Robust Cognitive–Flexible Filtering under Noisy Innovation Scores

Thanana Nuchkrua

Outline

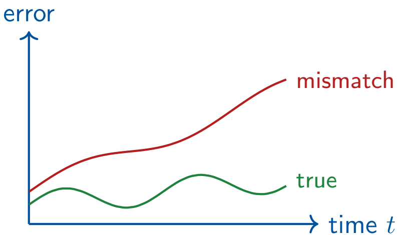
Bayesian Filtering: The Structure Problem

Classical Bayesian Filtering

- Assumes a **fixed** latent structure s
- Belief recursion: $\mathfrak{B}_{t+1} = \mathcal{F}_{\theta,s}(\mathfrak{B}_t, u_t, y_{t+1})$
- Works well when s matches reality

What if the structure is wrong?

- Posterior drifts from data-generating process
- **Persistent, biased** prediction errors
- Cannot be corrected within frozen model class



Structural mismatch $\rightarrow$ growing error

Existing Approaches & Their Limitations

Soft-mixing: IMM & variants

- Maintain distribution over model set
- **Problem:** requires pre-specified transition matrix
- No formal bound on switching rate under score perturbation

Hard-switching: GPB, MAP

- Select one structure at a time
- Rely on model-prior probabilities
- **No bound** on spurious transition rates

Cognitive Flexibility (CF) [14]

- Selects structure minimising *innovation score* $\Phi_t(s)$
- Leaves Bayesian recursion intact
- Guarantees under **exact** scores:
  1. Structural descent
  2. Finite switching
  3. Consistency

Critical Gap

What if scores are noisy?

(particle filters, learned predictors)

Noisy Innovation Score Model

True score (inaccessible in practice):

$$\Phi_t(s) := -\log \ell_{\theta,s}(y_{t+1}|\mathfrak{B}_t, u_t)$$

Observed noisy score:

$$\hat{\Phi}_t(s) = \Phi_t(s) + \epsilon_t(s)$$

Assumption 1

$\epsilon_t(s)$ is $\mathcal{I}_{t+1}$-measurable,

$$\mathbb{E}[\epsilon_t(s)|\mathcal{I}_t] = 0, \quad |\epsilon_t(s)| \le \bar{\varepsilon} \text{ a.s.}$$

Particle errors: $|\epsilon_t(s)| = O(N_p^{-1/2})$

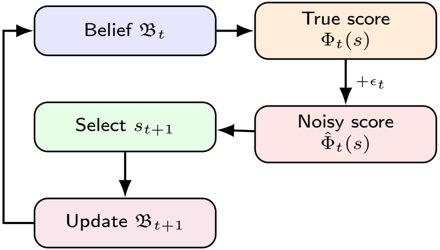


Fig. 1: Robust CF update cycle

Margin-Based Switching Rule

Greedy selector: $\hat{s}_t \in \arg\min_{s\in\mathcal{S}} \hat{\Phi}_t(s)$

Problem: Direct use induces *spurious switching* when $|\Phi_t(s_t) - \Phi_t(\hat{s}_t)| = O(\bar{\varepsilon})$

Proposed: Margin-based rule

$$s_{t+1} = \begin{cases} \hat{s}_t & \text{if } \hat{\Phi}_t(s_t) - \hat{\Phi}_t(\hat{s}_t) > \delta \\ s_t & \text{otherwise} \end{cases}$$

where **margin** $\delta > 2\bar{\varepsilon}$ (Assumption 3)

Key insight

Margin δ acts as a **noise gate**: filters out noise-driven transitions $\Rightarrow$ restores all 3 noiseless guarantees

Lock-in failure mode

If $\gamma \leq \delta - 2\bar{\varepsilon}$: algorithm may lock into suboptimal structure. **Requirement:** $\delta < \gamma + 2\bar{\varepsilon}$

Practical choice

$\delta = \alpha\bar{\varepsilon}$, $\alpha \in (2, 4]$
Paper uses: $\alpha = 2.5$

Estimate $\bar{\varepsilon}$ via: $\bar{\varepsilon} \leftarrow 2\max_s \hat{\sigma}_s$ from $W_0 = 50$ steps pilot window

Three Main Theorems

Lemma 1 (Margin descent)

Under Assms. 1–3, every switch satisfies $\Phi_t(\hat{s}_t) < \Phi_t(s_t)$
Proof: margin $\delta > 2\bar{\varepsilon}$ implies noisy ordering is consistent with true ordering

Theorem 1
Descent

$$\mathbb{E}[\Phi_t(s_{t+1})|\mathcal{I}_t] \leq \mathbb{E}[\Phi_t(s_t)|\mathcal{I}_t]$$

CF update = conditional descent map on innovation score

Theorem 2
Finite Switching

$$\mathbb{E}[N_T] \leq \frac{\sum_t \mathbb{E}[\Delta_t]}{\delta - 2\bar{\varepsilon}}$$

$\Delta_t = \Phi_t(s_t) - \min_s \Phi_t(s)$
Suboptimality drives switches

Theorem 3
Non-chattering

If $\liminf_t [\Phi_t(s_t) - \Phi_t(s^\star)] = 0$ a.s., then

$$\sum_{t=0}^{\infty} \mathbf{1}\{s_{t+1} \neq s_t\} < \infty \text{ a.s.}$$

All three noiseless guarantees are preserved under $\delta > 2\bar{\varepsilon}$

Proof Sketch: Key Steps

Lemma 1 (foundation of all proofs):

1. Switch requires $\hat{\Phi}_t(s_t) - \hat{\Phi}_t(\hat{s}_t) > \delta$
2. Expand via $\hat{\Phi}_t = \Phi_t + \epsilon_t$:

$$\Phi_t(s_t) - \Phi_t(\hat{s}_t) > \delta - \underbrace{[\epsilon_t(s_t) - \epsilon_t(\hat{s}_t)]}_{\leq 2\bar{\varepsilon} \text{ a.s.}}$$

3. Since $\delta > 2\bar{\varepsilon}$: $\Phi_t(\hat{s}_t) < \Phi_t(s_t)$ ✓

Theorem 2 (finite switching):

1. Each switch $\Rightarrow \Phi_t(s_t) - \min_s \Phi_t(s) > \delta - 2\bar{\varepsilon}$
2. Therefore:

$$(\delta - 2\bar{\varepsilon}) \cdot \mathbf{1}\{s_{t+1} \neq s_t\} \leq \Phi_t(s_t) - \min_s \Phi_t(s)$$

3. Sum over t, take $\mathbb{E}$:

$$\mathbb{E}[N_T] \leq \frac{\text{cumul. suboptimality}}{\delta - 2\bar{\varepsilon}}$$

Experimental Setup

Benchmark: Nonlinear stochastic growth model

$$z_{t+1} = \tfrac{1}{2} z_t + \tfrac{25 z_t}{1+z_t^2} + 8\cos(1.2t) + w_t$$
$$y_t = \tfrac{1}{20} z_t^2 + v_t$$

$w_t \sim \mathcal{N}(0, 10)$, $v_t \sim \mathcal{N}(0, 1)$, $z_0 \sim \mathcal{N}(0, 5)$

Candidate structures: $\mathcal{S} = \{s_{\mathrm{nl}}, s_{\mathrm{lin}}\}$

- s_{nl}: true dynamics (correct)
- s_{lin}: $z_{t+1} = 0.5 z_t + w_t$ (wrong)

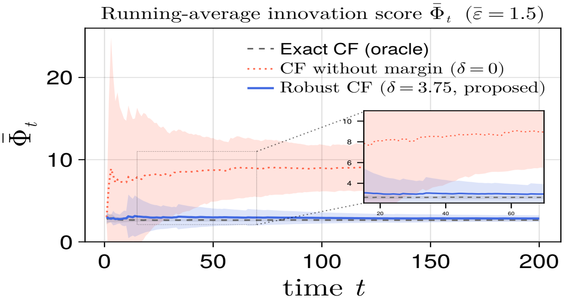
Compared methods

1. **Exact CF** (oracle): true $\Phi_t(s)$, $\delta = 2.5\bar{\varepsilon}$
2. **CF w/o margin** ($\delta = 0$): ablation baseline
3. **Robust CF** (proposed): $\hat{\Phi}_t(s)$, $\delta = 2.5\bar{\varepsilon}$

Parameters

$T = 200$, $M = 100$ runs, $N_p = 500$
$\bar{\varepsilon} \in \{0.5, 1.5, 3.0\}$
Score noise: Uniform$(-\bar{\varepsilon}, \bar{\varepsilon})$

Experiment 1: Descent in Expectation (Thm. 1)



Running-average $\bar{\Phi}_t$. Shaded: ± 1 s.d.

Observation

- **Robust CF** tracks oracle throughout
- **CF w/o margin** elevated by $\approx 3\times$ persistently
- Descent guarantee **fails** without margin condition

Validates Theorem 1

$$\mathbb{E}[\Phi_t(s_{t+1})|\mathcal{I}_t] \leq \mathbb{E}[\Phi_t(s_t)|\mathcal{I}_t]$$

iff $\delta > 2\bar{\varepsilon}$ ✓

Experiment 2: Bounded Expected Switching (Thm. 2)

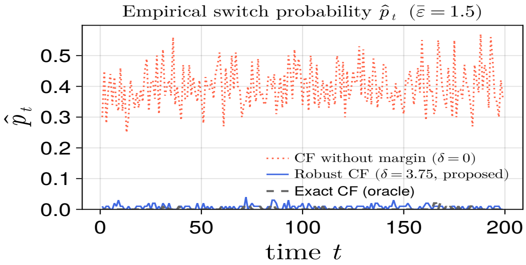
	$\bar{\varepsilon} = 0.5$		$\bar{\varepsilon} = 1.5$		$\bar{\varepsilon} = 3.0$	
Method	$\mathbb{E}[N_T]$	$\bar{\Phi}_T$	$\mathbb{E}[N_T]$	$\bar{\Phi}_T$	$\mathbb{E}[N_T]$	$\bar{\Phi}_T$
Exact CF	0.3	2.71	0.3	2.71	0.3	2.71
CF w/o margin	83.7	9.60	81.1	9.58	79.2	9.55
Robust CF	**0.3**	**2.83**	**1.3**	**2.95**	**7.9**	**3.48**
Thm. 2 bound	8.2	–	11.4	–	17.6	–

Key results

- Robust CF: $\mathbb{E}[N_T]$ **100**$\times$ **lower** than CF w/o margin
- $\bar{\Phi}_T$ close to oracle
- Bound tight as $\bar{\varepsilon}\uparrow$

Validates Theorem 2:
$\mathbb{E}[N_T] \leq \frac{\text{suboptimality}}{\delta - 2\bar{\varepsilon}}$

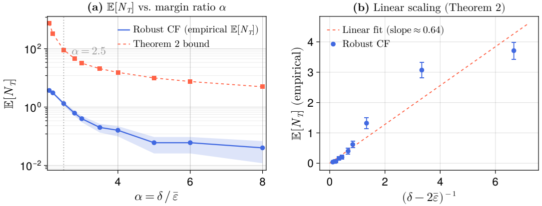
Experiment 3: Non-Chattering + Bound Tightness



Empirical switch probability $\hat{p}_t$ ($\bar{\varepsilon} = 1.5$)

Validates Theorem 3

Robust CF: $\hat{p}_t < 0.05$
CF w/o margin: $\hat{p}_t \approx 0.4$ persistently



$(\delta - 2\bar{\varepsilon})^{-1}$ *scaling ($\bar{\varepsilon} = 1.5$, $M = 100$)*

Bound tightness

Linear fit slope ≈ 0.64
confirms $(\delta - 2\bar{\varepsilon})^{-1}$ rate of Thm. 2

Conclusion & Future Work

Main result

A margin-based switching rule with $\delta > 2\bar{\varepsilon}$ restores the **complete guarantee set** of ideal CF under practical score noise, with **no modification** to the filter recursion.

Three guarantees established:

1. **Conditional score descent** (Thm. 1)
2. **Finite expected switching** $\sim (\delta - 2\bar{\varepsilon})^{-1}$ (Thm. 2)
3. **Almost-sure non-chattering** (Thm. 3)

Experiments confirm: up to **$100\times$ switching reduction** while tracking oracle performance

Future directions

- Time-varying score separation (relax Assm. 2)
- Finite-sample guarantees for online $\bar{\varepsilon}$ estimation
- Extension to learned-predictor scores (KalmanNet, etc.)

Reproducibility

Julia code available at:
`thanana.github.io/RobustCF.html`

Submitted to
IEEE Signal Processing Letters

Thank You

Questions?