A Unified Bayesian Framework for Data-Driven Smoothing, Prediction, and Control

Mingzhou Yin, Andrea Iannelli, Seyed Ali Nazari, Matthias A. Müller

Institute of Automatic Control
Leibniz University Hannover

IFAC 2026 BUSAN

August 25, 2026

Data-driven tasks with stochastic data

Various methods exist to extend deterministic data-driven formulations to stochastic data

- Recover low-dimensional structure: subspace identification (Van Overschee & De Moor, 2012), projection (Breschi et al., 2023), low-rank matrix approximation (Markovsky & Dörfler, 2022)
- Tailored methods for specific tasks: distributionally robust optimization (Coulson et al., 2022), polynomial chaos expansions (Pan et al., 2023), ARX kernel method (Chiuso et al., 2025), maximum likelihood estimation (Yin et al., 2023)

This work:

- A unified Bayesian framework that extends behavioral systems theory to stochastic data through maximum a posteriori (MAP) estimation
- Applies to general data-driven tasks: smoothing, prediction, & control
- Supports input errors & correlated errors

The trajectory estimation problem

- LTI system with input disturbance & output noise: $y_t^0 = G u_t^0$, $\underbrace{\begin{bmatrix} u_t \\ y_t \end{bmatrix}}_{z_t \in \mathbb{R}^n} = \underbrace{\begin{bmatrix} u_t^0 \\ y_t^0 \end{bmatrix}}_{z_t^0} + \underbrace{\begin{bmatrix} w_t \\ v_t \end{bmatrix}}_{\nu_t}$
- Instead of model G, the system is characterized by offline data:

$$\text{length-L signal: } \mathbf{z}_i^d := \begin{bmatrix} z_{t_i+1}^d \\ \vdots \\ z_{t_i+L}^d \end{bmatrix}, \quad \text{signal matrix: } H := \begin{bmatrix} \mathbf{z}_1^d & \mathbf{z}_2^d & \cdots & \mathbf{z}_M^d \end{bmatrix} = H^0 + \Delta_H$$

- Consider the problem of estimating length-L trajectory $\mathbf{z}^0 := \mathrm{col}\left(z_1^0, z_2^0 \ldots, z_L^0\right)$ from (partial) linear observation ζ & signal matrix H:

$$\hat{\mathbf{z}} = \hat{\mathbf{z}}(\zeta, H), \quad \zeta = \Phi \mathbf{z}^0 + \underbrace{\epsilon}_{\text{obsv. error}}, \quad \mathbf{z}^0 = H^0 \mathbf{g} = H\mathbf{g} - \underbrace{\Delta_H \mathbf{g}}_{\text{offline data unc.}}$$

  - Correctness (identifiability) guaranteed by $\mathrm{rank}\left(H^0\right) = \mathrm{rank}\left(\Phi H^0\right) = n_u L + n_x$ (Willems' lemma)

Examples

- **Smoothing:** the whole trajectory is known

$$\zeta = \mathrm{col}\left(z_1, z_2, \ldots, z_L\right), \quad \epsilon = \mathrm{col}\left(\nu_1, \nu_2, \ldots, \nu_L\right)$$

- **Prediction:** all inputs & first L_0 outputs are known ($L = L_0 + L'$)

$$\zeta = \mathrm{col}\left(z_1, \ldots, z_{L_0}, u_{L_0+1}, \ldots, u_L\right), \quad \epsilon = \mathrm{col}\left(\nu_1, \ldots, \nu_{L_0}, w_{L_0+1}, \ldots, w_L\right)$$

- **Control:** first L_0 inputs & outputs are known, last L' inputs & outputs are subject to quadratic control objective $\sum_{t=L_0+1}^{L}\left(\left\|u_t^0 - u_t^{\mathrm{ref}}\right\|_R^2 + \left\|y_t^0 - y_t^{\mathrm{ref}}\right\|_Q^2\right)$

$$\zeta = \mathrm{col}\left(z_1, \ldots, z_{L_0}, \zeta_{L_0+1}^{\mathrm{ctr}}, \ldots, \zeta_L^{\mathrm{ctr}}\right), \quad \epsilon = \mathrm{col}\left(\nu_1, \ldots, \nu_{L_0}, \epsilon^{\mathrm{ctr}}\right),$$

where $\zeta_t^{\mathrm{ctr}} = \mathrm{col}\left(u_t^{\mathrm{ref}}, y_t^{\mathrm{ref}}\right)$, ϵ^{ctr}: tracking error.

Institute of Automatic Control
Leibniz University Hannover

Uncertainty specification

- Consider zero-mean, stationary, Gaussian input disturbance & output noise:

$$\begin{bmatrix}\nu_t^d \\ \nu_{t+\tau}^d\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\Sigma_\nu^d(0) & \Sigma_\nu^d(\tau) \\ \Sigma_\nu^d(\tau) & \Sigma_\nu^d(0)\end{bmatrix}\right), \quad \begin{bmatrix}\nu_t \\ \nu_{t+\tau}\end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\Sigma_\nu(0) & \Sigma_\nu(\tau) \\ \Sigma_\nu(\tau) & \Sigma_\nu(0)\end{bmatrix}\right)$$

- Uncertainty from reference tracking encoded by $\Sigma^{\mathrm{ctr}} = \mathrm{blkdiag}\left(R^{-1}, Q^{-1}\right)$
- Characterize uncertainties in both online observations & offline data:

$$\zeta = \Phi \mathbf{z}^0 + \boldsymbol{\epsilon}, \quad \mathbf{z}^0 = H^0 \mathbf{g} = H\mathbf{g} - \Delta_H \mathbf{g}$$

  - Online observations: $\boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}_p, \Sigma_\epsilon\right)$. Let $\left(\Sigma_z\right)_{i,j} = \Sigma_\nu(|i-j|)$, $\Gamma = \begin{bmatrix}\mathbf{0} & \mathbb{I}\end{bmatrix}$.

    Smoothing: $\Sigma_\epsilon = \Sigma_z$, Prediction: $\Sigma_\epsilon = \Phi \Sigma_z \Phi^\top$, Control: $\Sigma_\epsilon = \Sigma_z + \Gamma^\top \left(\mathbb{I}_{L'} \otimes \Sigma^{\mathrm{ctr}}\right) \Gamma$

  - Offline data: $\Delta_H \mathbf{g} | \mathbf{g} \sim \mathcal{N}\left(\mathbf{0}_{nL}, \Sigma_g(\mathbf{g})\right)$

$$\left(\Sigma_g(\mathbf{g})\right)_{i,j} = \sum_{k=1}^{M} \sum_{m=1}^{M} g_k g_m \Sigma_\nu^d(|t_k + i - (t_m + j)|)$$

Institute of Automatic Control
Leibniz University Hannover

Empirical Bayes approach

- Now we have the following Bayesian model:

$$\zeta|\mathbf{z}^0 \sim \mathcal{N}\left(\Phi\mathbf{z}^0, \Sigma_\epsilon\right), \qquad \mathbf{z}^0|\mathbf{g} \sim \mathcal{N}\left(H\mathbf{g}, \Sigma_g(\mathbf{g})\right)$$

- The empirical Bayes approach is commonly used, i.e., 1) estimate $\mathbf{g}$ by maximum marginal likelihood (MML) as hyperparameters, 2) find MAP estimate / posterior mean conditioned on estimated $\hat{\mathbf{g}}$.
- MML step: $\boldsymbol{\zeta} = \Phi H\mathbf{g} - \Phi\Delta_H\mathbf{g} + \boldsymbol{\epsilon} \rightarrow \boldsymbol{\zeta}|\mathbf{g} \sim \mathcal{N}\left(\Phi H\mathbf{g}, \Phi\Sigma_g(\mathbf{g})\Phi^\top + \Sigma_\epsilon\right)$

$$\hat{\mathbf{g}} = \operatorname*{argmax}_{\mathbf{g}} p\left(\boldsymbol{\zeta}|\mathbf{g}\right) = \operatorname*{argmin}_{\mathbf{g}} \log\det\left(\Psi(\mathbf{g})\right) + \|\boldsymbol{\zeta} - \Phi H\mathbf{g}\|^2_{\Psi^{-1}(\mathbf{g})}, \ \Psi(\mathbf{g}) := \Phi\Sigma_g(\mathbf{g})\Phi^\top + \Sigma_\epsilon$$

- MAP step: $p\left(\mathbf{z}^0|\boldsymbol{\zeta}, \hat{\mathbf{g}}\right) \propto p\left(\boldsymbol{\zeta}|\mathbf{z}^0\right) p\left(\mathbf{z}^0|\hat{\mathbf{g}}\right)$

$$\hat{\mathbf{z}} = \operatorname*{argmax}_{\mathbf{z}^0} p\left(\boldsymbol{\zeta}|\mathbf{z}^0\right) p\left(\mathbf{z}^0|\hat{\mathbf{g}}\right) = \operatorname*{argmax}_{\mathbf{z}^0} \|\boldsymbol{\zeta} - \Phi\mathbf{z}^0\|^2_{\Sigma_\epsilon^{-1}} + \|\mathbf{z}^0 - H\hat{\mathbf{g}}\|^2_{\Sigma_g^{-1}(\hat{\mathbf{g}})}$$

Solutions and extensions

- MML step: an SQP approximation assuming Page construction of H & uncorrelated offline uncertainties

$$\hat{\mathbf{g}}_{k+1} = \operatorname*{argmin}_{\mathbf{g}} c(\hat{\mathbf{g}}_k) \|\mathbf{g}\|_2^2 + \|\boldsymbol{\zeta} - \Phi H \mathbf{g}\|^2_{\Psi^{-1}(\mathbf{g})}$$

- MAP step: estimate coincides with posterior mean:

$$\begin{bmatrix} \mathbf{z}^0 \\ \boldsymbol{\zeta} \end{bmatrix} \Bigg| \hat{\mathbf{g}} \sim \mathcal{N}\left(\begin{bmatrix} H\hat{\mathbf{g}} \\ \Phi H \hat{\mathbf{g}} \end{bmatrix}, \begin{bmatrix} \Sigma_g(\hat{\mathbf{g}}) & \Sigma_g(\hat{\mathbf{g}})\Phi^\top \\ \Phi\Sigma_g(\hat{\mathbf{g}}) & \Psi(\hat{\mathbf{g}}) \end{bmatrix} \right), \quad \mathbf{z}^0 | \boldsymbol{\zeta}, \hat{\mathbf{g}} \sim \mathcal{N}(\hat{\mathbf{z}}, \Omega)$$

- The Gaussian uncertainty assumption can be relaxed to elliptical distributions covering robust noise models, e.g., Student's t distribution.
- When part of the trajectory is noise-free, we have degenerate distributions and similar constrained problems.

Relations to existing algorithms

- Under further specialization: 1) no input errors & uncertainty correlation, 2) scalar variance & control weights, existing algorithms are recovered
- Prediction: regularized data-driven predictor (Lian et al., 2023)

$$\hat{\mathbf{g}}_1 = \underset{\mathbf{g}}{\operatorname{argmin}}\ \lambda \|\mathbf{g}\|_2^2 + \|\mathbf{y}_p - H_{yp}\mathbf{g}\|_2^2 \quad \text{s.t. } \mathbf{u} = H_u \mathbf{g}, \qquad \lambda := n_y L_0 \sigma_d^2$$

- Control: regularized DeePC (Coulson et al., 2019)

$$\hat{\mathbf{g}}_1 = \underset{\mathbf{g}}{\operatorname{argmin}}\ r \left\|\mathbf{u}^{\text{ref}} - H_{uf}\mathbf{g}\right\|_2^2 + \lambda_1 \left\|\mathbf{y}^{\text{ref}} - H_{yf}\mathbf{g}\right\|_2^2 + \lambda_2 \|\mathbf{y}_p - H_{yp}\mathbf{g}\|_2^2 + \lambda_3 \|\mathbf{g}\|_2^2 \quad \text{s.t. } \mathbf{u}_p = H_{up}\mathbf{g}$$

$$\lambda_1 := \left(\sigma_d^2 \|\hat{\mathbf{g}}_0\|_2^2 + 1/q\right)^{-1}, \quad \lambda_2 := \left(\sigma_d^2 \|\hat{\mathbf{g}}_0\|_2^2 + \sigma^2\right)^{-1}, \quad \lambda_3 := n_y \sigma_d^2 \left(L' \lambda_1 + L_0 \lambda_2\right)$$

Numerical simulation

- Monte Carlo simulations with $n_x = 10$, $n_u = n_y = 1$, $L = 40$, $L_0 = 10$, $M = 61$
- Three case studies:

Smoothing	no input errors & i.i.d. Student's t output errors
Prediction	i.i.d. Gaussian input & output errors
Control	no input errors & correlated Gaussian output errors

- Four methods:

Bayes	proposed nonconvex algorithm
Approx	proposed algorithm with 1-iteration SQP approximation
N4SID	model-based solution with subspace identification
Proj	low-dimensional approximation by projection (γ-DDPC, Breschi et al., 2023)

Numerical simulation

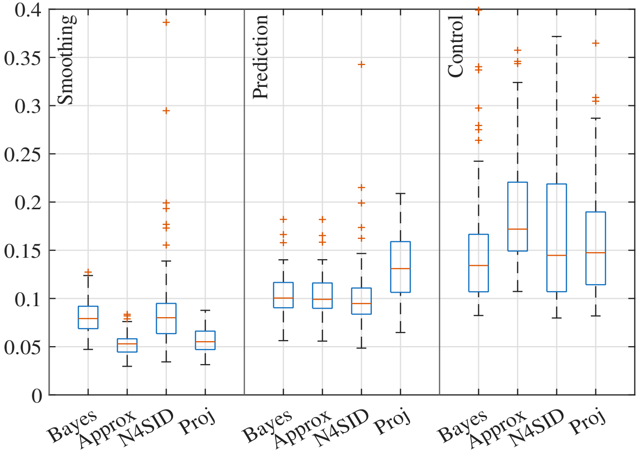

Figure: Boxplots of mean squared errors (smoothing & prediction) / normalized root control cost (control)

Unified Bayesian Framework for Data-Driven Smoothing, Prediction, and Control

- One Bayesian estimation problem for stochastic data-driven smoothing, prediction, & control
- Generalization to input errors & correlated uncertainties

Next steps:

- General convex optimization algorithm via semidefinite relaxation & MM algorithm
- Error quantification via hierarchical Bayesian approach

Institute of Automatic Control
Leibniz University Hannover

Mingzhou Yin
yin@irt.uni-hannover.de

Institute of Automatic Control
Leibniz University Hannover