Estimating Linear Time-Invariant Models of Nonlinear, Time-Varying Systems

Lennart Ljung
Linköping

Outline

1. What is the LTI approximation of a general system?
2. Why does LTI identification give models with vanishing uncertainty as the data length increases, even for a nonlinear system?
3. How to obtain a reliable uncertainty measure for the estimated model?
4. Can uncertain LTI models be used to handle nonlinear model errors?
5. How to estimate such an uncertain LTI model?

The LTI World

$$y(t) = G(q)u(t) + H(q)e(t)$$

- A hub in systems and control theory and practice.
- Yet an abstraction ...
- ... that works well:
  - Good LTI approximations often available
  - Feedback is forgiving model errors

The Paradigm of Estimating LTI Models

1. Try a model structure

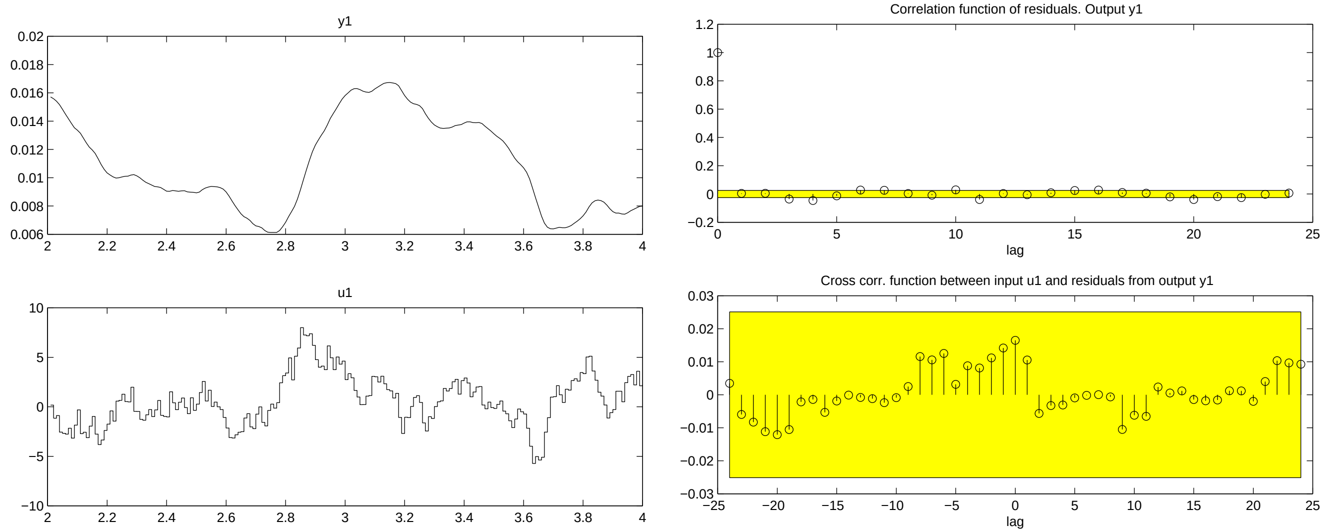
$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t)$$
$$\Longleftrightarrow$$
$$\hat{y}(t|\theta) = G(q,\theta)u(t) + (I - H^{-1}(q,\theta))(y(t) - G(q,\theta)u(t))$$

2. Estimate $\hat{\theta}_N$ and increase the model orders until the residuals $\varepsilon(t) = y(t) - \hat{y}(t|\hat{\theta}_N)$ pass a validation test.
3. Accept the estimate as an uncertain model with the uncertainty given by the standard statistical measures (parameter covariance matrix). For models for control design this could be depicted as a band in the Nyquist plot, or equivalent measures.
4. (Use the uncertain LTI model for robust linear control design.)

An example

A rotating rigid body: From torque to angular velocity

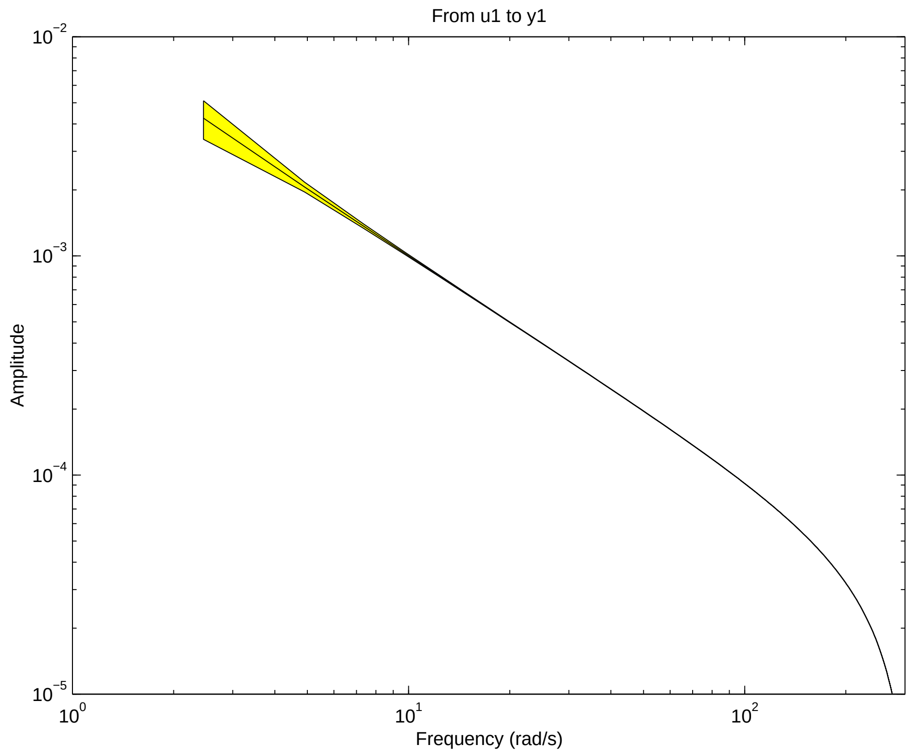
Data:



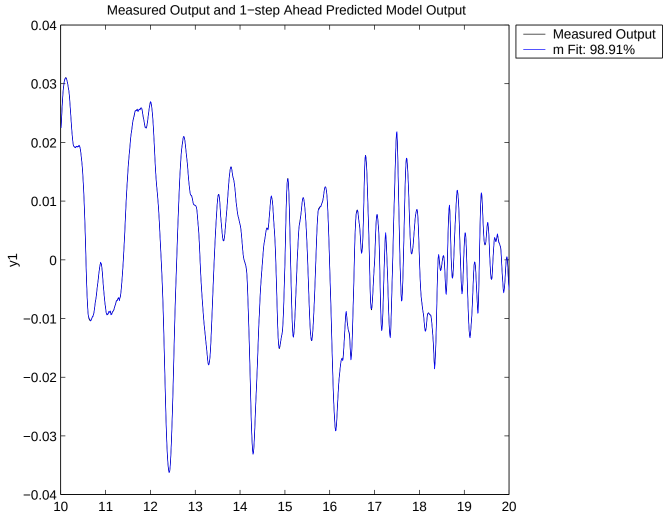
$$G(q) = \frac{5 + 0.01q^{-1} + 5q^{-2}}{1 + 1.9q^{-1} + 0.99q^{-2}} 10^{-5}$$

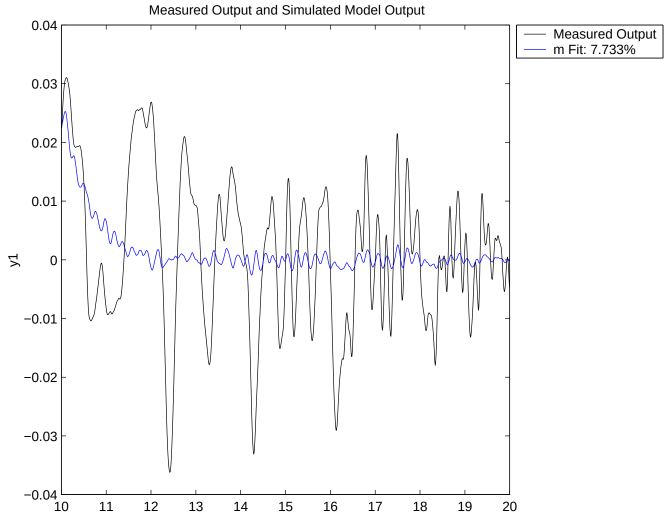
$$H(q) = \frac{1 + 0.5q^{-1} + 0.08q^{-2} + 0.02q^{-3}}{1 + 2.9q^{-1} + 2.8q^{-2} - 0.9q^{-3}}$$

Model and Uncertainty



Predicted and Simulated Output





What Happens in LTI modeling? A Naked Convergence Result

Some Theory

$x(t)$ is *quasistationary* if

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} x(t)x^T(t-\tau) = R_x(\tau) \quad \forall\tau$$

Suppose the *Spectral function*

$$\Phi_x(z) = \sum_{\tau=-\infty}^{\infty} R_x(\tau)z^{-\tau}$$

is well defined (with some regularity properties).

Define cross spectra analogously.

The Wiener Filter

Then the *Wiener filter* for predicting $x(t)$ from past x can be determined:

$$\boxed{\hat{x}(t|t-1) = W_x(q)x(t)}$$

where the strictly causal filter W_x is computed from $\Phi_x(z)$ in a well defined way.

The basic property is that the estimation error

$$\tilde{x}(t) = \hat{x}(t|t-1)$$

is such that the cross spectrum $\Phi_{x\tilde{x}}(z)$ is an *anticausal function* ("$\tilde{x}(t)$ is uncorrelated with past $x(s)$")

Input-Output data

So, given any quasistationary input/output data $z = \begin{pmatrix} y(t) \\ u(t) \end{pmatrix}$ we can define the Wiener filter for predicting $y(t)$ from past data:

$$\hat{y}(t|t-1) = W_y(q)y(t) + W_u(q)u(t) \qquad (*)$$

where the strictly causal functions W are computed from $\Phi_z(z)$ in a well defined way.

Introduce the notation

$$H_0(z) = (I - W_y(z))^{-1}, \quad G_0(z) = H_0(z)W_u(z)$$
$$e_0(t) = y(t) - \hat{y}(t|t-1)$$

Then the spectral function $\Phi_{e_0}(z)$ will be a constant λ_0, and the cross spectral function $\Phi_{ue_0}(z)$ will be anticausal.

Rearrange (*): $\boxed{y(t) = G_0(q)u(t) + H_0(q)e_0(t)}$

Prediction Error Identification Methods

- Given any quasistationary input output data set with spectral function $\Phi_z(z)$.
- Pick a model structure $y(t) = G(q,\theta)u(t) + H(q,\theta)e(t)$
- Estimate $\hat{\theta}$ by minimizing $\sum \|y(t) - \hat{y}(t|\theta)\|^2$
- Let λ_0, e_0, G_0 and H_0 be defined from Φ_z as on the previous slide.
- Define $\Phi_\zeta(z) = \begin{pmatrix} \Phi_u(z) & \Phi u e_0(z) \\ \Phi_{e_0 u}(z) & \lambda \end{pmatrix}$

Limiting Model

Then

$$\hat{\theta}_N \to \arg\min \int \| [\hat{G}_\theta(z) - G_0(z) \quad \hat{H}_\theta(z) - H_0(z)] \|^2_{\frac{\Phi_\zeta(z)}{H_\theta(z)}} dz$$

- A "naked" result: No stochastic assumptions, no system assumptions, other than data being quasistationary, no assumptions about feedback.
- Same expression as if data were generated by
  $y(t) = G_0(q)u(t) + H_0(q)e_0(t)$, $e_0(t)$ white noise (**)
- Second order methods cannot distinguish measured data from (**)
- Note: G_0, H_0 depend in general on Φ_u.

Consequences

- Given "any" data set, a LTI-model of sufficient complexity will always be unfalsified by the standard linear system identification machinery.
- The uncertainty region around this LTI-equivalent will decrease to zero as the number of observed data increases.
- It would seem more "realistic" if there were some "remaining uncertainty" in the model even when an arbitrary amount of data is available.

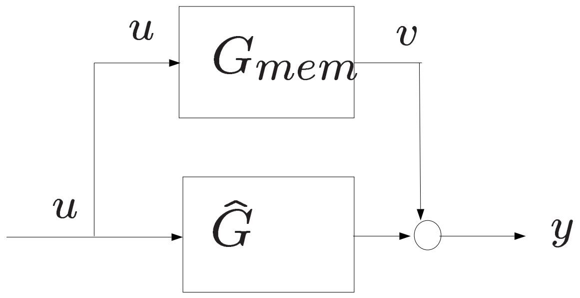
Can an Uncertain LTI Model Describe Nonlinear Model Errors?

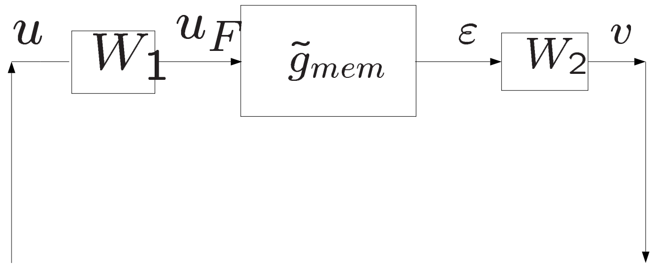
Idea # 1:

- Since the LTI-equivalent depends on the input spectrum, can we take the envelope of all LTI-equivalents as the uncertain LTI model?
- Does not work!

Idea #2: Model Error Models

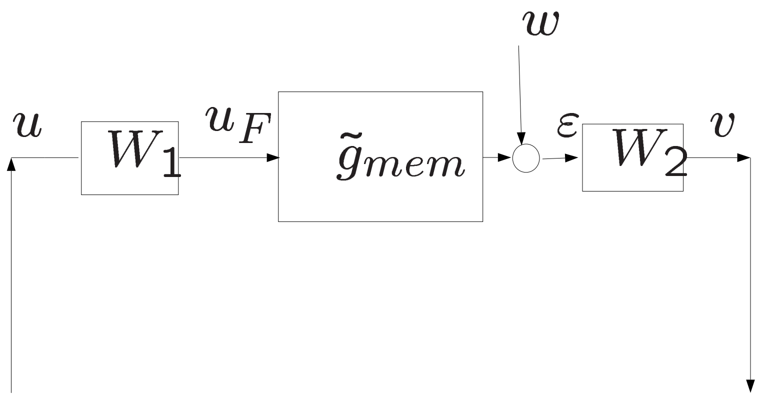
$$v = y - \hat{G}u \qquad \varepsilon = W_2^{-1}v,\ u_F = W_1 u$$





Linear $\tilde{g}_{mem} \iff$ Standard model validation

Model Error Model Size



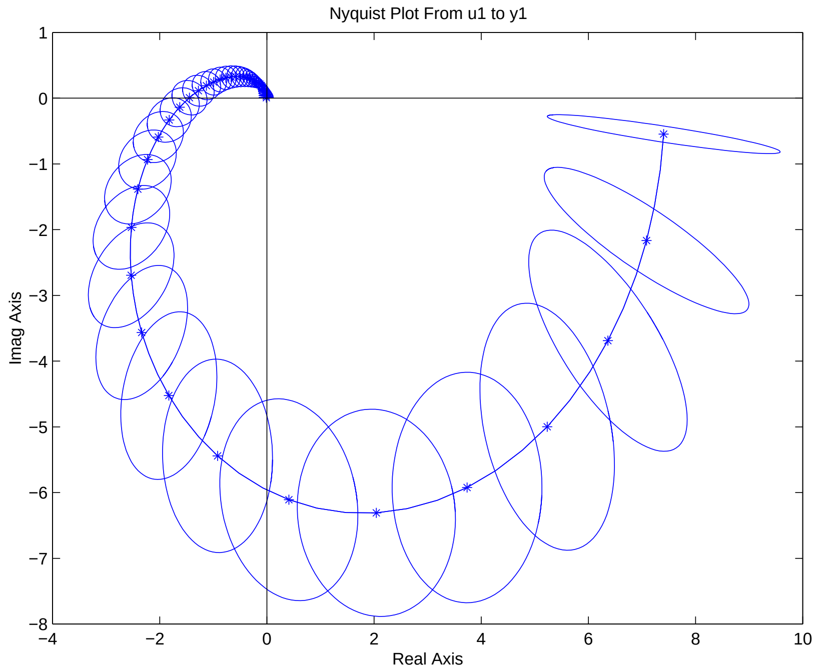
$$\varepsilon(t) = \tilde{g}_{mem}(u_F^{t-1}), \quad \|\varepsilon\| \leq \beta\|u_F\| + \alpha$$

(More precisely:

$$\int_0^T |\varepsilon(t)|^2 dt \leq \beta^2 \int_0^T |u_F(t)|^2 dt + T\alpha^2 \quad \forall T)$$

- $\hat{H}$ is a natural choice of W_2
- Model + Model Error Model: A band $\hat{G} \pm \beta W_1 W_2$

An Equivalent Uncertain LTI Model



$$\mathcal{G} = \widehat{G} \pm \beta W_1(e^{i\omega})W_2(e^{i\omega})$$

Does this work?

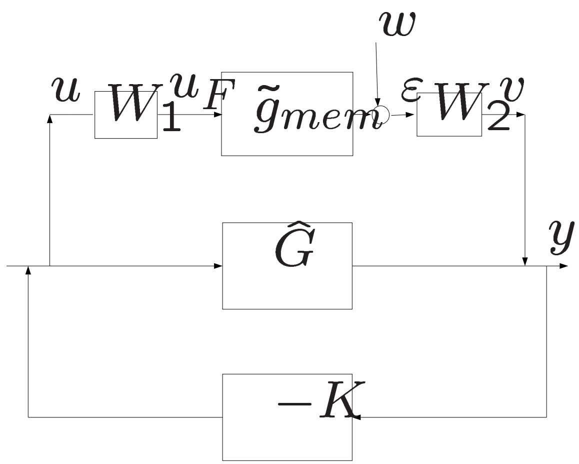
Control Design

Use standard robust LTI design to design a regulator K for this set of linear models:

- Shape the sensitivity $S = 1/(1 + K\hat{G})$ so that the disturbance at the output $W_2 S$ becomes small.
- At the same time make sure that the complementary sensitivity T is such it matches the relative model uncertainty:

$$\frac{K\hat{G}}{1 + K\hat{G}} = T < \frac{\hat{G}}{\beta W_1 W_2} \iff \beta \|T W_1 W_2 / \hat{G}\| < 1$$

The Non-linear Closed Loop System



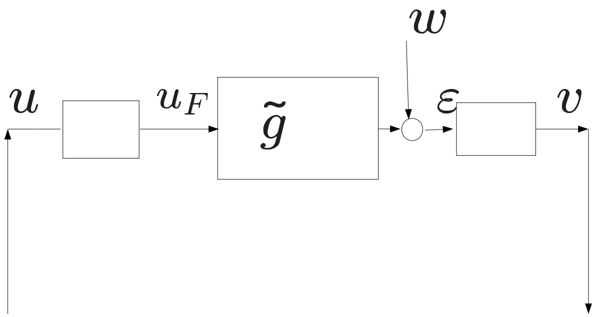
Feedback between $\tilde{g}_{mem}$ and $\frac{KW_1W_2}{1+K\hat{G}}$. Following the signals round the loop (recall the definition of affine power norm) gives

$$\|y\| \leq \|SW_2\| \frac{\alpha}{1 - \beta\|TW_1W_2/\hat{G}\|}$$

The linear robust design does the right thing!

Gain Estimation

Back to

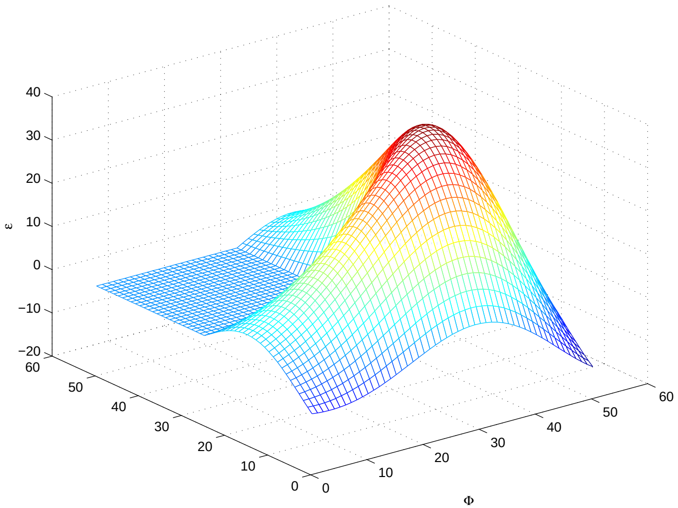


$$\varepsilon(t) = \tilde{g}_{mem}(u_F^{t-1}), \quad \|\varepsilon\| \leq \beta\|u_F\| + \alpha$$

How to estimate α and β?

- Make approximating assumption that $\varepsilon(t)$ only depends on the d past $u_F(s)$.
- Inspect the corresponding surface from $\mathcal{R}^d$ to $\mathcal{R}$

The Surface



- The floor is formed by the regressors φ, and the upright wall is the output ε.
- The gain of the system is bounded by $\sqrt{d}$ times the highest slope: $\sqrt{d}\max\frac{|\varepsilon|}{\|\varphi\|}$

How to Explore the Surface?

- At all feasible? ($d \approx 20, \dots$)
- Assume surface is a hyperplane (linear model)
- Assume surface has a low dimensional parameterization (sigmoidal NN)
- Use raw data: Radial basis NN, Local polynomial approximations, kernel methods, ...

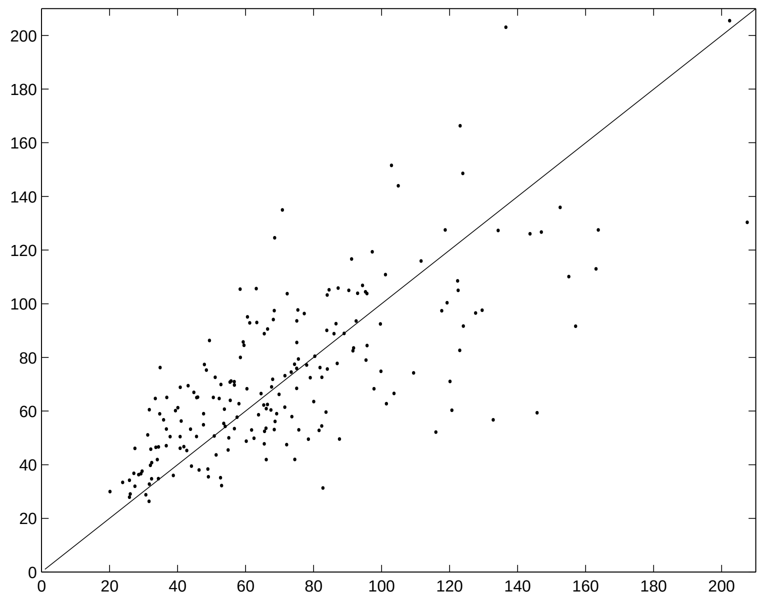
Feasibility Study: A Direct Method

Pick d as the length of the impulse response and use

$$\hat{\beta} = \sqrt{d} \max_t \frac{|\varepsilon(t)|}{\|\varphi(t)\|}$$

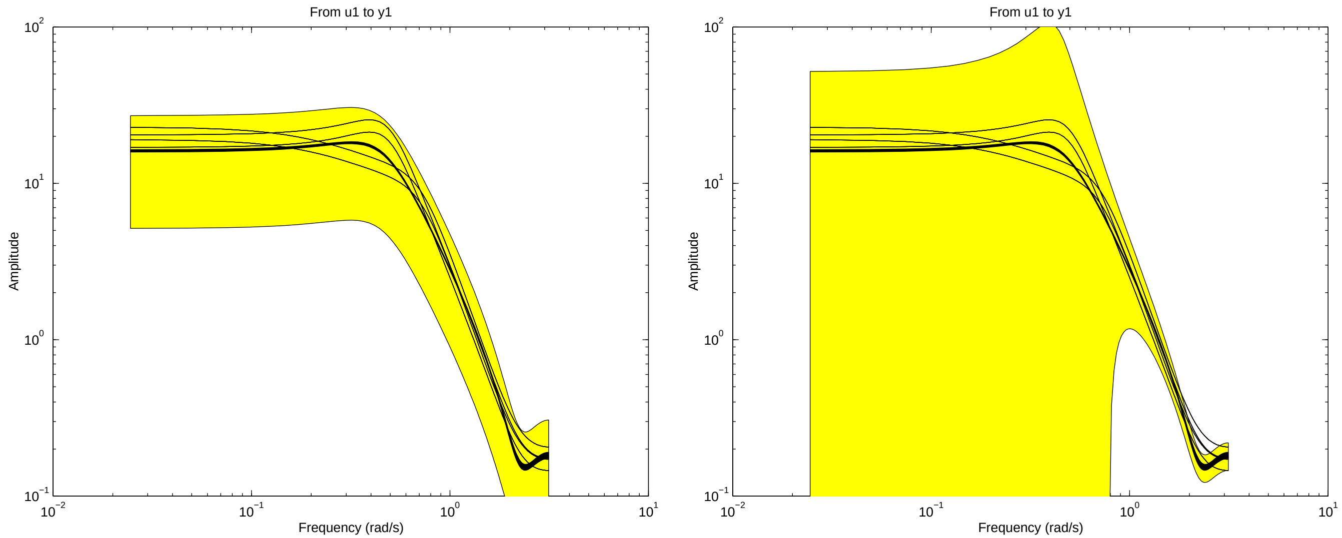
Does it work?

Tested on time-varying systems with input as well as output static non-linearities with a SNR of 10. 200 different systems tested:



x-axis: True gain, y-axis: Estimated gain

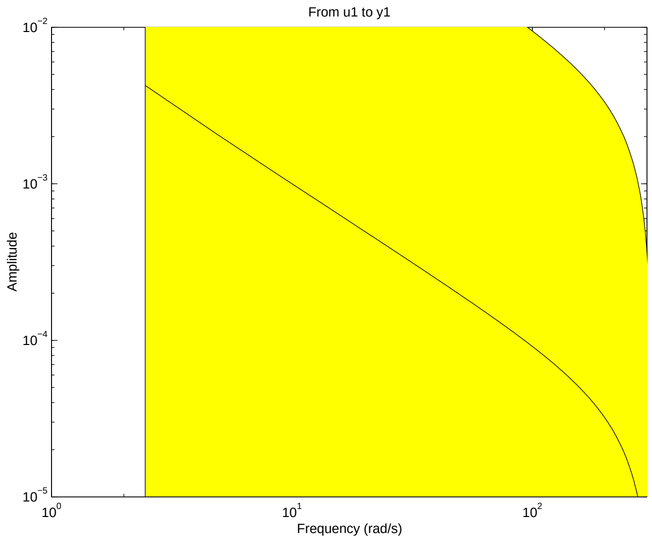
Some Plots of Remaining Uncertainty



Time varying non-linear system: “Thick” black curve: Conventional uncertainty region. Yellow region: Model error model uncertainty region. Left $W_1 = \hat{G}$, $W_2 = 1$. Right: $W_1 = \hat{G}\ W_2 = \hat{H}$

Back to the Rigid Body

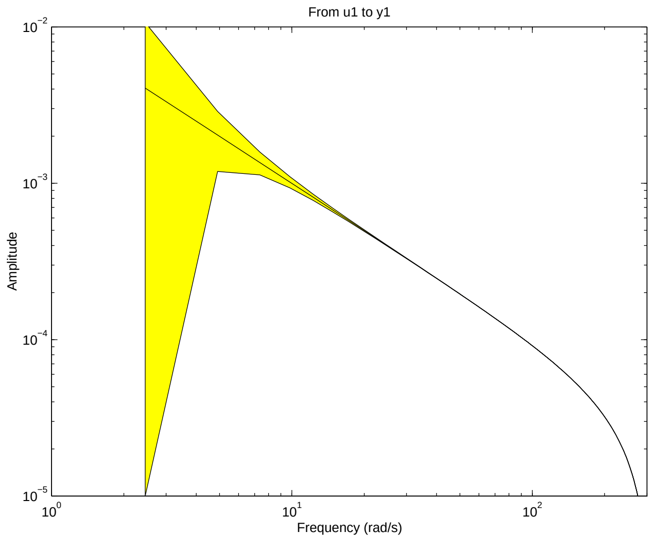
Bode plot with uncertainty:



No robust LTI design possible: “Too nonlinear”

Back to the Rigid Body

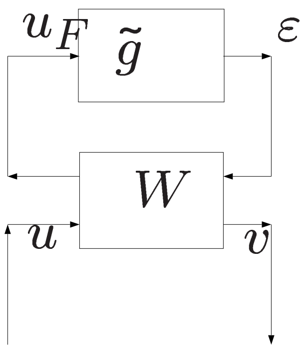
A case where the torque is more aligned with the principal axes of inertia ("more linear") Bode plot with uncertainty:



Robust LTI design should be possible in this case.

Remaining Issues

- Choice of weighting functions W.
- Choice of "weighting structure", like below, or IQC's



- Choice of W with respect to the interplay between control design requirements and obtaining small bounds. (Recall stability robustness depends only on W_1W_2.)
- More sophisticated gain estimation, dealing with noise in a better fashion etc.

Conclusions

- What does an estimated LTI model converge to?
- Why do we get unrealistic under-estimation of frequency function uncertainty?
- Can a (slightly) non-linear and time-varying system be described as an uncertain LTI model?
- Will that uncertainty decrease as we measure more data?
- How to estimate the gain of a general non-linear system?