Further Results on Probabilistic Model Validation in Wasserstein Metric

Abhishek Halder and Raktim Bhattacharya

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

51st IEEE CDC, Maui
December 12, 2012
Model validation problem: introduction

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?

- **Predictive modeling**
  - Systems biology
  - Atmospheric modeling in planetary entry-descent-landing (EDL)

- **Controller robustness verification**
  - Flight certification
  - Switching protocol synthesis

- **Fault detection**
  - Actuator failure in physical systems
  - Security breach in cyberphysical systems
Proposed formulation

Proposed framework: Valid if $W(t_j) \leq \gamma_j, \forall j = 1, 2, \ldots, M$

Step 1. Uncertainty propagation
Step 2. Distributional comparison (Focus today)
Step 3. Construction of validation certificates
Model validation problem: state-of-the-art

Linear Model Validation

- Robust control framework
  - Time domain
    [Poolla et al., 1994; Smith and Dullerud, 1996; 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)?
Outline

- Introduction
- Proposed formulation
- State-of-the-art
- Distributional comparison
- Comparison with barrier certificate method
- Model discrimination
- Conclusions
Distributional comparison in Wasserstein metric

Wasserstein metric

Definition: Let \( y(t) \in \mathcal{Y}_t, \hat{y}(t) \in \hat{\mathcal{Y}}_t \),

\[
W_2(\eta(y, t), \hat{\eta}(\hat{y}, t)) := \left[ \inf_{\rho \in \mathcal{M}_2(\eta, \hat{\eta})} \int_{\mathcal{Y}_t \times \hat{\mathcal{Y}}_t} \| y - \hat{y} \|^2 \rho(y, \hat{y}) \, dy \, d\hat{y} \right]^{1/2}.
\]

Interpretation: Minimum effort required to convert one shape to another

Advantage: Can handle \( \text{supp}(\eta) \neq \text{supp}(\hat{\eta}), \# \text{sample}(\eta) \neq \# \text{sample}(\hat{\eta}) \)

\( W \) computation \( \leadsto \) Monge-Kantorovich optimal transportation plan

minimize \( \sum_{i=1}^m \sum_{j=1}^n c_{ij} \varphi_{ij} \) subject to

\[
\sum_{j=1}^n \varphi_{ij} = \alpha_i, \quad \forall \, u_i \in U, \quad (C1)
\]

\[
\sum_{i=1}^m \varphi_{ij} = \beta_j, \quad \forall \, v_j \in V, \quad (C2)
\]

\[
\varphi_{ij} \geq 0, \quad \forall \, (u_i, v_j) \in U \times V. \quad (C3)
\]
Distributional comparison: computing Wasserstein distance

In standard LP form

\[
\begin{align*}
\text{minimize} \quad & \tilde{c}^\top x, \\
\text{subject to} \quad & Ax = b,
\end{align*}
\]

with \(\tilde{c}_{mn \times 1} = \text{vec}(c), \ x_{mn \times 1} = \text{vec}(\varphi), \ b_{(m+n) \times 1} = [\alpha_{m \times 1}; \beta_{n \times 1}]^\top,\) Let

\[
e_n = \begin{bmatrix} 1, 1, \ldots, 1 \end{bmatrix}^\top \quad \text{n times}\]

. Then fast construction of \(A_{(m+n) \times mn} = \begin{bmatrix} e_n^\top \otimes I_m \\
I_n \otimes e_m^\top\end{bmatrix}\).

Solver used

Large scale sparse LP solver: MOSEK®
Distributional comparison: computing Wasserstein distance

Sample complexity

- Rate-of-convergence of empirical Wasserstein estimate

\[ P \left( \left| W(\eta_m, \hat{\eta}_n) - W(\eta, \hat{\eta}) \right| > \epsilon \right) \leq K_1 \exp \left( -\frac{me^2}{32C_1} \right) + K_2 \exp \left( -\frac{ne^2}{32C_2} \right) \]

Runtime complexity

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

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 \( 2n(3n + d + 1) = O(n^2) \)
- Non-sparse storage complexity is \( 2n(n^2 + d + 1) = O(n^3) \)
Comparison with barrier certificate method

- **Model:** \( \dot{x} = -px^3, \)
- **Parameter:** \( p \in P = [0.5, 2], \)
- **Measurement data:** \( X_0 = [0.85, 0.95] \) at \( t = 0 \), and \( X_T = [0.55, 0.65] \) at \( t = T = 4 \),
- **Prajna’s Barrier certificate (from SOS optimization):**
  \[
  B(x, t) = B_1(x) + tB_2(x),
  \]
  \[
  B_1(x) = 8.35x + 10.40x^2 - 21.50x^3 + 9.86x^4,
  \]
  \[
  B_2(x) = -1.78 + 6.58x - 4.12x^2 - 1.19x^3 + 1.54x^4. 
  \]
- **Our approach:** Prove that the final measure
  \[
  \xi_T(x_T, p, T) \sim \mathcal{U}(x_T, p) = \frac{1}{\text{vol}(\tilde{X}_T)}
  \]
  is not reachable from the initial measure
  \[
  \xi_0(x_0, p) \sim \mathcal{U}(x_0, p) = \frac{1}{\text{vol}(\tilde{X}_0)} \text{ in } T = 4. 
  \]
Comparison with barrier certificate method

\[ \tilde{X}_T \]

\[ \tilde{X}_0 \]
Worst case initial PDF: model discrimination problem

Results for scalar linear systems

- Deterministic (continuous and discrete time): gap $\propto \sqrt{m_{20}}$
- Uniform $\rho_0 \not\sim \sup_{\rho_0} 2W_2(t)$.

Further Results

\[ \rho_0 \sim \text{BetaPDF}(\alpha, \beta) \text{ on } [-3, 3] \]
\[ m_{20} = \frac{1}{\alpha + \beta} \sqrt{(\alpha b + \beta a)^2 + \frac{\alpha \beta (b - a)^2}{\alpha + \beta + 1}} \]
\[ a_1 = -5, a_2 = -2, c_1 = 4, c_2 = 9 \]
\[ \alpha = \beta = \frac{1}{2}, m_{20} = \frac{9}{2} \]
\[ \alpha = \beta = 1, m_{20} = 3 \]
Conclusions

- Model validation in Wasserstein metric: review
- Computational performance guarantees: need sparse LP solver
- Can recover set-based model falsification results
- Initial results on model discrimination

Questions