

Identifying \mathcal{H}_∞ -Models: An LMI Approach

Gray C. Thomas and Luis Sentsis

Abstract—Practical application of \mathcal{H}_∞ robust control relies on system identification of a valid model-set, described by a norm-bounded differential inclusion, which explains all possible behavior for the control plant. This is usually approximated by measuring the plant repeatedly and finding a model that explains all observed behavior. Typical modern approaches must anticipate the uncertainty-shaping aspects of the final model in order to maintain tractability. This paper offers a linear matrix inequality constrained optimization for the MIMO model fitting problem that does not require such knowledge. We do this with a novel “Quadric Inclusion Program” which replaces the least squares problem in traditional model identification—however rather than linear equation models, it returns linear norm-bounded inclusion models. We prove several key properties of this algorithm and give a geometric interpretation for its behavior. While we stress that the models are outlier-sensitive by design, we offer a method to mitigate the effect of measurement noise. The paper includes an example of how the theory could be applied to frequency domain data. Time-domain data could also be used, provided a state vector is constructed from measured signals and their derivatives to use as regressors for a vector of maximal derivatives.

I. INTRODUCTION

SYSTEM identification of a valid \mathcal{H}_∞ plant model marks the first obstacle to applying the robust \mathcal{H}_∞ control theory of e.g. [1], [2], [3]. If this model is not believable then \mathcal{H}_∞ synthesis provides guesses rather than guarantees—with the parameters of the uncertainty acting as tunable knobs. In many cases this is an acceptable strategy, but in some cases uncertainty demands more accurate measurement. Guesswork is done conservatively, and conservatism in the uncertainty model can degrade performance. When uncertainty is the performance-limiting factor, we expect this uncertainty model to represent some sort of physical limit to the plant.

Most identification stems from the celebrated prediction error method [4] which produces high quality linear models complete with a measure of model certainty in the form of a model parameter covariance matrix. This parameter covariance, its implication for robust control, its improved value when using instrumental variables (or orthonormal basis function parameterizations), and the influence of weighting functions and closed loop identification controllers on it have all been extensively studied [5], [6], [7], [8], [9]. This confidence measure is often taken out of context, however, as it represents only the distribution of models which would result from the same identification process if the data were regenerated. Prediction error uncertainty is not capable of

representing the influence of a nonlinearity [10]. Moreover, with additional data the model parameter covariance will decrease even if the error variance is constant—a sought after property of consistency—but a property which clearly indicates that the parameter covariance is not a measure of any physical property.

A paradigm known as stochastic embedding [11], [10], has been proposed to work around this—adding an additional source of uncertainty to the computation of parameter covariance. By supposing that the model parameters are sampled from a distribution with pre-defined covariance, the stochastic embedding approach estimates the means of these parameter distributions rather than the parameters themselves—and returns a much more conservative covariance estimate. This covariance does not approach zero with more samples—instead it approaches the a priori covariance. Recent developments in Stochastic embedding parallel and predate the modeling assumption of this paper, in that they allow experimenters to gather data which represents a model-set rather than a specific model [12], [13].

The primary alternative to prediction error identification is broad-spectrum frequency-domain estimation [14]. This approach uses a ratio of the Fast Fourier Transform (FFT) spectra of the input to the output. To eliminate noise, the FFT data must be averaged in the frequency domain, often weighted by the magnitude of the input (or occasionally by the magnitude of the output)—making a ratio of cross spectrum to power spectrum. An uncertainty boundary can be obtained by repeatedly generating estimates of the transfer function and then drawing a bound around them numerically [2], but our method takes this further and simultaneously optimizes the model and the shape matrices. \mathcal{H}_∞ -oriented identification based on corrupted point-samples of the frequency response have been analyzed before in a single input, single output setting [15], [16], but there are no real shape matrices for SISO systems, and the approach assumes a unique true model (i.e. not a true model-set).

The popular domain of model validation (through lack of invalidation) tests a priori model-sets on time domain data [17]. This approach uses Kalman Yakubovitch Popov Lemma alchemy to commute frequency domain bounds to time domain bounds on “uncertainty” signals, and tests for the satisfiability of those bounds using convex optimization (linear programming) within a finite horizon. An elegant approach to be sure, but not one which identifies model-sets. Nor one which easily could, since adding flexibility in the model-set would make the problem non-convex.

In an effort to capture a physical component to the \mathcal{H}_∞ robust model, we assume the system is repeatable (though potentially corrupted by noise) if a condition vector [18]—comprised of the factors that cause the real system to deviate

This work was supported by NASA Space Technology Research Fellowship grant NNX15AQ33H, “Controlling Robots with a Spring in Their Step,” for which Gray is the Fellow and Luis is the Advising Professor. Authors are with—respectively—the Departments of Mechanical and Aerospace Engineering, University of Texas at Austin, Austin, TX 78712, USA. Send correspondence to gray.c.thomas@utexas.edu

from a unique linear model—is held constant. This includes input signal amplitudes and operating points for sinusoidal tests on nonlinear systems, and also exogenous signals like the pressure or Mach number for aircraft. By repeating stepped sine tests using various condition vectors we obtain a structured data set which represents not a single model, but a model-set. Averaging results with the same condition vector allows us to separate the effects of measurement noise from plant variability. With data representing plant variability, we can fit \mathcal{H}_∞ model-sets to include all observed variation—using convex programs inspired by the minimal bounding ellipsoid problem [19]. If the plant is non-linear and the condition vector includes the signal amplitudes, then the approach is conceptually similar to bounding the describing function [20] for some range of input magnitudes.

In this paper, Section II introduces the Quadric Inclusion Program (QIP), a convex program that identifies norm-bounded linear inclusions (a type of model-set with good scaling properties) from a regressor of inputs and a vector of outputs—much like the ordinary least squares algorithm identifies equality models from a similar starting point. Section VII discusses the general applicability of the QIP, and highlights some open problems. Section VI applies this QIP to identifying dynamical systems, using an output-based state-vector construction strategy from nonlinear and adaptive control theory. While this might suggest a time-domain strategy, Section VIII provides an example of how frequency domain, condition-group based data sets in the style of [18] can be used to fit a state-space uncertain model.

II. THE QUADRIC INCLUSION PROGRAM

In this section we consider the linear norm-bounded inclusion $y \in \{(A + B\Delta C)x : \Delta^T \Delta \preceq I\}$ with $y \in \mathbb{R}^{n_y}$ and $x \in \mathbb{R}^{n_x}$ vectors, A a real matrix of appropriate dimension, B and C real invertible square matrices, and Δ unknown but norm bounded¹: $\Delta^T \Delta \preceq I$. If the linear equation $y = Ax$ is geometrically analogous to a line, then this inclusion is analogous to a cone. In this section we present a convex program which finds the real-valued parameters of the inclusion, the elements of the A , B , and C matrices, based on a series of measurements of x and y (with y potentially noisy).

However, we must apply a lossless convexification, as the problem is not naturally amenable to convex optimization tools. And for this convexification we need to invoke an alternative form for this inclusion:

Proposition 1 (Quadratic Form). *A pair of input and output vectors (x, y) satisfies the inclusion*

$$y \in \{(A + B\Delta C)x : \|\Delta\| \leq 1\}, \quad (1)$$

with full rank B and C if and only if it satisfies the following quadratic form inequality:

$$\begin{pmatrix} y \\ x \end{pmatrix}^T \underbrace{\begin{pmatrix} -B^{-T}B^{-1} & B^{-T}B^{-1}A \\ A^T B^{-T}B^{-1} & C^T C - A^T B^{-T}B^{-1}A \end{pmatrix}}_Q \begin{pmatrix} y \\ x \end{pmatrix} \geq 0. \quad (2)$$

¹ $\Delta^T \Delta \preceq I \iff \|\Delta\| \leq 1$.

Proof. First, consider x and y which satisfy (1):

$$\Delta Cx = B^{-1}(y - Ax), \quad (3)$$

$$\|Cx\|^2 \geq \|\Delta Cx\|^2 = \|B^{-1}(y - Ax)\|^2, \quad (4)$$

$$0 \leq x^T C^T Cx - (y - Ax)^T B^{-T} B^{-1} (y - Ax). \quad (5)$$

Which is equivalent to (2). Conversely, for x and y which satisfy (2) we can choose $\Delta = B^{-1}(y - Ax)x^T C^T / (x^T C^T Cx)$ to satisfy both $y = Ax + B\Delta Cx$ (trivially) and $\Delta^T \Delta \preceq I$:

$$\Delta^T \Delta = \frac{Cx(y - Ax)^T B^{-T} B^{-1} (y - Ax)x^T C^T}{(x^T C^T Cx)^2} \quad (6)$$

$$= \gamma \frac{Cxx^T C^T}{(x^T C^T Cx)}, \quad (7)$$

with $\gamma = (y - Ax)^T B^{-T} B^{-1} (y - Ax) / (x^T C^T Cx) \leq 1$ as $x^T C^T Cx \geq (y - Ax)^T B^{-T} B^{-1} (y - Ax)$. This is rank 1, positive semi-definite, with $\gamma \leq 1$ as the only non-zero eigenvalue. This ensures that $\Delta^T \Delta \preceq I$. \square

In this form, the inclusion has become a linear inequality constraint on the elements of Q . But not all symmetric matrices will have the appropriate structure to be interpreted as Q for the purpose of backing out the inclusion matrices. Fortunately, we can re-parameterize around this issue by decomposing the matrix Q .

Definition 1 (SS-DD). The matrix Q can be expressed as the difference of two positive semi-definite matrices. Using linear matrix inequality constraints on four new matrix variables, we can construct a similar structure to Q , which we call the Split Semi-Definite Decomposition (SS-DD). The following three equations constrain the new SS-DD variables X_B , X_A , X_{AA} and X_C :

$$0 \preceq Q' \triangleq \begin{pmatrix} X_B & -X_A \\ -X_A^T & X_{AA} \end{pmatrix}, \quad (8)$$

$$0 \preceq X_C, \quad (9)$$

$$Q = \begin{pmatrix} -X_B & X_A \\ X_A^T & X_C - X_{AA} \end{pmatrix}. \quad (10)$$

However, the two structures are not exactly the same. All Quadratic form Q have an SS-DD form, but the converse is not true in general; the Q of the SS-DD will only match the Q structure in (2) in a special case:

$$X_{AA} = X_A^T X_B^{-1} X_A, \quad (11)$$

that is, if we can write the RHS of (8) as

$$\begin{pmatrix} X_B \\ -X_A^T \end{pmatrix} (X_B^{-1}) \begin{pmatrix} X_B & -X_A \end{pmatrix}, \quad (12)$$

then the rank of the RHS of (8) is the rank of X_B and the SS-DD structure matches that of (2):

$$\begin{cases} X_B & = & B^{-T} B^{-1} \\ X_A & = & B^{-T} B^{-1} A \\ X_{AA} & = & A^T B^{-T} B^{-1} A \\ X_C & = & C^T C \end{cases} \quad (13)$$

The SS-DD is a lossless convexification of the search space, and we will introduce a convex program which is formatted to follow the rules of disciplined convex programming [21]

to find the SS-DD. However, that program must also result in satisfaction of (11) for all optimal solutions if the variables are to be interpreted as an inclusion. This is guaranteed for our QIP, but not for other cost functions or additional constraints involving the SS-DD.

Problem 1 (Degenerate Quadric² Inclusion Program). The inclusion that minimizes a width-like cost³ while including a list of data point pairs $\xi_i = (x_i^T, y_i^T)$, $i = 1, \dots, N$ can be found by the following convex optimization program

$$\begin{aligned} & \text{maximize} && \log(\det(X_B)) \\ & \text{over} && Q, X_B, X_A, X_{AA}, X_C \\ & \text{subject to} && \text{SS-DD equations (8)-(10)} \\ & && 1 = \text{tr}[X_C] \\ & && 0 \leq \xi_i^T Q \xi_i \quad \forall i \in 1 \dots N \end{aligned} \quad (14)$$

Proposition 2. *The SS-DD satisfies (11) for all solutions to Prob. 1 with finite cost.*

Proof. In the general case where (11) does not hold, we can define an equation error matrix

$$\tilde{X}_{AA} \triangleq X_{AA} - X_A^T X_B^{-1} X_A \succeq 0, \quad (15)$$

which is p.s.d. since it is a Schur complement of Q' , the matrix in (8). Suppose \tilde{X}_{AA} has a non-zero eigenvalue $\lambda > 0$ and corresponding eigenvector v . Now consider another potential solution identical in all ways but one to the previous solution: $X'_{AA} = X_{AA} - v\lambda v^T$. Since the other matrices are the same, the new solution trivially satisfies all constraints which do not involve X_{AA} , leaving (8)—satisfied by Schur complement since $\tilde{X}_{AA} - v\lambda v^T \succeq 0$, and the inclusion inequalities, which the new solution relaxes to (recall that x_i is the lower part of ξ_i):

$$0 \leq \xi_i^T Q \xi_i + x_i^T v \lambda v^T x_i. \quad (16)$$

Relaxing constraints increases the objective function for our maximization problem. Assuming that the objective was previously bounded by the inclusion inequalities in all the various ways it could increase, at least one of them must have been relaxed by the change (albeit potentially through rearrangement of the X_C matrix), demonstrating that a higher objective solution must exist—a contradiction of the premise that optimal solutions can have non-zero equation error for equation (11). \square

Remark 1 (Trivial Solution). A trivial solution to the constraints always exists, with $Q' = 0$, $X_C = \lambda I : \text{tr}[X_C] = 1$. This is the worst possible solution, since it represents an inclusion with infinite uncertainty magnitude, and it has infinitely negative optimization function value.

Remark 2 (Linear Solutions). If there are insufficient data points, or if the data points share a perfect linear relationship, X_B will have an unbounded eigenvalue. In this situation, the

optimization function value will be infinite and the solution⁴ will not be unique.

The next three sections explain the cost function, prove consistency of estimation, and relax the degenerate quadric inclusion program to allow for Gaussian measurement noise in y .

III. A GEOMETRIC NOTION OF CONE WIDTH FOR DEGENERATE QUADRIC “CONES”

The set of all points (x, y) satisfying the quadratic inequality (2) has a geometric interpretation—a shape which is technically a degenerate quadric. Quadrics in 2D space are familiar to many: non-degenerate varieties include ellipses, hyperbola, and circles; but cones are degenerate. Degenerate quadrics are a general class of hyperdimensional shapes: they are described by a symmetric matrix quadratic form inequality relative to zero. Non-degenerate quadrics are similar, but with this inequality relative to some non-zero constant. With some positive and some negative eigenvalues (separated in the SS-DD), our degenerate quadrics have a useful analog in the simple 2D cone.

Definition 2 (Cone width for a toy cone). Consider the following 2D cone in real scalar x, y space: $|y - Ax| \leq r(x) = w|x|$. We call $r(x)$ the “radius” of the cone opening as a function of x , and w the “width” of this cone. We can equivalently define this width as $\sqrt{\mathbb{E}(r(x)^2)}$ if $x \sim \mathcal{N}(0, 1)$, where \mathbb{E} is expected value, and $\mathcal{N}(0, 1)$ is the normal distribution with mean 0 and variance 1.

Proof. $\sqrt{\mathbb{E}(r(x)^2)} = \sqrt{w^2 \mathbb{E}(|x|^2)} = w$. \square

Definition 3 (Characteristic radius of a cross section). Consider a special case SS-DD satisfying (10)–(13) such that it is equivalent to an inclusion. If we specify a particular input x , the space of included y can be interpreted as a geometric shape: a hyper-ellipsoid. The characteristic radius $R(x)$ of this cross section is defined as the radius of the hyper-ball that has equal hyper-volume to this hyper-ellipsoid. The hyper ellipsoid can be described as

$$(y - Ax)^T X_B (y - Ax) \leq G^2(x), \quad G(x) \triangleq \sqrt{x^T X_C x}; \quad (17)$$

and the characteristic radius,

$$R(x) = G(x) (\det(X_B^{-1}))^{\frac{1}{2n_y}}, \quad (18)$$

where $y \in \mathbb{R}^{n_y}$.

Definition 4 (Generalized Cone Width). The generalized cone width is the square root of the expected value of the squared characteristic radius given inputs drawn from the standard multivariate normal distribution. That is

$$W \triangleq \sqrt{\mathbb{E}(R^2(x))} \quad | \quad \mathbb{E}(xx^T) = I \quad (19)$$

Theorem 3 (Generalized Cone Width). *The generalized cone width of a special case SS-DD form satisfying (10)–(13)*

$$W = (\det(X_B^{-1}))^{\frac{1}{2n_y}} \sqrt{\text{tr}(\tilde{X}_C)}. \quad (20)$$

⁴if one is returned at all—data sets like this typically cause numerical solvers to fail.

²Named for the geometric shape explained in the next subsection.

³ $\text{GM}\{\sigma(B)\} \|C\|_{\text{Frobenius}}$, where GM denotes the geometric mean; $\sigma(B)$, the spectrum, or set of singular values. Cone-width analogy explained in the next subsection.

Proof. The width,

$$W = \sqrt{\mathbb{E}(R^2(x))} = (\det(X_B^{-1}))^{\frac{1}{2n_y}} \sqrt{\mathbb{E}(G^2(x))}; \quad (21)$$

$$\mathbb{E}(G^2(x)) = \mathbb{E}(x^T X_C x) \quad (22)$$

$$= \text{tr}[X_C \mathbb{E}(xx^T)] = \text{tr}[X_C]. \quad (23)$$

Substitution yields (20). \square

We use the following corollary to avoid having both X_C and X_B in the cost function of the QIP.

Corollary 3.1. *if $\text{tr}(X_C) = 1$,*

$$\log(W) = -\frac{1}{2n_y} \log(\det(X_B)) \quad (24)$$

Proof. $\log(W) = \log\left((\det(X_B^{-1}))^{\frac{1}{2n_y}} \sqrt{\text{tr}(X_C)}\right) = \frac{1}{2n_y} \log(\det(X_B^{-1})) + \frac{1}{2} \log(\text{tr}(X_C)) = \frac{-1}{2n_y} \log(\det(X_B)) + \frac{1}{2} \log(1)$ and $\log(1) = 0$. \square

The maximization objective $\log(\det(X_B))$ in Prob. 1 is a negative multiple of this expression for generalized cone width (and therefore minimizes it).

IV. PROOF OF CONSISTENT ESTIMATION

In this section we use properties of the generalized cone width to prove that, when data is generated from a norm-bounded linear inclusion, the estimates from the degenerate QIP converge, in a certain sense, to equivalence with the true inclusion. This is essentially a property of the choice of cost function, and could be otherwise stated ‘‘every norm-bounded inclusion is optimal for the data it produces’’.⁵

Norm-bounded linear inclusions are functionally equivalent (include the same points) up to an orthogonal pre-multiplication of C , an orthogonal post-multiplication of B , and reciprocal scaling of C and B , and the cone width does not to change due to any such alteration.

Corollary 3.2. *Generalized cone width is the product of the geometric mean of the singular values of B , and the 2-norm of the singular values (the Frobenius norm) of C*

Proof. Converting the model-set from standard form (1) to the special case of the SS-DD form (10)–(11) and applying Thm. 3,

$$W = (\det BB^T)^{\frac{1}{2n_y}} \sqrt{\text{tr}[C^T C]}, \quad (25)$$

$$= \prod_{\lambda \in \sigma(B)} \left(\lambda^{\frac{1}{n_y}}\right) \sqrt{\sum_{\gamma \in \sigma(C)} \gamma^2}. \quad (26)$$

Where the spectrum of a matrix $\sigma(\cdot)$ is the set of the singular values of that matrix (with repetition). \square

Multiplication by a orthogonal matrix cannot change the singular values of a matrix, and cannot influence the cone width.

⁵Without this property one could imagine a cost function that always opts for sphere-like models, scaled identity matrices for B and C : any information about the shape of the uncertainty in the data would be ignored.

Corollary 3.3. *The generalized cone width is invariant to scaling C and B by reciprocal values.*

Proof. $W' = \text{GM}\{\sigma(\alpha^{-1}B)\} \|\alpha C\|_{\text{Frobenius}} = \frac{\alpha}{\alpha} W$. \square

The cone width also satisfies an intuitive notion that a cone can only contain another cone if it is wider. In terms of inclusions, this geometric containment becomes a concept of implication: if one inclusion is implied by a second, this is equivalent to saying that the second inclusion is geometrically contained within the first.

Lemma 1. *An inclusion $y \in \{(A_o + B_o \Delta C_o)x : \Delta^T \Delta \preceq I\}$ (subscript o for outer) contains another inclusion $y \in \{(A_i + B_i \Delta C_i)x : \Delta^T \Delta \preceq I\}$ (subscript i for inner) if and only if*

$$\|\tilde{A} + \tilde{B} \Delta \tilde{C}\| \leq 1 \quad \forall \Delta \mid \|\Delta\| \leq 1, \quad (27)$$

with $\tilde{B} \triangleq B_o^{-1} B_i$, $\tilde{A} \triangleq B_o^{-1}(A_i - A_o) C_o^{-1}$, and $\tilde{C} \triangleq C_i C_o^{-1}$.

Proof. Geometric containment of inclusion shapes is logical implication of inclusion inequalities:

$$\left\{ \begin{array}{l} y = A_o x + B_o \Delta_o C_o x \\ \Delta_o^T \Delta_o \preceq I \end{array} \right\} \iff \left\{ \begin{array}{l} y = A_i x + B_i \Delta C_i x \\ \Delta^T \Delta \preceq I \end{array} \right\}. \quad (28)$$

The two equalities define a relationship between Δ and Δ_o , so we can equivalently state

$$\left\{ \begin{array}{l} \Delta_o^T \Delta_o \preceq I \iff \Delta^T \Delta \preceq I \\ A_o + B_o \Delta_o C_o = A_i + B_i \Delta C_i \end{array} \right. \quad (29)$$

This is because (28) holds for all x . By algebra, $\Delta_o = \tilde{A} + \tilde{B} \Delta \tilde{C}$. Re-stating the implication in (29) as $\|\Delta_o\| \leq 1 \quad \forall \Delta \mid \|\Delta\| \leq 1$ we get (27). \square

This eventually leads to a necessary condition for inclusion based on a cone width inequality. An intermediate necessary condition uses the singular values of the B and C matrices.

Lemma 2. *If $y \in \{(A_o + B_o \Delta C_o)x : \Delta^T \Delta \preceq I\}$ includes $y \in \{(A_i + B_i \Delta C_i)x : \Delta^T \Delta \preceq I\}$ then $\sigma_{\max}(\tilde{B}) \sigma_{\max}(\tilde{C}) \leq 1$.*

Proof. Assume the contrary ($\sigma_{\max}(\tilde{B}) \sigma_{\max}(\tilde{C}) > 1$) and construct the following:

$$\Delta = \tilde{B}^T \zeta_B \text{sign}(\zeta_B^T \tilde{A} \zeta_C) \zeta_C^T \tilde{C}^T, \quad (30)$$

where ζ_B and ζ_C are unit eigenvectors for $\tilde{B} \tilde{B}^T$ and $\tilde{C}^T \tilde{C}$ corresponding to their respective maximum eigenvalues: $\lambda_{\max}(\tilde{B} \tilde{B}^T) \zeta_B = \tilde{B} \tilde{B}^T \zeta_B$, $\lambda_{\max}(\tilde{C}^T \tilde{C}) \zeta_C = \tilde{C}^T \tilde{C} \zeta_C$, $\zeta_B^T \zeta_B = \zeta_C^T \zeta_C = 1$. This choice of Δ leads to:

$$\begin{aligned} 1 &\geq \|\tilde{A} + \tilde{B} \Delta \tilde{C}\| \geq |\zeta_B^T (\tilde{A} + \tilde{B} \Delta \tilde{C}) \zeta_C| \\ &= |\zeta_B^T (\tilde{A} + \tilde{B} \tilde{B}^T \zeta_B \text{sign}(\zeta_B^T \tilde{A} \zeta_C) \zeta_C^T \tilde{C}^T \tilde{C}) \zeta_C| \\ &= |\zeta_B^T \tilde{A} \zeta_C| + \lambda_{\max}(\tilde{B} \tilde{B}^T) \lambda_{\max}(\tilde{C}^T \tilde{C}), \end{aligned} \quad (31)$$

$$\geq (\sigma_{\max}(\tilde{B}) \sigma_{\max}(\tilde{C}))^2 > 1. \quad (32)$$

A contradiction, as desired. \square

Proposition 3. *If the outer inclusion has a generalized cone width W_o ; and the inner, W_i then $W_o \geq W_i$.*

Proof. Assume the contrary ($W_i > W_o$),

$${}^{2n_y} \sqrt{\det B_i B_i^T} \|C_i\|_F > {}^{2n_y} \sqrt{\det B_o B_o^T} \|C_o\|_F, \quad (33)$$

$$\frac{{}^{n_y} \sqrt{\det B_i B_i^T}}{{}^{n_y} \sqrt{\det B_o B_o^T}} \|C_i\|_F^2 > \|C_o\|_F^2, \quad (34)$$

$$\text{tr} \left({}^{n_y} \sqrt{\det(B_o^{-1} B_i B_i^T B_o^{-T})} C_i^T C_i - C_o^T C_o \right) > 0, \quad (35)$$

$$\text{tr} \left(C_o^T \left(\lambda_{\max}(\tilde{B}\tilde{B}^T) \tilde{C}^T \tilde{C} - I \right) C_o \right) > 0. \quad (36)$$

However by Lemma 2, the argument of trace in the above inequality is negative semi definite, so it can not have a positive trace—a contradiction. \square

The special case where the cone widths reach equality marks the residual set of a Lyapunov-like argument in the consistency proof. Conveniently, this residual set has only one element.

Proposition 4. *If $W_o = W_i$ then our two inclusions are equivalent in the sense that $A_o = A_i$, and $\exists \lambda > 0 : \lambda B_o B_o^T = B_i B_i^T$, $C_o^T C_o = \lambda C_i^T C_i$.*

Proof. When the two widths are equal, the derivation which produced (35) yields:

$$\text{tr} \left(C_o^T \left({}^{n_y} \sqrt{\det(\tilde{B}\tilde{B}^T)} \tilde{C}^T \tilde{C} - I \right) C_o \right) = 0. \quad (37)$$

Yet as before, singular values are limited by the inclusion constraint, and this guarantees (31) and a long series of matrix inequalities,

$$\begin{aligned} I &\succeq |\zeta_B^T \tilde{A} \zeta_C| I + \lambda_{\max}(\tilde{B}\tilde{B}^T) \lambda_{\max}(\tilde{C}^T \tilde{C}) I \\ &\succeq \lambda_{\max}(\tilde{B}\tilde{B}^T) \tilde{C}^T \tilde{C} \succeq {}^{n_y} \sqrt{\det(\tilde{B}\tilde{B}^T)} \tilde{C}^T \tilde{C} = I, \end{aligned} \quad (38)$$

with this last equality due to the combination of (37) and the last inequality above (which can be extended to

$$C_o^T \left({}^{n_y} \sqrt{\det(\tilde{B}\tilde{B}^T)} \tilde{C}^T \tilde{C} - I \right) C_o \preceq 0, \quad (39)$$

another negative semi-definite matrix) ultimately forcing the inner matrix difference to be zero (as it is both negative semi-definite and has trace zero).

With both the first and last element identity, (38) is actually a long chain of equalities. This gives

$$\lambda \triangleq \lambda_{\max}(\tilde{B}\tilde{B}^T) = {}^{n_y} \sqrt{\det(\tilde{B}\tilde{B}^T)}, \quad (40)$$

$$|\zeta_B^T \tilde{A} \zeta_C| = 0, \quad \lambda \tilde{C}^T \tilde{C} = I, \quad (41)$$

from which it follows that $C_o^T C_o = \lambda C_i^T C_i$. When the geometric mean of the eigenvalues is equal to the largest eigenvalue, all the eigenvalues must be equal; thus $\tilde{B}\tilde{B}^T = \lambda I$ (or equivalently $\lambda B_o B_o^T = B_i B_i^T$). Since both $\tilde{B}\tilde{B}^T$ and $\tilde{C}^T \tilde{C}$ have only one eigenvalue with high multiplicity, the eigenvectors ζ_B and ζ_C can be any unit vectors. This in turn guarantees $\tilde{A} = 0$, that is, $A_o = A_i$, completing the conditions necessary for the two model-sets to be equivalent. \square

Using these preliminaries, we can prove the following notion of estimation consistency, noting that without some knowledge of how frequently the true inclusion generates

extreme data—data on the very edge of the inclusion—it is impossible to claim any rate of convergence.

Theorem 4 (Estimation Consistency). *Consider an infinite list of input output data $\xi_i = [y_i^T, x_i^T]^T \forall i \in \mathbb{N}$ points from the inclusion $y_i \in \{(A_T + B_T \Delta C_T)x_i : \Delta^T \Delta \preceq I\}$ with generalized cone width W_T (subscript T for true) in the sense that any possible output will eventually be produced within a non-zero tolerance. Suppose that inclusion estimates $y \in \{(A_n + B_n \Delta C_n)x : \Delta^T \Delta \preceq I\}$ with generalized cone width W_n are calculated via Prob. 1 using the subset of data indexes $i = 1, \dots, n$ as n increases towards infinity.*

Then $W_{n+1} \geq W_n \forall n \in \mathbb{N}$, and $W_n \leq W_T$.

Most importantly, the identification procedure is consistent in the sense that if $\exists n'$ such that the inclusion width stops changing, $W_n = W_{n'} \forall n \geq n'$, then the n^{th} result inclusion must be equivalent to the generating inclusion.

Proof. The first claim follows from the nature of the maximization: more constraints can only reduce the objective, this objective is proportional to the negative log of the width, and log is monotonic. The second is a consequence of the true inclusion being a feasible solution to the optimization problem: the optimal solution has the maximal objective over all feasible solutions. As for the third, suppose to the contrary that the inclusions are distinct. The n^{th} result inclusion cannot contain the true inclusion because it has lesser or equal cone width and is not the same (by supposition). There must be points within the true inclusion and outside the n^{th} result inclusion. And these points, which will eventually occur for some $n > n'$, will not satisfy the inclusion inequalities with the n^{th} result inclusion—contradicting the notion that the estimates could stop changing without reaching the true inclusion. \square

Convergence to a non-trivial inclusion is an important distinguishing aspect of this style of identification. Inclusions which are built on the error estimates in a least-squares fit [4], [8], [22] notably lack this property—converging instead towards a unique model (one element inclusion) as the estimated parameter covariance vanishes with additional samples.

V. A NON-DEGENERATE QUADRIC FOR NOISY MEASUREMENTS

We now consider the notion that the measurement \hat{y} is a deterministic function of a condition vector c that includes x , corrupted by stochastic zero-mean measurement noise η (from a potentially c -dependent distribution): $\hat{y} = f(c) + \eta$. These assumptions allow us to take advantage of the central limit theorem: we average multiple samples \hat{y} to significantly reduce the effect of noise. As the number of averaged samples grows, the distribution of the average approaches a Gaussian distribution and its covariance shrinks towards zero. With many samples of \hat{y} , we can find both the sample mean, $\bar{y} = \sum_{n=1}^N \hat{y}_n / N$, and sample covariance—which gives us an estimate of the distribution of this mean, $\Sigma_\eta = \sum_{n=1}^N (\hat{y}_n - \bar{y})(\hat{y}_n - \bar{y})^T / (N^2 - N)$. We therefore expect that the noisy measurement case approaches the noiseless case as the number of averaged samples increases.

Yet the practical limits on the number of samples that can be obtained forces us to consider the intermediate case where noise is small and Gaussian, but not entirely eliminated. That is, data $\bar{y} = f(c) + \bar{\eta}$, with $\bar{\eta} \sim \mathcal{N}(0, \Sigma_{\eta})$. In this scenario, our degenerate quadric model is inflexible near zero input: a zero x must produce a zero y , and a near-zero x must produce a near-zero y unless the cone is preposterously wide. But \bar{y} can take on a non-zero value even for zero y . To address this issue, we introduce the idea of fitting a non-degenerate quadric relaxation of the model.

The derivation starts with a noise-corrupted version of (2):

$$\xi^T Q \xi = \left(\bar{\xi} - \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix} \right)^T Q \left(\bar{\xi} - \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix} \right) \geq 0; \quad (42)$$

$$\bar{\xi}^T Q \bar{\xi} \geq 2\xi^T Q \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix} - \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix}^T Q \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix}, \quad (43)$$

for which we can offer a high-likelihood lower-bound on each of the two RHS terms. First, $2\xi^T Q \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix}$ is normal, with zero mean, covariance $4\xi^T Q \begin{pmatrix} \Sigma_{\eta} & 0 \\ 0 & 0 \end{pmatrix} Q \xi$, and standard deviation $2 \left\| \begin{pmatrix} \Sigma_{\eta}^{1/2} & 0 \\ 0 & 0 \end{pmatrix} Q \xi \right\|_2 \leq 2 \left\| \begin{pmatrix} \Sigma_{\eta}^{1/2} & 0 \\ 0 & 0 \end{pmatrix} Q \xi \right\|_1$; the standard deviation is bounded by a linear-programming compatible expression in Q . Since the inverse survival function of a normal distribution is linear in the standard deviation, this expression can be used to lower bound the first term to arbitrary likelihood. However this term works to make the model less conservative in doing so, and does not greatly influence the behavior near the origin, where the effects of noise are the most problematic. It is therefore reasonable to take a conservative approach and ignore this term.

The second term is more helpful, and is bounded in magnitude by a chi-square distributed value and an expression which is linear in X_B :

$$\begin{aligned} - \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix}^T Q \begin{pmatrix} \bar{\eta} \\ 0 \end{pmatrix} &= \bar{\eta}^T X_B \bar{\eta}, \\ &\leq \|\Sigma_{1/2} X_B \Sigma_{1/2}\| \|\Sigma_{1/2}^{-1} \bar{\eta}\|^2, \\ &\leq \text{tr}[\Sigma_{\eta} X_B] \nu, \quad \nu \sim \chi_{n_y}^2. \end{aligned} \quad (44)$$

Choosing a constant threshold α based on the survival function of $\chi_{n_y}^2$, the inclusion's quadratic form inequality threshold is shifted to provide an arbitrarily low chance of feasibility problems with low-magnitude inputs:

Problem 2 ((Non-Degenerate) Quadric Inclusion Program).

$$\begin{aligned} &\text{maximize} \quad \log(\det(X_B)) \\ &\text{over} \quad Q, X_B, X_A, X_{AA}, X_C \\ &\text{subject to} \quad \text{SS-DD equations (8)-(10)} \\ &1 = \text{tr}[X_C] \\ &0 \leq \alpha \text{tr}[X_B \Sigma_y] + \xi_i^T Q \xi_i \quad \forall i \in 1 \dots N \end{aligned} \quad (45)$$

Note that, since this modification changes the constraints of the problem, it requires re-examining Prop. 2. Fortunately, (16) does not gain any terms which would invalidate the proof as a result of this noise modification, and we can therefore accept it safely.

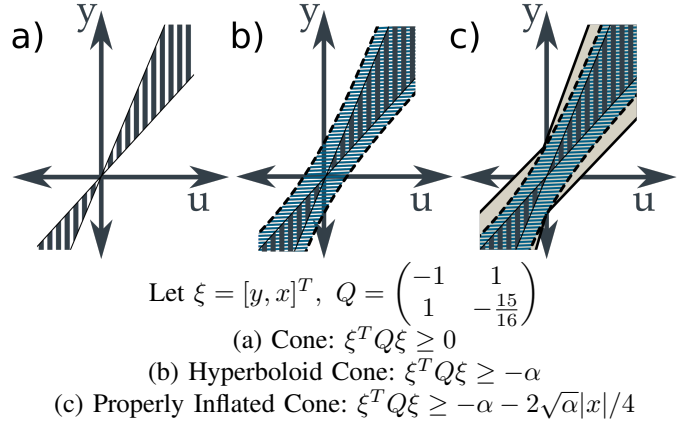


Fig. 1. Illustration of a hyperboloid cone. (a) a (degenerate quadric) cone in 2D. Note that $\xi^T A \xi \geq 0 \iff |u|/4 > |y - u|$. (b) a hyperboloid (non-degenerate quadric) cone is overlaid over the original cone. (Conic section $\xi^T A \xi = -\alpha$ is a hyperbola.) (c) a properly (Minkowski sum) inflated cone ($|u|/4 + \sqrt{\alpha} > |y - u|$) is overlaid on the previous two cones.

As shown in Fig. 1, the non-degenerate quadric can be visualized as a hyper-dimensional (multi-axially) revolved hyperbola, which asymptotically approaches the original, degenerate, quadric at large amplitudes. As noise-magnitude is reduced through averaging repeated measurements, the deviation between this approximate “hyperboloid” relaxation and a potentially more accurate Minkowski-sum style noise-relaxation becomes less significant. The QIP is not consistent in the sense that the degenerate QIP is, but the non-degenerate QIP approaches the degenerate one as noise is averaged away.⁶

VI. IDENTIFICATION OF DYNAMICAL SYSTEMS

The QIP identifies inclusions in the form $y \in \{(A + B\Delta C)x \mid \|\Delta\|_{\infty} \leq 1\}$, reminiscent of the ordinary⁷ least squares identification of the equality $y = Ax$. As in standard identification theory, there are several ways to contrive x and y to identify the system.

While a discrete time prediction error setup may be the classic example of system identification using least-squares fitting, our aim is robust control, and an inclusion-based version of a prediction error model is not convenient for this purpose. Instead, our aim is the continuous time norm bounded linear differential inclusion model which can be used in the feedback synthesis schemes of, for example, [19]:

$$\dot{\xi} = \mathbf{A}\xi + \mathbf{B}u + \mathbf{B}'p, \quad (46)$$

$$q = \mathbf{C}'\xi + \mathbf{D}'u, \quad (47)$$

$$\|p\| \leq \|q\|. \quad (48)$$

This is already in the correct form to fit an inclusion, so long as $\dot{\xi}$ and ξ are known. More realistically, the only available state measurement is an output $w = \mathbf{C}\xi$, and we must choose a state-space representation that allows us to back out the state and its derivative from these measurements. The classic choice

⁶Consistency in the presence of deterministic noise may be possible as an extension of the degenerate QIP, however.

⁷or weighted by the covariance meta-data associated with x and y

is to define ξ based on derivatives of w , according to the relative degree ([23]) of each output.

Based on the (vector) relative degree, (r_0, r_1, \dots, r_p) , of a MIMO system, there exists a representation of the system's state based on the system outputs and their derivatives:

$$\xi \triangleq (w_0, \dot{w}_0, \dots, w_0^{[r_0-1]}, \dots, w_p, \dot{w}_p, \dots, w_p^{[r_p-1]})^T. \quad (49)$$

This simply represents each output as a perfect r^{th} -order integrator (integrating an unknown quantity that must involve the inputs). This, along with the inputs is the QIP regressor x , and the QIP regressand y is the vector of maximal order output derivatives:

$$z \triangleq \begin{pmatrix} w_0^{[r_0]} & w_1^{[r_1]} & \dots & w_p^{[r_p]} \end{pmatrix}^T. \quad (50)$$

Representing z as a linear inclusion in all lower output derivatives and the inputs,

$$z \in \{(A + B\Delta C) \begin{pmatrix} \xi \\ u \end{pmatrix} \mid \|\Delta\|_\infty \leq 1\}, \quad (51)$$

we arrive at the desired norm bounded differential inclusion structure.

Defined this way, the vector equation (46) has many known rows, corresponding to the equations which define integrator behavior. The rest of the rows (and all of the C' and D' matrices) will consist of elements from the A , B , and C matrices. Borrowing [23]'s notation,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \end{pmatrix} = (\text{diag}(A_0, A_1, \dots, A_p) \quad 0) + SA, \quad (52)$$

$$\mathbf{B}' = SB, \quad (53)$$

$$\begin{pmatrix} \mathbf{C}' & \mathbf{D}' \end{pmatrix} = C, \quad (54)$$

with a row selector matrix that singles out the rows corresponding to the relative degree derivatives of the outputs,

$$S = \text{diag}(S_0, S_1, \dots, S_p), \quad (55)$$

$$S_i = (0 \quad 0 \quad \dots \quad 1)^T \in \mathbb{R}^{r_i \times 1}, \quad (56)$$

and independent integrator blocks,

$$A_i = \begin{pmatrix} 0 & 1 & & \\ & 0 & 1 & \\ & & \ddots & \ddots \\ & & & 0 \end{pmatrix} \in \mathbb{R}^{r_i \times r_i}. \quad (57)$$

Thus, so long as the outputs can be reliably differentiated, a QIP regression can be constructed to identify the elements of the system matrices directly. It is conceivable that if some pre-existing state estimator uses some more complex mechanism to convert measurements to state history, that too could be used to make the regressor. Orthonormal basis functions [22], [24] are another possible source of regressors.

Non-linear regressors are also a possibility if, as is the case in robotics, a suite of unknown and potentially varying parameters can be made to appear linearly with a matrix of well known and well-measured nonlinear coefficients. Differentiating the non-linear coefficient matrix with respect to deviation states would convert this model into the linear differential inclusion required by \mathcal{H}_∞ controller synthesis, albeit with a bounded disturbance term (from everything else).

VII. DISCUSSING THE LIMITATIONS OF QIP

Fig. 2 shows the QIP fit for a standard statistics data set [25]. Unlike Least Squares, the QIP is outlier-sensitive, as it bounds worst-case behavior. Only two data points in each plot lie on the non-degenerate quadric boundry, and these data determine the final result. Hence the caution with which we urge averaging repeatable tests to ensure that each point in the QIP has as little noise as possible.

This sensitivity also suggests the possibility of more efficient testing if QIP fitting were paired with a machine learning system to select the next test condition (perhaps estimating the parameter-realization maps of [26]), rather than collecting all data before hand.

A less obvious sensitivity is that the QIP tests the limits of the numerical interior point solvers. As demonstrated in Prop. 2, the QIP maintains a large positive semi-definite Q' matrix with more than half of its eigenvalues zero. Also, the log of determinant cost function, while technically concave, is near the edge of what interior point solvers can reliably handle—and the feature is not supported by many solver packages. This numerical complexity manifests as a surprisingly unreliable software system for a problem which is mathematically guaranteed to have a solution. It may take an effort in special purpose solver design to reach the full potential of the QIP for general purpose use.

Another numerical hurdle comes with data that contain a linear equality behavior (a subspace of y which is actually linear in x) in addition to non-linear (that is, linear inclusion) behavior. For the QIP to solve such a problem, at least one eigenvalue of X_B must go to infinity. Obviously this will cause the (floating-point) numerical solvers to fail. A proper solution might define some maximum eigenvalue for X_B , and explicitly remove the linear relationship if this constraint becomes active. Or these linear relationships might be identified and removed as a pre-processing step.

Analyzing the benefits of closed loop model identification also poses a challenge. It has been shown that performing closed loop identification can shape the identification of nominal models [27] with \mathcal{H}_∞ error bounds much the same way it shapes the amplification of system noise—Identifying a closed loop system should identify a linear fractional uncertainty for the open loop system, and potential exploitation of closed loop identification is very interesting for this reason.

A related concept, for feedback involving uncertain outputs, is the “uncertainty D-term” in the sense of $y \in \{Ax + Bp; p = \Delta q; q = Cx + Dp \mid \|\Delta\| \leq 1\}$. It remains unclear if systems with D-terms can be identified using QIP-like identification strategies. Since the QIP does not account for a D-term, an uncertain feedback arrangement of an identified system will have a different structure than would be identified by QIP, and this makes it unlikely that QIP will arrive at models which are expressed using a D-term, like passivity models.

Another open problem is handling noise in more flexible ways, allowing noise in the regressor in addition to noise in the regressand, allowing a notion of consistency despite the noise, and softening the geometric constraints to reduce the sensitivity to outliers.

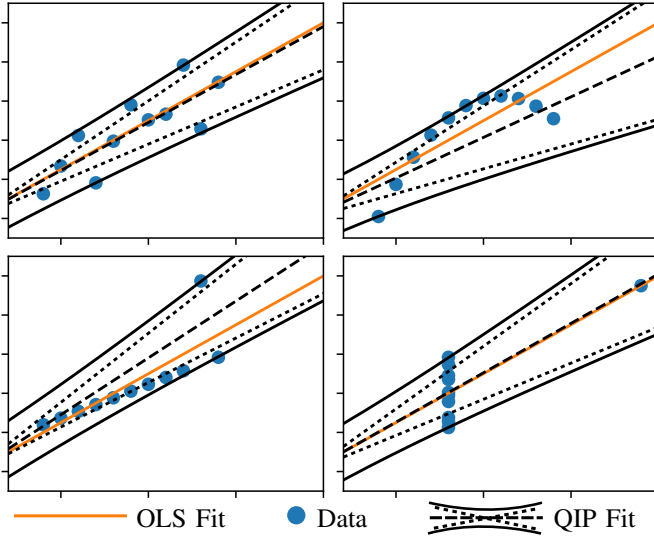


Fig. 2. Anscombe’s Quartet [25], a set of four lists of x-y points that have all the same statistics, yet have very different underlying data. While the OLS fit (and its parameter certainty) is the same for all four plots, the QIP fit (with $\alpha\Sigma_\eta = 2$) responds differently to each shape—in a way which could be said to be dominated by outliers representing the worst case behavior. The data is shifted from Anscombe’s original (-3 in y) so that the least squares fit lines intersect the origin. The QIP Fit features the nominal model, the degenerate quadric asymptote, and the (non-degenerate) quadric (which must include the data).

VIII. A FREQUENCY-DOMAIN IDENTIFICATION EXAMPLE

Suppose we have a system that follows the norm-bounded linear differential inclusion model

$$\begin{pmatrix} \dot{w}_0 \\ \ddot{w}_0 \\ \dot{w}_1 \\ \ddot{w}_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & -1 & -0.1 \end{pmatrix} \begin{pmatrix} w_0 \\ \dot{w}_0 \\ w_1 \\ \dot{w}_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ .125 & 0 \\ 0 & 0 \\ 0 & .125 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}, \quad (58)$$

$$\begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \end{pmatrix} = \begin{pmatrix} \sqrt{2}/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{2}/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} w_0 \\ \dot{w}_0 \\ w_1 \\ \dot{w}_1 \\ u_0 \\ u_1 \end{pmatrix}, \quad (59)$$

or (46) and (47) in vector notation, with

$$w = \mathbf{C}\xi, \quad \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (60)$$

$$p = \Delta q, \quad \|\Delta\| \leq 1. \quad (61)$$

Suppose that sinusoidal input, sinusoidal output tests in the style of [18] are performed using this model, producing pairs of input and output phasor vectors. Repetitions of the same input share a condition vector and can therefore be used to identify meaningful average and variance statistics for the condition group. For 20 frequencies log-spaced between 10^{-5} and 10^{-5} rad/sec, we sample 200 condition groups, and

each condition group represents averages of 100 statistically independent steady state measurements of the output phasor vector.

The distribution over which we randomly sample Δ is important for the convergence rate of the algorithm: we generate a random matrix with independent and normally distributed elements, calculate the singular value decomposition of this matrix, replace the singular values with new ones chosen from a uniform distribution between -1 and 1, and re-assemble the matrix (which now has all of its singular values less than or equal to one). If only extreme deltas with singular values equal to one were tested, then the rigidly geometric QIP would find a model close to the true model more quickly. But the worst case Δ for any input only occurs when q avoids the null space of Δ , so even this strategy would be rather inefficient. Thus our strategy of choosing random Δ has an exceedingly low chance of finding the type of outlier which reaches the outer limit of the true model.

Condition groups are heteroscedastic, and to simulate this we generate a new random 2 by 2 matrix, $\Sigma_{1/2}$, with normally distributed elements (mean 0, standard deviation .1) to represent a decomposition of the output covariance matrix $\Sigma_\eta = \Sigma_{1/2}\Sigma_{1/2}^T$ for each condition group. We assume that the noise covariance matrix does not have an imaginary component.

For each individual test then, the output phasor \hat{w} is thus

$$\hat{w} = C[j\omega I - \mathbf{A} - \mathbf{B}'\Delta\mathbf{C}']^{-1}(\mathbf{B} + \mathbf{B}'\Delta\mathbf{D}')\hat{u} + \Sigma_{1/2}\eta \quad (62)$$

with complex η having both imaginary and real parts standard normal distributed.

For each condition group we construct a state vector phasor using the relative degree vector (2,2):

$$\hat{\xi} = (\hat{w}_0 \quad j\omega\hat{w}_0 \quad \hat{w}_1 \quad j\omega\hat{w}_1)^T, \quad (63)$$

and we additionally construct the phasor vector of the relative degree terms (accelerations, in this case)

$$\hat{z} = (-\omega^2 w_0 \quad -\omega^2 w_1). \quad (64)$$

Both real and imaginary parts must independently satisfy (51), since $A+B\Delta C$ is real, so a condition group is equivalent to two QIP data points.

The estimated covariance of \hat{z} is based on the sample covariance of the condition group average \hat{w} , Σ_w : $\Sigma_z = \omega^4 \Sigma_w$. Ignoring the cross-coupling between real and imaginary noise terms, we use the real part of Σ_z for the purpose of relaxing the degenerate quadric.

After shuffling the data set, we learn the model on the first 200 points and use this model to cull inliers from the first 800 points, and continue in this fashion—re-learning, expanding the pool of data, and culling un-interesting points (which slow down the optimization). A model estimate is returned is slightly less than half a minute, representing 8,000 data points (with associated output sample covariance measurements).

The resulting model, (65) and (66), closely matches the true \mathbf{A} and \mathbf{B} matrices. The estimated \mathbf{B}' should be interpreted by its image space, since this matrix is not unique up to an orthogonal post multiplication. In this case it is clear that by

$$\begin{pmatrix} \dot{w}_0 \\ \dot{w}_1 \\ \ddot{w}_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1.000 & -0.099 & 1.000 & -0.000 \\ 0 & 0 & 0 & 1 \\ 0.999 & -0.005 & -0.996 & -0.103 \end{pmatrix} \begin{pmatrix} w_0 \\ \dot{w}_0 \\ w_1 \\ \dot{w}_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -0.003 & 0.997 \\ 0 & 0 \\ 1.003 & 0.001 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -0.016 & -0.096 \\ 0 & 0 \\ -0.105 & 0.015 \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} \quad (65)$$

$$\begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \end{pmatrix} = \begin{pmatrix} -0.000 & 0.000 & -0.000 & 0.000 & -0.000 & -0.000 \\ -0.000 & -0.000 & -0.000 & 0.000 & -0.000 & 0.000 \\ -0.006 & 0.057 & 0.111 & 0.071 & -0.053 & 0.038 \\ 0.037 & 0.109 & -0.093 & 0.142 & 0.106 & -0.002 \\ 0.317 & -0.015 & -0.194 & -0.034 & -0.211 & 0.410 \\ -0.570 & 0.030 & -0.107 & -0.045 & 0.140 & 0.459 \end{pmatrix} \begin{pmatrix} w_0 \\ \dot{w}_0 \\ w_1 \\ \dot{w}_1 \\ u_0 \\ u_1 \end{pmatrix} \quad (66)$$

TABLE I
RESULTANT MODEL FOR THE FREQUENCY DOMAIN IDENTIFICATION EXAMPLE.

flipping the order and sign of p_0 and p_1 the matrix is roughly 4/5 the true width in both dimensions—which indicates a dearth of extreme points. Finally, the matrix in (66) shows that the estimated model’s uncertainty is mostly a function of the first state and second input, as is the case in the true model. But there is also evidence that the optimization has exploited a relationship between the first and second outputs⁸ within the frequency range of 10^{-5} to 10^5 rad/sec.⁹

IX. CONCLUSION

When people use \mathcal{H}_∞ control they expect a guarantee of performance, a responsibility which \mathcal{H}_∞ control delegates to the system model-set. Due to the importance of this guarantee, practitioners will estimate uncertainty which is large enough to make the system work—sacrificing performance. It was our aim to extract the best possible performance from a system, and so we sought leaner, more aggressive model-sets.

This led us to visualize the model-set as a high dimensional degenerate quadric in the space of inputs and outputs. We introduced the QIP as a lossless convexification for the problem of fitting a minimal quadric around a list of observed data points. This new machinery appears to be somewhat more general our context of identification for robust control, since it offers a geometry-based alternative to the nearly universal least squares problem. Even within system identification, there are many approaches which use least squares and could potentially identify robust models using a QIP.

Potential theoretical impact aside, the identification procedure works towards taking the guesswork out of employing robust control. And this is particularly important in domains for which no strong intuition is to be had. Consider the problem of designing robustness into a system which is build upon an unfathomably deep tower of assumptions. Examples abound within our specialty, robotics: the design of footstep planners built on whole body robot controllers [29], or another feedback linearization scheme e.g. [30]; the tracking of unstable center of mass abstractions including the capture point [31] or divergent component of motion [32]—especially in the presence of series elasticity [33]; or even the impedance control of badly-modeled tendon driven fingers with compliant

actuators [34]. In situations like these, we can now bypass deep understanding of (or accurate guesses for) the likely uncertainty shape simply by measuring it. This could push the innovative ideas in [35], which applies robust control to flexible actuators, and [36], which applies it to whole body control of a quadruped, towards higher performance controller designs.

REFERENCES

- [1] K. Zhou, J. C. Doyle, K. Glover *et al.*, *Robust and optimal control*. Prentice hall New Jersey, 1996.
- [2] K. Zhou and J. C. Doyle, *Essentials of robust control*. Prentice hall Upper Saddle River, NJ, 1998.
- [3] G. E. Dullerud and F. Paganini, *A course in robust control theory: a convex approach*. Springer, 2013, vol. 36.
- [4] L. Ljung, Ed., *System Identification (2nd Ed.): Theory for the User*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.
- [5] Z. Zang, R. R. Bitmead, and M. Gevers, “Iterative weighted least-squares identification and weighted LQG control design,” *Automatica*, vol. 31, no. 11, pp. 1577–1594, 1995.
- [6] R. G. Hakvoort and M. J. V. den Hof, “Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors,” *IEEE Transactions on Automatic Control*, vol. 42, no. 11, pp. 1516–1528, Nov 1997.
- [7] U. Forsell and L. Ljung, “Closed-loop identification revisited,” *Automatica*, vol. 35, no. 7, pp. 1215–1241, 1999.
- [8] P. Albertos and A. Sala, *Iterative identification and control: advances in theory and applications*. Springer, 2002.
- [9] X. Bombois, M. Gevers, G. Scorletti, and B. D. Anderson, “Robustness analysis tools for an uncertainty set obtained by prediction error identification,” *Automatica*, vol. 37, no. 10, pp. 1629–1636, 2001.
- [10] S. Tøffner-Clausen, *System identification and robust control: A case study approach*. Springer, 1996.
- [11] G. C. Goodwin and M. E. Salgado, “A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models,” *International Journal of Adaptive Control and Signal Processing*, vol. 3, no. 4, pp. 333–356, 1989.
- [12] L. Ljung, G. C. Goodwin, and J. C. Agero, “Stochastic embedding revisited: A modern interpretation,” in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 3340–3345.
- [13] L. Ljung, G. C. Goodwin, J. C. Agero, and T. Chen, “Model error modeling and stochastic embedding,” *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 75 – 79, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2405896315027275>
- [14] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach*. John Wiley & Sons, 2012.
- [15] A. J. Helmicki, C. A. Jacobson, and C. N. Nett, “Identification in \mathcal{H}_∞ : a robustly convergent, nonlinear algorithm,” in *1990 American Control Conference*, May 1990, pp. 386–391.
- [16] —, “Control oriented system identification: a worst-case/deterministic approach in \mathcal{H}_∞ ,” *IEEE Transactions on Automatic Control*, vol. 36, no. 10, pp. 1163–1176, Oct 1991.
- [17] K. Poolla, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal, “A time-domain approach to model validation,” *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 951–959, May 1994.

⁸If given the choice of two signals that move together, the Frobenius-norm-bounded matrix which yields the largest magnitude q is one that uses both inputs equally.

⁹Experiments with larger bands are difficult due to as-of-yet unexplained failures of the CVXOPT, [28], interior point solver.

- [18] G. C. Thomas and L. Sentis, "MIMO identification of frequency-domain unreliability in SEAs," in *American Control Conference (ACC), 2017*. IEEE, 2017.
- [19] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994, vol. 15.
- [20] R. J. Kochenburger, "A frequency response method for analyzing and synthesizing contactor servomechanisms," *Transactions of the American Institute of Electrical Engineers*, vol. 69, no. 1, pp. 270–284, 1950.
- [21] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [22] P. S. Heuberger, P. M. van den Hof, and B. Wahlberg, *Modelling and identification with rational orthogonal basis functions*. Springer Science & Business Media, 2005.
- [23] A. Isidori, *Nonlinear control systems*. Springer Science & Business Media, 2013.
- [24] P. M. Van Den Hof, P. S. Heuberger, and J. Bokor, "System identification with generalized orthonormal basis functions," *Automatica*, vol. 31, no. 12, pp. 1821–1834, 1995.
- [25] F. J. Anscombe, "Graphs in statistical analysis," *The American Statistician*, vol. 27, no. 1, pp. 17–21, 1973.
- [26] R. Tóth, *Modeling and identification of linear parameter-varying systems*. Springer, 2010.
- [27] T. Oomen and O. Bosgra, "System identification for achieving robust performance," *Automatica*, vol. 48, no. 9, pp. 1975–1987, 2012.
- [28] D. Andersen, J. Dahl, and L. Vandenberghe, "Cvxopt: Python software for convex optimization," 2013.
- [29] D. Kim, Y. Zhao, G. Thomas, B. R. Fernandez, and L. Sentis, "Stabilizing series-elastic point-foot bipeds using whole-body operational space control," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1362–1379, 2016.
- [30] A. D. Ames, P. Tabuada, A. Jones, W.-L. Ma, M. Rungger, B. Schrmann, S. Kolathaya, and J. W. Grizzle, "First steps toward formal controller synthesis for bipedal robots with experimental implementation," *Nonlinear Analysis: Hybrid Systems*, pp. –, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1751570X1730002X>
- [31] J. Pratt, T. Koolen, T. De Boer, J. Rebula, S. Cotton, J. Carff, M. Johnson, and P. Neuhaus, "Capturability-based analysis and control of legged locomotion, part 2: Application to m2v2, a lower body humanoid," *The International Journal of Robotics Research*, p. 0278364912452762, 2012.
- [32] J. Engelsberger, C. Ott, and A. Albu-Schaffer, "Three-dimensional bipedal walking control based on divergent component of motion," *Robotics, IEEE Transactions on*, vol. 31, no. 2, pp. 355–368, 2015.
- [33] M. A. Hopkins, R. J. Griffin, A. Leonessa, B. Y. Lattimer, and T. Furukawa, "Design of a compliant bipedal walking controller for the darpa robotics challenge," in *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*. IEEE, 2015, pp. 831–837.
- [34] P. Rao, G. C. Thomas, L. Sentis, and A. D. Deshpande, "Analyzing achievable stiffness control bounds of robotic hands with compliantly coupled finger joints," in *IEEE International Conference on Robotics and Automation (ICRA), 2017*.
- [35] K. Haninger, J. Lu, and M. Tomizuka, "Robust impedance control with applications to a series-elastic actuated system," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 5367–5372.
- [36] F. Farshidian, E. Jelavić, A. W. Winkler, and J. Buchli, "Robust whole-body motion control of legged robots," *arXiv preprint arXiv:1703.02326*, 2017.



Luis Sentis (S04—M07) received the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, where he developed leading work in theoretical and computational methods for the compliant control of humanoid robots. He is currently an Associate Professor in Aerospace Engineering at The University of Texas at Austin (UT Austin), Austin, TX, USA, where he directs the Human Centered Robotics Laboratory. He was UT Austins Lead for DARPA's Robotics Challenge entry with the NASA Johnson Space Center in 2013. His research focuses on foundations for the compliant control of humanoid robots, algorithms to generate extreme dynamic locomotion, and building robots for educating students in mechatronics.



Gray Cortright Thomas (S'12,15,17) was born in The United States of America in 1989. He received his B.S.E. degree in 2012 from Olin College of Engineering in Needham Massachusetts—a self-designed major in robotics engineering. He joined the Florida Institute for Human and Machine Cognition, where he eventually participated in the DARPA virtual robotics challenge, helping IHMC win. Since the fall of 2013, he has been with the Human Centered Robotics Lab at the University of Texas at Austin, pursuing a Ph.D. in Mechanical Engineering. Since

2015 he has been a NASA Space Technology Research Fellow, collaborating with Johnson Space Center's Robotics Team. He studies identification and control for compliant bipedal robots, with a decidedly mathematical bent.