Duality-Based Stochastic Policy Optimization for Estimation with Unknown Noise Covariances

Presenter: Shahriar Talebi[1]

Collaborators: Amirhossein Taghvaei[2] and Mehran Mesbahi[2]

[1]Harvard University
[2]University of Washington

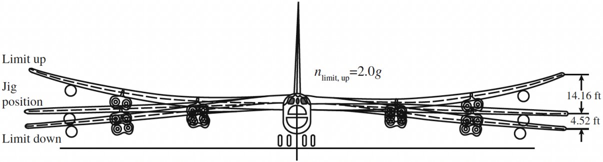
10th November 2023

a motivating example

Flexible wing aircraft:

- Boeing 787 Dreamliner

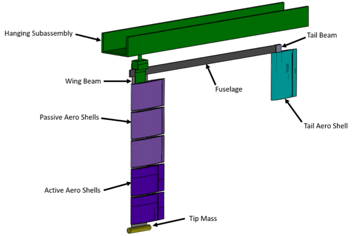
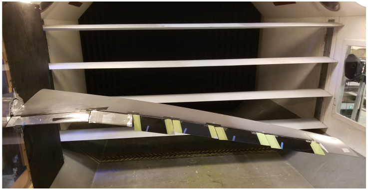
- B-52 [1]



[1] Vas and Farokhi, "Introduction to Transonic Aerodynamics"

a motivating example

- (Large) Model for Aeroelastic Response to Gust Excitation

MARGE[2]	LARGE[3]
	

- Goal: **estimate** aeroelastic modal displacements with **approximate** model and **unknown noise covariances**

[2] [Quenzer, et. al. '19]
[3] [Nguyen, et. al. '20]

a few estimation approaches

- **full identification**: [Zhang, et. al. '20] IEEE access
  - [Odelson '03], [Rajamani, et. al. '09] Autocovariance LS
  - [Matisko, et. al. '10], [Akesson, et. al. '08], [Dunik, et. al. '09],...
  - [Hinson, et. al. '22] applied to LARGE

a few estimation approaches

- **full identification**: [Zhang, et. al. '20] IEEE access
  - [Odelson '03], [Rajamani, et. al. '09] Autocovariance LS
  - [Matisko, et. al. '10], [Akesson, et. al. '08], [Dunik, et. al. '09],...
  - [Hinson, et. al. '22] applied to LARGE

- **completely unknown**: no knowledge of dynamic/noise under open-loop stability:
  - [Lale, et. al. '20], [Tsiamis, et. al. '23] → logarithmic regret
  - [Umenberger, et. al. '22] → sublinear convergence

a few estimation approaches

- **full identification**: [Zhang, et. al. '20] IEEE access
  - [Odelson '03], [Rajamani, et. al. '09] Autocovariance LS
  - [Matisko, et. al. '10], [Akesson, et. al. '08], [Dunik, et. al. '09],...
  - [Hinson, et. al. '22] applied to LARGE

- **completely unknown**: no knowledge of dynamic/noise under open-loop stability:
  - [Lale, et. al. '20], [Tsiamis, et. al. '23] → logarithmic regret
  - [Umenberger, et. al. '22] → sublinear convergence

- **our approach**: middle ground→ known dynamics, unknown noise

how to learn optimal estimation policy from output data?

estimation-control **Duality** →

a few estimation approaches

- **full identification**: [Zhang, et. al. '20] IEEE access
  - [Odelson '03], [Rajamani, et. al. '09] Autocovariance LS
  - [Matisko, et. al. '10], [Akesson, et. al. '08], [Dunik, et. al. '09],...
  - [Hinson, et. al. '22] applied to LARGE

- **completely unknown**: no knowledge of dynamic/noise under open-loop stability:
  - [Lale, et. al. '20], [Tsiamis, et. al. '23] → logarithmic regret
  - [Umenberger, et. al. '22] → sublinear convergence

- **our approach**: middle ground→ known dynamics, unknown noise

how to learn optimal estimation policy from output data?

estimation-control **Duality** →

control related works: [Fazel, et. al. '18], [Bu, et. al. '19], [Mohammadi, et. al. '21], [Zhao, et. al. '21], [Tang, et. al. '21], [Talebi, et. al. '22]...

problem formulation

Consider

$$
\begin{aligned}
x(t+1) &= Ax(t) + \xi(t) \\
y(t) &= Hx(t) + \omega(t)
\end{aligned}
$$

$$
\mathbb{E}\xi(t)\xi(t)^\intercal = Q, \quad \mathbb{E}\omega(t)\omega(t)^\intercal = R, \quad \mathbb{E}x(0)x(0)^\intercal = P_0, \quad \xi(t) \perp \omega(s)
$$

(A, H) **known** (possibly **unstable**) and Q, R, P_0 **unknown**

problem formulation

Consider

$$x(t+1) = Ax(t) + \xi(t)$$
$$y(t) = Hx(t) + \omega(t)$$
$$\mathbb{E}\xi(t)\xi(t)^\intercal = Q, \quad \mathbb{E}\omega(t)\omega(t)^\intercal = R, \quad \mathbb{E}x(0)x(0)^\intercal = P_0, \quad \xi(t) \perp \omega(s)$$

(A, H) **known** (possibly **unstable**) and Q, R, P_0 **unknown**

Problem: given observations $\mathcal{Y}(t) = \{y(0), y(1), \dots, y(t-1)\}$, find

$$\hat{x}(t) = \underset{\hat{x} \in \sigma(\mathcal{Y}(t))}{\arg\min} \, \mathbb{E}\|y(t) - H\hat{x}\|^2$$

problem formulation

Consider

$$\begin{aligned} x(t+1) &= Ax(t) + \xi(t) \\ y(t) &= Hx(t) + \omega(t) \end{aligned}$$

$$\mathbb{E}\xi(t)\xi(t)^\intercal = Q, \quad \mathbb{E}\omega(t)\omega(t)^\intercal = R, \quad \mathbb{E}x(0)x(0)^\intercal = P_0, \quad \xi(t) \perp \omega(s)$$

(A, H) **known** (possibly **unstable**) and Q, R, P_0 **unknown**

Problem: given observations $\mathcal{Y}(t) = \{y(0), y(1), \ldots, y(t-1)\}$, find

$$\hat{x}(t) = \underset{\hat{x} \in \sigma(\mathcal{Y}(t))}{\arg\min}\, \mathbb{E}\|y(t) - H\hat{x}\|^2$$

- optimal MSE estimate $\hat{x}_L$ is Kalman filter:

$$\hat{x}_L(t+1) = A\hat{x}_L(t) + L(t)(y(t) - H\hat{x}_L(t))$$

with Kalman gain $L(t) = AP(t)H^\intercal(HP(t)H^\intercal + R)^{-1}$ and

$$P(t+1) = \text{Riccati}(P(t), Q, R), \quad P(0) = P_0$$

estimation-control duality

- minimize an **LQR cost** over the **backward adjoint** dynamics

$$\min_{u(t)} \;\leftarrow J_T(u_t) = z^\intercal(0) P_0 z(0) + \sum_{t=1}^{T} z^\intercal(t) Q z(t) + u^\intercal(t) R u(t)$$

$$s.t.\;\; z(t) = A^\intercal z(t+1) - H^\intercal u(t+1), \quad z(T) = a$$

estimation-control duality

- minimize an **LQR cost** over the **backward adjoint** dynamics

$$\min_{u(t)} \leftarrow J_T(u_t) = z^\intercal(0) P_0 z(0) + \sum_{t=1}^{T} z^\intercal(t) Q z(t) + u^\intercal(t) R u(t)$$

$$s.t. \;\; z(t) = A^\intercal z(t+1) - H^\intercal u(t+1), \quad z(T) = a$$

- estimation$\leftarrow$ $\boxed{\mathbb{E}|a^\intercal x(T) - a^\intercal \hat{x}_L(T)|^2 \overset{\text{Duality}}{=} J_T(u_t = Lx_t)}$ $\rightarrow$ control

estimation-control duality

- minimize an **LQR cost** over the **backward adjoint** dynamics

$$\min_{u(t)} \leftarrow J_T(u_t) = z^\intercal(0) P_0 z(0) + \sum_{t=1}^{T} z^\intercal(t) Q z(t) + u^\intercal(t) R u(t)$$

$$s.t. \;\; z(t) = A^\intercal z(t+1) - H^\intercal u(t+1), \quad z(T) = a$$

- estimation $\leftarrow$ $\boxed{\mathbb{E}|a^\intercal x(T) - a^\intercal \hat{x}_L(T)|^2 \overset{\text{Duality}}{=} J_T(u_t = Lx_t)}$ $\rightarrow$ control

- the reparameterized problem: $J(L) := \lim_{T\to\infty} J_T(L)$

$$\min_{L} \leftarrow J(L) := \operatorname{tr}\left[X_{(L)} H^\intercal H\right],$$
$$\text{s.t.} \quad X_{(L)} = A_L X_{(L)} A_L^\intercal + Q + LRL^\intercal$$
$$L \in \mathcal{S} := \{L \in \mathbb{R}^{n\times m} : \rho(A - LH) < 1\}$$

estimation-control duality

- minimize an **LQR cost** over the **backward adjoint** dynamics

$$\min_{u(t)} \leftarrow J_T(u_t) = z^\intercal(0) P_0 z(0) + \sum_{t=1}^{T} z^\intercal(t) Q z(t) + u^\intercal(t) R u(t)$$

$$s.t.\ \ z(t) = A^\intercal z(t+1) - H^\intercal u(t+1), \quad z(T) = a$$

- estimation $\leftarrow$ $\boxed{\mathbb{E}|a^\intercal x(T) - a^\intercal \hat{x}_L(T)|^2 \overset{\text{Duality}}{=} J_T(u_t = Lx_t)}$ $\rightarrow$ control

- the reparameterized problem: $J(L) := \lim_{T\to\infty} J_T(L)$

$$\boxed{\begin{aligned} \min_L \leftarrow J(L) &:= \mathrm{tr}\left[X_{(L)} H^\intercal H\right], \\ \text{s.t.} \quad X_{(L)} &= A_L X_{(L)} A_L^\intercal + Q + LRL^\intercal \\ L \in \mathcal{S} &:= \{L \in \mathbb{R}^{n\times m} : \rho(A - LH) < 1\} \end{aligned}}$$

- define squared estimation error $\mathtt{SE}(L, \mathcal{Y}_T) := \|y(T) - H\hat{x}_L(T)\|^2$

$$J(L) \overset{\text{Duality}}{=} \lim_{T\to\infty} \mathbb{E}[\mathtt{SE}(L, \mathcal{Y}_T)]$$

Stochastic Gradient Descent for estimation

the algorithm:

$$[\text{Feed observations to the built stochastic oracle}] \quad \mathcal{Y}_T \mapsto \nabla_L \mathtt{SE}(L, \mathcal{Y}_T)$$

$$[\text{Approximate a biased noisy gradient}] \quad \nabla \widehat{J}_T(L) = \frac{1}{M} \sum_{i=1}^{M} \nabla_L \mathtt{SE}(L, \mathcal{Y}_T^i)$$

$$[\text{Run } M\text{-batch SGD}] \qquad L_{k+1} \leftarrow L_k - \eta_k \nabla \widehat{J}_T(L)$$

[4] **[Fazel; et. al. '18] and similar LQR control works.**

Stochastic Gradient Descent for estimation

the algorithm:

[Feed observations to the built stochastic oracle] $\mathcal{Y}_T \mapsto \nabla_L \texttt{SE}(L, \mathcal{Y}_T)$

[Approximate a biased noisy gradient] $\nabla \widehat{J}_T(L) = \frac{1}{M} \sum_{i=1}^{M} \nabla_L \texttt{SE}(L, \mathcal{Y}_T^i)$

[Run M-batch SGD] $L_{k+1} \leftarrow L_k - \eta_k \nabla \widehat{J}_T(L)$

- comparison: control vs optimal estimation

Problem	Parameters			Constraints	Gradient Oracle	
	cost value	Q and R	A and H	stability $\mathcal{S}$	model	biased
[4]	known	known	unknown	yes	$\mathbb{E}J(L + r\Delta)\Delta$ $\Delta \sim U(\mathbb{S}^{mn})$	yes
Estimation (this work)	unknown	unknown	known	yes	$\mathbb{E}\nabla\texttt{SE}(L, \mathcal{Y})$ $\mathcal{Y} \sim$ output data	yes
Vanila SGD	*	*	*	no	$\mathbb{E}\nabla\texttt{SE}(L, \mathcal{Y})$ $\mathcal{Y} \sim$ data dist.	no

[4] **[Fazel; et. al. '18] and similar LQR control works.**

Convergence [Informal]

Consider observable (A, H), bounded noise $\xi(t), \omega(t)$, and small stepsize. Then (w.p. $\geqslant 1 - \delta$), SGD converges linearly and globally to ϵ-optimal Kalman gain if

- # of trajectories $\geqslant O(\ln(1/\delta)/\epsilon^2)$
- trajectory length $\geqslant O(\ln(1/\epsilon))$

guarantees

Convergence [Informal]

Consider observable (A, H), bounded noise $\xi(t), \omega(t)$, and small stepsize. Then (w.p. $\geqslant 1-\delta$), SGD converges linearly and globally to ϵ-optimal Kalman gain if

- # of trajectories $\geqslant O(\ln(1/\delta)/\epsilon^2)$
- trajectory length $\geqslant O(\ln(1/\epsilon))$

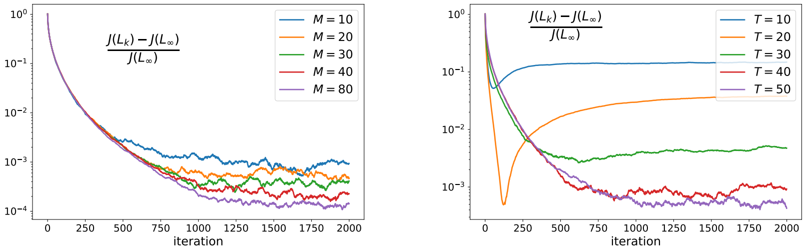
—SGD convergence with:

biased gradient, **locally Lipschitz** and **stability constraint**

—statistical bounds: with high prob

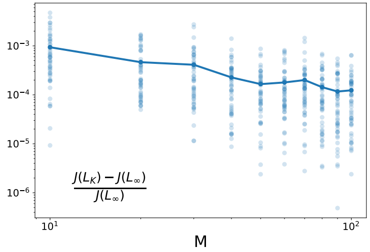
$$\|\nabla \widehat{J}_T(L) - \nabla J(L)\| \leqslant \underbrace{\|\nabla \widehat{J}_T(L) - \nabla J_T(L)\|}_{\text{concentration}} + \underbrace{\|\nabla J_T(L) - \nabla J(L)\|}_{\text{truncation}\to\text{bias}}$$

- The **average optimality gap** (average over 40 simulations) for different (a) batch-size M and (b) trajectory length T.

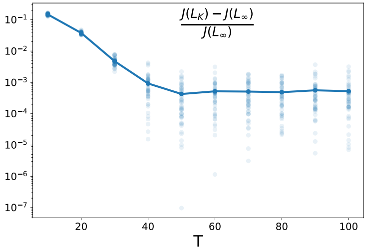


(a) (b)

- The **optimality gap at final iteration** of every simulation as a function of (a) batch-size M and (b) trajectory length T.



(a)



(b)

- Future directions: "optimal" data usage, and perturbed systems parameters