Uncertainty–Shape Identification in the Frequency Domain

Gray C. Thomas and Luis Sentis

***Abstract*—Practical application of robust control relies on validated robust model-sets. Additive uncertainty model-sets are defined by a nominal linear model and an unstructured uncertainty which is shaped by two weighting transfer matrices. Today's identification tools assume they know this shape beforehand. But our new tool can identify all three matrices simultaneously at a single frequency. We do this with a novel, lossless transformation between the robust model-set and a quadratic-form inequality—a form amenable to convex optimization. The identification has a geometric interpretation—fitting a minimum "cone width" quadric "cone" around complex data points in high dimensional space. We prove that the identification consistently returns the true model-set under certain conditions. Simulations demonstrate the algorithm and demonstrate implementation tricks. In the presence of noise we explore a modification based on statistical survival functions which has geometric interpretation as a non-degenerate quadric or "hyperboloid cone". We also suggest a statistical test to differentiate between data-sets bearing uncertainty information and uninformative ones.**

I. The Identification of Robust Model-Sets

IT is widely acknowledged that robust $\mathcal{H}_\infty$ control theory, widely available in e.g. [1], [2], [3], relies on its assumption of an $\mathcal{H}_\infty$ plant model. If this model is not believable then $\mathcal{H}_\infty$ 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 phase lag beyond the known model order, nor can it represent 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 doesn't 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 (specifically, the condition groups from Sec. II), 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. $\mathcal{H}_\infty$–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

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. Both authors are with the Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712, USA. Send correspondence to gray.c.thomas@utexas.edu

would make the problem non-convex.

In an effort to capture a physical component to the $\mathcal{H}_\infty$ robust model, we assume the system is essentially repeatable (though potentially corrupted by noise) if a vector of experimental conditions is held constant. By repeating stepped sine tests under many conditions 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 $\mathcal{H}_\infty$ model-sets to include all observed variation—using convex programs inspired by the minimal bounding ellipsoid problem [18]. 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 [19] for some range of input magnitudes.

In this paper Sections II and VII cover material we originally presented at ACC [20] but in archival detail and without restricting attention to series elastic actuators. Our new "cone parameterization" for $\mathcal{H}_\infty$ model-sets (at a single frequency) in III hardly resembles our previous attempt, and it forms the core theoretical contribution of the paper. The convex optimization results in IV, the simulation results in V, and the noise modification in VI are all novel outgrowths of this core idea.

II. Basic Models and Notation

We denote the commonly used linear frequency domain model as,

$$y(j\omega) = P(j\omega)u(j\omega), \tag{1}$$

with $P(j\omega)$ the nominal "plant" transfer matrix, $y(j\omega)$ the vector of output signals, and $u(j\omega)$ the vector of input signals. To denote the additive uncertainty model-set from $\mathcal{H}_\infty$ control theory we write

$$y(j\omega) = P(j\omega)u(j\omega) + E(j\omega)\Delta(j\omega)J(j\omega)u(j\omega), \tag{2}$$

which adds a bounded model uncertainty (or variability) term $E(j\omega)\Delta(j\omega)J(j\omega)$ to the nominal plant $P(j\omega)$. $E(j\omega)$ and $J(j\omega)$ are the "weighting" transfer matrices which represent the "shape" of the uncertainty as well as its scale. Both are square. The unstructured uncertainty object $\Delta(j\omega)$ is a construct from $\mathcal{H}_\infty$ control theory—traditionally defined by the constraint $\|\Delta(j\omega)\|_\infty \leq 1$.

This is equivalent to saying that the matrix $\Delta(j\omega)$ has singular values less than or equal to one—and so we can equivalently write $\Delta^*\Delta \preceq I$, where $\cdot^*$ is the complex conjugate transpose operator and $\preceq$ indicates the cone inequality for the positive semi-definite cone: $A \succeq B \iff A - B$ is p.s.d. or positive semi-definite. We often let the dimension of matrices (particularly the identity) be inferred from context.

The problem considered in this paper is optimal learning of the $\mathcal{H}_\infty$ additive uncertainty model-set matrices **evaluated at a single frequency** from phasor vector input and output data pairs. This restricted scope (learning only a single frequency-slice of the model-set) allows us to focus the paper on the geometry of the identification problem, temporarily sidestepping the thorny issue of guaranteeing realizability and lack of pole cancellation in the estimates. In this simpler case, the model-set and data are complex matrices and vectors, and we write them without the functional argument $(j\omega)$ to indicate this.

We consider a system which is repeatable, but depends on a **condition vector** c such that

$$y = F(c, u) \triangleq Pu + E\Delta(c)Ju, \tag{3}$$

and $\Delta(c)^*\Delta(c) \preceq I \;\forall\; c \in \mathcal{C}$ (where $\mathcal{C}$ represents e.g. realistic conditions, or conditions of interest). Note that this model variability concept is related to the Linear Parameter Varying (LPV) Model [21], [22] used to explain gain scheduling controllers, but differs in that it guarantees a bounded deviation and does not require a known (or even worth learning) mapping $c \mapsto \Delta(c)$. When identifying the model-set we write Δ without argument, treating it simply as an indeterminate complex matrix satisfying the singular value bound.

We also consider the case where experiments provide imperfect measurements of this ideal system, corrupted by Gaussian noise:

$$y = F(c, u) + \eta, \tag{4}$$

for a multivariate standard complex normal noise vector η of covariance Σ. Consider the prototypical experiment in which the system is excited to steady state oscillation by a mono-frequency input (represented by phasor vector u), and a single period of the output is cast to the closest sinusoid, to fit a phasor vector y. In experiments like this, the key to overcoming measurement noise is to take many independent measurements and average them—but only for experiments which represent the same condition vector. We introduce the indexing schemes $g \in 1, \ldots, N_G$ to represent the index of the **condition group**—the set of experiments having the same condition vector c_g; and $i \in 1, \ldots, N_g$ to index the independent measurements within the gth condition group.

Our favored experimentation strategy falls under the broad category of stepped-sine, or "swept sine" in industry [23]. Under this category of approaches the excitation signals carry only a single sinusoidal excitation at a time, and this sinusoid's frequency is then swept through a range in discrete increments. This measures the steady state sinusoid response of the system directly, and avoids the frequency corruption issues of broad spectrum input signals [14], especially important for the more non-linear systems. For a more practice-oriented description of condition group data collection see [20].

III. A Novel Parameterization of the $\mathcal{H}_\infty$ Model-Set at ω

The following alternative expression for (3) is a lossless convexification of the space of (single frequency) model-sets which can explain all previously observed data points.

Theorem 1 (Cone Parameterization)**.** *A pair of input and output phasor vectors (u, y) satisfies the single frequency additive uncertainty model-set*

$$\begin{cases} y = Pu + E\Delta Ju \\ \Delta^*\Delta \preceq I \end{cases} \tag{5}$$

(with full rank uncertainty shape matrices E and J) if and only if it satisfies the following quadratic form inequality

$$\begin{pmatrix} y \\ u \end{pmatrix}^* \underbrace{\begin{pmatrix} -E^{-*}E^{-1} & E^{-*}E^{-1}P \\ P^*E^{-*}E^{-1} & J^*J - P^*E^{-*}E^{-1}P \end{pmatrix}}_{\triangleq M} \begin{pmatrix} y \\ u \end{pmatrix} \geq 0 \quad (6)$$

Here we name the matrix M the **Cone Parameterization** of the model-set—a nonlinear function of the model-set matrices: $M = \mathrm{Cone}(P, E, J)$. Which can be inverted up to post-multiplication of E by a unitary matrix, and pre-multiplication of J by a unitary matrix. We additionally name lists of the inequality (6) for some list of (y,u) points **inclusion inequalities**, since they represent these points being included in the geometric shape represented by the model-set—these inclusion inequalities form the basis of the convex optimization approach to learning model-sets in Sec. IV.

Proof. First, consider u and y which satisfy (5).

$$\Delta J u = E^{-1}(y - Pu) \quad (7)$$
$$\|Ju\|^2 \geq \|\Delta J u\|^2 = \|E^{-1}(y - Pu)\|^2 \quad (8)$$
$$0 \leq u^*J^*Ju - (y - Pu)^*E^{-*}E^{-1}(y - Pu) \quad (9)$$

Which is equivalent to (6).

For u and y which satisfy (6),

$$u^*J^*Ju \leq (y - Pu)^*E^{-*}E^{-1}(y - Pu) \quad (10)$$

with both sides positive or zero by construction. Let

$$\gamma = \frac{u^*J^*Ju}{(y - Pu)^*E^{-*}E^{-1}(y - Pu)}, \quad (11)$$

so γ is in the inclusive range from zero to one. Consider the matrix

$$\Delta = \frac{E^{-1}(y - Pu)u^*J^*}{\|Ju\|^2}, \quad (12)$$

which is designed to satisfy

$$\Delta J u = E^{-1}(y - Pu). \quad (13)$$

The singular values of Δ are the eigenvalues of

$$\Delta^*\Delta = \frac{Ju(y - Pu)E^{-*}E^{-1}(y - Pu)u^*J^*}{(u^*J^*Ju)^2} \quad (14)$$
$$= \gamma \frac{Juu^*J^*}{u^*J^*Ju} \quad (15)$$

which is rank 1, positive semi-deifinite, with $\gamma \leq 1$ as the only non-zero eigenvalue. This ensures that

$$\Delta^*\Delta \preceq I, \quad (16)$$

completing (5).

□

Definition 1 (SS-DD)**.** The cone parameterization can be written as the subtraction of a positive definite matrix of rank n_y from a positive definite matrix of rank n_u. We call this the **Split Semi-Definite Decomposition (SS-DD)** of the cone parameterization, and use the following three equations to define the variables X_E, X_P, X_{PP} and X_J:

$$M = \begin{pmatrix} -X_E & X_P \\ X_P^* & X_J - X_{PP} \end{pmatrix}, \quad (17)$$
$$0 \preceq \begin{pmatrix} X_E & -X_P \\ -X_P^* & X_{PP} \end{pmatrix}, \quad (18)$$
$$0 \preceq X_J. \quad (19)$$

In general this is not a unique decomposition. However if in addition

$$X_{PP} = X_P^* X_E^{-1} X_P, \quad (20)$$

that is, if we can write the RHS of (18) as

$$\begin{pmatrix} X_E \\ -X_P^* \end{pmatrix} \left(X_E^{-1}\right) \begin{pmatrix} X_E & -X_P \end{pmatrix}, \quad (21)$$

then the rank of the RHS of (18) is the rank of X_E and

$$\begin{cases} X_E &= E^{-*}E^{-1} \\ X_P &= E^{-*}E^{-1}P \\ X_{PP} &= P^*E^{-*}E^{-1}P \\ X_J &= J^*J \end{cases} \quad (22)$$

This ultimately allows separate costs for the positive and negative component of the cone decomposition M in convex optimization problems.

The set of all points satisfying the quadratic inequality (6) has a geometric interpretation—a shape which is technically a degenerate quadric. This is a very general class of hyper-dimensional shapes, and within this class we would like to point out the similarity of our degenerate quadrics (with their full rank, roughly half negative definite and half positive definite) to simple 2D cones—hence the name. While equations (6) and (5) both represent the same shape, the former shares the quadratic form structure which is typically used to describe quadrics in 2D space—better known as ellipses, hyperbola, circles (non-degenerate), and cones (degenerate).

When 2D cones are described by matrices—e.g. $(x, y)A(x, y)^T > 0$ in x, y Cartesian space—they are symmetric, with one positive and one negative eigenvalue. The positive eigenvalue works to inflate the cone, and its eigenvector points along the cone axis. The negative eigenvalue works to reduce deviation from the cone axis—if the cone axis was a linear model, the negative eigenvalue would penalize model error. Its eigenvector points perpendicular to the cone axis. We'll come back to this interpretation in Fig. 3 when we introduce the hyperboloid cones (non-degenerate quadrics) we use to handle noisy measurements.

The cone parameterization, and the inequality constraint on cone inclusion, is linear in the elements of the matrix. It trivializes the proofs for several useful properties.

Proposition 1 (Addition preserves cone-inclusion)**.** *If the two cone parameterization matrices for two different additive uncertainty model-sets, M_1 and M_2, both satisfy a list of quadratic form inequalities*

$$\xi_i^* M_1 \xi_i \geq 0 \quad \textit{and} \quad \xi_i^* M_2 \xi_i \geq 0 \qquad \forall\ i \in 1, \ldots, N \quad (23)$$
$$\textit{then} \quad \xi_i^* (M_1 + M_2) \xi_i \geq 0 \qquad \forall\ i \in 1, \ldots, N. \quad (24)$$

Proof. Trivial by linearity. □

Proposition 2 (Scaling is irrelevant)**.** *If M is a cone parameterization matrix for an additive uncertainty model-set, and λ is a positive scalar, then any point $\xi = \begin{pmatrix} y \\ u \end{pmatrix}$ satisfies $\xi^* M \xi \geq 0$ if and only if the inequality holds for the scaled matrix λM, that is, if $\xi^*(\lambda M)\xi \geq 0$.*

Proof. Again, trivial by linearity. □

Proposition 3 (Convex Combination)**.** *If M_1 and M_2 satisfy a list of quadratic form inequalities and some linear constraints, then for any positive $0 \leq \alpha \leq 1$, the convex combination $M_\alpha = \alpha M_1 + (1-\alpha)M_2$ also satisfies those inequalities and constraints.*

Proof. Yet again, trivial by linearity. □

However, Prop. 3 would not have been straightforward in the original parameterization. Consider the complexity of the equations below, which describe (constraints on the three matrix expression for) the model-set M_α (as defined in Prop. 3) using the original parameterization (5) to describe all three model-sets:

$$M_z = \mathrm{Cone}(P_z, E_z, J_z), \quad \forall\ z \in \{1, 2, \alpha\} \implies \tag{25}$$

$$E_\alpha E_\alpha^* = \left[\alpha E_1^{-*} E_1^{-1} + (1-\alpha) E_2^{-*} E_2^{-1}\right]^{-1}, \tag{26}$$

$$P_\alpha = E_\alpha E_\alpha^* \left[\alpha E_1^{-*} E_1^{-1} P_1 + (1-\alpha) E_2^{-*} E_2^{-1} P_2\right] \tag{27}$$

$$\begin{aligned} J_a^* J_a = {} & \alpha J_1^* J_1 + (1-\alpha) J_2^* J_2 \\ & - \alpha P_1^* E_1^{-*} E_1^{-1} (P_1 - P_\alpha), \\ & - (1-\alpha) P_2^* E_2^{-*} E_2^{-1} (P_2 - P_\alpha). \end{aligned} \tag{28}$$

IV. Convex Identification with the Cone Parameterization

With this lossless convexification of the search space, convex programming is a viable means of identifying the uncertainty shape to fit the data. Before getting to the optimization process however, we need some geometric insight into the nature of cones and the measurement of their "width".

Definition 2 (Cone width for a toy cone)**.** Consider the following 2D cone in real scalar u, y space: $|y - pu| \leq r(u) = w|u|$. We call w the "width" of this cone, and it measures the slope of $r(u)$ as a function of u for positive u. However it can also be indirectly measured by $\sqrt{\mathrm{E}(r(u)^2)}$ if $u \sim \mathcal{N}(0,1)$ (E is expected value, and $\mathcal{N}(0,1)$ is the normal distribution with mean 0 and variance 1).

Proof. $\sqrt{\mathrm{E}\,(r(u)^2)} = \sqrt{w^2 \mathrm{E}(|u|^2)} = w.$ □

Definition 3 (Characteristic radius of a cross section)**.** Consider a model-set in SS-DD form satisfying (17)–(22). If we specify a particular input u, the space of possible y that will be accepted by the model-set can be interpreted as a geometric shape: a hyper-ellipsoid. The **characteristic radius** $R(u)$ 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 (in $\mathbb{C}^{n_y}$ where $y \in \mathbb{C}^{n_y}$) can be described as

$$(y - Pu)^* X_E (y - Pu) \leq G^2(u), \quad G(u) \triangleq \sqrt{u^* X_J u}. \tag{29}$$

And the characteristic radius is easily calculated:

$$R(u) = G(u)\ (\det(X_E^{-1}))^{\frac{1}{2n_y}}. \tag{30}$$

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

$$W \triangleq \sqrt{\mathrm{E}\,(R^2(u))} \quad | \quad \mathrm{E}\,(uu^*) = I \tag{31}$$

Theorem 2 (Generalized Cone Width)**.** *The generalized cone width of a model-set in SS-DD form satisfying* (17)–(22)

$$W = (\det\left(X_E^{-1}\right))^{\frac{1}{2n_y}} \sqrt{\mathrm{tr}\,(X_J)} \tag{32}$$

Proof. The width

$$W = \sqrt{\mathrm{E}\,(R^2(u))} = (\det(X_E^{-1}))^{\frac{1}{2n_y}} \sqrt{\mathrm{E}(G^2(u))} \tag{33}$$

$$\mathrm{E}\left(G^2(u)\right) = \mathrm{E}\,(u^* X_J u) \tag{34}$$

$$= \mathrm{tr}[X_J \mathrm{E}\,(uu^*)] = \mathrm{tr}[X_J]. \tag{35}$$

Substitution yields (32). □

Corollary 2.1. *if* $\mathrm{tr}(X_J) = 1$,

$$\log(W) = -\frac{1}{2n_y} \log(\det(X_E)) \tag{36}$$

Proof. $\log(W) = \log\left((\det\left(X_E^{-1}\right))^{\frac{1}{2n_y}} \sqrt{\mathrm{tr}\,(X_J)}\right) = \frac{1}{2n_y} \log\left(\det\left(X_E^{-1}\right)\right) + \frac{1}{2}\log\left(\mathrm{tr}\,(X_J)\right) = \frac{-1}{2n_y}\log\left(\det\left(X_E\right)\right) + \frac{1}{2}\log(1)$ and $\log(1) = 0$. □

Corollary 2.2. *The generalized cone width is a function of the eigenvalues of J and E.*

Proof. Convert the model-set from standard form (5) to SS-DD cone form (17)–(20) and apply Thm. 2

$$W = (\det EE^*)^{\frac{1}{2n_y}} \sqrt{\mathrm{tr}[J^* J]} \tag{37}$$

$$= (\det EE^*)^{\frac{1}{2n_y}}\ \|J\|_{\text{Frobenius}} \tag{38}$$

$$= \prod_{\lambda \in \sigma(E)} \left(|\lambda|^{\frac{1}{n_y}}\right) \sqrt{\sum_{\gamma \in \sigma(J)} |\gamma|^2}. \tag{39}$$

Where the spectrum of a matrix $\sigma(X)$ is the set of the eigenvalues of that matrix (with repetition). □

Singular values are just the absolute values of eigenvalues, so we also can re-state this: *Generalized cone width is the product of the geometric mean of the singular values of E, and the 2-norm of the singular values of J.*

Corollary 2.3. *The generalized cone width is invariant to scaling the cone parameterization, and invariant to scaling J and E by reciprocal values.*

Proof. Consider $J' \triangleq \alpha J$ and $E' \triangleq \alpha^{-1} E$ for any $\alpha \neq 0$. The generalized cone width for this new model-set

$$W' = \prod_{\lambda \in \sigma(E)} \left(|\alpha^{-1}\lambda|^{\frac{1}{n_y}}\right) \sqrt{\sum_{\gamma \in \sigma(J)} |\alpha\gamma|^2} \tag{40}$$

$$= |\alpha|^{\frac{-n_y}{n_y}} \prod_{\lambda \in \sigma(E)} \left(|\lambda|^{\frac{1}{n_y}}\right) |\alpha| \sqrt{\sum_{\gamma \in \sigma(J)} |\gamma|^2} = W \tag{41}$$

□

Definition 5 (model-set inclusion)**.** A model-set (P_o, E_o, J_o) (subscripts o for outer) includes another model-set (P_i, E_i, J_i) (subscripts i for inner) if

$$\xi^* \mathrm{Cone}(P_o, E_o, J_o)\xi \geq 0 \iff \xi^* \mathrm{Cone}(P_i, E_i, J_i)\xi \geq 0 \tag{42}$$

Proposition 4. *A model-set (P_o, E_o, J_o) includes another model-set (P_i, E_i, J_i) if and only if*

$$\|\tilde{P} + \tilde{E}\Delta\tilde{J}\| \leq 1 \quad \forall\ \Delta \mid \|\Delta\| \leq 1 \tag{43}$$

with $\tilde{E} \triangleq E_o^{-1}E_i$, $\tilde{P} \triangleq E_o^{-1}(P_i - P_o)J_o^{-1}$, and $\tilde{J} \triangleq J_i J_o^{-1}$.

Proof. Starting first with (42), we go backwards through Thm. 1 to return to an equivalent statement using the standard model-set

$$\begin{cases} y = P_o u + E_o \Delta_o J_o u \\ \Delta_o^* \Delta_o \preceq I \end{cases} \iff \begin{cases} y = P_i u + E_i \Delta J_i u \\ \Delta^* \Delta \preceq I \end{cases}. \tag{44}$$

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

$$\begin{cases} \Delta_o^* \Delta_o \preceq I \iff \Delta^* \Delta \preceq I \\ P_o + E_o \Delta_o J_o = P_i + E_i \Delta J_i \end{cases}. \tag{45}$$

This is because (44) holds for all u ($Au = Bu\ \forall\ u \iff A = B$—consider u selected from columns of the identity matrix.) By algebra, $\Delta_o = \tilde{P} + \tilde{E}\Delta\tilde{J}$. Re-stating the implication in (45) as $\|\Delta_o\| \leq 1\ \forall\ \Delta \mid \|\Delta\| \leq 1$ we get (43). □

Lemma 1. *If (P_o, E_o, J_o) includes another model-set (P_i, E_i, J_i) then $\sigma_{\max}(\tilde{E})\sigma_{\max}(\tilde{J}) \leq 1$.*

Proof. Assume the contrary ($\sigma_{\max}(\tilde{E})\sigma_{\max}(\tilde{J}) > 1$) and construct the following

$$\Delta = \tilde{E}^* \zeta_E \operatorname{sign}(\zeta_E^* \tilde{P} \zeta_J) \zeta_J^* \tilde{J}^* \tag{46}$$

where ζ_E and ζ_J are unit eigenvectors for $\tilde{E}\tilde{E}^*$ and $\tilde{J}^*\tilde{J}$ corresponding to their respective maximum eigenvalues: $\lambda_{\max}(\tilde{E}\tilde{E}^*)\zeta_E = \tilde{E}\tilde{E}^*\zeta_E$, $\lambda_{\max}(\tilde{J}^*\tilde{J})\zeta_E = \tilde{J}^*\tilde{J}\zeta_E$, $\zeta_E^*\zeta_E = \zeta_J^*\zeta_J = 1$. This choice of Δ leads to:

$$\begin{aligned} 1 \geq \|\tilde{P} + \tilde{E}\Delta\tilde{J}\| &\geq |\zeta_E^*(\tilde{P} + \tilde{E}\Delta\tilde{J})\zeta_J| \\ &= |\zeta_E^*(\tilde{P} + \tilde{E}\tilde{E}^*\zeta_E \operatorname{sign}(\zeta_E^*\tilde{P}\zeta_J)\zeta_J^*\tilde{J}^*\tilde{J})\zeta_J| \\ &= |\zeta_E^*\tilde{P}\zeta_J| + \lambda_{\max}(\tilde{E}\tilde{E}^*)\lambda_{\max}(\tilde{J}^*\tilde{J}) \qquad (47) \\ &\geq (\sigma_{\max}(\tilde{E})\sigma_{\max}(\tilde{J}))^2 > 1, \quad \lightning \qquad (48) \end{aligned}$$

A contradiction ($\lightning$) as desired. □

Proposition 5. *If a model-set (P_o, E_o, J_o), width W_o, includes another model-set (P_i, E_i, J_i), width W_i, then $W_o \geq W_i$.*

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

$$\sqrt[2n_y]{\det E_i E_i^*}\|J_i\|_{\mathrm{F}} > \sqrt[2n_y]{\det E_o E_o^*}\|J_o\|_{\mathrm{F}} \tag{49}$$

$$\frac{\sqrt[n_y]{\det E_i E_i^*}}{\sqrt[n_y]{\det E_o E_o^*}}\|J_i\|_{\mathrm{F}}^2 > \|J_o\|_{\mathrm{F}}^2 \tag{50}$$

$$\mathrm{tr}\left(\sqrt[n_y]{\det(E_o^{-1}E_i E_i^* E_o^{-*})}J_i^* J_i - J_o^* J_o\right) > 0 \tag{51}$$

$$\mathrm{tr}\left(J_o^*\left(\lambda_{\max}(\tilde{E}\tilde{E}^*)\tilde{J}^*\tilde{J} - I\right)J_o\right) > 0 \tag{52}$$

However by Lemma 1, the argument of trace in the above inequality is negative semi definite, so it cannot have a non-negative trace. $\lightning$ □

Proposition 6. *If a model-set (P_o, E_o, J_o), width W_o includes (P_i, E_i, J_i), width W_i, and $W_o = W_i$ then the model-sets are equivalent in the sense that $P_o = P_i$, and $\exists\ \lambda > 0\ : \lambda E_o E_o^* = E_i E_i^*,\ J_o^* J_o = \lambda J_i^* J_i$ (Recall Prop. 2 on scale invariance).*

Proof. When the two widths are equal, the derivation which produced (51) yields:

$$\mathrm{tr}\left(J_o^*\left(\sqrt[n_y]{\det(\tilde{E}\tilde{E}^*)}\tilde{J}^*\tilde{J} - I\right)J_o\right) = 0, \tag{53}$$

Yet as before, singular values are limited by the inclusion constraint, and this guarantees (47) and a long series of matrix inequalities

$$\begin{aligned} I &\succeq |\zeta_E^*\tilde{P}\zeta_J|I + \lambda_{\max}(\tilde{E}\tilde{E}^*)\lambda_{\max}(\tilde{J}^*\tilde{J})I \\ &\succeq \lambda_{\max}(\tilde{E}\tilde{E}^*)\tilde{J}^*\tilde{J} \succeq \sqrt[n_y]{\det(\tilde{E}\tilde{E}^*)}\tilde{J}^*\tilde{J} = I, \qquad (54) \end{aligned}$$

with this last equality due to the combination of (53) and the last inequality above (which can be extended to

$$J_o^*\left(\sqrt[n_y]{\det(\tilde{E}\tilde{E}^*)}\tilde{J}^*\tilde{J} - I\right)J_o \preceq 0, \tag{55}$$

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).

Note that with both the first and last element identity, (54) is actually a long chain of equalities. This gives

$$\lambda \triangleq \lambda_{max}(\tilde{E}\tilde{E}^*) = \sqrt[n_y]{\det(\tilde{E}\tilde{E}^*)}, \tag{56}$$

$$|\zeta_E^*\tilde{P}\zeta_J| = 0, \quad \lambda\tilde{J}^*\tilde{J} = I, \tag{57}$$

from which it follows that $J_o^* J_o = \lambda J_i^* J_i$. When the geometric mean of the eigenvalues is equal to the largest eigenvalues all the eigenvalues must be equal, with $\tilde{E}\tilde{E}^* = \lambda I$ (or equivalently $\lambda E_o E_o^* = E_i E_i^*$). Since both $\tilde{E}\tilde{E}^*$ and $\tilde{J}^*\tilde{J}$ have only one eigenvalue with high multiplicity, the eigenvectors ζ_E and ζ_J can be any unit vectors. This in turn guarantees $\tilde{P} = 0$, that is, $P_o = P_i$, completing the conditions necessary for the two model-sets to be equivalent. □

With these properties of our cost function established, we introduce a convex optimization problem which is formatted to follow the rules of disciplined convex programming [24].

Convex Problem 1 (Cone Identification)**.** Given data in the form of N_G noiseless measurements (representing distinct condition groups) ξ_g, $g \in 1 \ldots N_G$ of the system, each with a different condition vector c_g

$$\begin{aligned} \text{minimize} \quad & -\frac{1}{2n_y}\log(\det(X_E)) \\ \text{over} \quad & M,\ X_E,\ X_P,\ X_{PP},\ X_J \\ \text{subject to} \quad & \text{the SS-DD equations: (17)–(19)} \\ & 1 = \mathrm{tr}[X_J] \\ & 0 \leq \xi_g^* M \xi_g \quad \forall \quad g \in 1 \ldots N_G \end{aligned}$$

(This last set of inequalities are the *inclusion inequalities*.)

Lemma 2. *The uniqueness of SS-DD (equation* (20)*) is guaranteed for solutions to Prob. 1 with finite cost.*

Proof. In the general case where (20) does not hold, we can define an equation error matrix

$$\tilde{X}_{PP} \triangleq X_{PP} - X_P^* X_E^{-1} X_P \succeq 0, \tag{58}$$

which is p.s.d. since it is the Schur complement of the matrix in (18). Suppose $\tilde{X}_{PP}$ 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: $X'_{PP} = X_{PP} - v\lambda v^*$. Since the other matrices are the same, the new solution trivially satisfies all constraints which do not involve X_{PP}, leaving (18)—satisfied by Schur complement since $\tilde{X}_{PP} - v\lambda v^* \succeq 0$, and the inclusion inequalities, which the new solution relaxes to (recall that u_g is the lower part of ξ_g):

$$0 \leq \xi_g^* M \xi_g + u_g^* v \lambda v^* u_g \tag{59}$$

Relaxing constraints reduces cost. Assuming that the cost was previously bounded by the inclusion inequalities in all the various ways it could increase then at least one of them must have been relaxed by the change (albeit potentially through re-arrangement of the X_J matrix), demonstrating that a lower cost solution must exist. ↯ □

Proposition 7. *The model-set extracted from Prob. 1's cone parameterization M (via Thm. 1) is the minimum cone width model-set which includes every data point. (Assuming the data points justify a non-singular E model-set)*

Proof. Applying Lemma 2, (17)–(22) all hold, so we can apply Corr. 2.1, using the constraint that $\mathrm{tr}(X_J) = 1$, and establish that the minimized cost function is the log of the cone width. Log is monotonic, so the generalized cone width is also minimized. The inclusion inequalities guarantee the model-set includes every measured point. (Cases where Lemma 2 does not hold due to non-finite cost are degenerate cases with insufficient data points to prevent a singular E matrix.) □

Theorem 3 (Estimation Consistency)**.** *Consider an infinite list of input output data $\xi_i = [y_i^T, u_i^T]^T \ \forall \ i \in \mathbb{N}$ points generated from a noiseless true model-set*[1] *(P_T, J_T, E_T) in the sense that any possible output will eventually be produced—that is, a model-set is only the true model-set if no sub-model-set could have generated the data. Suppose that a sequence of model-set estimates (P_n, J_n, E_n) are calculated via Prob. 1 using the subset of data indexes $i = 1, \ldots, n$.*

Then $W(P_{n+1}, J_{n+1}, E_{n+1}) \geq W(P_n, J_n, E_n) \ \forall \ n \in \mathbb{N}$, and $W(P_n, J_n, E_n) \leq W_T \triangleq W(P_T, J_T, E_T)$.

Furthermore if $\exists \ n'$ such that $(P_n, J_n, E_n) \cong (P_{n'}, J_{n'}, E_{n'}) \ \forall \ n \geq n'$ (where $\cong$ means equivalent in the sense of Prop. 6), then $(P_{n'}, J_{n'}, E_{n'}) \cong (P_T, J_T, E_T)$

Proof. The first claim follows from the nature of optimization: more constraints can only increase the cost, cost is the log of the width, and log is monotonic. The second is a consequence of the true model-set being a feasible solution to the optimization problem: the optimal solution has the least cost over all feasible solutions. As for the third, suppose the contrary: $(P_{n'}, J_{n'}, E_{n'}) \ncong (P_T, J_T, E_T)$. The n' model-set cannot contain (P_T, J_T, E_T) because it has lesser cone width and is not the same model-set (by supposition), there must be points within the true model-set and outside the n' model-set. And these points, which will eventually occur for some $n > n'$, will not satisfy the inclusion inequalities with the n' model-set. ↯ □

[1]Note that this is a restrictive structural assumption on the true model set.

V. Simulation-Based Example

In this section we analyze the performance of the estimator on data generated by a true model-set, while, of course, acknowledging that there are no true model-sets in the real world (and certainly none exactly satisfying our particular quadric model-set structure). The implementation we use is available free of charge (Apache License) from the first author's GitHub page [25].

To generate data, we begin with a model-set of the form (3). This model-set is then sampled by randomly generating a valid Δ_g and u_g and combining $(P + E\Delta_g J)u_g$ into F_g, which represents $F(c_g, u_g)$—the deterministic and repeatable output. We assume that it isn't possible to test multiple inputs with the same condition vector, since the input could be a part of the condition vector.

The distribution over which we randomly sample Δ has a huge impact over the behavior of the algorithm. We have implemented several including a *unitary* distribution which always picks Δ to be unitary and an *even* distribution which picks random complex numbers for each element of Δ and throws away samples with singular values greater than one. Note that this type of data generation is still unrealistic because it matches so exactly to the shape of the model-set—every possible result can happen—but it does nicely illustrate Thm. 3. Fig. 1 shows a mixture of two distributions: a 50–50 split between even and unitary.

A. Implementation with Real Number Solvers

The convex optimization problem we described includes positive semi-definite hermitian constraints. However, our optimizer framework [26] does not handle complex variables. To get around this we map complex matrices to real matrices of twice their dimension:

$$\phi(A + Bj) \mapsto \begin{pmatrix} A & -B \\ B & A \end{pmatrix}. \tag{60}$$

Most of these implementation details are trivial, but a few are non-obvious. Keeping the block structure of the cone parameterization makes it easier to handle the logarithm of determinant cost:

$$\begin{aligned} &\phi\left(\begin{pmatrix} -X_E & X_P \\ X_P^* & X_J - X_{PP} \end{pmatrix}\right) \\ &\mapsto \begin{pmatrix} \phi(-X_E) & \phi(X_P) \\ \phi(X_P)^T & \phi(X_J) - \phi(X_{PP}) \end{pmatrix} \end{aligned} \tag{61}$$

since

$$|\det(X_E)|^2 = \det(X_E)\overline{\det(X_E)} \tag{62}$$
$$= \det(\Re X_E + j\Im X_E)\det(\Re X_E - j\Im X_E)$$
$$= \det((\Re X_E)^2 + (\Im X_E)^2) \tag{63}$$
$$= \det(\phi(X_E)) \tag{64}$$

Where $\Re$ is the real part and $\Im$ is the imaginary part operators respectively. Also important is that positive definiteness is preserved through ϕ. Consider complex vectors, mapped $\phi(\xi) \mapsto [\Re\xi^T, \Im\xi^T]^T$ and $\phi(\xi^*) \mapsto [\Re\xi^T, -\Im\xi^*] = \phi(\xi)^T$. then if $X = X^*$ (equivalently $\phi(X) = \phi(X)^T$) the inner product $\phi(\xi^* X\xi) = \phi(\xi)^T\phi(X)\phi(\xi)$, with $\phi(x) = x$ for scalars. And so $X \succeq 0 \iff \phi(X) \succeq 0$. All of this is with some minor abuse of notation, since the mapping ϕ clearly depends on the dimension of the input, treating 2×1 matrices differently from 2–row column vectors, and scalars differently from either 1–row column vectors or 1×1 matrices.

Convex Problem 2 (Real Number Implementation)**.**

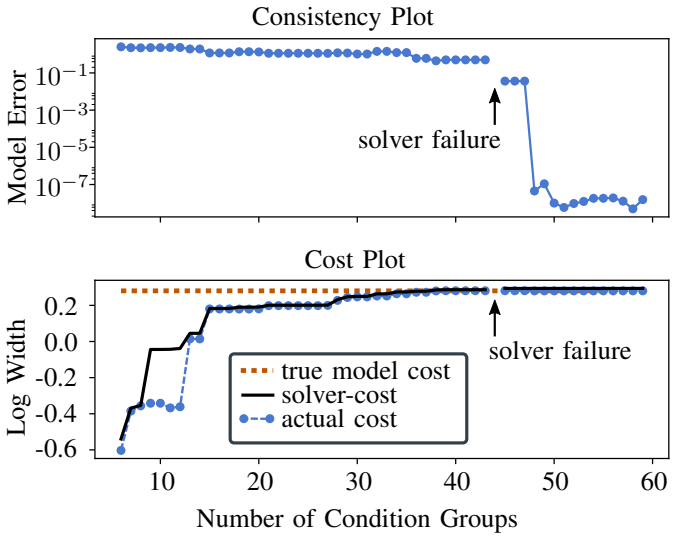
$$\begin{aligned}
\text{minimize} \quad & -\frac{1}{4n_y}\log(\det(\phi(X_E))) \\
\text{over} \quad & \phi(M),\ \phi(X_E),\ \phi(X_P),\ \phi(X_{PP}),\ \phi(X_J) \\
\text{subject to} \quad & 2 = \mathrm{tr}[\phi(X_J)] \\
& 0 \le \phi(\xi_g)^T\phi(M)\phi(\xi_g), \\
& \quad \forall \quad g \in 1\dots N_G \\
& 0 \preceq \begin{pmatrix} \phi(X_E) & -\phi(X_P) \\ -\phi(X_P)^T & \phi(X_{PP}) \end{pmatrix} \\
& 0 \preceq \phi(X_J) \\
& \phi(M) = \begin{pmatrix} -\phi(X_E) & \phi(X_P) \\ \phi(X_P)^T & \phi(X_J) - \phi(X_{PP}) \end{pmatrix} \\
& \text{and, } \forall\ X \in \{X_E, X_P, X_{PP}, X_J\}: \\
& \quad \begin{pmatrix} I & 0 \end{pmatrix}\phi(X)\begin{pmatrix} I \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & I \end{pmatrix}\phi(X)\begin{pmatrix} 0 \\ I \end{pmatrix} \\
& \quad \begin{pmatrix} I & 0 \end{pmatrix}\phi(X)\begin{pmatrix} 0 \\ I \end{pmatrix} = -\begin{pmatrix} 0 & I \end{pmatrix}\phi(X)\begin{pmatrix} I \\ 0 \end{pmatrix}
\end{aligned}$$

B. Numerical Issues in Converting Back to P, E, J Matrices

Numerical convex optimization results are fraught with rounding errors, and they return solutions which do not strictly satisfy their constraints. Rather their constraints are close to satisfaction, within some pre-defined tolerance. Matrices which should be positive semidefinite can have negative eigenvalues, matrices which should be symmetric are not symmetric, skew symmetry of the imaginary parts of hermitian matrices can be violated, et cetera. In order to guarantee valid answers we cast approximate solutions to valid solutions. We replace the "symmetric" matrix X with $0.5(X + X^T)$, the "skew-symmetric" X with $0.5(X - X^T)$, a matrix "$\phi(A+Bj)$" which, due to error, is actually $\begin{pmatrix} A_1 & -B_1 \\ B_2 & A_2 \end{pmatrix}$ is corrected to

$$\frac{1}{2}\begin{pmatrix} A_1 + A_2 & -B_1 - B_2 \\ B_1 + B_2 & A_1 + A_2 \end{pmatrix}. \tag{65}$$

Given that the model-set may subtly change in this process, Fig. 1 shows both the solver-cost claimed by the solver using the inaccurate constraint satisfaction, and the actual cost of the returned model-set. Note that the solver cost exceeds the true cost of the model-set, but the actual cost more accurately obeys this theoretical upper bound. Fig. 1 also shows that a measure of distance between the true model-set and the estimated model-set ($d((P_1,E_1,J_1),(P_2,E_2,J_2)) \triangleq \|P_1 - P_2\|_F^2 + \|E_1E_1^* - E_2E_2^*\|_F^2 + \|J_1^*J_1 - J_2^*J_2\|_F^2$) reaches zero (within numerical precision). The behavior of the actual cost relative to the solver cost is incompletely understood at the moment.



$$E = \begin{pmatrix} 1.2 & 0 \\ 0 & 1 \end{pmatrix} \quad P = \begin{pmatrix} 30 & 0 \\ 0 & 30 \end{pmatrix} \quad J = \begin{pmatrix} 1.1 & 0.2 \\ 0 & 0.5 \end{pmatrix}$$

Fig. 1. Simulation results: 2×2 system with true model-set as above, no measurement noise, Δ sampled from either the unitary distribution (with 50% probability) or from the even distribution (50% probability). This amalgamated distribution helps highlight the sudden bursts of cost when the past model-set is invalidated (typically when unitary Δ are selected). Since the cost can only increase with more data points, the cost is bounded from above by the cost of the true model-set, the model-set will eventually be forced to contain the true model-set. By Prop. 6, we can conclude that the only equilibrium behavior is for the model-set estimate to be exactly the true model-set. (However, if we use a less convenient distribution for Δ this may never occur.) In this example there is only a single solver failure, but they happen frequently in practice, especially when the refinements solver parameter is reduced.

The solvers are also surprisingly unreliable, given that the problem is convex. Our best results use CVXOPT [27] (abstol=1.0e-9, reltol=1.0e-9, feastol=1.0e-9, refinement=5, 20, 100, or 300 (switching up whenever it fails), kktsolver="robust", max_iters=200). The policy of attempting first with refinement=5 and then increasing from there is purely motivated by time considerations. Since this problem uses both semi-definite cones and exponential cones, there are only two solvers within the framework of CVXPY [26] capable of solving such problems, and SCS [28] didn't seem to work (though it did work much faster for several other combinations of constraint and cost which are not discussed here).

It is often the case that estimation algorithms get better with more data. However the traditional notion of consistent estimators is inappropriate in our case, since we want a model-set and not a single model. Our estimation algorithm can converge to the true model-set in a finite number of samples,

but these samples need to represent the extrema of Δ. Much the way three points can define a circle, our algorithm has a rigidly geometric interpretation. As shown in Fig. 1, each new data point will either inflate the model-set, or make no mark.

In reality there are never true model-sets, and even for model-sets that have deterministic bounds, the bounds are not necessarily achieved in finitely many samples. You can't invalidate a model-set based on an extreme result you have yet to see. However, this consistency result suggests that the generalized cone width is a natural choice for defining the best model-set that includes all data captured so far.

VI. Accounting for Noise

Up to this point our analysis has neglected the measurement noise term η in (3). To add it back in, we first need to be clear about the relationship between noisy measurements and conservatism in model-set choice. Robust control seeks to conservatively bound the true model-set, so one could reasonably expect that we seek to find a statistical worst case—that we would treat every deviant data point as an imperfect measurement of an even more deviant point. In fact we do the opposite. We use the data to prove that the model-set must be wide. Deviant points are treated as bad measurements of less-deviant points.

The decision is unavoidable, as the naïve choice leads to contradiction. Consider an observed (y, u) point with non-zero y but zero u. This type of point exists because of measurement error: $y = \eta$. And this type of point cannot be explained by our idealized model-set structure. An equally or more deviant point would have larger y, and so would remain inexplicable. This would contradict the model-set structure, and lead to infeasible constraints.

With that in mind we introduce the following:

Proposition 8 (Noisy Cone Parameterization)**.** *A triple of input, output, and noise phasor vectors (u, y, η) satisfies*

$$\begin{cases} y = Pu + E\Delta Ju + \eta \\ \Delta^*\Delta \preceq I \end{cases} \tag{66}$$

(with full rank uncertainty shape matrices E and J) if and only if it satisfies the following quadratic form inequality

$$\xi^* M \xi \geq \eta^* X_E \eta - 2\eta^* X_E (y - Pu) \tag{67}$$

where M is the cone parameterization of the model-set and $X_E = E^{-*}E^{-1}$, the negative of the top left block of M.

Proof. The proof structure does not meaningfully change from Theorem 1. New terms introduced by noise are simply collected on the right side of the inequality. □

The critical η is that which pushes y away from Pu in the same direction as $E\Delta Ju$. Other (y, u) points will naturally fall inside the tent propped up by the critical ones.

A. Analysis at the origin

We can reformulate the RHS of (67) as

$$\xi^* M \xi \geq -\eta^* X_E \eta - 2\eta^* X_E (y - Pu - \eta), \tag{68}$$

so that when u is zero and therefore $y = \eta$, the second term drops out:

$$\xi^* M \xi \geq -\eta^* X_E \eta. \tag{69}$$

Proposition 9. *If η is multivariate standard complex normal noise with $\mathrm{E}(\eta) = 0$ and $\Sigma \triangleq \mathrm{E}(\eta\eta^*)$ (its covariance matrix) then $\eta^* X_E \eta$ is generalized chi-square distributed. More specifically,*

$$\eta^* X_E \eta = \sum_{\lambda \in \sigma(S^* X_E S)} \lambda z_\lambda \qquad z_\lambda \sim \chi^2_2/2, \tag{70}$$

where $\sigma(A)$ is the spectrum, or set of eigenvalues of A (treating repeats as unique for indexing purposes), and S is the Cholesky decomposition of Σ, i.e. $\Sigma = SS^*$.

Proof. S can be used to represent η as a transformation of a multivariate standard complex normal distributed ν—which is a vector of complex numbers where each element has real and the imaginary parts drawn independently from a central normal distribution with variance $1/2$. This ν satisfies $\mathrm{E}[\nu\nu^*] = I$, and $\eta = S\nu$. $\mathrm{E}[\eta\eta^*] = \mathrm{E}[S\nu\nu^*S^*] = S\mathrm{E}[\nu\nu^*]S^* = SIS^* = \Sigma$. This means we can express

$$\eta^* X_E \eta = \nu^* S^* X_E S \nu \tag{71}$$

$$= \nu^* U^* \Lambda U \nu \tag{72}$$

Where real positive diagonal Λ and unitary U^* are together the eigendecomposition of $S^* X_E S$ (which exists for all complex positive definite matrices).

Since U is unitary, $\nu' = U\nu$ is also a vector of independent standard complex normals. $\mathrm{E}[\nu'\nu'^*] = U\mathrm{E}[\nu\nu^*]U^* = UIU^* = I$.

$$\eta^* X_E \eta = \sum_{\lambda \in \mathrm{diag}[\Lambda] = \sigma(S^* X_E S)} \|\nu'_\lambda\|^2 \lambda \tag{73}$$

And since for all λ, ν'_λ is complex and both the real and imaginary part are independently unit normal distributed with variance $\frac{1}{2}$, its squared magnitude is distributed $\sim \chi^2_2/2$, the chi-squared distribution with two degrees of freedom scaled by a factor of $\frac{1}{2}$ to account for the non-unity variance. □

Proposition 10. *If we constrain M such that $\mathrm{tr}[X_E\Sigma] = 1$ then the expected value of $\eta^* X_E \eta$ is one, and its variance is at most 1.*

Proof. By the properties of trace manipulation, $\mathrm{tr}[X_E\Sigma] = \mathrm{tr}[S^* X_E S]$. The trace is the sum of the eigenvalues, (70) holds, and $\mathrm{E}[\chi^2_2/2] = 1$, so

$$\mathrm{E}[\eta^* X_E \eta] = \sum_{\lambda \in \sigma(S^* X_E S)} \lambda \mathrm{E}[z_\lambda] = 1. \tag{74}$$

The variance of the $\chi^2_2/2$ distribution is 1, and a weighted average of two uncorrelated unit expected value random variates x_1 and x_2, $\alpha x_1 + (1-\alpha)x_2$ with $0 \leq \alpha \leq 1$, will have at most the larger of their two variances (Σ_1 and Σ_2 respectively):

$$\mathrm{E}[(\alpha x_1 + (1-\alpha)x_2 - 1)^2] = \\ \alpha^2\Sigma_1 + 0 + (1-\alpha)^2\Sigma_2 \leq \max(\Sigma_1, \Sigma_2) \tag{75}$$

since this variance is convex w.r.t. α.

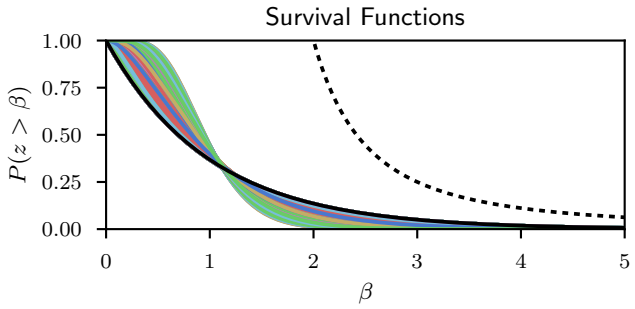


Fig. 2. Monte Carlo demonstration: the survival function of $\chi^2_2/2$ overbounds the survival function of $\eta^* X_E \eta$ for randomly chosen X_E such that $\mathrm{tr}[S^* X_E S] = 1$ for β greater than a threshold $\beta_{\min} \approx 1.25$. Survival functions in color are generated by sampling an integer order n from the uniform distribution between 1 and 9, sampling n positive numbers from the uniform distribution from 0.02 to 1.02 to represent eigenvalues of $S^* X_E S$, normalizing by their sum, and applying an analytical solution for the survival function of a weighted average of $\chi^2_2/2$ distributions. The solid black line is $e^{-\beta}$, the survival function of the $\chi^2_2/2$ distribution. The dotted black line is the bound due to Chebyshev's Inequality.

Now we extend this result by induction. Consider the N $\chi^2_2/2$ distributed variates z_i for $i = 1, \ldots, N$, and suppose WOLOG that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. We will prove by induction that after n of the $\lambda_i z_i$ terms have been combined,

$$V(n) = \mathrm{var}\left[\frac{\sum_{i=1}^{n} \lambda_i z_i}{\sum_{i=1}^{n} \lambda_i}\right] \leq \max_{i=1}^{n}(\mathrm{var}[z_i]). \tag{76}$$

$$V(n+1) = \mathrm{var}\left[V(n)\alpha_n + (1-\alpha_n) z_{n+1}\right],$$

$$0 < \alpha_n = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{n+1} \lambda_i} \leq 1 \tag{77}$$

$$V(n+1) \leq \max(V(n), \mathrm{var}[z_{n+1}]) \leq \max_{i=1}^{n+1}(\mathrm{var}[z_i]) \tag{78}$$

Thanks to our ordering assumption we avoid dividing by zero. Using $n = 1$ as a trivial base case, we complete the induction. By this logic

$$\mathrm{var}[\eta^* X_E \eta] \leq \mathrm{var}[\chi^2_2/2] = 1 \tag{79}$$

□

In [29], expressions are provided for higher order moments of the generalized chi square distribution (and an approximation is provided). And [30] shows that the generalized chi square distribution can be computed using differential equations. We have employed a similar strategy to generate Fig. 2.

Conjecture 1 (Empirical Survival Function Bound)**.** *For any X_E satisfying $\mathrm{tr}[S^* X_E S] = 1$, the survival function for $\eta^* X_E \eta$—a function mapping a threshold value β to the probability that $\eta^* X_E \eta > \beta$—is bounded from above by the survival function of the $\chi^2_2/2$ distribution for all $\beta > \beta_{\min}$, where $\beta_{\min} \approx 1.25$. This can be observed if Fig. 2, which plots the survival functions for randomly chosen valid X_E, and the survival function for $\chi^2_2/2$, $P(z > \beta \mid z \sim \chi^2_2/2) = e^{-\beta}$.*

Remark 4 (Chebyshev's Inequality bound). Since we know the variance, we are free to employ a more rigorous, but far less useful bound: $P(\eta^* X_E \eta > \beta \mid \mathrm{tr}[S^* X_E S] = 1) \leq 1/\beta^2 \ \forall \beta > 1$.

Proof. Using Chebyshev's inequality: $P(|X - \mu| \geq k\sigma) \leq 1/k^2$, where X is the variate, μ is its mean, and σ is its standard deviation. Since $\sigma \leq 1$ and $\mu = 1$ for our distribution, this is simply the worst case bound. □

We suggest use of the empirical bound in practice, since it leads to better model-sets. However there is an even more practical result we can employ when each condition group has many samples: the central limit theorem. Assuming the variance of individual samples matches our worst case variance bound from Prop. 10, then in the limit as $N_g \to \infty$, $\mathrm{P}(\frac{1}{N_g}\sum_{i=1}^{N_g} \eta^*_{g,i} X_E \eta_{g,i} > \alpha) \to \mathrm{P}(x > \alpha,\ x \sim \mathcal{N}(1, 1/N_g))$ giving us a much more aggressive inverse survival function bound to employ if we have the luxury of many samples.

B. The neglected term

Ultimately our strategy for identifying cone model-sets in the presence of noise is to amend the many exact inequality constraints

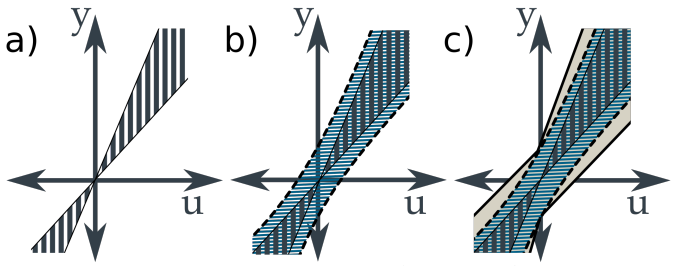
$$0 \leq \xi^*_g M \xi_g \quad \forall \quad g \in 1 \ldots N_G \tag{80}$$

from Prob. 1 by adding a bias term ($\alpha \cdot \mathrm{tr}[X_E \Sigma]$) and averaging samples within a condition group

$$0 \leq \alpha \cdot \mathrm{tr}[X_E \Sigma] + \frac{1}{N_g}\sum_{i=1}^{N_g} \xi^*_{g,i} M \xi_{g,i} \quad \forall \quad g \in 1 \ldots N_G \tag{81}$$

where the bias scale α is chosen so that $P(\eta^* X_E \eta > \alpha) < \epsilon$ for some small likelihood ϵ under the previously analyzed assumption that $\mathrm{tr}[X_E \Sigma] = 1$. Since that constraint isn't met (recall $\mathrm{tr}[X_J] = 1$), we simply scale the whole equation by $\mathrm{tr}[X_E \Sigma]/\mathrm{tr}[X_J]$. It is important to note that this strategy handles noise near the origin in an elegant way, but is a mere approximation elsewhere, as the second RHS term in (68), $-2\eta^* X_E (y - Pu - \eta)$, has been neglected.

In Fig. 3 we provide a 2D analogy to help visualize the influence of this term. Our original model-set is like a shape in high dimensional complex space. Points inside the shape are explained by the model-set. By adding the bias α, we are changing our model-set's shape to be wider, especially near the origin. If we were to rigorously inflate the original shape perfectly we would use a statistical bound on the survival function of both terms in (68). The resulting model-set would be, geometrically, a Minkoski sum. Sadly, this "inflated cone" shape does not preserve the properties we need for a tractable, convex optimization based approach. In compromise, we employ a "hyperboloid" approximation, which accurately inflates the cone near the origin, but which tapers off and asymptotes to the true cone at high values of the input magnitude. Since the origin is the region in which the default cone model-set has the most difficulty we feel this handles the most critical aspect of a noise relaxation.



Let $\xi = [y, u]^T$, $A = \begin{pmatrix} -1 & 1 \\ 1 & -\frac{15}{16} \end{pmatrix}$

(a) Cone: $\xi^T A \xi \geq 0$

(b) Hyperboloid Cone: $\xi^T A \xi \geq -\alpha$

(c) Inflated Cone: $\xi^T A \xi \geq -\alpha - 2\sqrt{\alpha}|u|/4$

Fig. 3. Cartoon illustration of the hyperboloid cone. (a) a cone in 2D. Note that $\xi^T A \xi \geq 0 \iff |u|/4 > |y-u|$. (b) a hyperboloid cone is overlaid over the original cone. We call it hyperboloid because the conic section $\xi^T A \xi = -\alpha$ is a hyperbola, and we name our high dimensional complex shape a hyperboloid cone by extension, even though its boundary is no longer so simple. (c) a properly inflated cone ($|u|/4 + \sqrt{\alpha} > |y-u|$) is overlaid on the previous two cones. This is a metaphor for our noise handling scheme: the hyperboloid cone approximates the inflated cone. It does a good job at the origin, but asymptotically approaches the pure cone at higher values of u. However, it avoids the radical term from the properly inflated cone.

VII. Statistical Test For Justifying a Model-Set

In this section we provide an analysis of variance (ANOVA) (originally in [20]) which can be used to justify the choice of a robust additive uncertainty model-set in place of a linear model. In question is the meaningfulness of the deviation between condition group averages and the prediction of a linear model. Could these errors arise simply because of measurement noise? Is a model-set even supported by the data? This test answers these questions.

In the language of statistics, our null hypothesis is that a single linear model adequately explains all our data. And we will show this model to be inconsistent by constructing two conflicting estimates of its noise variance. Specifically we check to see if the variance of the condition group averages is too wide to be explained by the variance the data within a condition group demonstrates. If the condition groups are tightly clustered around their averages, and if these condition group averages are loosely clustered about the values predicted by the linear model, then the linear model is highly unlikely to have produced the data.

To find a linear model, we perform least squares fitting, choosing the matrix P_{lin} which minimizes the sum of squared residuals $\sum_{g=1}^{N_G} \sum_{i=1}^{N_g} \epsilon_{g,i}^2$ where $\epsilon_{g,i} = y_{g,i} - P_{\text{lin}} u_{g,i}$ with double indexes representing condition group g's ith measurement. (Recall that experiments are within the same condition group if they have the same parameter vector c_g, and thus they differ only due to different realizations of η.) Here N_G is the number of condition groups measured, and N_g is the number of individual measurements in the gth condition group. We will consider the complex vector case, with $y_{g,i} \in \mathbb{C}^{n_y}$, and $u_{g,i} \in \mathbb{C}^{n_u}$.

The null hypothesis claims that while the condition groups will have average residuals which are not quite zero, the average

$$\varepsilon_g \triangleq \sum_{i=1}^{N_g} \epsilon_{g,i}/N_g \in \mathbb{C}^{n_y} \tag{82}$$

of the N_g samples from condition group g should be drawn from a zero-centered multi-complex-variate gaussian with covariance matrix $\frac{1}{N_g}$ times the covariance of the samples themselves. This means that the averages of each condition group can be used to estimate the covariance of the samples as

$$\Sigma_{\text{“group avg.”}} \triangleq \frac{\sum_{g=1}^{N_G} \varepsilon_g \varepsilon_g^* N_g^2}{\left(\sum_{g=1}^{N_G} N_g\right) - n_u}. \tag{83}$$

But this can also be estimated using the deviation of the samples within their respective groups from their group average

$$\Sigma_{\text{“within group”}} \triangleq \frac{\sum_{g=1}^{N} \sum_{i=1}^{N_g} (\epsilon_{g,i} - \varepsilon_g)(\epsilon_{g,i} - \varepsilon_g)^*}{\sum_{g=1}^{N} (N_g - 1)}. \tag{84}$$

Let us define the following distance metric between hermitian matrices:

$$\Sigma_1/\Sigma_2 = \text{tr}\left[K^{-1}\Sigma_1 K^{-*}\right]/\text{tr}[I], \;\; | \;\; KK^* = \Sigma_2. \tag{85}$$

With I the identity matrix of the same size as Σ_1 and Σ_2, so that $\text{tr}[I]$ is just n_y for our matrices in $\mathbb{C}^{n_y \times n_y}$. Note that such a K could be the Cholesky decomposition of Σ_2, but another choice (post-multiplication of K by a unitary matrix) won't influence the resulting trace. Taking the trace of this matrix and then dividing it by the trace of the identity matrix is a construction to find the average eigenvalue.

Following the logic of the null hypothesis, the residuals are normally distributed and we expect the distance between the two sample covariance estimates

$$F \triangleq \Sigma_{\text{“group avg.”}}/\Sigma_{\text{“within group”}} \tag{86}$$

$$\approx \frac{\sum_{g=1}^{N_G} \varepsilon_g^* \Sigma^{-1} \varepsilon_g N_g^2 \Big/ \left(\sum_{g=1}^{N_G} (N_g) \;\; - n_u\right)}{\sum_{g=1}^{N_G} \sum_{i=1}^{N_g} (\epsilon_{g,i} - \varepsilon_g)^* \Sigma^{-1} (\epsilon_{g,i} - \varepsilon_g) \Big/ \sum_{g=1}^{N_G} (N_g - 1)} \tag{87}$$

to be near one. This is only approximately f-distributed since it represents of a ratio of generalized χ^2 variates. We can say more about the scalar ratios of diagonal elements:

$$F_i \triangleq \Sigma_{(i,i),\text{“group avg.”}}/\Sigma_{(i,i),\text{“within group”}} \tag{88}$$

should be f-distributed with $2N_G$ and $2N_G(N_g - 1)$ degrees of freedom when all N_g are equal. The approximation (87) (which is true for $\Sigma = \Sigma_{\text{“within group”}}$) represents the behavior when a huge number of experiments are conducted. Basic Monte-Carlo simulations we conducted with equally sized, large condition groups have suggested that F's distribution is similar to the f-distribution with $2N_g n_y$ and $2n_y(N_G N_g - N_g)$

degrees of freedom under these circumstances. Prior hardware experience from [20] suggests that this test statistic tends to be so astronomically large as to be unambiguous, despite the nuances of its actual distribution.

Statistical hypothesis are refuted by showing that the observed result has an extremely low P-value according to the expected distribution defined in the hypothesis. And if the null hypothesis produces low P-values then we can rest assured that our data-driven uncertainty shapes represent more than noise in an ultimately linear system.

As an added benefit of performing this analysis, we point out that the matrix $K^{-1}\Sigma_1 K^{-*}$ acts as a sort of proxy for the signal to noise ratio in our identification of uncertainty. Larger values roughly associate with wider model-sets, and small values (relative to unity) indicate that the noise-modification will likely result in an unconstrained optimization.

VIII. Discussion

When people use $\mathcal{H}_\infty$ control they expect a guarantee of performance, a responsibility which $\mathcal{H}_\infty$ 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 consider the fundamental issues of the robust model-set—what exactly does it claim—and to visualize the model-set as a geometric shape: a high dimensional degenerate quadric in the complex space of all frequency domain inputs and outputs. We realized we could express this shape linearly in the parameters of a matrix: the cone parameterization. Switching from a degenerate to a non-degenerate quadric was a simple follow up step, and allowed our identification process some robustness to sensor noise.

With this linear form we could easily define a measure of tightness and optimize to find the tightest model-set which would not be invalidated by the data. This new optimization problem bridges between control theory and real systems in a new and interesting way, standing apart from methods which require more a-priori shape assumptions and methods which merely check for the invalidation of a previously generated model-set.

The apparent simplicity of the cone parameterization also leads us to wonder if there exist any other connections to uncertain modeling or feedback stability. The inclusion inequality is linear not only in the cone parameterization but also in the rank one outer-product matrix $\xi\xi^*$, representing a covariance and cross-covariance between input and output. There have been investigations into the re-expression of linear control concepts using the input output covariance and cross-covariance matrix [31], and this matrix formulation can be used to re-derive many of the classic $\mathcal{H}_\infty$ results [32]. Also influenced by the re-formulation is the computation of worst case $\mathcal{H}_\infty$ disturbances to the L_2 norm of a system [33]. Together these results suggest a way of thinking about $\mathcal{H}_\infty$ control which centers around the cross-covariance matrix of the inputs and outputs, and we wonder if our cone parameterization has such potential as well. We wonder if the cone parameterization is a re-usable analytical trick.

There are also several ideas in recent identification theory research which might be re-usable in connection with our work. It has been shown that performing closed loop identification can shape the identification of nominal models [34] with $\mathcal{H}_\infty$ error bounds much the same way it shapes the amplification of system noise—Identifying an additive uncertainty model-set for 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. Our model-set description also resembles a linear parameter varying (LPV) system in which we don't identify the parameter–realization maps of [22], but some effort in this direction might be used to design experiments and speed up convergence. Finally, orthonormal basis functions [35], [36] could potentially parameterize not the nominal plant, but the cone parameterization itself—porting its convexity properties to continuous-frequency optimization problems.

Potential theoretical impact aside, the identification procedure works to take 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 [37], or another feedback linearization scheme e.g. [38]; the tracking of unstable center of mass abstractions including the capture point [39] or divergent component of motion [40]—especially in the presence of series elasticity [41]; or even the impedance control of badly-modeled tendon driven fingers with compliant actuators [42]. In situations like these, we can now bypass deep understanding of (or accurate guesses for) the likely uncertainty shape simply by measuring it at some sample frequencies. This could push the innovative ideas in [43], which applies robust control to flexible actuators, and [44], 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. Forssell 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 $\mathcal{H}_\infty$;: 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 $\mathcal{H}_\infty$," *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.

[18] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994, vol. 15.

[19] 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.

[20] G. C. Thomas and L. Sentis, "Mimo identification of frequency-domain unreliability in seas," in *American Control Conference (ACC), 2017*. IEEE, 2017.

[21] J. S. Shamma and M. Athans, "Guaranteed properties of gain scheduled control for linear parameter-varying plants," *Automatica*, vol. 27, no. 3, pp. 559 – 564, 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/000510989190116J

[22] R. Tóth, *Modeling and identification of linear parameter-varying systems*. Springer, 2010.

[23] D. Y. Abramovitch, "Trying to keep it real: 25 years of trying to get the stuff i learned in grad school to work on mechatronic systems," in *2015 IEEE Conference on Control Applications (CCA)*. IEEE, 2015, pp. 223–250.

[24] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[25] G. C. Thomas, "ShapeIdentification: A free implementation of the uncertainty shape identification algorithm," http://github.com/GrayThomas/ShapeIdentification.git, 2017.

[26] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[27] M. S. Andersen, J. Dahl, and L. Vandenberghe, "Cvxopt: A python package for convex optimization, version 1.1.9," *Available at cvxopt.org*, 2013.

[28] B. ODonoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *Journal of Optimization Theory and Applications*, vol. 169, no. 3, pp. 1042–1068, 2016.

[29] S. Gabler and C. Wolff, "A quick and easy approximation to the distribution of a sum of weighted chi-square variables," *Statistische Hefte*, vol. 28, no. 1, pp. 317–325, 1987. [Online]. Available: http://dx.doi.org/10.1007/BF02932611

[30] A. W. Davis, "A differential equation approach to linear combinations of independent chi-squares," *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 212–214, 1977. [Online]. Available: http://www.jstor.org/stable/2286940

[31] A. Gattami, "Generalized linear quadratic control," *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 131–136, 2010.

[32] S. You, A. Gattami, and J. C. Doyle, "Primal robustness and semidefinite cones," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 6227–6232.

[33] A. Gattami and B. Bamieh, "Simple covariance approach to $\mathcal{H}_\infty$ analysis," *IEEE Transactions on Automatic Control*, vol. 61, no. 3, pp. 789–794, 2016.

[34] T. Oomen and O. Bosgra, "System identification for achieving robust performance," *Automatica*, vol. 48, no. 9, pp. 1975–1987, 2012.

[35] P. S. Heuberger, P. M. van den Hof, and B. Wahlberg, *Modelling and identification with rational orthogonal basis functions*. Springer Science & Business Media, 2005.

[36] 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.

[37] 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.

[38] 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

[39] 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.

[40] J. Englsberger, 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.

[41] 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.

[42] 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.

[43] 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.

[44] F. Farshidian, E. Jelavić, A. W. Winkler, and J. Buchli, "Robust whole-body motion control of legged robots," *arXiv preprint arXiv:1703.02326*, 2017.

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's 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.

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 DARPAs 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.