Free Boundaries and Finite Elements
in One Dimension

By William W. Hager and Gilbert Strang*

Abstract. Two problems in control theory, one with state constraints and the other with control constraints, have been approximated by the finite element method. This discretization has been applied to both the primal and the dual formulation, in order to make a number of observations and comparisons:

1. The rate of convergence as the grid interval h is decreased, for polynomial elements of different degrees.

2. The presence or absence of a boundary layer in the error, concentrated at the "contact points" where the constraints change between binding and nonbinding.

3. The advantages of simpler constraints in the dual formulation, and the disadvantages of replacing strict convexity by ordinary convexity.

4. The numerical efficiency of each possible variation in achieving an approximate solution of reasonable accuracy.

We concluded that in our model problems, linear elements and the dual method provide the most efficient combination.

1. Introduction. Apparently no one is sure of the best way to approximate a continuous problem in quadratic programming. We decided to experiment with the finite element method, and to start with one-dimensional problems, even though this is not the setting in which finite elements have become famous. They are not regarded as optimal for two-point boundary value problems, and very probably this is still true when there are inequality constraints, although these constraints so alter the problem that all the accepted opinions (including also the arguments favorable to finite elements) have to be reconsidered.

We have chosen two special cases of the following quadratic problem in control theory:

$$(1) \qquad \text{Minimize } \frac{1}{2}\int_0^1 x(t)^T Q x(t) + u(t)^T R u(t)\, dt$$

subject to $\dot{x}(t) = Ax(t) + Bu(t)$, $x(0) = x_0$, $K_s x(t) + b_s \leqslant 0$, $K_c u(t) + b_c \leqslant 0$. We write h for the finite element grid interval, (x^h, u^h) for the "Ritz-Trefftz approximation" to the dual problem, and (x^*, u^*) for the solution to the continuous problem. Hager has previously proved [1] that the errors $u^h - u^*$ and $x^h - x^*$, measured in L^2, are bounded by $ch^{3/2}$ for piecewise polynomial spaces. And unless grid points are moved especially close to every "free boundary" where the constraint becomes binding (we shall call these *contact points*) this estimate is optimal [2].

Using techniques similar to those in [1], the same error bounds can be proved for the primal problem when the control u is restricted to a subspace. On the other hand, for the case where $A = Q = 0$ and $B = R = 1$, Strang [3] notes a similar bound for the H^1 error when the approximate state is required to be a piecewise polynomial.

This paper examines the control error $u^h - u^*$ in more detail. It is shown to exhibit a boundary layer, with most of the error concentrated at the contact points. This behavior occurred even if the approximating space possessed more continuity than the exact solution to the problem. The rate of decay of the error in the boundary layer depended on the ratio of diagonal to off-diagonal elements in the Hessian matrix corresponding to the cost function. When the state rather than the control was restricted to a subspace, the Hessian matrix lost diagonal dominance and the boundary layer faded away.

Finally, we measured the efficiency of each algorithm by the number of multiplications required to reduce the error to a given magnitude. The dual method using piecewise linear elements proved to be the most practical. This method also enjoys an important programming advantage over the primal approach: the constraints are simpler.

2. Problem Description. The following two problems were studied:

(S)
$$\text{Minimize} \quad \tfrac{1}{2}\int_0^1 u(t)^2\, dt$$
$$\text{subject to } \dot{x}(t) = u(t), \quad x(0) = 0,$$
$$x(t) \geqslant \sin(\pi t) + a,$$
$$a = \pi\sqrt{3}/12 - .5,$$

(C)
$$\text{Minimize} \quad \tfrac{1}{2}\int_0^1 u(t)^2\, dt$$
$$\text{subject to } \dot{x}(t) = u(t), \quad x(0) = 0,$$
$$x(1) = b = 1/6 + \sqrt{3}/\pi,$$
$$u(t) \geqslant \sin(\pi t).$$

The problem (S) has a *state constraint*, (C) has a *control constraint*, and u and x are real valued.

The solution to (S) is given by $u_s(t) = 6a + 3$ for $0 \leqslant t \leqslant 1/6$, $u_2(t) = \pi\cos(\pi t)$ for $1/6 \leqslant t \leqslant 1/2$, and $u_s(t) = 0$ for $t > 1/2$. On $[0, .5]$, this is identical to the solution of the obstacle problem where $x(1)$ also vanishes. Removing the constraint at the right end makes u_s vanish after the peak of the obstacle at $t = 1/2$. The problem (C) is similar in structure to the variational formulation of a problem in plasticity theory in which a sphere is subjected to external pressure loading; u would represent the stress rate. The constant b was chosen so that the solution to (C) is $u_c(t) = 1/2$ for $0 \leqslant t \leqslant 1/6$ or $5/6 \leqslant t \leqslant 1$ and $u_c(t) = \sin(\pi t)$ for $1/6 \leqslant t \leqslant 5/6$.

According to the theory developed in [4], the Lagrangian duals of (S) and (C) are:

(DS)
$$\text{Maximize} \quad -\tfrac{1}{2}\int_0^1 v(t)^2\, dt + \int_0^1 (\sin(\pi t) + a)\, dv,$$
$$\text{subject to } v(1) = 0, \ v \text{ nondecreasing,}$$

(DC)
$$\text{Maximize} \quad \int_0^1 \left[-\tfrac{1}{2}(v(t) - q)^2 + v(t)\sin(\pi t) \right] dt - bq,$$
$$\text{subject to } v(t) \geqslant 0, \ q \text{ unconstrained.}$$

Piecewise Polynomials	Convergence Rate in L^2	Convergence Rate at $t = 1/3$	Convergence Rate at $t = 0$
Constants	1.0	~ 1	1.95
Continuous Linear	1.5	~ 2	2.05
Continuous Quadratic	1.5	~ 3	1.98
Continuous Cubic	1.5	~ 4	1.98
Cubic Hermite	1.5	~ 4	1.98

TABLE I. *Convergence of $u^h - u^*$ in Problem* (DS)

The variable q is a real number. One reason for studying these two problems was to analyze the effect of losing strict concavity in the cost functional; in general, the dual problem is only concave even though the primal problem is strictly convex. In our case (DS) is strictly concave and (DC) is only concave. The solutions are related to u_s and u_c, the solutions to (S) and (C) respectively, by $v_s = -u_s$, $v_c - q_c = u_c$, and $q_c = -1/2$. These relations also hold for the approximations u_c^h, v_c^h, q_c^h, u_s^h, and v_s^h generated by the Ritz-Trefftz method.

The finite element approximation replaces a continuous problem by a discrete one; u or v is restricted to lie in a finite-dimensional subspace. The following subspaces were analyzed: piecewise constant functions, continuous piecewise linear, quadratic, and cubic polynomials, and the Hermite cubic polynomials—for which both function value and slope are continuous at the grid points.

3. Convergence Results. In [1] it was observed that the L^2 norm of the error $u^h - u$ depended on the distance between a contact point and the nearest grid point. In order to eliminate this dependence from the error, the grid points were always chosen so that the contact points occurred very near the center of a grid interval.

Note that in problem (DS), v^h is constrained to be nondecreasing. For piecewise constant and linear spaces, this constraint is easily maintained by the requirement $v^h(t_k) - v^h(t_{k+1}) \leq 0$ and for piecewise quadratic functions, by the constraints $3v^h(t_k) - 4v^h(t_{k+1/2}) + v^h(t_{k+1}) \leq 0$ and $-v^h(t_k) + 4v^h(t_{k+1/2}) - 3v^h(t_{k+1}) \leq 0$ for k an integer. For higher order spaces, however, the constraint $\dot{v}^h \geq 0$ imposes a nonlinear condition on the nodal parameters. Hence the computations could be simplified if the monotonicity requirement could be replaced by $v^h(t_k) - v^h(t_{k+1}) \leq 0$ at the mesh points. A study of the structure of the necessary conditions for (DS) revealed that this replacement can be made without destroying the order of the convergence.

An analogous result also held for the finite element approximations to the other problems analyzed herein. In each case, the continuous linear constraint leads to a non-
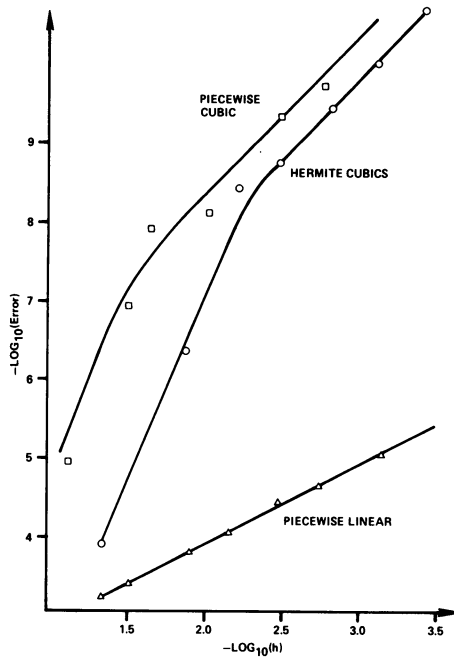
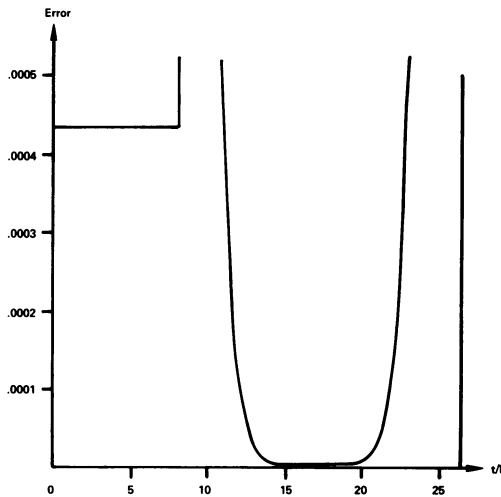**FIGURE 1.** *Problem* (DS): *Error at* $t = 1/3$ *as a Function of* h

FIGURE 2. *Pointwise Error for* (DS) *Using Hermite Cubics*: $h = 1/51$

linear condition in the finite-dimensional problem. However, there is no loss in rate of convergence by replacing the nonlinear constraint with a linear restriction on the nodal parameters. In problem (S), above, the constraint $\int_0^t u^h(s)\, ds \geqslant \sin(\pi t) + a$ is replaced by $\int_0^{t_k} u^h(s)\, ds \geqslant \sin(\pi t_k) + a$ for $k = 0, \ldots, N$, where $\{t_k\}$ are evenly spaced on $[0, 1]$ and the dimension of the subspace containing u^h is $N + 1$. Similarly, the constraint $v^h(t) \geqslant 0$ in (DC) is replaced by $v^h(t_k) \geqslant 0$; and in another formulation of (S) given below, the constraint $x^h(t) \geqslant (\sin(\pi t) + a)$ is replaced by $x^h(t_k) \geqslant (\sin(\pi t_k) + a)$.

(a) *The State Constrained Problem.* The state constrained problem was first solved using the dual approach and the convergence results are given in Table I. It was observed that the convergence rate for $v_s^h(0) - v_s(0)$ was $O(h^2)$ for all the spaces except the piecewise constants. A plot of the convergence data for three of the spaces
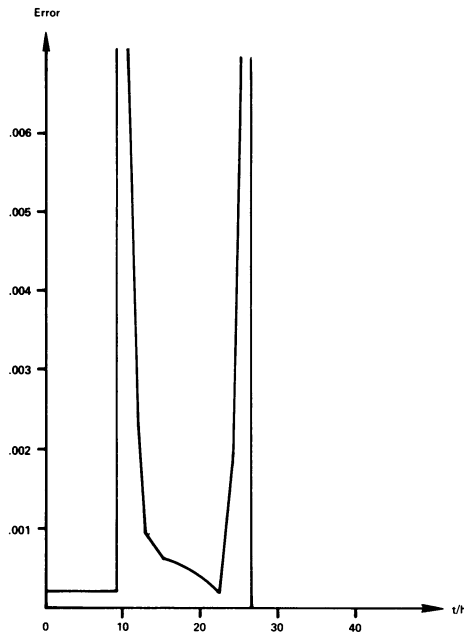
FIGURE 3. *Pointwise Error for* (DS) *Using Linear Elements*: $h = 1/51$

listed in Table I is given in Fig. 1. For the quadratic and the cubic spaces, the grid interval must be on the order of 1/300 before the pointwise error reaches the asymptotic range. Figures 2 and 3 plot the pointwise error in the finite element approximation using Hermite cubic and piecewise linear elements. Note that the Hermite cubics have a continuous derivative and the error in the approximation exhibits a boundary layer, with most of the error concentrated at the contact points $t = 1/6$ and $t = 1/2$ where the derivative of the exact solution is discontinuous.

The L^2 convergence rates given in Table I are exactly as expected, since [2] proved that the convergence estimates in the L^2 norm were tight. The convergence rate for $|v_s^h(0) - v_s(0)|$, on the other hand, can be rigorously proved by adding together the equalities in the Kuhn-Tucker conditions corresponding to the derivative with respect to the nodal variables on the interval $[0, 1/6]$.

The interesting behavior in Table I and Figures 1, 2, and 3 is the convergence rate away from the boundary layer (at $t = 1/3$ for example). The reason for the decay in the error as we leave the contact points is that the diagonal elements in the Hessian matrix corresponding to the cost function in (DS) are much larger than the off-diagonal elements. This is illustrated for the piecewise linear space: Let $d_k = v_s^h(t_k) - v_s(t_k)$ where t_k is the kth grid time. If e denotes the vector consisting of the components of d that correspond to grid points inside the interval $[1/6, 1/2]$, then by a Taylor expansion, it can be shown that e satisfies

$$(2) \qquad He = (O(h), O(h^2), O(h^2), \ldots, O(h^2), O(h))^T,$$

where H is a matrix with 2/3 along the main diagonal and 1/6 for its super- and subdiagonal entries.

Decompose the vector on the right side of (2) into three terms: $O(h)p + O(h)q + O(h^2)r$ where $p^T = (1, 0, 0, \ldots, 0)$, $q^T = (0, 0, \ldots, 0, 1)$, and $r^T = (0, 1, 1, \ldots,$

1, 0). Hence the solution to (2) is the sum of three terms: $e_p + e_q + e_r$. We shall compute

$$(3) \qquad \|H^{-1}\| = \max(\|y\|/\|Hy\|)$$

in the maximum norm, or l_∞ norm. Let $\|y\| = 1$, and $y_k = 1$ for some k. Since $|y_j| \leqslant 1$ for all j, we have

$$\|Hy\| \geqslant \left| \frac{1}{6} y_{k-1} + \frac{2}{3} y_k + \frac{1}{6} y_{k+1} \right| \geqslant \frac{2}{3} - \frac{2}{6} = \frac{1}{3} \quad \text{and} \quad \|H^{-1}\| \leqslant 3.$$

Thus $\|e_r\| \leqslant O(h^2) \|r\| \|H^{-1}\| = O(h^2)$.

Now solve for the e_q term by Gaussian elimination. As the sub-diagonal elements in H are eliminated, the kth row converges to

$$(4) \qquad .622 e_k + .167 e_{k+1} = 0.$$

If H is of order m, the last row is $.622 e_m = O(h)$. Solving by back-substitution,

$$(5) \qquad |e_k| \sim .269 |e_{k+1}| = \frac{.268^{m-k}}{.621} O(h).$$

Thus the contribution of $e_p + e_q$ to the total error e decreases roughly by a factor of .268 on each grid interval as one moves away from the first and last component of e. The thickness of the boundary layer is approximated by the smallest k such that $(.268)^k/(.622) \sim h$. When 50 grid intervals are used and $h = .02$, then $k \sim 3$, which agrees with Fig. 3.

A similar analysis holds for the other piecewise polynomial spaces. At each stage in the elimination of the lower triangular terms in H, we divide a row by a diagonal element and multiply by an off-diagonal element. For the piecewise linear space the ratio of diagonal to off-diagonal elements is .25, which agrees well with the decay of .268. A proof of this decay property for the solution of a diagonally dominant linear system can be found in the appendix.

Next the problem (S) was rewritten, eliminating the control u and leaving an "obstacle problem" for the state x:

$$(S') \qquad \text{Minimize } \int_0^1 \dot{x}(t)^2 \, dt \text{ subject to } x(0) = 0, \quad x(t) \geqslant \sin(\pi t) + a.$$

The solution was exact at the grid points in the binding constraint region while the error in slope on the interval [0, 1/6] was $O(h^2)$ for all the subspaces studied: continuous piecewise linear, quadratic, cubic, and quintic spaces. Some of the convergence data is plotted in Figs. 4 and 5. Two questions arise: Why didn't the error exhibit a boundary layer, and why is the convergence second order even in piecewise linear spaces? The second order convergence resulted from a fortunate combination of circumstances; the constraints are simple and the exact solution on [0, 1/6] lies in the piecewise linear subspace. By changing the cost functional slightly (for example, with a linear term as in (DS)), or by making the constraints a little more complicated (see below), the convergence in derivative will only be $O(h)$ for piecewise linear spaces and at best $O(h^{3/2})$ for higher order spaces.

The loss of the boundary layer resulted from the loss of diagonal dominance in the Hessian matrix corresponding to the cost functional. For example, the Hessian ma-
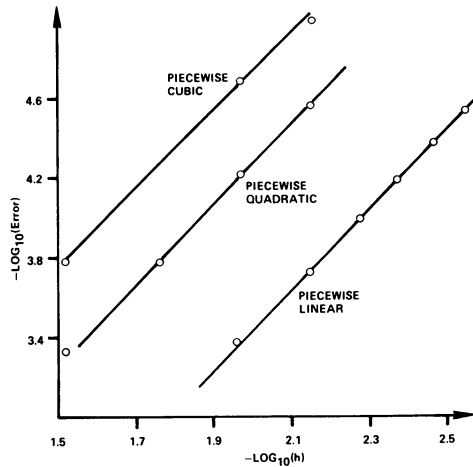
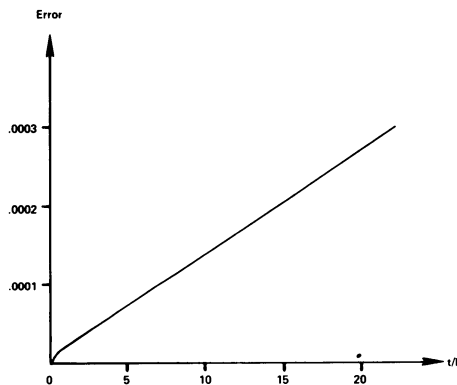FIGURE 4. *Problem* (S'): *Error* $x^* - x^h$ *on* [0, 1/6]

FIGURE 5. *Piecewise Linear Approximation of* (S') *with* $h = 1/141$

trix corresponding to the piecewise linear subspace has 2's on the diagonal and -1's for sub- and super-diagonal elements.

By choosing an appropriate subspace, however, the diagonal dominance can be restored. One possibility is to expand $\dot{x}$ or u, rather than x, in a piecewise polynomial subspace. The quadratic part of the cost function becomes exactly the same as for (DS), and the boundary layer reappears. Unfortunately, the constraint $\int_0^t \dot{x}(s)\, ds \geqslant \sin \pi t + a$ is no longer in band form, and the quadratic programming algorithm consumes much more time. The efficiency of all these methods for solving the state constrained problem is discussed in Section 5.

(b) *The Control Constrained Problem.* The convergence rates for problem (DC) using the finite element method are given in Table II, and the error using a piecewise linear subspace is plotted in Fig. 6. The primary difference between the convergence behavior for (DC) and (DS) is that the error is $O(h^2)$ for all polynomial spaces beyond the piecewise constants on the interior [1/6, 5/6], where the control constraint is binding, and the dual constraint is nonbinding. Recall that in (DS), the error was concentrated in a boundary layer on the edges of this nonbinding region. As shown below, the reason for the slower convergence in (DC) away from the contact points is that the cost functional is only semidefinite instead of negative definite.

Piecewise Polynomials	Convergence Rate in L^2	Convergence Rate at $t = 1/2$	Convergence Rate at $t = 0$
Constants	1.0	1.1	1.0
Continuous Linear	1.5	2.0	2.0
Continuous Quadratic	1.5	2.0	2.0

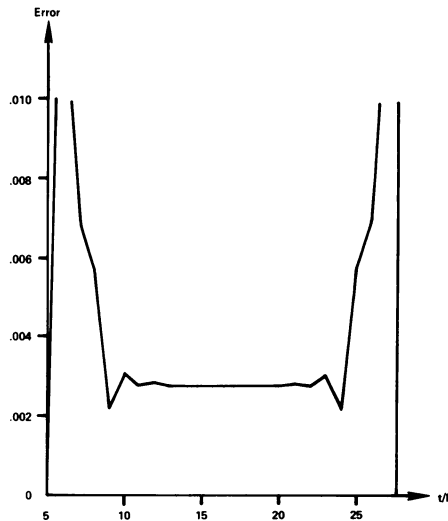TABLE II. *Convergence Rate for $u^h - u^*$ in Problem* (DC)

FIGURE 6. *Pointwise Error in* (DC) *with Linear Elements*: $h = 1/33$

Again consider the space of piecewise linear polynomials and let e be the vector consisting of components $v_c(t_j) - v_c^h(t_j)$ corresponding to grid points inside the interval $[1/6, 5/6]$. Also define $f = q_c - q_c^h$. By a Taylor series expansion of v_c, the following relations hold:

$$(6) \qquad He - gf = O(h)p + O(h)q + O(h^2)r,$$

$$(7) \qquad hg^Te - f = O(h^2),$$

where H, p, q, and r were defined earlier and $g^T = (1, 1, \ldots, 1)$. Solving (6) for e and inserting the result into (7) yields:

$$(8) \qquad (hg^TH^{-1}g - 1)f = O(h^2) + hg^TH^{-1}[O(h)p + O(h)q + O(h^2)r].$$

It was shown above that $\|H^{-1}\| \leqslant 3$. Therefore, $|g^TH^{-1}p| \leqslant \|p\|_{l_1} \|H^{-1}g\|_{l_\infty} \leqslant 3$ and

$$|g^TH^{-1}r| \leqslant \|g\|_{l_1} \|H^{-1}r\|_{l_\infty} \leqslant (1/h)\|H^{-1}\|_{l_\infty} \|r\|_{l_\infty} = 3/h;$$

and the right side of (8) is $O(h^2)$. By direct computation (or rigorously by a Gaussian elimination argument) it can be shown that $hg^TH^{-1}g$ is bounded away from 1 and hence $f = O(h^2)$ from (8). Thus by (6),

$$(9) \qquad e = H^{-1}gf + H^{-1}[O(h)p + O(h)q + O(h^2)r],$$

where $\|H^{-1}g\| \leqslant 3$, $\|H^{-1}r\| \leqslant 3$, and as shown above $H^{-1}p$ and $H^{-1}q$ yield an error term which decays by .269 over each grid interval as one moves toward the middle of the vector e. Hence these "middle components" of e are $O(h^2)$.

Now suppose that higher order polynomials are employed. The error on the right side of (7) is given by

$$(10) \qquad \int_0^1 [v^I(t) - v_c(t)]\, dt,$$

where v^I is the interpolate of v_c in the piecewise quadratic subspace. Since $\dot{v}_c(1/6)$ and $\dot{v}_c(5/6)$ are discontinuous, (10) is at best $O(h^2)$ for all piecewise polynomial spaces unless grid points are placed exactly at $t = 1/6$ and $t = 5/6$. Thus $f = q_c - q_c^h = O(h^2)$, and hence by (9), gf contributes $O(h^2)$ to all components of e. The error then is $O(h^2)$ everywhere in the nonbinding region.

4. Numerical Algorithm. These problems were solved by both gradient projection and conjugate gradient projection. From the numerical experiments, the conjugate gradient algorithm appeared to be very efficient both in determining the binding constraints and in solving the quadratic programming problem in the tangent plane corresponding to these constraints.

In projecting the gradient or conjugate gradient onto the tangent plane corresponding to the binding constraints, our approach was to determine the projection matrix $(-I + A^T(AA^T)^{-1}A)$; A is the matrix consisting of the rows of binding constraints. Note that if the rows of A are taken in the natural order, AA^T is a band matrix in problem (DS) and is the identity matrix in (DC) and (S'). Thus AA^T can be stored in Cholesky's factored form LDL^T. Also note that for the Hermite space, with $r - 1$ continuous derivatives, we may group all rows of A which are identical after a translation and obtain at most r groups of rows. Thus AA^T does not have to be determined by column-row multiplication, but a $2r \times 2r$ table can be computed initially for the products of rows in the r different groups; when an element of AA^T is needed in the LDL^T factorization routine, it is accessed by a simple "look up" procedure.

Care should be taken not to introduce unnecessary constraints. For example, the space of Hermite cubics has dimension $2N + 1$, with $h = 1/N$ and $v_s^h(1) = 0$. Therefore, no more than $2N + 1$ monotone constraints of the form $v_s^h(t_k) - v_s^h(t_{k+1}) \leqslant 0$ can be binding at one time without some subset of the constraints being dependent. And when some rows of A are dependent, the projection matrix is not given by the simple form above since AA^T is no longer invertible.

5. Conclusions. We want to compare the efficiency of the finite element procedures described above for the state constrained problem. Once the binding constraint set has been determined in the quadratic programming problem generated by the finite element method, the number of conjugate gradient iterations required to solve the quadratic programming problem in the tangent plane of the binding constraint set is essentially the number of variables in the quadratic program. Similarly, the cost of each iteration can be estimated by counting the number of multiplications required. The "complexity" of a quadratic programming problem will mean the number of multiplications required to solve the program once the binding constraint set is determined. Figures 7, 8, and 9, respectively, plot the errors as a function of the logarithm of the com-
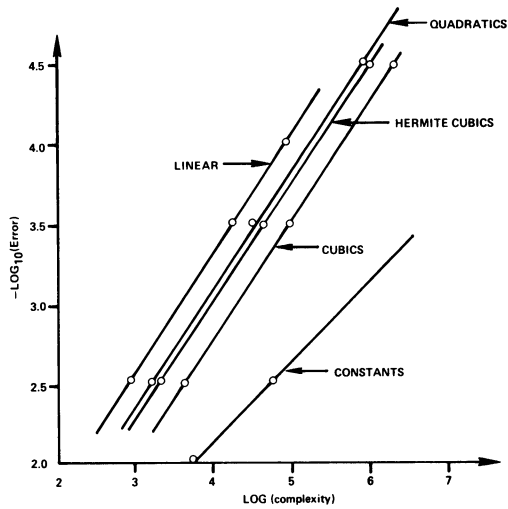
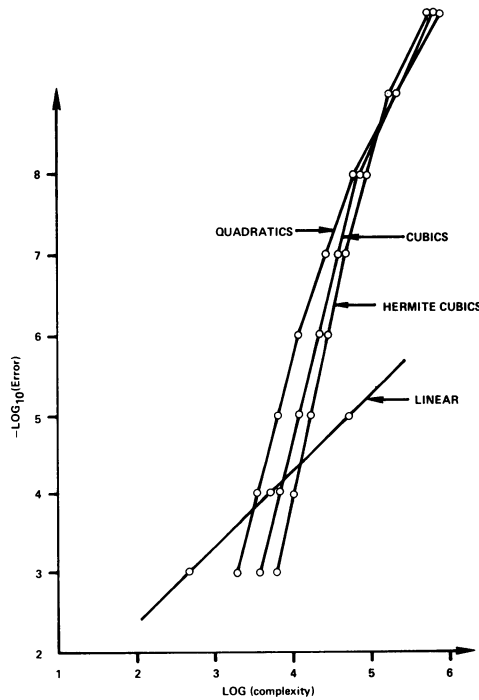FIGURE 7. *The L^2 Error in* (DS) *as a Function of Complexity*

FIGURE 8. *Problem* (DS): *Error at* $t = 1/3$

plexity, as follows: Fig. 7 gives the L^2 error in (DS), Fig. 8 the error at $t = 1/3$ (the farthest point from the contacts), and Fig. 9 the error at $t = 1/6$ in $\dot{x}$ using the finite element approximation of the control in (S), the state in (S'), and the dual variable in (DS).

In terms of the L^2 error, the piecewise linear space was the most efficient for all grid intervals. Recall that convergence in the L^2 norm was first order for the piecewise constants and $O(h^{3/2})$ for higher order spaces. The increase to 3/2 with linear elements was worth the extra computation involved.
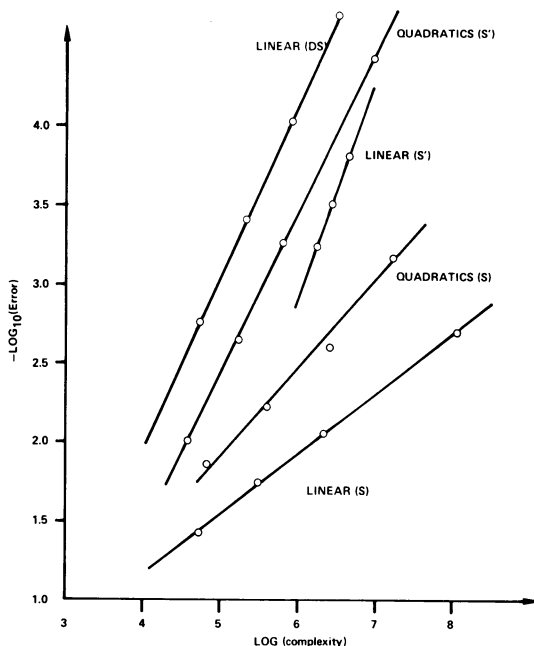
FIGURE 9. *The Error in* $\dot{x}$ *at* $t = 1/6$ *for the State Constrained Problem*

In Fig. 8, recall that $t = 1/3$ is inside the boundary layer and the full convergence rate can be achieved. That rate continues to increase with the degree of the piecewise polynomials. On the other hand, the figure shows that when the desired accuracy is less than 10^{-4}, the piecewise linear space is still the most efficient; *only for very high accuracy will the higher degree spaces be superior.* The reason lies in the slow approach of the convergence plots in Fig. 1 to the asymptotic range.

Finally, it is seen in Fig. 9 that the piecewise linear elements and the dual approach provide the most efficient estimate of $\dot{x}_s$ at $t = 1/6$. Note, however, that the efficiency plot for (S) appearing in Fig. 9 is based on the error at $t = 1/6$ and not the error inside the boundary layer. Inside that layer, the efficiency of quadratic elements is very close to the efficiency of the linear elements in (DS).

Appendix: Decay of the Solution to Diagonally Dominant Linear Systems. We now formally state and prove the decay property mentioned in Section 3 for the solution of a diagonally dominant linear system.

THEOREM. *Suppose E is a matrix with $E_{ij} = 0$ for $|i - j| > \Delta$. Let $j_0 \in \{1, 2, \ldots, n\}$, let f be the vector with all entries zero except for entry j_0 which contains a one, and assume that $E_{j_0 j_0} = 1$ and*

$$\sum_{j \neq i} |E_{ij}|/|E_{ii}| \leqslant r < 1 \quad \text{for all } i.$$

Then the solution to the linear system $Ew = f$ satisfies $|w_j| \leqslant r^{m+1}/(1 - r)$ for all integers $m \geqslant -1$ and j such that $|j - j_0| > m\Delta$.

Proof. Defining $N = D^{-1}(D - E)$ where D is the diagonal of E, the equation $Ew = f$ can be rewritten in the form $(I - N)w = D^{-1}f = f$ where I is the identity matrix and the last equality follows since $D_{j_0 j_0} = E_{j_0 j_0} = 1$ and $f_j = 0$ for $j \neq j_0$. Let

$\|\cdot\|$ denote the l_∞ norm and recall that the l_∞ norm of a matrix is the maximum absolute row sum; then the diagonal dominance condition in the theorem's statement implies that $\|N\| \leqslant r < 1$. Thus

$$(I - N)^{-1} = \sum_{k=0}^{\infty} N^k \quad \text{and} \quad w = \sum_{k=0}^{\infty} N^k f = \sum_{k=0}^{m-1} N^k f + \sum_{k=m}^{\infty} N^k f = \xi^m + \eta^m.$$

Notice that

$$|(\eta^m)_j| \leqslant \|\eta^m\| \leqslant \sum_{k=m}^{\infty} \|N\|^k \|f\| \leqslant \sum_{k=m}^{\infty} r^k = r^m/(1-r).$$

Furthermore $(\xi^m)_j = 0$ for $|j - j_0| > (m-1)\Delta$ since $N_{ij} = 0$ for $|i - j| > \Delta$. Thus the bound on $w_j = (\xi^m)_j + (\eta^m)_j$ in the theorem follows immediately.

Department of Mathematics
University of South Florida
Tampa, Florida 33620

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

1. W. W. HAGER, "The Ritz-Trefftz method for state and control constrained optimal control problems," *SIAM J. Numer. Anal.* (To appear.)

2. W. W. HAGER, *Rates of Convergence for Discrete Approximations to Problems in Control Theory*, Ph. D. Thesis, M.I.T., June 1974.

3. G. STRANG, "The finite element method—linear and nonlinear applications," *Proceedings of the International Congress of Mathematicians* (Vancouver, Canada, 1974), Amer. Math. Soc., Providence, R. I., 1975.

4. W. W. HAGER & S. K. MITTER, "Lagrange duality theory for convex control problems," *SIAM J. Control.* (To appear.)

5. W. W. HAGER, "Rates of convergence for discrete approximations to unconstrained control problems," *SIAM J. Numer. Anal.* (To appear.)