Model Validation: A Probabilistic Formulation

Abhishek Halder and Raktim Bhattacharyya

Department of Aerospace Engineering
Texas A&M University
College Station, TX 77843-3141

50th IEEE CDC-ECC, Orlando, Florida
December 12, 2011
Given (i) a candidate model, and (ii) experimentally observed measurements of the physical system at times \( \{t_j\}_{j=1}^M \), how well does the model replicate the experimental measurements?
Given (i) a candidate model, and (ii) experimentally observed measurements of the physical system at times \( \{t_j\}_{j=1}^M \), how well does the model replicate the experimental measurements?

- **Model invalidation**
  [Smith and Doyle, 1992; Poolla et. al., 1994; Prajna, 2006]

  “The best model of a cat is another cat, or better yet, the cat itself”.
  
  — *Norbert Wiener*

- **Binary invalidation oracle**

Q1. Is this overly conservative?
Q2. Can we compute the “degree of (in)validation”?
Model validation problem: state-of-the-art

Linear Model Validation

- Robust control framework
  - Time domain
    [Poolla et. al., 1994; Chen and Wang, 1996]
  - Frequency domain
    [Smith and Doyle, 1992; Steele and Vinnicombe, 2001; Gevers et. al., 2003]
  - Mixed domain
    [Xu et. al., 1999]
- Statistical setting
  - Correlation analysis
    [Ljung and Guo, 1997]
  - Bayesian conditioning
    [Lee and Poolla, 1996]

Nonlinear Model Validation

- Barrier certificate method
  [Prajna, 2006]
- Polynomial chaos method
  [Ghanem et. al., 2008]
Model validation problem: state-of-the-art

Linear Model Validation

- Robust control framework
  - Time domain
    [Poolla et. al., 1994; Chen and Wang, 1996]
  - Frequency domain
    [Smith and Doyle, 1992; Steele and Vinnicombe, 2001; Gevers et. al., 2003]
  - Mixed domain
    [Xu et. al., 1999]
- Statistical setting
  - Correlation analysis
    [Ljung and Guo, 1997]
  - Bayesian conditioning
    [Lee and Poolla, 1996]

Nonlinear Model Validation

- Barrier certificate method
  [Prajna, 2006]
- Polynomial chaos method
  [Ghanem et. al., 2008]

“For the general case of nonparametric (uncertainty) models, the situation is significantly more complicated”
– [Lee and Poolla, 1996]

Q3. Nonlinear model validation in the sense of nonparametric statistics (aleatoric uncertainty)?
Our approach: intuitive idea

What to compare for nonlinear systems?

- Our proposal: compare shapes of the output PDFs at \( \{t_j\}_{j=1}^M \)
- Why PDFs instead of trajectories?
- Why shapes?
- Why supports?
- Why moments?

Should work for any nonlinearity, any uncertainty, both discrete and continuous time, computationally tractable, validation certificate.

Abhishek Halder (TAMU)
Our approach: intuitive idea

What to compare for nonlinear systems?

- Our proposal: compare shapes of the output PDFs at $\{t_j\}^{M}_{j=1}$
- Why PDFs instead of trajectories? supports? moments?
- Why shapes?

Should work for

- any nonlinearity
- any uncertainty
- both discrete and continuous time
- computationally tractable
- validation certificate
Outline

- Introduction
- State-of-the-art
- Intuitive idea
- Problem formulation
- Uncertainty propagation
- Distributional comparison
- Construction of validation certificates
- Examples
- Conclusions
Problem formulation

Proposed framework

Step 1. Uncertainty propagation
Step 2. Distributional comparison
Step 3. Construction of validation certificates

Abhishek Halder (TAMU)
Uncertainty propagation

### Continuous-time deterministic model

- **Model**
  \[ \dot{x} = f(x, t, p) \Rightarrow \dot{\tilde{x}} = \tilde{f}(\tilde{x}, t), \]
  \[ y = h(\tilde{x}, t) \]

- **Liouville equation**
  \[ \frac{\partial \hat{\xi}}{\partial t} = - \sum_{i=1}^{n_s} \frac{\partial}{\partial x_i} \left( \hat{\xi} f_i \right), \]
  \[ \hat{\eta}(y, t) = \sum_{j=1}^{\nu} \frac{\hat{\xi}(\tilde{x}^*_j, t)}{|\det (J_h(\tilde{x}^*_j, t))|} \]

- **Method-of-characteristics**
  \[ \frac{d \hat{\xi}}{dt} = - \hat{\xi} \nabla f, \hat{\xi}(\tilde{x}(0), 0) = \xi_0 \]

### Continuous-time stochastic model

- **Model**
  \[ d\tilde{x} = \tilde{f}(\tilde{x}, t) \, dt + g(\tilde{x}, t) \, dW, \]
  \[ y = h(\tilde{x}, t) + V \]

- **Fokker-Planck equation**
  \[ \frac{\partial \hat{\xi}}{\partial t} = - \sum_{i=1}^{n_s} \frac{\partial}{\partial x_i} \left( \hat{\xi} f_i \right) + \]
  \[ \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \frac{\partial^2}{\partial x_i \partial x_j} \left( (gQg^T)_{ij} \hat{\xi} \right), \]
  \[ \hat{\eta}(y, t) = \left( \sum_{j=1}^{\nu} \frac{\hat{\xi}(\tilde{x}^*_j, t)}{|\det (J_h(\tilde{x}^*_j, t))|} \right) \ast \phi_V \]

- **Karhunen-Loève + MOC**
  \[ \hat{x} = \tilde{f}(\tilde{x}, t) + g(\tilde{x}, t) \, KL_N \]
  \[ KL_{\infty} \, m.s. = \sqrt{2} \sum_{i=1}^{\infty} \zeta_i(\omega) \cos \left( \left( i - \frac{1}{2} \right) \frac{\pi t}{T} \right) \]
### Uncertainty propagation

#### Discrete-time deterministic model

- **Model**
  \[ \ddot{x}_{k+1} = T(\ddot{x}_k), \ y_k = h(\ddot{x}_k) \]

- **Perron-Frobenius operator**
  \[ \hat{\xi}_{k+1} = \mathcal{L} \hat{\xi}_k = \frac{\hat{\xi}_k(T^{-1}(x_{k+1}))}{|\det(\mathcal{J}_T(x_{k+1}))|}, \]

  \[ \hat{\eta}_k = \sum_{j=1}^{\nu} \frac{\hat{\xi}_k(\dddot{x}_j^*, t)}{|\det(\mathcal{J}_h(\dddot{x}_j^*, t))|} \]

- **Cell-to-cell mapping**
  
  Transition probability matrix
  \[ P_{ij} := \frac{n_{ij}}{n} \]

#### Discrete-time stochastic model

- **Model**
  \[ \ddot{x}_{k+1} = S(\ddot{x}_k) + w_k, \]
  \[ \ddot{x}_{k+1} = w_kS(\ddot{x}_k), \]
  \[ y_k = h(\ddot{x}_k) + v_k \]

- **Stochastic transfer operator**
  \[ \hat{\xi}_{k+1} = \mathcal{L}_{\text{add}} \hat{\xi}_k = \]
  \[ \int_{\mathbb{R}^{n_s}} \hat{\xi}_k(y) \phi_w(x_{k+1} - S(y)) \, dy, \]
  \[ \hat{\xi}_{k+1} = \mathcal{L}_{\text{mul}} \hat{\xi}_k = \]
  \[ \int_{\mathbb{R}^{n_s}} \hat{\xi}_k(y) \frac{1}{S(y)} \phi_w \left( \frac{x_{k+1}}{S(y)} \right) \, dy, \]
  \[ \hat{\eta}_k = \left( \sum_{j=1}^{\nu} \frac{\hat{\xi}_k(\dddot{x}_j^*, t)}{|\det(\mathcal{J}_h(\dddot{x}_j^*, t))|} \right) * \phi_v \]
Distributional comparison: axiomatic approach

Candidates for validation distance

- Kullback-Leibler divergence $D_{KL}(\rho_1||\rho_2) := \int_{\mathbb{R}^d} \rho_1(x) \log \left( \frac{\rho_1(x)}{\rho_2(x)} \right) dx$
- Symmetric KL divergence $D_{KL}^{symm}(\rho_1||\rho_2) := \frac{1}{2} (D_{KL}(\rho_1||\rho_2) + D_{KL}(\rho_2||\rho_1))$
- Wasserstein distance $pW_q(\mu_1, \mu_2) := \left[ \inf_{\mu \in \mathcal{M}_2(\mu_1, \mu_2)} \int_{\Omega} \|x - y\|^q d\mu(x, y) \right]^{1/q}$

<table>
<thead>
<tr>
<th>What we want</th>
<th>$D_{KL}$</th>
<th>$D_{KL}^{symm}$</th>
<th>$W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\geq 0$</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Symmetry</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Triangle inequality</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>$\text{supp}(\eta) \neq \text{supp}(\hat{\eta})$</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>$\text{dim}(\text{supp}(\eta)) \neq \text{dim}(\text{supp}(\hat{\eta}))$</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>$#\text{sample}(\eta) \neq #\text{sample}(\hat{\eta})$</td>
<td>×</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Convexity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Finite range</td>
<td>$[0, \infty)$</td>
<td>$[0, \infty)$</td>
<td>$[0, \text{diam}(\Omega)]$</td>
</tr>
</tbody>
</table>
Distributional comparison: axiomatic approach

Wasserstein distance in validation context

\[ p W_q (\mu_1, \mu_2) = \left( \inf_{\mu \in \mathcal{M}_2(\mu_1, \mu_2)} \mathbb{E} \left[ \| x - y \|_p^q \right] \right)^{1/q} \]

- Minimum effort required to convert one shape to another
- We choose \( p = q = 2 \), and denote \( 2W_2 \) as \( W \)
- Parametric interpretation: \( W \) depends on shape difference but not on shape i.e. for \( e_r := \| m_r - \hat{m}_r \|_2 \), \( W = W (\left\{ e_r \right\}_{r \geq 1}) \)

When can we write \( W \) in closed-form

- Single output case:
  \[ p W_q^q (\eta, \hat{\eta}) = \int_{\mathbb{R}} \| F (x) - G (x) \|_p^q \, dx = \int_0^1 \| F^{-1} (u) - G^{-1} (u) \|_p^q \, du \]

- Multivariate Normal case (comparing Linear Gaussian systems):
  \[ W \left( (A, C) ; (\hat{A}, \hat{C}) \right) = W (\eta, \hat{\eta}) = W (\mathcal{N} (\mu_1, \Sigma_1), \mathcal{N} (\mu_2, \Sigma_2)) = \sqrt{\| \mu_1 - \mu_2 \|_2^2 + \text{tr} (\Sigma_1) + \text{tr} (\Sigma_2) - 2 \text{tr} \left( (\sqrt{\Sigma_1 \Sigma_2} \sqrt{\Sigma_1})^{1/2} \right)} \]
Distributional comparison: computing Wasserstein distance

$W$ computation $\leadsto$ Monge-Kantorovich optimal transportation plan

- At each time $\{t_j\}_{j=1}^M$, we have two sets of colored scattered data
- Construct complete, weighted, directed bipartite graph $K_{m,n} (U \cup V, E)$ with $\# (U) = m$ and $\# (V) = n$
- Assign edge weight $c_{ij} := \| u_i - v_j \|_{\ell_2}^2$, $u_i \in U$, $v_j \in V$
- minimize $\sum_{i=1}^m \sum_{j=1}^n c_{ij} \varphi_{ij}$ subject to
  \[
  \sum_{j=1}^n \varphi_{ij} = \alpha_i, \quad \forall u_i \in U, \quad (C1)
  \]
  \[
  \sum_{i=1}^m \varphi_{ij} = \beta_j, \quad \forall v_j \in V, \quad (C2)
  \]
  \[
  \varphi_{ij} \geq 0, \quad \forall (u_i, v_j) \in U \times V. \quad (C3)
  \]
- Necessary feasibility condition: $\sum_{i=1}^m \alpha_i = \sum_{j=1}^n \beta_j$
**Distributional comparison: computing Wasserstein distance**

### Sample complexity

- **Rate-of-convergence of empirical Wasserstein estimate**
  \[ P \left( \left| W(\eta_m, \hat{\eta}_n) - W(\eta, \hat{\eta}) \right| > \epsilon \right) \leq K_1 \exp \left( -\frac{m\epsilon^2}{32C_1} \right) + K_2 \exp \left( -\frac{n\epsilon^2}{32C_2} \right) \]

### Runtime complexity

- An LP with \( mn \) unknowns and \( (m + n + mn) \) constraints
- For \( m = n \), runtime is \( \mathcal{O}(dn^{2.5} \log n) \)

### Storage complexity

- For \( m = n \), constraint is a binary matrix of size \( 2n \times n^2 \)
- Each row has \( n \) ones. Total \# of ones = \( 2n^2 \)
- At a given snapshot, sparse storage complexity is \( 2n(3n + d + 1) = \mathcal{O}(n^2) \)
- Non-sparse storage complexity is \( 2n(n^2 + d + 1) = \mathcal{O}(n^3) \)
Construction of validation certificates: PRVC

How robust is the inference?

- Set of admissible initial densities: \( \Psi := \{\xi_0^{(1)}, \xi_0^{(2)}, \ldots, \xi_0^{(N)}\} \)

- At time step \( k \), validation probability is \( p(\gamma_k) := \mathbb{P}(W(\eta_k, \hat{\eta}_k) \leq \gamma_k) \)

- Let \( V_k^i := \{\hat{\eta}_k^{(i)}(y) : W(\eta_k^i, \hat{\eta}_k^i) \leq \gamma_k\} \)

- Empirical validation probability is \( \hat{p}_N(\gamma_k) := \frac{1}{N} \sum_{1}^{N} 1_{V_k^i} \)

- (Chernoff bound) For any \( \epsilon, \delta \in (0, 1) \), if \( N \geq N_{ch} := \frac{1}{2\epsilon^2} \log \frac{2}{\delta} \), then
\[
\mathbb{P}(|p(\gamma_k) - \hat{p}(\gamma_k)| < \epsilon) > 1 - \delta
\]

Abhishek Halder (TAMU)
A Probabilistic Formulation
50th IEEE CDC-ECC, 2011
Algorithm 1 Construct PRVC

Require: $\epsilon, \delta \in (0, 1)$, $n$, experimental data $\{\eta_k(y)\}_{k=1}^{M}$, model, tolerance vector $\{\gamma_k\}_{k=1}^{M}$

1: $N \leftarrow N_{ch}(\epsilon, \delta)$
2: Draw $N$ random functions $\xi_0^{(1)}(\tilde{x}), \xi_0^{(2)}(\tilde{x}), \ldots, \xi_0^{(N)}(\tilde{x})$
3: for $k = 1$ to $T$ do
4: \hspace{1em} for $i = 1$ to $N$ do
5: \hspace{2em} for $j = 1$ to $\nu$ do
6: \hspace{3em} Propagate states using dynamics
7: \hspace{3em} Propagate measurements
8: \hspace{2em} end for
9: \hspace{1em} Propagate state PDF
10: \hspace{1em} Compute instantaneous output PDF
11: \hspace{1em} Compute $2W_2\left(\eta_k^{(i)}(y), \hat{\eta}_k^{(i)}(y)\right)$ \hspace{1em} $\triangleright$ Distributional comparison by solving LP
12: \hspace{1em} sum $\leftarrow 0$
13: \hspace{1em} if $2W_2\left(\eta_k^{(i)}(y), \hat{\eta}_k^{(i)}(y)\right) \leq \gamma_k$ then
14: \hspace{2em} sum $\leftarrow$ sum + 1
15: \hspace{1em} else
16: \hspace{2em} do nothing
17: \hspace{1em} end if
18: \hspace{1em} end for
19: $\hat{p}_N(\gamma_k) \leftarrow \frac{\text{sum}}{N}$ \hspace{1em} $\triangleright$ Construct PRVC vector of length $M \times 1$
20: end for
Example: Continuous-time model

▶ Truth: \( \ddot{x} = -ax - b \sin 2x - c \dot{x} \), 
\( a = 0.1, b = 0.5, c = 1 \).

▶ Five equilibria

▶ Model: Linearization about origin

▶ \( \xi_0 = \mathcal{U}([-4, 6] \times [-4, 6]) \)

▶ We plot time history of 
\[ W := \frac{W(\eta_k, \hat{\eta}_k)}{\text{diam}(\Omega_k)} \in [0, 1] \]
Example: Continuous-time model: $\bar{W}$ vs. $t$
Example: Continuous-time model: $\overline{W}$ vs. $t$
Example: Continuous-time model: $\overline{W}$ vs. $t$

- $\xi_0^{(i)} = \mathcal{N}(0, \sigma_{0i}^2 I)$

- $\text{PRVC}_{25 \times 1} = \begin{bmatrix} 1, 1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{7}{8}, \frac{3}{4}, \ldots, \frac{3}{4} \end{bmatrix}^T$

Abhishek Halder (TAMU)  
A Probabilistic Formulation  
50th IEEE CDC-ECC, 2011
Conclusions

- Unifying framework for nonlinear model validation
- Transport-theoretic Wasserstein distance as (in)validation measure
- Computable probabilistic validation certificate
- Can guide to model refinement