JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 19, No. 5, September–October 1996

Discrete-Time, Mixed-Norm Control Synthesis Applied to Aircraft Terrain Following

David R. Jacques,* D. Brett Ridgely,† and Robert A. Canfield†
U.S. Air Force Institute of Technology, Wright–Patterson Air Force Base, Ohio 45433-7765

A fixed-order, mixed-norm control synthesis technique for discrete-time linear systems is demonstrated. The method allows the control-system designer to combine the H_2, l_1, and H_∞ norms of dissimilar transfer functions into a single constrained optimization problem. Any number or combination of constraints can be added to the problem, and the method constrains the norms directly without reliance on upper bounds to the norms. The method is applied to a terrain-following problem involving a multi-input multi-output model of a fighter aircraft. An H_2–l_1 problem is formulated and solved to provide altitude and flight-path angle tracking in the presence of wind gusts and sensor noise. Finally, an H_∞ constraint on output complementary sensitivity is added to improve robust stability.

Introduction

RECENTLY, there has been a great deal of interest in formulating and solving multiple-objective optimal control problems capable of handling different classes of input and output signals. A typical multiobjective control problem would be to determine the tradeoff between noise rejection H_2 and some unstructured perturbation H_∞ that embodies desired performance and robustness at either the input or output of the plant (or some combination). Concerns for control and rate saturation also make it desirable to be able to bound the magnitude of an output in response to a specific input (l_∞) or a class of bounded-magnitude inputs (l_1). Because a real system is never limited to a single class of inputs or outputs, it is desirable to handle several classes of inputs and outputs within the framework of a single optimal control problem.

Walker and Ridgely[1] and Jacques et al.[2] developed a numerical algorithm for the general H_2/H_∞ problem with output feedback. Their algorithm produces a suboptimal solution with the controller order fixed at some specifiable level. Their method has no requirements that the constraints be nonsingular, and it allows for multiple H_∞ constraints.[3] The problem of adding a continuous-time L_1 constraint was formulated by Walker[4] and later solved by Spillman.[5] The method described here extends the Walker–Ridgely method for H_2–H_∞ optimization in that it applies to discrete-time systems, and it accommodates l_1 norm constraints. As in the work by Walker,[4] the method allows any number of transfer-function norms to be minimized or constrained directly, as opposed to limiting conservative upper bounds or approximations to the norms. Unlike the continuous-time problem, however, the discrete-time problem has no restrictions on the closed-loop D term of the H_2 compensator. Although the method is demonstrated with the compensator order equal to that of the underlying H_2 problem, the method has been applied without modification to develop either higher- or lower-order compensators for the same problem. The method carries no claim of uniqueness or global optimality over all compensators; however, it has proven to be effective in balancing performance and robustness design constraints for a variety of problems tested to date.

Problem Setup

The system considered for the general mixed-norm control problem is shown in Fig. 1. It contains three sets of exogenous inputs and controlled outputs. In general, no relationship is assumed between r, d, and w, or m, e, and z. The input d is assumed to be a signal of unknown but bounded energy with $\|d\|_2 \leq 1$, and the input r is assumed to be a signal of unknown but bounded magnitude with $\|r\|_\infty \leq 1$. The input w is assumed to be the discrete-time equivalent of zero-mean white Gaussian noise of unit intensity. The state space of P is formed by wrapping the stable weights of an H_2 problem from w to z, the stable weights of an l_1 problem from r to m, and the stable weights of an H_∞ problem from d to e around the original plant. The order of the individual problems is, in general, less than the order of P. An expanded state–space realization of P can be written as

$$\begin{aligned} x_2(k+1) &= A_2 x_2(k) + B_w w(k) + B_{u_2} u(k) \\ z(k) &= C_z x_2(k) + D_{zw} w(k) + D_{zu} u(k) \\ y_2(k) &= C_{y_2} x_2(k) + D_{yw} w(k) + D_{yu} u(k) \end{aligned} \tag{1}$$

$$\begin{aligned} x_1(k+1) &= A_1 x_1(k) + B_r r(k) + B_{u_1} u(k) \\ m(k) &= C_m x_1(k) + D_{mr} r(k) + D_{mu} u(k) \\ y_1(k) &= C_{y_1} x_1(k) + D_{yr} r(k) + D_{yu} u(k) \end{aligned} \tag{2}$$

$$\begin{aligned} x_\infty(k+1) &= A_\infty x_\infty(k) + B_d d(k) + B_{u_\infty} u(k) \\ e(k) &= C_e x_\infty(k) + D_{ed} d(k) + D_{eu} u(k) \\ y_\infty(k) &= C_{y_\infty} x_\infty(k) + D_{yd} d(k) + D_{yu} u(k) \end{aligned} \tag{3}$$

where x_2, x_1, and x_∞ may have some or all states in common. The different y represent the same measurement, but each has only the exogenous input associated with its subproblem. In the absence of disturbances, $y_2 = y_1 = y_\infty$.

The objective of mixed H_2–l_1–H_∞ control is to minimize the 2-norm of the closed-loop transfer function T_{zw}, while constraining the 1-norm of the transfer function T_{mr} and the ∞-norm of the transfer function T_{ed} to be less than some specified levels. A practical application might be to use the ∞-norm constraint to guarantee a certain level of stability robustness, whereas acceptable tracking error could be ensured by enforcing a 1-norm constraint on the sensitivity, and the 2-norm could be used to optimize noise rejection

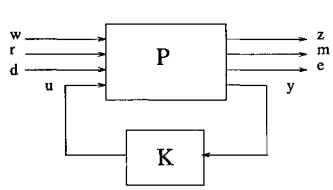

Fig. 1 General mixed-norm optimization problem.

Received Aug. 21, 1995; revision received May 1, 1996; accepted for publication May 2, 1996. This paper is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

*Graduate Student, Aeronautics and Astronautics Department; currently Controls Engineer, Armament Directorate, Wright Laboratories, WL/MNAG, 101 West Eglin Boulevard, Suite 346B, Eglin Air Force Base, FL 32542-6810. Member AIAA.

†Assistant Professor, Aeronautics and Astronautics Department, 2950 P Street, Building 640. Senior Member AIAA.

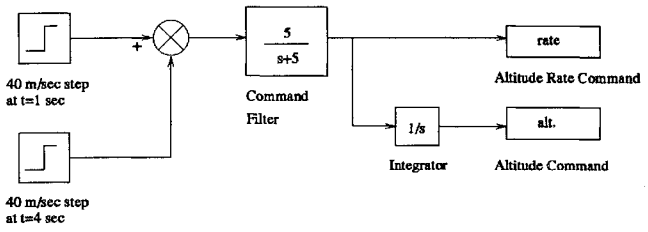

Fig. 2 Command generator used for simulation.

while still meeting the other constraints. Mathematically, the problem can be stated as follows: For the system shown in Fig. 2, find a stabilizing controller $K(z)$ of a specified order that achieves

$$\inf_{K_{\text{admissible}}} \|T_{zw}\|_2 \tag{4}$$

subject to $\|T_{mr}\|_1 \leq \nu$ and $\|T_{ed}\|_\infty \leq \gamma$.

The following discussion assumes a fixed controller order equal to or greater than the minimal order of the H_2 subproblem; however, extensions of the method for reduced order design are presented. The following assumptions are made:

1) (A_2, B_{u_2}) stabilizable, (C_{y_2}, A_2) detectable
2) (A_1, B_{u_1}) stabilizable, (C_{y_1}, A_1) detectable
3) (A_∞, B_{u_∞}) stabilizable, (C_{y_∞}, A_∞) detectable
4) $D_{zu}^T D_{zu}$ full rank, $D_{yw} D_{yw}^T$ full rank
5) $\begin{bmatrix} A_2 - e^{j\theta} I & B_{u2} \\ C_z & D_{zu} \end{bmatrix}$ full column rank $\forall \theta \in [0, 2\pi)$
6) $\begin{bmatrix} A_2 - e^{j\theta} I & B_w \\ C_{y_2} & D_{yw} \end{bmatrix}$ full row rank $\forall \theta \in [0, 2\pi)$
7) $D_{yu} = 0$

Conditions 1–3 ensure the existence of stabilizing controllers, and if only stable weighting functions are used, the existence of a K that stabilizes the H_2 problem has been shown to be necessary and sufficient for K stabilizing the H_∞ and l_1 problems.[4] Conditions 4–6 ensure that the H_2 problem considered separately has a nonsingular solution. A nonsingular H_2 subproblem is not necessary for the numerical solution, but it provides a convenient starting location and an easily defined anchor point for the Pareto-optimal surface. Numerically, it is only necessary that the overall mixed-norm problem be nonsingular, to avoid solutions that require infinite control power or result in open-loop response characteristics. Condition 7 also is not necessary, but a strictly causal P_{yu} or K is sufficient for well-posedness of the problem. A strictly causal P_{yu} is often a mild assumption when actuator dynamics are considered. Further, it frees us from making a similar assumption on the compensator, and it eases the notation required for this development. The relaxation of this final assumption is discussed at the end of this section.

Under certain conditions, the numerical method presented here is applicable to reduced-order control synthesis problems. Specifically, we need to add the assumption that there exist reduced-order stabilizing controllers that satisfy the constraints. The difficulty that this presents is that we must be able to find an initial stabilizing controller with the desired order. Note also that the regularity assumptions on the H_2 subproblem are no longer appropriate for reduced-order controllers. This second concern is not a problem for the numerical solution because, as stated above, it is only necessary that the overall mixed-norm problem be nonsingular.

The desired compensator K can be written in the form

$$\begin{aligned} x_c(k+1) &= A_c x_c(k) + B_c y(k) \\ u(k) &= C_c x_c(k) + D_c y(k) \end{aligned} \tag{5}$$

where A_c, B_c, C_c, and D_c are to be determined from the optimization problem. Note that, for the continuous-time problem, D_c cannot be used as a design variable. This is because it is uniquely determined by

$$D_{zw} + D_{zu} D_c D_{yw} = 0 \tag{6}$$

which is necessary in order for $\|T_{zw}\|_2$ to be finite.[1] This is not the case for discrete time.

Using Eq. (5), the closed-loop state–space equations can be written as

$$\begin{aligned} \boldsymbol{x}_2(k+1) &= \boldsymbol{\mathcal{A}}_2 \boldsymbol{x}_2(k) + \boldsymbol{\mathcal{B}}_w w(k) \\ z(k) &= \boldsymbol{\mathcal{C}}_z \boldsymbol{x}_2(k) + \boldsymbol{\mathcal{D}}_{zw} w(k) \end{aligned} \tag{7}$$

$$\begin{aligned} \boldsymbol{x}_1(k+1) &= \boldsymbol{\mathcal{A}}_1 \boldsymbol{x}_1(k) + \boldsymbol{\mathcal{B}}_r r(k) \\ m(k) &= \boldsymbol{\mathcal{C}}_m \boldsymbol{x}_1(k) + \boldsymbol{\mathcal{D}}_{mr} r(k) \end{aligned} \tag{8}$$

$$\begin{aligned} \boldsymbol{x}_\infty(k+1) &= \boldsymbol{\mathcal{A}}_\infty \boldsymbol{x}_\infty(k) + \boldsymbol{\mathcal{B}}_d d(k) \\ e(k) &= \boldsymbol{\mathcal{C}}_e \boldsymbol{x}_\infty(k) + \boldsymbol{\mathcal{D}}_{ed} d(k) \end{aligned} \tag{9}$$

where

$$\boldsymbol{x}_2 = \begin{bmatrix} x_2 \\ x_c \end{bmatrix}, \qquad \boldsymbol{x}_1 = \begin{bmatrix} x_1 \\ x_c \end{bmatrix}, \qquad \boldsymbol{x}_\infty = \begin{bmatrix} x_\infty \\ x_c \end{bmatrix} \tag{10}$$

and the various $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{B}}$, $\boldsymbol{\mathcal{C}}$, and $\boldsymbol{\mathcal{D}}$ matrices are formed by closing the loop around the compensator for each of the subproblems.

The assumption of $D_{yu} = 0$ can be relaxed via a simple change of variables as described by Safonov et al.[6] The remaining assumption is one of well-posedness, namely $(I + D_c' D_{yu})$ must be invertible. With this relaxation, the only remaining restrictions for numerical solution of the full-order problem are stabilizability, detectability, well posedness, and an overall mixed problem that is nonsingular. This is a minimal set of assumptions that are met easily in any meaningful control problem. Note that, if we wish to solve for reduced-order controllers, we must also assume the existence of, and be able to find, at least one reduced-order stabilizing controller.

Numerical Approach to the Solution

Canfield et al.[7] used a sequential quadratic programming (SQP) algorithm to solve the continuous-time H_2–H_∞ problem. Applying their method to the current problem, the objective (f) and constraints (g) are as follows:

$$f(\kappa) := \xi_2 \|T_{zw}\|_2^2 \tag{11}$$

$$g_1(\kappa) := \xi_1 (\|T_{mr}\|_1 - \nu) \tag{12}$$

$$g_\infty(\kappa) := \xi_\infty (\|T_{ed}\|_\infty - \gamma) \tag{13}$$

where κ represents a vectorized compensator and the ξ are scaling parameters. A modal form is assumed for the compensator to reduce the number of variables. Although this method does not allow repeated eigenvalues in the compensator, in practice it has proven sufficient. If it is necessary to allow for repeated eigenvalues, a block-Jordan form or fully populated state space could be used instead. Finally, because SQP searches over both feasible and infeasible solutions, a stability constraint and an exterior penalty function were added to keep the algorithm from getting lost in an unstable region. The stability constraint is stated as

$$g_s(\kappa) = \xi_s \left(\max_i \left\{ |\lambda_i(\boldsymbol{\mathcal{A}}_2)|^2 \right\} - 1 \right) \tag{14}$$

where $\lambda_i(\boldsymbol{\mathcal{A}}_2)$ denotes the ith eigenvalue of the closed-loop system and $g_s(\kappa)$ is constrained to be less than zero. The penalty function added to the objective function is simply the square of the stability constraint, thus providing continuous derivatives at the stability boundary.

Computing Gradients of the 2-Norm

We begin by defining n_2 as the number of states, n_w as the number of exogenous inputs, and n_z as the number of controlled outputs of the H_2 subproblem, and n_c as the number of states of the compensator. The square of the 2-norm for stable discrete-time systems can be calculated as

$$\|T_{zw}\|_2^2 = \text{trace}\left[\boldsymbol{\mathcal{D}}_{zw}^T \boldsymbol{\mathcal{D}}_{zw} + \boldsymbol{\mathcal{C}}_z Q_2 \boldsymbol{\mathcal{C}}_z^T \right] \tag{15}$$

where Q_2 is a solution to the Lyapunov equation

$$\boldsymbol{\mathcal{A}}_2 Q_2 \boldsymbol{\mathcal{A}}_2^T + \boldsymbol{\mathcal{B}}_w \boldsymbol{\mathcal{B}}_w^T = Q_2 \tag{16}$$

The dual Lyapunov equation

$$\boldsymbol{\mathcal{A}}_2^T X_2 \boldsymbol{\mathcal{A}}_2 + \boldsymbol{\mathcal{C}}_z^T \boldsymbol{\mathcal{C}}_z = X_2 \tag{17}$$

has a real, symmetric, positive semidefinite solution X_2, which is guaranteed to exist if $\mathcal{A}_2$ is stable. With this we can define

$$X_2 = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix}, \qquad Q_2 = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix} \tag{18}$$

Following the derivation for a similar cost function by Mukhopadhyay,[8] the remainder of the gradients now can be expressed as

$$\frac{\partial J}{\partial A_c} = 2\big[X_{12}^T A_2 Q_{12} + X_{12}^T B_{u_2} C_c Q_{22} + X_{22} B_c C_{y_2} Q_{12} + X_{22} A_c Q_{22} + X_{12}^T B_{u_2} D_c C_{y_2} Q_{12}\big] \tag{19}$$

$$\frac{\partial J}{\partial B_c} = 2\big[X_{12}^T A_2 Q_{11} C_{y_2}^T + X_{12}^T B_{u_2} C_c Q_{12}^T C_{y_2}^T + X_{22} B_c C_{y_2} Q_{11} C_{y_2}^T + X_{22} A_c Q_{12}^T C_{y_2}^T + X_{12}^T B_w D_{yw}^T + X_{22} B_c D_{yw} D_{yw}^T + X_{12}^T B_{u_2} D_c C_{y_2} Q_{11} C_{y_2}^T + X_{12}^T B_{u_2} D_c D_{yw} D_{yw}^T\big] \tag{20}$$

$$\frac{\partial J}{\partial C_c} = 2\big[D_{zu}^T C_z Q_{12} + D_{zu}^T D_{zu} C_c Q_{22} + B_{u_2}^T X_{11} A_2 Q_{12} + B_{u_2}^T X_{11} B_{u_2} C_c Q_{22} + B_{u_2}^T X_{12} B_c C_{y_2} Q_{12} + B_{u_2}^T X_{12} A_c Q_{22} + D_{zu}^T D_{zu} D_c C_{y_2} Q_{12} + B_{u_2}^T X_{11} B_{u_2} D_c C_{y_2} Q_{12}\big] \tag{21}$$

$$\frac{\partial J}{\partial D_c} = 2\big[D_{zu}^T D_{zw} D_{yw}^T + D_{zu}^T D_{zu} D_c D_{yw} D_{yw}^T + D_{zu}^T C_z Q_{11} C_{y_2}^T + D_{zu}^T D_{zu} D_c C_{y_2} Q_{11} C_{y_2}^T + D_{zu}^T D_{zu} C_c Q_{12}^T C_{y_2}^T + B_{u_2}^T X_{11} A_2 Q_{11} C_{y_2}^T + B_{u_2}^T X_{11} B_{u_2} D_c C_{y_2} Q_{11} C_{y_2}^T + B_{u_2}^T X_{11} B_{u_2} C_c Q_{12}^T C_{y_2}^T + B_{u_2}^T X_{12} B_c C_{y_2} Q_{11} C_{y_2}^T + B_{u_2}^T X_{12} A_c Q_{12}^T C_{y_2}^T + B_{u_2}^T X_{11} B_w D_{yw}^T + B_{u_2}^T X_{11} B_{u_2} D_c D_{yw} D_{yw}^T + B_{u_2}^T X_{12} B_c D_{yw} D_{yw}^T\big] \tag{22}$$

The penalty function associated with the objective makes it unlikely that a 2-norm gradient calculation for an unstable system will be necessary. However, in the event that it is required, the algorithm detects the unstable closed-loop system and switches to a finite-difference calculation for the gradient of the stable and antistable projections of the objective transfer function.

Computing Gradients of the 1-Norm

The state–space expression for the 1-norm of a single-input single-output (SISO) transfer function is

$$\|T_{mr}\|_1 = \sum_{k=0}^{\infty} \left|\mathcal{C}_m \mathcal{A}_1^k \mathcal{B}_r\right| + |\mathcal{D}_{mr}| \tag{23}$$

The partial derivatives with respect to the closed-loop state–space can be expressed as

$$\frac{\partial \|T_{mr}\|_1}{\partial \mathcal{A}_1} = \sum_{k=0}^{\infty} \left\{ \operatorname{sgn}\left(\mathcal{C}_m \mathcal{A}_1^k \mathcal{B}_r\right) \sum_{l=0}^{k-1} \left[\left(\mathcal{A}_1^T\right)^l \mathcal{C}_m^T \mathcal{B}_r^T \left(\mathcal{A}_1^T\right)^{k-l-1} \right] \right\} \tag{24}$$

$$\frac{\partial \|T_{mr}\|_1}{\partial \mathcal{B}_r} = \sum_{k=0}^{\infty} \operatorname{sgn}\left(\mathcal{C}_m \mathcal{A}_1^k \mathcal{B}_r\right) \left[\left(\mathcal{A}_1^T\right)^k \mathcal{C}_m^T \right] \tag{25}$$

$$\frac{\partial \|T_{mr}\|_1}{\partial \mathcal{C}_m} = \sum_{k=0}^{\infty} \operatorname{sgn}\left(\mathcal{C}_m \mathcal{A}_1^k \mathcal{B}_r\right) \left[\mathcal{B}_r^T \left(\mathcal{A}_1^T\right)^k \right] \tag{26}$$

$$\frac{\partial \|T_{mr}\|_1}{\partial \mathcal{D}_{mr}} = \operatorname{sgn}(\mathcal{D}_{mr}) \tag{27}$$

where $\operatorname{sgn}(\cdot)$ is 1, −1, or 0 depending on the sign of the argument. From these, we can obtain the gradients with respect to the compensator state space by employing the chain rule for the close-loop plant matrices in conjunction with Eqs. (24–27). Naturally, the infinite summations cannot be carried out; the current implementation truncates the summation at some specified number of time steps. Once a solution is found for a given truncation level, the tolerance for the 1-norm is checked by simply recalculating the norm using a significantly higher truncation level. If the error is greater than a specified tolerance, the truncation level is increased and the optimization is repeated, starting with the last computed compensator. The method above works well (albeit slowly) for truncation levels up to approximately 1000, which is often more than sufficient for systems with closely spaced modes and sample rates set at 5–10 times the highest mode. For systems with widely separated poles, a sampling rate fast enough for the highest mode often resulted in truncation levels of 5000–10,000 to capture the lower modes. These truncation levels resulted in prohibitively high run times and/or numerical instability. For systems such as this, a modification to the subroutines providing l_1 norm and gradient information was developed by Spillman[5] and is described subsequently.

For now, we again limit the discussion to SISO systems. If $\mathcal{A}_1$ is nondefective, and denoting the ith left and right eigenvectors as L_i and R_i, respectively, then the partial derivative of $\mathcal{A}_1^k$ with respect to any element of $\mathcal{A}_1$ is given by

$$\frac{\partial \mathcal{A}_1^k}{\partial a_{pq}} = \frac{\partial [R \Lambda^k R^{-1}]}{\partial a_{pq}} = \frac{\partial R}{\partial a_{pq}} \Lambda^k R^{-1} + R \frac{\partial \Lambda^k}{\partial a_{pq}} R^{-1} - R \Lambda^k R^{-1} \frac{\partial R}{\partial a_{pq}} R^{-1} \tag{28}$$

where Λ is a diagonal matrix of eigenvalues,

$$\frac{\partial R_i}{\partial a_{pq}} = \sum_{j=1, j\neq i}^{n_1} c_{ij} R_j + c_i R_i = V_i + c_i R_i \tag{29}$$

$$c_{ij} = \frac{L_j^T [R_i (\partial \lambda_i / \partial a_{pq}) - (\partial \mathcal{A}_1 / \partial a_{pq}) R_i]}{\lambda_j - \lambda_i}, \qquad i \neq j \tag{30}$$

$$c_i = -\Re\left(R_i^H V_i\right) \tag{31}$$

$$\frac{\partial \Lambda^k}{\partial a_{pq}} = \sum_{i=1}^{n_1} \frac{\partial \Lambda^k}{\partial \lambda_i} L_i^T \frac{\partial \mathcal{A}_1}{\partial a_{pq}} R_i \tag{32}$$

and where n_1 is the number of states, the symbol $\Re$ signifies the real part, and the eigenvectors are normalized such that $R_i^H R_i = 1$. The partial derivative of the 1-norm with respect to $\mathcal{A}_1$ can now be found element-wise from

$$\frac{\partial \|T_{mr}\|_1}{\partial \mathcal{A}_{1_{pq}}} = \sum_{k=0}^{N} \operatorname{sgn}\left(\mathcal{C}_m \mathcal{A}_1^k \mathcal{B}_r\right) \mathcal{C}_m \frac{\partial \mathcal{A}_1^k}{\partial a_{pq}} \mathcal{B}_r \tag{33}$$

The sign of $\mathcal{C}_m \mathcal{A}_1^k \mathcal{B}_r$ can be stored as part of the norm calculation. This eliminates a costly part of the gradient calculation, because this factor no longer needs to be recomputed for $k = 0, \ldots, N$. With this, the only remaining part that requires evaluation is the diagonal matrix Λ^k, at each index k, and this only requires the calculation of the diagonal elements. Once these terms have been calculated, the gradient expressions can be evaluated quickly.

Currently, the new method is limited to nondefective matrices with no repeated eigenvalues in the closed-loop $\mathcal{A}_1$ matrix. If the repeated roots stem from weighting functions, it is usually possible to perturb these functions to avoid repeated roots, without any adverse effects on the overall design. To a lesser degree, it is also possible to perturb the plant model to avoid having repeated roots. Although extensions for the case of repeated eigenvalues may be possible, the current method switches to a finite difference calculation in the event that they occur. Although this involves a degradation in the efficiency of the gradient algorithm, the accuracy of the finite

difference gradient is sufficient to maintain numerical stability and convergence.

For multi-input/multi-output (MIMO) transfer functions, the 1-norm is determined by the maximum row sum of SISO transfer-function norms. To implement this, the row where the maximum occurs can be determined first, and SISO norms can be computed for each transfer function in the row. However, the nature of the l_1 optimal solution is such that the maximum row sum may occur over more than one row, leading to discontinuous derivatives. For this reason, the current implementation is to append each row sum as a separate multi-input/single-output transfer-function constraint, with the same constraint level used for each row sum.

Computing Gradients of the Infinity-Norm

The approach for taking gradients of the ∞-norm is based on the singular value sensitivity analysis of Giesy and Lim,[9] with some modification required to apply it to a discrete-time problem. If we assume that the maximum singular value of T_{ed} evaluated at κ has a single peak for $\theta \in [0, 2\pi)$, the derivative of $\|T_{ed}\|_\infty$ can be written as

$$\frac{\partial \|T_{ed}\|_\infty}{\partial \kappa_i} = \Re\left\{u_1^H\left[\frac{\mathrm{d}T_{ed}(e^{j\theta_0})}{\mathrm{d}\kappa_i}\right]v_1\right\} \tag{34}$$

where u_1 and v_1 are the singular vectors associated with the maximum singular value of T_{ed}, θ_0 is the phase angle where the singular value reaches its maximum, and κ is the vectorized compensator. The derivative of T_{ed} can be determined from

$$\begin{aligned}\frac{\mathrm{d}T_{ed}(e^{j\theta_0})}{\mathrm{d}\kappa_i} &= \left[\frac{\mathrm{d}\mathcal{C}_e}{d\kappa_i}\left(e^{j\theta_0}I - \mathcal{A}_\infty\right)^{-1}\mathcal{B}_d\right.\\ &+ \mathcal{C}_e\left(e^{j\theta_0}I - \mathcal{A}_\infty\right)^{-1}\frac{\mathrm{d}\mathcal{B}_d}{\mathrm{d}\kappa_i} + \mathcal{C}_e\left(e^{j\theta_0}I - \mathcal{A}_\infty\right)^{-1}\\ &\left.\times\frac{\mathrm{d}\mathcal{A}_\infty}{\mathrm{d}\kappa_i}\left(e^{j\theta_0}I - \mathcal{A}_\infty\right)^{-1}\mathcal{B}_d + \frac{\mathrm{d}\mathcal{D}_{ed}}{\mathrm{d}\kappa_i}\right]\end{aligned} \tag{35}$$

The actual implementation of this method uses a banded search to locate any number of singular-value peaks. Each peak is then treated as a separate constraint with the same constraint value. Although this may result in carrying inactive constraints in the optimization problem, it avoids the problem of discontinuous derivatives, which can occur if gradient information is obtained from a single peak. If a new peak develops while the optimization problem is being solved, the problem is restarted with a composite of new and old frequency bands, thus avoiding the problem of bouncing back and forth between two different peaks.

Computing Stability Gradients

The stability constraint was defined by Eq. (14). Define

$$\lambda_m = \sigma_m + j\omega_m \equiv \arg(\max|\lambda_i(\mathcal{A}_2)|) \tag{36}$$

so that we can write the gradient expression as

$$\frac{\partial g_s}{\partial \kappa_i} = 2\xi_s\left\{\sigma_m\Re\left(l_m^T\frac{\partial \mathcal{A}_2}{\partial \kappa_i}r_m\right) + \omega_m\Im\left(l_m^T\frac{\partial \mathcal{A}_2}{\partial \kappa_i}r_m\right)\right\} \tag{37}$$

where l_m and r_m denote the left and right eigenvectors, respectively, associated with the maximum eigenvalue and the partial of $\mathcal{A}_2$ with respect to κ_i can be easily evaluated componentwise.

The analytical gradient expressions are valid for the discretized plant at whatever frequency the plant is discretized. For the following example, the analytical gradients were validated by finite difference calculations. If the sample rate were changed, of course, the discretized plant model also would change, and the gradient expressions would reflect this change. Naturally, if the plant's sample rate is too slow, dynamic instability can occur just as it can for any discrete-time synthesis method. This error manifests itself only when the discrete-time controller is implemented with the continuous-time plant in a sampled data mode. The mixed-norm method makes no attempt to protect the user against bad choices for sample rate. It is up to the user to choose a sample rate consistent with the dynamics of the system.

MIMO Aircraft Terrain-Following Example

Problem Overview

This section demonstrates H_2–l_1–H_∞ control synthesis for a general multiblock plant. The example used is a longitudinal controller design for a fighter aircraft in a terrain-following mode. The aircraft is an A-4 operating at sea level, Mach 0.85 (Ref. 10). There are five states in the basic plant: vertical velocity in the aircraft axes (w, meters/second), altitude perturbation (h, meters), pitch rate (q, radians/second), pitch angle (θ, radians), and forward velocity perturbation in the aircraft axes (u, meters/second). The control inputs are elevator deflection (δ_e, radians) and throttle (δ_T). A throttle control was specifically included in this example to help maintain forward velocity during a commanded change in altitude. The throttle response was assumed to be more sluggish than that of the elevator, and this makes it less attractive for controlling the states that vary quickly. However, natural forward velocity variations are mainly attributable to the slowly varying phugoid mode, and the throttle provides excellent control authority over this state. Measurements are assumed to be available from a combination of air data and some type of inertial measurement unit, and measurements consist of altitude rate, altitude, normal acceleration felt by the pilot (n_z), pitch angle, and forward velocity perturbation. Commanded inputs are a combination of flight-path angle (or altitude rate) and altitude. Controlled outputs are various combinations of the measurements. The goal is to enable the aircraft to follow flight-path angle and altitude commands quickly and accurately, maintain low overshoot and reasonable settling times, keep control usage and pilot g-levels within reasonable bounds, and have the steady-state flight be relatively insensitive to wind gusts and measurement noise. The continuous-time plant is given by

$$\begin{Bmatrix}\dot{w}\\ \dot{h}\\ \dot{q}\\ \dot{\theta}\\ \dot{u}\end{Bmatrix} = \begin{bmatrix}-2.23 & 0 & 289.13 & 0 & -0.1134\\ -1 & 0 & 0 & 289.13 & 0\\ -0.15513 & 0 & -4.1708 & 0 & 0.0041218\\ 0 & 0 & 1 & 0 & 0\\ -0.0368 & 0 & 0 & -9.8066 & -0.0308\end{bmatrix}$$

$$\times\begin{Bmatrix}w\\ h\\ q\\ \theta\\ u\end{Bmatrix} + \begin{bmatrix}-57.302 & 0\\ 0 & 0\\ -63.754 & 0.000243\\ 0 & 0\\ 0 & 0.002344\end{bmatrix}\begin{Bmatrix}\delta_e\\ \delta_T\end{Bmatrix} \tag{38}$$

$$\begin{Bmatrix}\dot{h}\\ h\\ n_z\\ \theta\\ u\end{Bmatrix} =$$

$$\begin{bmatrix}-1 & 0 & 0 & 289.13 & 0\\ 0 & 1 & 0 & 0 & 0\\ -0.17918 & 0 & 30.780 & 0 & -0.01284\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{bmatrix}\begin{Bmatrix}w\\ h\\ q\\ \theta\\ u\end{Bmatrix} \tag{39}$$

Actuator dynamics are assumed for both the elevator and the thrust and are modeled as follows:

$$\begin{aligned}\delta_e(s) &= [20/(s+20)]\delta_{e_c}(s)\\ \delta_T(s) &= [5/(s+5)]\delta_{T_c}(s)\end{aligned} \tag{40}$$

As mentioned previously, the throttle was assumed to be more sluggish than the elevator to discourage its use for controlling the faster-varying modes of the system.

The basic plant includes the short-period mode ($\omega_{sp} = 7.35$ rad/s), phugoid mode ($\omega_p = 0.0696$ rad/s), and an additional pole

at the origin resulting from the altitude state. The wide separation in frequency between the modes presents a challenging problem; a sample rate fast enough to accommodate the short period and actuator modes will result in high truncation levels to capture the phugoid and integrator states. A sample rate of 20 Hz was chosen, based on the actuator modes ($6.3 \times \omega_{\delta_e}$), and a zero-order hold was used for discretization of the transfer functions. Simulations of the various closed-loop systems used a continuous-time plant and discrete-time compensators. Further, to provide a more realistic reference command (as opposed to a discrete step input) with which to evaluate the controllers, a command generator was used in the closed-loop simulations (see Fig. 2). The command generator provides a commanded increase in altitude of 120 m, which takes place over 3 s. The command filter with $\omega_c = 5$ rad/s was chosen to model either pilot lag or command smoothing on the part of an automatic terrain-following mode. As a pilot model, the command filter represents a fast pilot, thus presenting a more challenging design problem.

H_2 Subproblem

The H_2 subproblem is based on a steady-state linear quadratic Gaussian design. A wind disturbance, which enters the plant as an angle-of-attack disturbance, is modeled as the output of a first-order shaping filter driven by zero-mean white Gaussian noise, and is included in both the design model and the simulation model. Intensity levels and time constants are chosen for sea-level flight in cumulus clouds as used for the control designs on the NASA F-8C.[11]

Controlled outputs consist of the measurements with weightings of 10, 1000, 1000, 10, and 1 on $\dot{h}$, h, n_z, θ, and u, respectively. To enable nonzero set-point tracking, a pseudointegral state on altitude error was added to the plant as follows:

$$i_h(s) = \frac{1000}{s + 0.001}[h(s) - h_c(s)] \tag{41}$$

and a weight of 0.1 was used for the output of the pseudointegral state. Control weights of 0.1 and 1 were used for δ_e and δ_T, respectively. All of the weights were chosen subjectively and iteratively, based on closed-loop simulations, with a goal of maintaining steady forward velocity during a commanded change in altitude, maintaining g-levels below 4 g for simulated commands, and accurately tracking altitude and altitude-rate commands. The higher weight on δ_T as opposed to that on δ_e is again meant to discourage the use of thrust over elevator commands. Thrust is primarily intended to maintain forward velocity.

The measurements are assumed to be corrupted by zero-mean, white Gaussian noises with the strength initially set to 10^{-4} rad^2-s. After several iterations, the noise strength on the θ measurement was dropped to 10^{-5} to reflect the smaller unit dimensions (radians, as compared to meters, and meters/second). The H_2 solution was very sensitive to the value of the noise strength placed on the altitude measurements, and this is shown in Fig. 3 for noise strengths of 10^{-4} and 10^{-5}. As shown, the 10^{-5} setting results in a much quicker response at the expense of higher g-loading (not shown) and much wider fluctuations in pitch response. The final design opted for the quicker tracking associated with a noise strength of 10^{-5} rad^2-s. It was decided to allow for the high overshoot and large pitch variations in the H_2 portion of the problem in hopes that it could be reduced significantly using an l_1 constraint. Overall, the H_2 subproblem provides quick response to altitude and altitude-rate errors, and is generally unaffected by wind gusts and measurement noise. The noise levels used for the simulation were generally higher than those used to synthesize the controller. This was done to highlight differences between the controllers. The final H_2 plant for this example is ninth order, which also will be the order of the H_2 optimal compensator. This example does not attempt to examine reduced-order compensators, and so, the mixed-norm compensators also will be ninth order.

l_1 Subproblem

The l_1 subproblem was used to reduce the maximum magnitude of the altitude error, thereby reducing the large overshoot resulting from the H_2 subproblem. The bounded magnitude inputs were taken as the wind gust and commanded altitude and altitude rate (or flight-path angle). We wish to be able to reject the wind gust while tracking the altitude and altitude-rate commands. A more general sensitivity constraint could have been used as an alternative, but the H_2 subproblem provided adequate state regulation, and we were exclusively interested in constraining altitude error. The controlled altitude error was band-limited to reduce the susceptibility to measurement noise,

$$h_o(s) = [4/(s + 4)][h(s) - h_c(s)] \tag{42}$$

and a weight of 1 was used. The value of the weight for the l_1 constraint is unimportant for this problem because it is a single output constraint. The bandwidth of the frequency weight was based on the pilot model, and was set to 4 rad/s so as not to coincide with the open-loop pole from the throttle dynamics. This still represents relatively fast pilot commands for the design model, and it allows the l_1 gradient algorithm to take advantage of the analytical formulations.

H_2–l_1 Results

Table 1 shows norm values for several of the H_2–l_1 compensators, and Fig. 4 shows the corresponding simulation results for a 40-m/s climb held for 3 s. As shown, the l_1 constraint can significantly reduce the overshoot without sacrificing rise time and overall speed of response. The large overshoot associated with the higher levels of the 1-norm constraint is especially troublesome, because it represents increased vulnerability of the aircraft. It also represents a potentially hazardous situation for the pilot if a decrease in altitude were commanded during low-level flight.

An additional drawback of the large overshoot associated with the H_2 solution is evident in the pitch-angle response shown in Fig. 5. Neglecting angle of attack, a 40-m/s climb represents an 8-deg pitch angle. Note that the H_2 solution (case 1) has a pitch response that peaks at close to 15 deg, whereas the H_2–l_1 solutions tend to peak closer to 10 deg. Also, a large negative pitch angle (−7 deg) in case 1 results from the rapid decrease in altitude as the

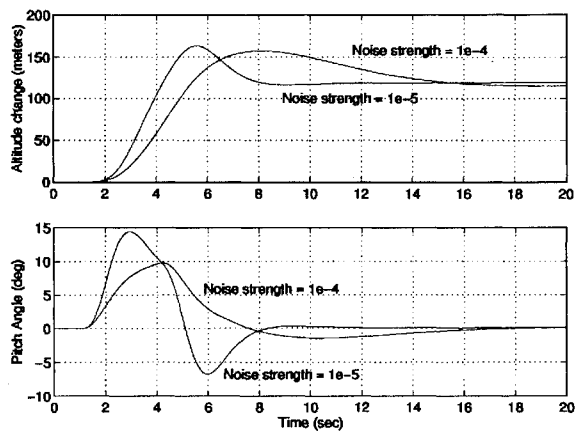

Fig. 3 Sensitivity of H_2 performance to altitude noise strength.

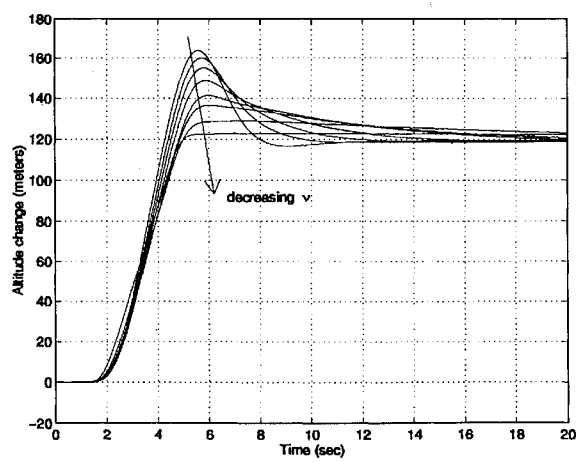

Fig. 4 Response to altitude change for H_2–l_1 designs, cases 1–8.

Table 1 Mixed H_2–l_1 control results

Case no.	$\|T_{zw}\|_2$ α	$\|T_{mr}\|_1$ ν
1 (H_2 optimal)	0.15840	3.086
2	0.15840	2.800
3	0.15841	2.600
4	0.15844	2.400
5	0.15851	2.200
6	0.15861	2.000
7	0.15907	1.800
8	0.16164	1.600

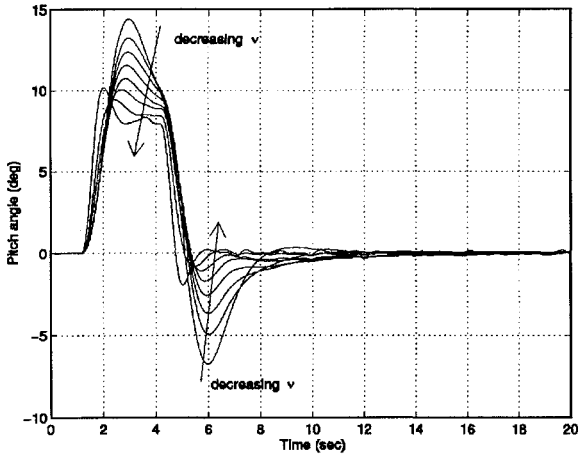

Fig. 5 Pitch-angle response for H_2–l_1 designs, cases 1–8.

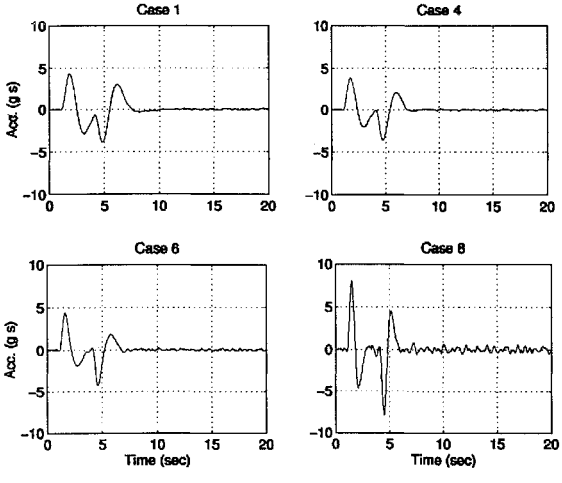

Fig. 6 Normal acceleration for H_2–l_1 designs.

compensator attempts to damp out the large error associated with the overshoot. Most of the H_2–l_1 systems show negative pitch angles of only −2.5 deg or less for the same commanded inputs. These are clearly desirable features of the H_2–l_1 systems. Drawbacks of the H_2–l_1 systems are increased g-loading on the pilot and greater susceptibility to measurement noise as shown in Fig. 6. Although all of the systems demonstrate relatively high g-levels associated with the command response, the desired g-limit of 4 for the simulated command has been met for $\nu = 2.0$ or higher. Furthermore, a slightly reduced bandwidth on the simulation command filter (down to 2–3 rad/s) decreases the peak g-level to below 3. The steady-state portion of the responses shown in Fig. 6 indicate that the pilot gets an increasingly bumpy ride because of measurement noise as the l_1 constraint level is decreased. The modeled flight condition results in sufficient actuator authority, with the peak commanded actuator deflection being less than 5 deg for all cases (not shown), and actuator rates (also not shown) are within 20 deg/s for all cases.

The value of $\|T_{mr}\|_1$ is so conservative for this example that the physical significance is almost meaningless. For instance, $\|T_{mr}\|_1 = 2$ indicates that there could be a worst-case altitude error

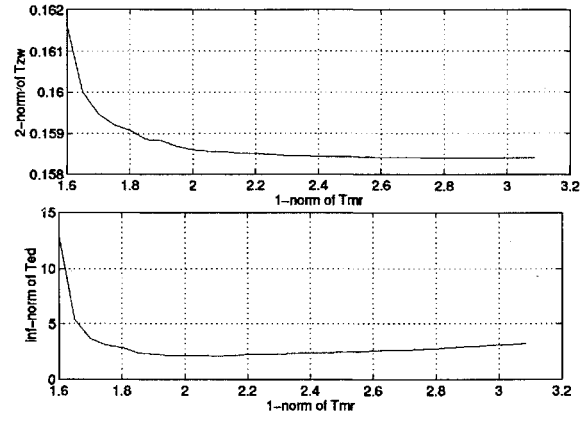

Fig. 7 Resulting norm values for H_2–l_1–h_∞ designs.

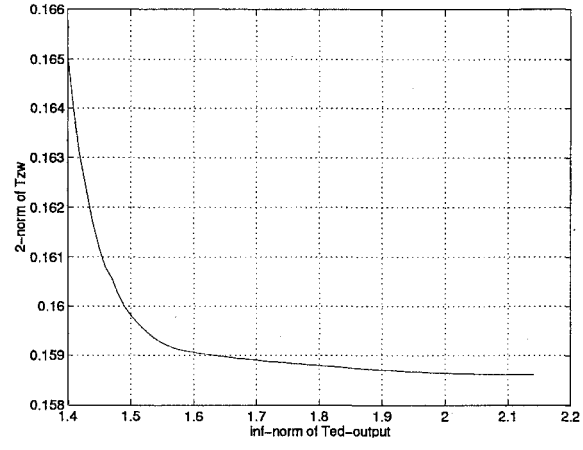

Fig. 8 Pareto-optimal H_2–l_1–H_∞ curve, $\nu = 2.0$.

of 200 m resulting from some combination of a 100-m/s (or less) altitude rate command and 100-m (or less) altitude command. Although this sounds extremely bad, the associated time response for case 6 (Figs. 6 and 7) shows a relatively good response to a combined altitude and altitude-rate command in the presence of wind gusts and measurement noise. For this reason, the actual values of the l_1 norm probably should be used only for comparison purposes in most problems.

H_2–l_1–H_∞ Results

For the H_2–l_1 cases discussed previously, no attempt was made to monitor or improve robust stability levels for the closed-loop system; these concerns were temporarily set aside for the H_∞ subproblem. Figure 8 shows the H_2/l_1 Pareto-optimal curve and the resulting ∞-norms of the output complementary sensitivity (T_{ed}), used here as a measure of robust stability in the presence of output multiplicative uncertainty.[12] The ∞-norm and 1-norm constraints are not competing until $\|T_{mr}\|_1$ is pushed well below 2.0, after which the value of $\|T_{ed}\|_\infty$ increases significantly. On the basis of these curves, and the time responses shown previously, a 1-norm constraint level of $\nu = 2.0$ was chosen as the desired level of tracking. Holding this 1-norm constraint level, $\|T_{ed}\|_\infty$ was then reduced from a starting value of $\gamma = 2.14$ using H_2–l_1–H_∞ optimization, and the resulting Pareto-optimal curve is shown in Fig. 8. An unweighted complementary sensitivity was used for T_{ed}, because the typical high-frequency weight for unmodeled dynamics would take effect beyond the Nyquist sampling frequency for this system. As shown, $\|T_{ed}\|_\infty$ can be reduced further without a significant increase in $\|T_{zw}\|_2$, until γ is pushed below 1.6. The effect of reducing the singular-value peaks of T_{ed} can be seen clearly in Fig. 9, and the appearance of multiple active peaks for the lower γ levels is evident.

As a second measure of robust stability, the independent gain and phase margins are shown in Table 2. The gain margins for MIMO systems cannot be taken as the algebraic union of the sensitivity (IGM_S) and complementary sensitivity (IGM_T) gain margins; therefore, both limits are shown. Modest improvements can be made in

Table 2 Mixed optimal control margins

Case no.	$\|T_{zw}\|_2$	$\|T_{ed}\|_\infty$	IGM_S, dB		IGM_T, dB		IPM, deg
1	0.15840	3.207	[−2.07	2.74]	[−3.26	2.36]	18.0
2	0.15840	2.741	[−2.31	3.15]	[−3.96	2.71]	21.1
3	0.15841	2.526	[−2.43	3.38]	[−4.39	2.90]	22.9
4	0.15844	2.379	[−2.53	3.58]	[−4.75	3.05]	24.3
5	0.15851	2.224	[−2.64	3.81]	[−5.20	3.23]	26.0
6	0.15861	2.140	[−2.73	4.01]	[−5.49	3.34]	27.1
7	0.15907	2.843	[−2.63	3.79]	[−4.96	3.14]	25.1
8	0.16164	12.942	[−1.67	2.08]	[−2.19	1.75]	12.8
9	0.15864	2.000	[−2.85	4.26]	[−6.04	3.53]	29.0
10	0.15879	1.800	[−3.02	4.68]	[−7.07	3.85]	32.3
11	0.15906	1.600	[−3.23	5.19]	[−8.56	4.23]	36.5
12	0.16507	1.400	[−3.46	5.84]	[−10.96	4.70]	42.0

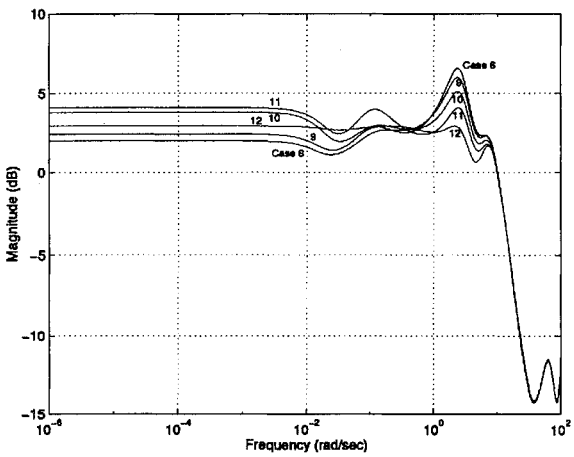

Fig. 9 Output complementary sensitivity for H_2–l_1–H_∞ designs, $\nu = 2.0$.

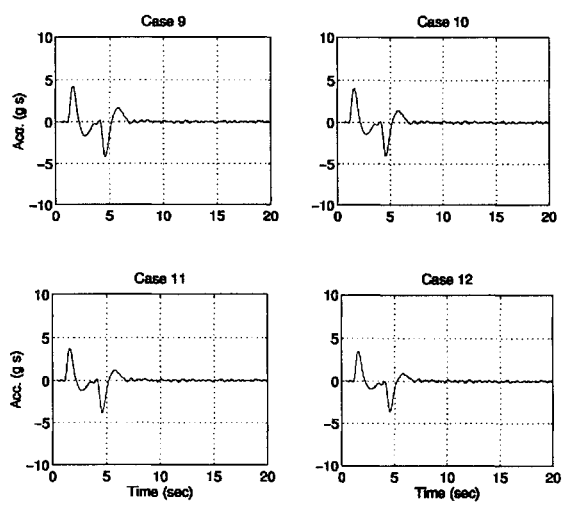

Fig. 10 Normal acceleration for H_2–l_1 designs.

the gain and phase margins with only slight increases in $\|T_{zw}\|_2$. Acceleration (g) levels for cases 9–12 are shown in Fig. 10. An unexpected bonus of improving the robust stability for this particular problem is that g-levels actually decreased. The target level of $4g$ has been met, and elevator commands (again not shown) are very low (2–3 deg) for the simulated commands.

Conclusions

As demonstrated by the terrain-following example, the proposed method is capable of combining the H_2, l_1, and H_∞ norms of different transfer functions into a single, constrained-optimization problem. The method has great practical potential because it yields compensators that are of a specifiable fixed order. This feature of the method eliminates the need for further model reduction. Although the method makes no claims of convexity, global optimality, or convergence, it has proven to be effective with a variety of problems. Conceptually, one drawback of the method is that it provides no information as to how far away the fixed-order solution is from the global (free-order) solution. Although this is an important theoretical issue, it may prove to be less significant in practice. Tradeoff curves showing objective vs controller order for given constraint levels can be generated by exercising the algorithm at increasingly higher (or lower) controller orders.

Using l_1 constraints to achieve tight tracking shows great promise; however, note that the actual value of the 1-norm can be so conservative that its physical significance becomes meaningless. This is important because l_1 optimization often has been mentioned as a method to put absolute magnitude bounds on errors, control usage, and control rates. Although l_1 optimization does an excellent job of limiting the magnitude of controlled outputs, the achieved values of the outputs for actual inputs are usually far less than might be indicated by the value of the norm. For specific inputs, a much less conservative method for limiting error magnitude would be to apply output l_∞ constraints.

Acknowledgment

The research was supported in part by the U.S. Air Force Office of Scientific Research.

References

[1]Walker, D. E., and Ridgely, D. B., "Reduced Order Mixed H_2/H_∞ Optimization with a Singular H_∞ Constraint," *Proceedings of the 1994 American Control Conference* (Baltimore, MD), Inst. of Electrical and Electronics Engineers, Piscataway, NJ, 1994, pp. 1128–1132.

[2]Jacques, D. R., Ridgely, D. B., Canfield, R. A., and Spillman, M. A., "A MATLAB Toolbox for Fixed-Order, Mixed-Norm Control Synthesis," *Proceedings of the 4th IEEE Conference on Control Applications* (Albany, NY), Inst. of Electrical and Electronics Engineers, Piscataway, NJ, 1995, pp. 470–475; also *Control Systems Magazine* (to be published).

[3]Ullauri, J., Walker, D. E., and Ridgely, D. B., "Reduced Order Mixed H_2/H_∞ Optimization with Multiple H_∞ Constraints," *Proceedings of the AIAA Guidance, Navigation, and Control Conference* (Scottsdale, AZ), AIAA, Washington, DC, 1994, pp. 1051–1060 (AIAA Paper 94-3658).

[4]Walker, D. E., "H_2 Optimal Control with H_∞, μ, and L_1 Constraints," Ph.D. Thesis, Dept. of Aeronautics and Astronautics, U.S. Air Force Inst. of Technology, Wright–Patterson AFB, OH, June 1994.

[5]Spillman, M., "Applications of l_1 and Mixed H_2/l_1 Optimization," M.S. Thesis, Dept. of Aeronautics and Astronautics, U.S. Air Force Inst. of Technology, Wright–Patterson AFB, OH, Dec. 1994.

[6]Safonov, M. G., Limebeer, D. J. N., and Chiang, R. Y., "Simplifying the H_∞ Theory via Loop-Shifting, Matrix-Pencils and Descriptor Concepts," *International Journal of Control*, Vol. 50, No. 6, 1989, pp. 2467–2488.

[7]Canfield, R., Ridgely, D. B., and Smith, L., "Numerical Solution to the Fixed-Order, General Mixed H_2/H_∞ Optimal Control Problem," *International Journal of Robust and Nonlinear Control* (submitted for publication).

[8]Mukhopadhyay, V., "Digital Control Law Synthesis Using Constrained Optimization," *Journal of Guidance, Control, and Dynamics*, Vol. 12, No. 2, 1989, pp. 175–181.

[9]Giesy, D. P., and Lim, K. B., "H_∞ Norm Sensitivity Formula with Control System Design Applications," *Journal of Guidance, Control, and Dynamics*, Vol. 16, No. 6, 1993, pp. 1138–1145.

[10]McRuer, D., Ashkenas, I., and Graham, D., *Aircraft Dynamics and Automatic Control*, Princeton Univ. Press, Princeton, NJ, 1973, pp. 700–706.

[11]Athans, M., Castañon, D., Dunn, K.-P., Greene, C. S., Lee, W. H., Sandell, N. R., Jr., and Willsky, A. S., "The Stochastic Control of the F-8C Aircraft Using a Multiple Model Adaptive Control (MMAC) Method—Part 1: Equilibrium Flight," *IEEE Transactions on Automatic Control*, Vol. AC-22, No. 5, 1977, pp. 768–780.

[12]Ridgely, D. B., and Banda, S. S., "Introduction to Robust Multivariable Control," U.S. Air Force Wright Aeronautical Labs., AFWAL-TR-85-3102, Wright–Patterson AFB, OH, 1986.