfied, one can show that sharp error estimates like (3.4) are impossible.

THEOREM 18. *Let* (0.3) *and* (0.6) *hold, and let* $\{S_h^1 \times S_h^2\}_h$ *be a discrete (or continuous) one-parameter family of closed subspaces of* $H_{0n}^1 \times L_m^2$ *satisfying* (3.1)–(3.3) *(or* (1.15)*). If* (H) *is not satisfied, then there cannot exist nonnegative error estimation functions* $E_1(\varepsilon, \hat{x}, \hat{u})$ *and* $E_2(h, \hat{x}, \hat{u})$ *satisfying*

$$\lim_{\varepsilon \downarrow 0} E_1(\varepsilon, \hat{x}, \hat{u}) = 0, \qquad \lim_{h \downarrow 0} E_2(h, \hat{x}, \hat{u}) = 0 \tag{3.23}$$

such that

$$\|\hat{x}_h^\varepsilon - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_h^\varepsilon - \hat{u}\|_{L_m^2} \leq E_1(\varepsilon, \hat{x}, \hat{u}) + E_2(h, \hat{x}, \hat{u}). \tag{3.24}$$

Proof. Assume the contrary that (3.23) and (3.24) hold. By Corollary 6, we have

$$\begin{aligned} \|\hat{x}_h - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_h - \hat{u}\|_{L_m^2} &= \lim_{\varepsilon \downarrow 0} \left[\|\hat{x}_h^\varepsilon - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_h^\varepsilon - \hat{u}\|_{L_m^2}\right] \\ &\leq \lim_{\varepsilon \downarrow 0} \left[E_1(\varepsilon, \hat{x}, \hat{u}) + E_2(h, \hat{x}, \hat{u})\right] = E_2(h, \hat{x}, \hat{u}). \end{aligned}$$

Hence

$$\lim_{h \downarrow 0} \left[\|\hat{x}_h - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_h - \hat{u}\|_{L_m^2}\right] \leq \lim_{h \downarrow 0} E_2(h, \hat{x}, \hat{u}) = 0. \tag{3.25}$$

Let

$$w_h^1 \equiv \dot{\hat{x}}_h - A\hat{x}_h \in V_h^1, \qquad w_h^2 \equiv -B\hat{u}_h \in V_h^2.$$

For any $f \in L_n^2$, (1.26) gives

$$\mathbf{P}_{V_h^1 + V_h^2} f = w_h^1 + w_h^2,$$

and (3.25) gives

$$\lim_{h \downarrow 0} \left[\|w_h^1\|_{L_n^2}^2 + \|w_h^2\|_{L_n^2}^2\right] = \|\dot{\hat{x}} - A\hat{x}\|_{L_n^2}^2 + \|B\hat{u}\|_{L_n^2}^2 < \infty.$$

This means that (H) is satisfied, a contradiction. □

From the proofs of Theorems 12 and 18, we conclude

COROLLARY 19. *Let* (0.3), (0.6), *and* (3.1)–(3.3) *hold. Let* $(\hat{x}_i^\varepsilon, \hat{u}_i^\varepsilon)$ *and* $(\hat{x}, \hat{u})$ *denote, respectively, the solutions of* (1.16) *and* (0.1), (0.2), *and let* $\mathfrak{L}_i$ *be defined as in* (1.25). *Then the following conditions are equivalent:*

(1) *Condition* (H).

(2) $\sup_{i \in \mathbf{Z}^+} \{\|\mathfrak{L}_i\|\} < \infty$.

(3) *There exist* $C_1(\hat{x}, \hat{u}) > 0$ *depending on* $(\hat{x}, \hat{u})$ *only and* $C_2 > 0$ *independent of* $\hat{x}$, $\hat{u}, f, h_i$, *and* ε *such that, for all* $h_i > 0$ *and* $\varepsilon > 0$,

$$\|\hat{x}_i^\varepsilon - \hat{x}\|_{H_{0n}^1} + \|\hat{u}_i^\varepsilon - \hat{u}\|_{L_m^2} \leq C_1(\hat{x}, \hat{u})\varepsilon + C_2\left[\inf_{y \in S_i^1} \|\hat{x} - y\|_{H_{0n}^1} + \inf_{v \in S_i^2} \|\hat{u} - v\|_{L_m^2}\right].$$

(4) *There exist two error estimation functions* $E_1(\varepsilon, \hat{x}, \hat{u})$ *and* $E_2(h_i, \hat{x}, \hat{u})$ *such that*

$$\lim_{\varepsilon \downarrow 0} E_1(\varepsilon, \hat{x}, \hat{u}) = 0, \qquad \lim_{h_i \downarrow 0} E_2(h_i, \hat{x}, \hat{u}) = 0,$$

and

$$\|\hat{x}_i^\varepsilon - \hat{x}\|_{H^1_{0n}} + \|\hat{u}_i^\varepsilon - \hat{u}\|_{L^2_m} \leqslant E_1(\varepsilon, \hat{x}, \hat{u}) + E_2(h_i, \hat{x}, \hat{u}). \quad \square$$

Theorem 18 tells us that, without (H), ε and h appearing in error estimates must be coupled in general. The following gives one of the simplest type of such errors.

THEOREM 20. *Let* (0.3) *and* (0.6) *hold. Let* $(\hat{x}_\varepsilon, \hat{u}_\varepsilon)$ *solve* (1.1) *and* $(\hat{x}, \hat{u})$ *solve* (0.1), (0.2). *Assume that* $(\hat{x}_\varepsilon, \hat{u}_\varepsilon)$ *converges to* $(\hat{x}, \hat{u})$ *in* $(H_n^{l+1} \cap H^1_{0n}) \oplus H^l_m$ *for some* $l \in \mathbf{Z}^+$. *Let* $S_h^1 \subseteq H^1_{0n}$ *and* $S_h^2 \subseteq L^2_m$ *be* $(r_1, 1)$*- and* $(r_2, 0)$*-systems, respectively. Then there exist constants* $C_1 > 0$ *and* $C_2 > 0$ (*both depending on* $(\hat{x}, \hat{u})$) *such that, for* $\varepsilon > 0$ *sufficiently small,*

$$\|\hat{x}_h^\varepsilon - \hat{x}\|_{H^1_{0n}} + \|\hat{u}_h^\varepsilon - \hat{u}\|_{L^2_m} \leqslant C_1\varepsilon + \frac{C_2 h^\mu}{\sqrt{\varepsilon}}\left[\|\hat{x}\|_{H_n^{\mu+1}} + \|\hat{u}\|_{H^\mu_m}\right], \tag{3.26}$$

where $\mu = \min(l-1, r_1 - 1, r_2)$.

Proof. We use the triangle inequality

$$\begin{aligned} &\|\hat{x}_h^\varepsilon - \hat{x}\|_{H^1_{0n}} + \|\hat{u}_h^\varepsilon - \hat{u}\|_{L^2_m} \\ &\leqslant \|\hat{x}_h^\varepsilon - \hat{x}_\varepsilon\|_{H^1_{0n}} + \|\hat{x}_\varepsilon - \hat{x}\|_{H^1_{0n}} + \|\hat{u}_h^\varepsilon - \hat{u}_\varepsilon\|_{L^2_m} + \|\hat{u}_\varepsilon - \hat{u}\|_{L^2_m}. \end{aligned} \tag{3.27}$$

From Theorem 2, we have

$$\|\hat{x}_\varepsilon - \hat{x}\|_{H^1_{0n}} + \|\hat{u}_\varepsilon - \hat{u}\|_{L^2_m} \leqslant C_1\varepsilon \quad \forall \varepsilon > 0, \tag{3.28}$$

for some constant C_1 depending on $(\hat{x}, \hat{u})$.

We use the bilinear form $a_\varepsilon(\cdot, \cdot)$ in (1.13). It satisfies

$$\left|a_\varepsilon\left(\begin{bmatrix} x_1 \\ u_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ u_2 \end{bmatrix}\right)\right| \leqslant \frac{\overline{C}_2}{\varepsilon}\|(x_1, u_1)\|_{H^1_{0n}\times L^2_m}\|(x_2, u_2)\|_{H^1_{0n}\times L^2_m}$$

for all $(x_1, u_1), (x_2, u_2) \in H^1_{0n} \times L^2_m$. Since $(\hat{x}_h^\varepsilon, \hat{u}_h^\varepsilon)$, $(\hat{x}_\varepsilon, \hat{u}_\varepsilon)$ satisfy the variational equation

$$a_\varepsilon\left(\begin{bmatrix} \hat{x}_\varepsilon - \hat{x}_h^\varepsilon \\ \hat{u}_\varepsilon - \hat{u}_h^\varepsilon \end{bmatrix}, \begin{bmatrix} y_h \\ v_h \end{bmatrix}\right) = 0 \quad \forall (y_h, v_h) \in S_h^1 \times S_h^2,$$

it follows from [1] that

$$\begin{aligned} &\|\hat{x}_\varepsilon - \hat{x}_h^\varepsilon\|_{H^1_{0n}} + \|\hat{u}_\varepsilon - \hat{u}_h^\varepsilon\|_{L^2_m} \\ &\leqslant \left(\frac{\overline{C}_3}{\varepsilon}\right)^{1/2}\left[\inf_{y_h \in S_h^1}\|\hat{x}_\varepsilon - y_h\|_{H^1_{0n}} + \inf_{v_h \in S_h^2}\|\hat{u}_\varepsilon - v_h\|_{L^2_m}\right] \\ &\leqslant \frac{\overline{C}_4^{1/2}}{\sqrt{\varepsilon}}\left[h^{\mu_1}\|\hat{x}_\varepsilon\|_{H_n^{l+1}} + h^{\mu_2}\|\hat{u}_\varepsilon\|_{H^l_m}\right]. \end{aligned} \tag{3.29}$$

Letting $\mu = \min(\mu_1, \mu_2)$, choosing C_2 properly from $\overline{C}_4^{1/2}$, and combining (3.27)–(3.29), we get (3.26). $\square$

4. Applications. We apply the theorems in Section 3 to several types of problem (0.1), (0.2) and its penalized finite element approximation (1.16). We produce a class of problems and approximating subspaces for which condition (H) (and Theorem 12) holds, and specific examples for which it does not.

The approximating subspaces we consider are piecewise polynomial spaces. Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a uniform parition of $[0, T]$ with mesh length $h = T/N = t_{i+1} - t_i$. Let $S_h^{(r,s)} = \{p \in C^s[0, T]$: p is a polynomial of degree $r - 1$ on each subinterval $[t_i, t_{i+1}]$, $i = 0, \ldots, N-1\}$. In the approximation (1.17), S_h^1 and S_h^2 will be the (r, s)-systems of n-fold and m-fold products of subspaces of $S_h^{(r,s)}$, respectively. If locally supported B-spline bases for $S_h^{(r,s)}$ are used, then the matrix equation resulting from (1.17) will have the symmetric block banded structure

$$\left[\begin{array}{c|c} A_{nn} & A_{nm} \\ \hline A_{mn} & A_{mm} \end{array}\right], \tag{4.1}$$

where each

$$A_{kl} = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1l} \\ B_{21} & B_{22} & & \vdots \\ \vdots & & \ddots & \\ B_{k1} & \cdots & & B_{kl} \end{bmatrix}$$

and B_{ij} are banded matrices according to the choice of $S_h^{(r,s)}$.

Example 1. Consider the optimal control problem

$$\begin{cases} \dot{x}(t) = Bu(t) + f(t), \qquad 0 \leqslant t \leqslant T, f \in L_n^2(0, T), B = \text{a constant } n \times m \text{ matrix}, \\ x(0) = 0 \in \mathbf{R}^n, \\ \displaystyle\operatorname*{Min}_{(x,u) \in H_{0n}^1 \times L_m^2} \quad J(x, u) = \int_0^T [\langle \dot{x}, N_1 \dot{x}\rangle + \langle x, N_2 x\rangle + \langle u, Mu\rangle]\, dt. \end{cases}$$

We let $S_h^1 = \prod_{i=1}^n \{\varphi \in S_h^{(k+1,1)} \mid \varphi(0) = 0\}$ and $S_h^2 = \prod_{i=1}^m S_h^{(k,0)}$. All hypotheses of Theorem 12 are evident except condition (H). To show this let $V_h^1 = \{\dot{\varphi} \mid \varphi \in S_h^1\}$, $V_h^2 = \{B\varphi \mid \varphi \in S_h^2\}$. We claim $V_h^1 \cap V_h^2 = V_h^2$. Then $\tilde{V}_h^2 = V_h^2 \ominus (V_h^1 \cap V_h^2) = \{0\}$, and $\mu = \cos(V_h^1, \tilde{V}_h^2) = 0$ giving condition (H).

For each vector-valued function $\bar{\psi} \in V_h^2$, any of its components is a scalar function ψ which can be represented piecewise as

$$\psi = \sum_{j=0}^{k-1} a_{ij}(t - t_i)^j, \qquad t_i \leqslant t \leqslant t_{i+1}, i = 0, 1, \ldots, N-1, \tag{4.2}$$

and

$$\psi(t_i^+) = \psi(t_i^-). \tag{4.3}$$

We wish to find a vector-valued function $\bar{\varphi} \in S_h^1$ such that $\dot{\bar{\varphi}} = \bar{\psi}$.

Any component φ of $\bar{\varphi}$ satisfies

$$\varphi = \sum_{j=0}^{k} b_{ij}(t - t_i)^j, \qquad t_i \leqslant t \leqslant t_{i+1}, i = 0, \ldots, N-1,$$

$$\varphi(t_0) = \varphi(0) = 0, \tag{4.4}$$

$$\varphi(t_i^-) = \varphi(t_i^+), \qquad i = 1, \ldots, N-1, \tag{4.5}$$

$$\varphi'(t_i^-) = \varphi'(t_i^+), \qquad i = 1,\ldots,N-1. \tag{4.6}$$

From (4.4), (4.5), and (4.6), we get, respectively,

$$b_{0,0} = 0, \tag{4.7}$$

$$\sum_{j=0}^{k} b_{ij} h^j = b_{i+1,0}, \qquad i = 0,\ldots,N-2, \tag{4.8}$$

$$\sum_{j=1}^{k} j b_{ij} h^{j-1} = b_{i+1,1}, \qquad i = 0,\ldots,N-2. \tag{4.9}$$

In order that $\dot\varphi = \psi$, we must have

$$\sum_{j=1}^{k} j b_{ij}(t-t_i)^{j-1} = \sum_{j=0}^{k-1} a_{ij}(t-t_i)^j, \qquad t_i \leqslant t \leqslant t_{i+1};$$

or

$$\sum_{j=0}^{k-1} (j+1) b_{i,j+1}(t-t_i)^j = \sum_{j=0}^{k-1} a_{ij}(t-t_i)^j.$$

From the linear independence of polynomials over each subinterval, we must have

$$(j+1) b_{i,j+1} = a_{ij}, \qquad j = 0,\ldots,k-1,\ i = 0,\ldots,N-2. \tag{4.1}$$

Substituting (4.10) into (4.8) and (4.9) gives

$$b_{i+1,0} = b_{i,0} + \sum_{j=1}^{k} \frac{1}{j} a_{i,j-1} h^j, \qquad i = 0,\ldots,N-2, \tag{4.11}$$

$$\sum_{j=1}^{k} a_{i,j-1} h^{j-1} = a_{i+1,0}, \qquad i = 0,\ldots,N-2. \tag{4.12}$$

Note that in (4.11), the expression can be also made valid even for $i = 0$ by choosing

$$\varphi(t) = \int_0^t \psi(\tau)\, d\tau, \qquad t \in [0, t_1]. \tag{4.13}$$

Relations (4.7), (4.10) and (4.11) (with (4.13)) define b's in terms of a's. Relation (4.12) is automatically satisfied because it is just (4.3). Therefore b's can be determined from a's in a unique way.

Therefore for any $\bar\psi \in V_h^2$ there exists a $\bar\varphi \in V_h^1$ such that $\dot{\bar\varphi} = \bar\psi$. That is, $V_h^2 \subset V_h^1$. Hence Theorem 12, Corollary 17 and estimate (3.21) hold with $\mu_1 = \mu_2 = k$.

Computational solutions of (1.17) were obtained for the specific problem

$$\begin{cases} \dot x = u + \cos t, & 0 \leqslant t \leqslant 1, \\ x(0) = 0, \end{cases} \tag{4.14}$$

$$\operatorname*{Min}_{(x,u) \in H_{01}^1 \times L_1^2} \int_0^1 [\dot x^2(t) + u^2(t)]\, dt, \tag{4.15}$$

with $S_h^1 = \{\varphi \in S_h^{(4,1)}\colon \varphi(0) = 0\}$, $S_h^2 = S_h^{(3,0)}$ (C^1-cubics and C^0-quadratics) [12]. The dimension of both S_h^1 and S_h^2 is $2N+1$ and hence the matrix (4.1) is of order $4N+2$.

Table 1 and Figure 1 give the errors between $(\hat x_h^\varepsilon, \hat u_h^\varepsilon)$ and the exact solution $(\hat x(t), \hat u(t)) = (\frac{1}{2}\sin t, -\frac{1}{2}\cos t)$ for various h and ε. Note that the slopes obtained in Figure 1 indicate the sharpness of Corollary 17.

TABLE 1

$\|\hat{x}_h^\varepsilon - \hat{x}\|_{H^1} + \|\hat{u}_h^\varepsilon - \hat{u}\|_{L^2}$ *error for Example* 1, (4.14), (4.15)

h \ ε	10^{-2}	10^{-4}	10^{-5}	10^{-6}
$\frac{1}{2}$	$.461 \times 10^{-2}$	$.263 \times 10^{-3}$	$.259 \times 10^{-3}$	$.252 \times 10^{-3}$
$\frac{1}{4}$	$.461 \times 10^{-2}$	$.607 \times 10^{-4}$	$.394 \times 10^{-4}$	$.350 \times 10^{-4}$
$\frac{1}{8}$	$.461 \times 10^{-2}$	$.466 \times 10^{-4}$	$.475 \times 10^{-5}$	$.472 \times 10^{-5}$
$\frac{1}{16}$	$.461 \times 10^{-2}$	$.463 \times 10^{-4}$	$.468 \times 10^{-5}$	$.612 \times 10^{-6}$

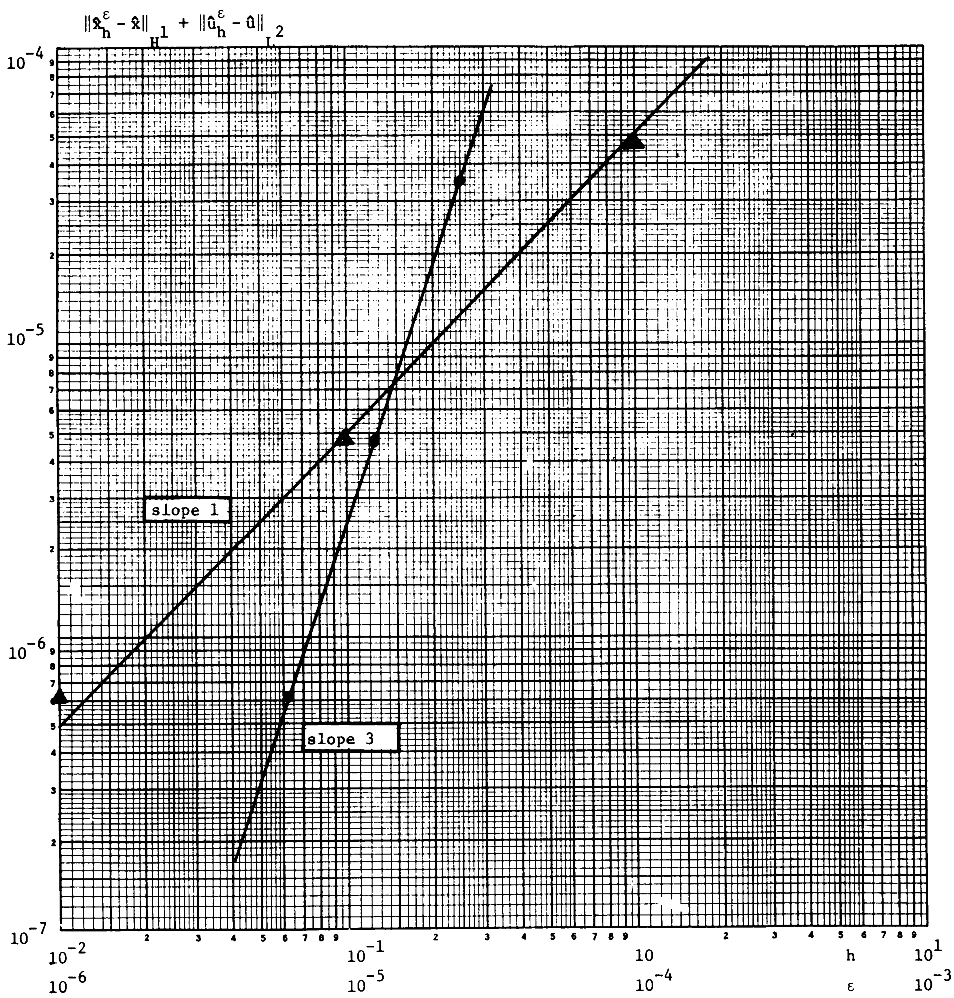

FIGURE 1

Example 1 *error vs* $h(\varepsilon = 10^{-6})$ ○ ○ ○

Example 1 *error vs* $\varepsilon(h = 1/16)$ △ △ △

Example 2. Consider the same optimal control problem as in Example 1, with the matrix B invertible ($m = n$). But, we now choose

$$S_h^1 = \prod_{i=1}^{n} \{\varphi \in S_h^{(k+1,0)} \mid \varphi(0) = 0\}, \qquad S_h^2 = \prod_{i=1}^{n} S_h^{(k+1,0)}, \qquad k \in \mathbf{Z}^+, k > 1.$$

In this case,

$$V_h^1 = \{\dot{\varphi} \mid \varphi \in S_h^1\}, \qquad V_h^2 = S_h^2.$$

Arguing in the same manner as in Example 1, we can verify that

$$V_h^1 \cap V_h^2 = \prod_{i=1}^{n} S_h^{(k,0)},$$

thus $\tilde{V}_h^2 \neq \{0\}$. In order to be able to apply the theory in Section 3, we must find $\tilde{V}_h^2$, the orthogonal complement of $V_h^1 \cap V_h^2$ in V_h^2. Unfortunately, *there does not seem to be any simple representation for basis elements in* $\tilde{V}_h^2$. *This points out a general difficulty in applying the theory of Section* 3. *If there exist convenient representations* (*e.g.*, *as piecewise polynomials*) *for basis elements in* $\tilde{V}_h^2$, *then* $\cos(V_h^1, \tilde{V}_h^2)$ *can be computed by optimization techniques.* □

Example 3. Let the optimal control problem be

$$\begin{cases} \dot{x}(t) = (1 + \cos t)x(t) + u(t) + f(t), \qquad 0 \leq t \leq T, f \in L_n^2(0, T), \\ x(0) = 0 \in \mathbf{R}^n, \\ \operatorname*{Min}_{(x,u)} J(x, u) = \int_0 [\langle \dot{x}, N_1\dot{x}\rangle + \langle x, N_2 x\rangle + \langle u, Mu\rangle]\, dt. \end{cases}$$

Let

$$(4.16) \quad S_h^1 \equiv S_h^{(k_1,j_1)} \cap H_{0n}^1, \qquad S_h^2 \equiv S_h^{(k_2,j_2)}, \qquad k_1, k_2, j_1, j_2 \in \mathbf{N}, k_1, k_2 > 1.$$

Then

$$V_h^1 = \{\dot{\varphi} - (1 + \cos t)\varphi \mid \varphi \in S_h^1\}, \qquad V_h^2 = S_h^2.$$

It is a simple exercise to verify that $V_h^1 \cap V_h^2 = \{0\}$. Therefore Corollary 9 holds. In this case, no asymptotic sharp error estimate like (3.4) is possible.

If the governing equation is changed to

$$\dot{x}(t) = x(t) + (2 + \cos t)\cdot u(t) + f(t),$$

for example, the same conclusion also holds. □

Example 4. Let the control system dynamics be autonomous:

$$\dot{x}(t) = Ax(t) + Bu(t) + f(t),$$

where A and B are constant matrices, and let S_h^1, S_h^2 be chosen as in (4.16). Consider $V_h^1 \cap V_h^2$. If $\varphi_h \in S_h^1$ and $\psi_h \in S_h^2$ satisfies

$$(4.17) \qquad \dot{\varphi}_h - A\varphi_h = B\psi_h \in V_h^1 \cap V_h^2, \qquad \varphi_h(0) = 0,$$

then

$$(4.18) \qquad \varphi_h(t) = \int_0^t e^{-A(t-s)}\psi_h(s)\, ds.$$

If the constant matrix A has only simple eigenvalues which are nonzero, then all the entries of $\exp(At)$ consist of (scalar) linear combinations of exponential functions

$\sum c_i e^{\lambda_i t}$ (λ_i = an eigenvalue). We see that (4.17) and (4.18) can hold only for φ_h and ψ_h which are constant vectors. Hence the dimension of $V_h^1 \cap V_h^2$ cannot exceed a bounded integer. It is not hard to see that condition (H) fails.

If A has zero eigenvalues with high multiplicity, then some entries of $\exp(At)$ are polynomials. Thus, the dimensions of $V_h^1 \cap V_h^2$ may increase somewhat, but in general are still bounded by a fixed integer. Hence condition (H) fails. □

In fact, Examples 3 and 4 indicate that except in very special circumstances (i.e., special choices of A and B) such as Example 1, sharp error estimates do not hold and ε, h must be coupled.

Acknowledgement. We wish to thank F. Deutsch and W. Hager for motivating discussions.

Department of Mathematics
Pennsylvania State University
University Park, Pennsylvania 16802

Department of Mathematics
Pennsylvania State University
University Park, Pennsylvania 16802

Department of Mathematics
Sichuan University
Chengdu, Sichuan
People's Republic of China

Martin-Marietta Company
Denver, Colorado 80200

1. I. Babuska & A. K. Aziz, *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations* (A. K. Aziz, ed.), Academic Press, New York, 1972.

2. A. V. Balakrishnan, "On a new computing technique in optimal control," *SIAM J. Control.*, v. 6, 1968, pp. 149–173.

3. W. E. Bosarge & O. G. Johnson, "Error bounds of high order accuracy for the state regulator problem via piecewise polynomial approximation," *SIAM J. Control*, v. 9, 1971, pp. 15–28.

4. G. Chen & W. H. Mills, "Finite elements and terminal penalization for quadratic cost optimal control problems governed by ordinary differential equations," *SIAM J. Control Optim.*, v. 19, 1981, pp. 744–764.

5. F. Deutsch, "The alternating method of von-Neumann," in *Multivariate Approximation Theory* (W. Schempp and K. Zeller, eds.), Birkhäuser Verlag, Basel, 1979.

6. R. S. Falk, "A finite element method for the stationary Stokes equations using trial functions which do not have to satisfy div $v = 0$," *Math. Comp.*, v. 30, 1976, pp. 698–702.

7. R. S. Falk & J. T. King, "A penalty and extrapolation method for the stationary Stokes equation," *SIAM J. Numer. Anal.*, v. 13, 1976, pp. 814–829.

8. I. C. Gohberg & M. G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators*, Transl. Math. Monographs, vol. 18, Amer. Math. Soc., Providence, R. I., 1969.

9. J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer-Verlag, Berlin and New York, 1971.

10. B. T. Polyak, "The convergence rate of the penalty function method," *Zh. Vychisl. Mat. i Mat. Fiz.*, v. 11, 1971, pp. 3–11. (Russian)

11. D. L. Russell, *Mathematics of Finite Dimensional Control Systems, Theory and Design*, Marcel Dekker, New York, 1979.

12. G. Strang & G. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, N. J., 1973.