Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prediction-centred active inference with stability guarantees for nonlinear delayed feedback

Mingyue Ji
Northwestern Polytechnical University
Dan Zhao
Potsdam Institute for Climate Impact Research
Quan Pan
Northwestern Polytechnical University
Yang Lyu
lyu.yang@nwpu.edu.cn
Northwestern Polytechnical University

Research Article

Keywords: Active inference, Predictor feedback, Input delay, Mean-square stability, Networked control

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11433367/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Prediction-centred active inference with stability guarantees for nonlinear delayed feedback

Mingyue Ji[a], Dan Zhao[b], Quan Pan[a], Yang Lyu[a,*]

[a]*School of Automation, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an, 710129, Shaanxi, China*
[b]*Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Telegrafenberg A56, Potsdam, 14412, Germany*

Abstract

State reconstruction and delay compensation are central to nonlinear feedback in which commands are delayed and measurements arrive intermittently. Conventional active inference (AIF) centres the hidden-state prior on the reference, while predictor feedback takes its state from a fixed-gain observer outside the control objective, so the estimate either inherits the tracking error or ignores the reliability of the evidence. Predictor-feedback delay-aware active inference (PF-DAIF) removes this coupling. In its factorized generative model, an input-history predictor centres the state prior, time-aligned measurements with bounded adaptive precision form the sensory likelihood, and the reference enters only a projected action evaluated at a forecast of the arrival state. The stationary belief error becomes a matrix-weighted convex combination of the predictor and sensing residuals, independent of the reference, and the inference dynamics contract the belief, keep the precision bounded and dissipate the free energy. An augmented model of the sampled loop yields an exact spectral stability test for fixed delay, linear matrix inequalities for mean-square stability under switching delay and packet loss, and a common Lyapunov condition over the precision range, while a local Razumikhin-type condition covers the nonlinear plant. Simulations of a delayed linear benchmark and an underactuated quadrotor with time-varying delay, packet loss, gusts and sensor bias compare PF-DAIF with target-centred AIF and a Smith-predictor PID controller. The forecast action preserves stability over delay ranges in which present-state feedback develops a growing oscillation, the estimation error stays well below that of AIF and is insensitive to reference shifts, and the tracking accuracy matches Smith-predictor control at moderate delays.

Keywords: Active inference, Predictor feedback, Input delay, Mean-square stability, Networked control

1. Introduction

Feedback through delayed and intermittent channels is a basic setting of nonlinear control. Actuation delay means that each command acts on a future state, and intermittent sensing means

*Corresponding author.
Email address: lyu.yang@nwpu.edu.cn (Yang Lyu)

that this state must be reconstructed from evidence that is both late and incomplete. Input delay can destabilize a loop that is well tuned without delay [1]. In delayed second-order feedback, a pair of characteristic roots crossing into instability produces stability switches, Hopf bifurcations and growing oscillations [2]. When the loop is closed over a network, variable transmission delay and packet loss become the central difficulties [3], and their joint effect on closed-loop performance remains an active topic [4]. Packet dropouts between sensor and estimator also degrade state estimation [5]. A controller for this setting therefore has to reconstruct the current state from delayed and intermittent data, compensate the actuation delay and weight measurement channels of varying reliability, with stability conditions that hold for the sampled loop.

Predictor feedback is the classical remedy for input delay. The Smith predictor compensates dead time with an internal model [6], and the Artstein reduction removes a known input delay from the nominal loop [7]. Nonlinear and adaptive extensions are well developed [8]. They cover time-varying input delay [9], unknown delay [10], approximate predictors [11], prescribed-time stabilization [12], inexact predictors for distinct input delays [13], stochastic delay [14] and learning-based optimal control of time-delay systems [15]. In flight control, networked quadrotors are a representative case [16]: delay-robust controllers combine delay compensation with disturbance estimation [17], and formation controllers of UAV swarms account for communication delays explicitly [18]. The resulting delay-dependent certificates usually rest on Lyapunov–Krasovskii functionals [19], and new delay-dependent functionals reduce their conservatism [20].

Despite this progress, two structural limitations remain in delayed networked loops with intermittent sensing. First, the predicted state is supplied by a separate observer with fixed gains, so estimation and control are designed apart, and the reliability of each sensor channel is not adapted online when packets are delayed or lost. Second, Lyapunov–Krasovskii conditions are sufficient only. They bound the admissible delay conservatively and do not locate the delay at which the sampled loop actually loses stability, nor the oscillation that it then develops.

Active inference (AIF) offers a unified perspective in which estimation and control minimise one variational objective. Belief and action descend a free energy [21], which yields a process theory of perception and action [22]. Perceptual inference minimises the variational free energy to update the belief over latent states [23], and the precision of each prediction error sets its influence and can be adapted online [24]. Under linear and Gaussian assumptions, AIF contains PID control as a special case [25]. Reported applications include joint state and input observation for linear systems under coloured noise [26], adaptive control of robot manipulators [27], multisensory torque control [28] and state estimation of a quadcopter flying in wind [29].

Most engineering realisations of AIF, however, place the desired state at the mean of the hidden-state prior, so one belief carries both evidence and preference. When evidence is delayed or missing, this prior pulls the belief toward the reference, and the tracking error enters the stationary estimate. Removing the hidden-state prior yields an unbiased estimate when fresh measurements arrive at every step [30], but it also discards the prediction that a delayed or intermittently observed loop requires between packets. Centring the prior on a predicted state retains this prediction. The construction was introduced for input delay [31] and extended to delayed sensing [32].

The two lines of work therefore each resolve only half of the problem. Predictor feedback compensates the actuation delay but keeps the estimator outside its objective. Predictor-centred AIF fuses prediction and evidence with adaptive weights but evaluates the preferred action at the

reconstructed current state, so the actuation delay remains in the physical loop. In addition, the sensory precision and the commanded input are not confined to admissible sets, and no stability condition covers the sampled loop under switching delay, packet loss and adaptive precision.

Motivated by these observations, predictor-feedback delay-aware active inference (PF-DAIF) separates state reconstruction from control preference through a factorized generative model. An input-history predictor centres the state prior, time-aligned measurements with bounded adaptive precision form the sensory likelihood, and the reference enters only a projected control likelihood evaluated at a forecast of the arrival state. Compared with predictor feedback, the state estimate and its channel weights follow from the objective that generates the action, so precision adaptation requires no separate estimator. Compared with earlier predictor-centred AIF, the forecast action removes the actuation delay from the nominal loop. Compared with Lyapunov–Krasovskii certificates, an augmented model of the sampled loop yields an exact spectral stability test for fixed delay. The numerical studies cover a delayed linear benchmark and an underactuated quadrotor whose position is regulated through the attitude of an inner loop [33], under time-varying delay, packet loss, gusts and sensor bias.

The main contributions are as follows.

1. A fixed-point characterization of the stationary belief under target-centred and predictor-centred priors. The predictor-centred belief error is a convex combination of predictor and sensing residuals, whereas reweighting a target-centred prior only exchanges the reference offset for sensing error.

2. A factorized PF-DAIF controller that combines input-history prediction, aligned delayed sensing, sensory precision confined to an admissible interval and a projected action evaluated at a forecast of the arrival state.

3. An exact spectral test for fixed delay, mean-square conditions under switching delay and packet loss, and a common Lyapunov condition over the adaptive precision, all derived from an augmented model of the sampled loop, together with a local Razumikhin-type condition for the nonlinear plant.

Section 2 formulates the problem, and Section 3 analyzes the target-centred prior. Section 4 presents PF-DAIF, and Section 5 establishes its closed-loop properties. Section 6 specializes the analysis to the quadrotor. Section 7 reports the simulation studies, Section 8 collects the linearization and the reduced sampled model, and Section 9 gives the conclusions.

Notation: $\|\cdot\|$ denotes the Euclidean norm, and $\|x\|_P^2 = x^\top P x$ for a symmetric positive definite matrix P. $\rho(X)$ denotes the spectral radius of a square matrix X; for a symmetric matrix, $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote its smallest and largest eigenvalues. $\mathrm{diag}(\cdot)$ denotes a diagonal or block diagonal matrix, and $\mathrm{col}(\cdot)$ stacks its arguments vertically. $\mathscr{K}$ and $\mathscr{KL}$ are the standard classes of comparison functions, and a subscript k denotes the value at $t = kT_s$. A hat marks a predicted, estimated or nominal quantity, a tilde an aligned quantity or a mismatch, a bar an upper bound and a star a stationary value; the only exception is the averaged Jacobian $\bar{H}$ of Theorem 5.3. All the symbols and their descriptions are summarized in Table 1.

Table 1: Summary of all the symbols and their descriptions.

Symbol	Description	Symbol	Description
x, u, d, w	state, input, disturbance, process noise	μ, a	belief, latent action
y_τ, $\tilde{y}$	delayed and aligned measurements	$\hat{x}_p$, $\hat{x}_p^+$	predictor and its forecast
$\mathscr{A}_y$	alignment operator	τ_f, N_f	forecast horizon, in samples
τ_u, τ_y	actuation and sensing delays	$\hat{u}_p$, K, u_d	preferred action, feedback gain, feedforward
$\hat{\tau}_u$, $\tilde{\tau}_u$	delay estimate and mismatch	$\mathscr{U}$, Δ_a	admissible input set, projection residual
x_d, e	reference, tracking error	Π_μ, Π_y, Π_a	precisions of state prior, sensory and control likelihoods
e_p, δ_y	predictor and sensing residuals	Π_x	precision of the target-centred prior
e_μ, e_y, e_a	prediction errors of the three factors	γ_y, ζ_γ	log sensory precision and its rate
F	variational free energy	T_s, N	sampling time, delay in samples
W, θ_i	$(\Pi_\mu + \Pi_y)^{-1}$, predictor weight of channel i	ξ_k, $\mathbb{A}_N$, $\mathbb{G}_N$	augmented state and closed-loop matrices
η_μ, η_a	belief relaxation factor, action step	Γ_μ, Γ_a, Λ_a	belief and action gains, $\eta_a\Pi_a$
p_{loss}	packet-loss probability	$\mathscr{T}$	second-moment map

2. Problem formulation

This section introduces the delayed plant and its aligned observation, recalls the target-centred construction of active inference, states the design problem through four required properties and collects the standing assumptions.

2.1. Delayed plant and aligned observation

The plant is the time-delay system

$$\dot{x}(t) = f\big(x(t), u(t-\tau_u(t)), d(t)\big) + w(t), \tag{1}$$

$$y_\tau(t) = h\big(x(t-\tau_y(t))\big) + \varphi_y(t) + \nu_y(t), \tag{2}$$

with state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^{n_u}$, disturbance $d \in \mathbb{R}^{n_d}$, process noise w, delayed measurement $y_\tau \in \mathbb{R}^{n_y}$, sensor fault φ_y and measurement noise ν_y. The map f is locally Lipschitz and h is continuously differentiable, so solutions exist and are unique [34]. The fault φ_y is an unknown bounded offset on a subset of the measured channels. The actuation delay τ_u and the sensing delay τ_y are bounded. The controller uses causal estimates $\hat{\tau}_u$ and $\hat{\tau}_y$, for instance from packet timestamps, with mismatches $\tilde{\tau}_u = \tau_u - \hat{\tau}_u$ and $\tilde{\tau}_y = \tau_y - \hat{\tau}_y$.

The raw sample y_τ refers to a past state. An alignment operator $\mathscr{A}_y$ propagates it over $\hat{\tau}_y$ with

the inputs applied in that window,

$$\begin{aligned}\tilde{y}(t) &= \mathscr{A}_y\big[y_\tau, \hat{\tau}_y, u_{[t-\hat{\tau}_y,t]}\big] \\ &= h\big(x(t)\big) + \delta_y(t),\end{aligned} \tag{3}$$

where the residual δ_y collects noise, fault, sensing-delay mismatch and alignment error. For full-state linear sensing, $\mathscr{A}_y$ is a finite-horizon propagation of the nominal model. When no packet arrives, the sensory term is omitted. All likelihoods over the current state use $\tilde{y}$.

2.2. Active inference and the target-centred prior

AIF specifies a generative density $p(\mu,\tilde{y},a)$ over hidden state, observation and action, together with an approximate posterior $q(\mu)$, and minimizes the variational free energy

$$\begin{aligned}F &= \mathbb{E}_q[\ln q(\mu) - \ln p(\mu,\tilde{y},a)] \\ &= -\ln p(\tilde{y},a) + D_{\mathrm{KL}}(q(\mu) \,\|\, p(\mu \mid \tilde{y},a)),\end{aligned}$$

an upper bound on the negative log evidence that is tight when q equals the posterior [23]. Under the Laplace approximation, q is Gaussian with mean μ, the belief, and F reduces up to constants to precision-weighted squared prediction errors plus log-determinants of the precisions [24]. Belief and action follow the gradient of F. The conventional construction sets the state prior to $p(\mu) = \mathscr{N}(\mu; x_d, \Pi_x^{-1})$ with the reference x_d, so a single belief carries both evidence and preference.

2.3. Control objective

Let $e = x - x_d$ be the tracking error, $\hat{x}_p$ the predictor carried by the controller, $e_p = \hat{x}_p - x$ its residual, a the latent action and $\Delta_a = u - a$ the residual of the projection of a onto the admissible set $\mathscr{U}$.

Consider the plant of Eqs. (1) and (2) with the aligned observation $\tilde{y}$, the delay estimates $\hat{\tau}_u$, $\hat{\tau}_y$ and a bounded reference x_d. The objective is to design the dynamics of the belief μ, the latent action a and the sensory precision Π_y, with the applied input $u = \mathrm{Proj}_{\mathscr{U}}(a)$, such that four requirements are met. First, for given $\hat{x}_p$ and $\tilde{y}$, the stationary belief is independent of x_d, and its deviation from x is a matrix-weighted sum of e_p and δ_y. Second, the precision Π_y remains in a prescribed interval, and a storage function built from F is non-increasing along the autonomous inference dynamics. Third, the linear sampled closed loop admits a computable stability test for fixed delay and verifiable mean-square stability conditions under switching delay and packet loss. Fourth, for bounded disturbances and residuals and for initial histories in a compact operating envelope, the tracking error satisfies

$$\|e(t)\| \le \beta\big(\big\|e_{[-\bar{\tau}_u,0]}\big\|_\infty, t\big) + \chi\big(\bar{d} + \bar{w} + \bar{\delta}_y + \bar{e}_p + \bar{\Delta}_a + \bar{\upsilon}_d\big), \qquad \beta \in \mathscr{KL},\ \chi \in \mathscr{K}, \tag{4}$$

where $e_{[-\bar{\tau}_u,0]}$ is the initial error history and $\bar{\upsilon}_d$ bounds the reference and the derivatives used by the preferred action.

Theorem 3.1 establishes the first requirement, Lemma 5.4 and Theorem 5.5 the second, and Theorems 5.6 and 5.9 the third. Theorem 5.13 establishes the bound of Eq. (4) under the local

envelope and small-gain assumptions of Sections 5.4.5 and 6. The complete stochastic quadrotor loop is evaluated numerically in Section 7.

2.4. Standing assumptions

Assumption 2.1 (Delays)**.** $0 \le \tau_u(t) \le \bar{\tau}_u$ and $0 \le \tau_y(t) \le \bar{\tau}_y$. Wherever τ_i is differentiable, $\sup_t \dot{\tau}_i(t) < 1$, $i \in \{u, y\}$. The estimates satisfy $0 \le \hat{\tau}_i(t) \le \bar{\tau}_i$ and $|\tilde{\tau}_u(t)| \le \bar{\tilde{\tau}}_u$.

Assumption 2.2 (Exogenous signals)**.** $\|d(t)\| \le \bar{d}$, $\|w(t)\| \le \bar{w}$ and $\left\|\delta_y(t)\right\| \le \bar{\delta}_y$, where $\bar{\delta}_y$ includes the fault bound $\bar{\varphi}_y$. In the mean-square statements, $\sup_t \mathbb{E}\|w(t)\|^2 \le \bar{w}_2^2$ and $\sup_t \mathbb{E}\left\|\delta_y(t)\right\|^2 \le \bar{\delta}_{y,2}^2$. The reference and the derivatives used by the preferred action are bounded.

Assumption 2.3 (Admissible precisions)**.** Π_μ and Π_a are constant, symmetric and positive definite. Π_y is diagonal with $\ln \Pi_{y,ii} \in [\underline{\gamma}, \bar{\gamma}]$. Hence there exist $0 < \underline{\pi}_i \le \bar{\pi}_i < \infty$ with $\underline{\pi}_i I \preceq \Pi_i \preceq \bar{\pi}_i I$ for $i \in \{\mu, y, a\}$.

Assumption 2.4 (Lipschitz envelope)**.** On a compact operating envelope $\mathscr{X}_c$, the nominal predictor field $\hat{f}$ satisfies

$$\left\|f(x,u,d) - \hat{f}(\hat{x},u,0)\right\| \le L_x \|x - \hat{x}\| + L_d \|d\| + \Delta_f$$

with finite constants L_x, L_d and a bound Δ_f on the model residual, and the nominal predictor remains in $\mathscr{X}_c$ under the projected input. In the linear case, (A,B) is stabilizable and (A,C) detectable.

These assumptions keep the error channels separate: model and delay mismatch enter through e_p and Δ_f, sensing through δ_y, and actuator limits through Δ_a.

3. Offset of the target-centred prior

This section compares the stationary beliefs produced by a target-centred and a predictor-centred state prior. The comparison identifies the reference offset of conventional AIF and proves that reweighting the precisions exchanges this offset for sensing error instead of removing it.

Consider a Gaussian state prior with centre c and precision Π, together with a linear sensory likelihood of precision Π_y,

$$F_c(\mu) = \tfrac{1}{2}(\mu - c)^\top \Pi (\mu - c) + \tfrac{1}{2}(\tilde{y} - C\mu)^\top \Pi_y (\tilde{y} - C\mu). \tag{5}$$

Conventional AIF takes $c = x_d$ and $\Pi = \Pi_x$, and PF-DAIF takes $c = \hat{x}_p$ and $\Pi = \Pi_\mu$. Both objectives are strongly convex in μ. Delay enters through the arguments: $\tilde{y}$ is aligned across the sensing delay, and $\hat{x}_p$ is reconstructed across the actuation delay.

Theorem 3.1 (Stationary beliefs of the two priors)**.** *The stationary beliefs of Eq.* (5) *are*

$$\mu^\star_{\mathrm{AIF}} = (\Pi_x + C^\top \Pi_y C)^{-1}(\Pi_x x_d + C^\top \Pi_y \tilde{y}),$$
$$\mu^\star_{\mathrm{PFD}} = (\Pi_\mu + C^\top \Pi_y C)^{-1}(\Pi_\mu \hat{x}_p + C^\top \Pi_y \tilde{y}).$$

If $C = I$ and $\tilde{y} = x + \delta_y$, then, with $W = (\Pi_\mu + \Pi_y)^{-1}$,

$$\mu^\star_{\mathrm{PFD}} - x = W\Pi_\mu e_p + W\Pi_y \delta_y, \tag{6}$$

$$\mu^\star_{\mathrm{AIF}} - x = (\Pi_x + \Pi_y)^{-1}\Pi_x(x_d - x) + (\Pi_x + \Pi_y)^{-1}\Pi_y\delta_y, \tag{7}$$

and

$$\|\mu^\star_{\mathrm{PFD}} - x\|_{W^{-1}} \le \left\|e_p\right\|_{W^{-1}} + \left\|\delta_y\right\|_{W^{-1}}. \tag{8}$$

Proof. See Appendix A.1 for details.

Both stationary errors contain a sensing term. Only Eq. (7) contains a term proportional to the tracking error, and this term vanishes only on the reference. In Eq. (6), the predictor residual e_p takes its place, so the stationary belief has no explicit dependence on x_d for given $\hat{x}_p$ and $\tilde{y}$. The reference reaches the belief only through the predictor and the applied inputs.

Lemma 3.2 (Channel-wise weights)**.** *Let $\Pi_\mu = \mathrm{diag}(\pi_{\mu,i})$, $\Pi_x = \mathrm{diag}(\pi_{x,i})$ and $\Pi_y = \mathrm{diag}(\pi_{y,i})$, and define $\theta_i = \pi_{\mu,i}/(\pi_{\mu,i} + \pi_{y,i})$ and $\theta^x_i = \pi_{x,i}/(\pi_{x,i} + \pi_{y,i})$. Then*

$$\begin{aligned} \mu^\star_{\mathrm{PFD},i} - x_i &= \theta_i e_{p,i} + (1-\theta_i)\delta_{y,i}, \\ \mu^\star_{\mathrm{AIF},i} - x_i &= \theta^x_i(x_{d,i} - x_i) + (1-\theta^x_i)\delta_{y,i}, \end{aligned} \tag{9}$$

and $|\mu^\star_{\mathrm{PFD},i} - x_i| \le \max\{|e_{p,i}|, |\delta_{y,i}|\}$.

Proof. See Appendix A.1 for details.

The weights on reference offset and sensing error sum to one. Lowering $\pi_{x,i}$ to shrink the offset therefore transfers the sensing error into the belief, and the two errors are exchanged rather than removed. With fresh evidence at every step, deleting the state prior sets $\theta^x_i = 0$ and removes the offset [30]. Under sensing delay, however, the aligned evidence carries the alignment error, and under packet loss no evidence is available, so the prior must be retained. Between packets, the target-centred belief then relaxes toward the reference. Under the predictor-centred prior, the same convex combination acts on $e_{p,i}$ and $\delta_{y,i}$, both residuals of the design, and the belief error is bounded by the larger of the two. The adaptive precision of Section 4.3 moves θ_i within fixed bounds.

4. Predictor-feedback delay-aware active inference

This section presents PF-DAIF in three steps: the prediction of the current state from the input history, the factorized generative model with its free energy, and the dynamics of the belief, the action and the sensory precision together with their sampled implementation.

PF-DAIF places prediction, sensory evidence and control preference in separate factors of the generative model. Conditioned on $\hat{x}_p$ and x_d,

$$p(\mu, \tilde{y}, a \mid \hat{x}_p, x_d) = p(a \mid \hat{x}_p, x_d)\, p(\tilde{y} \mid \mu)\, p(\mu \mid \hat{x}_p), \tag{10}$$

so $\hat{x}_p$ defines the prior on the current state, $\tilde{y}$ supplies aligned evidence, and x_d enters only the control likelihood. Fig. 1 contrasts this structure with target-centred AIF.

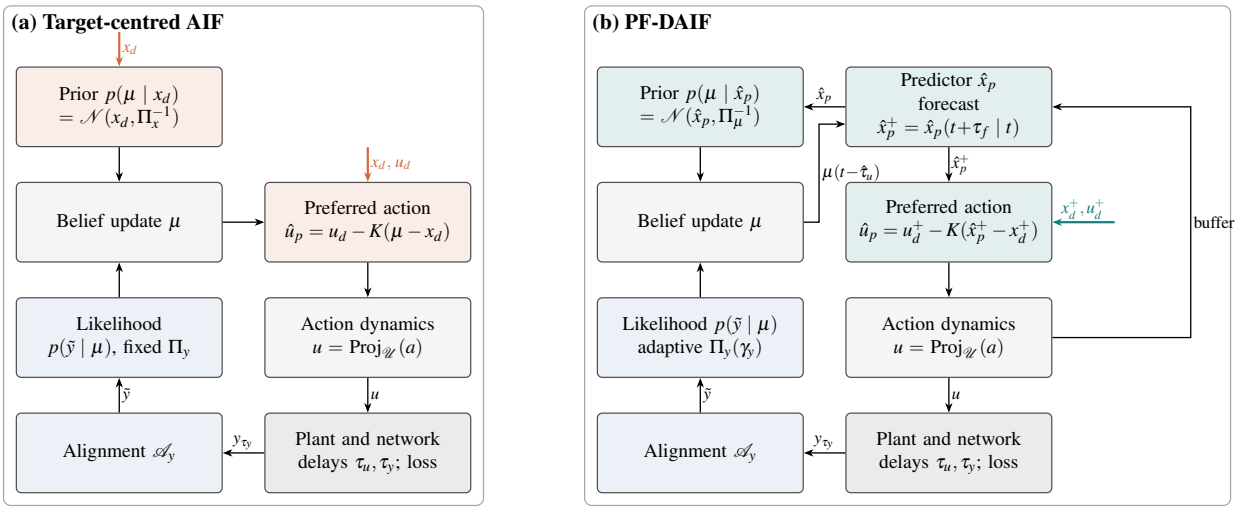


Figure 1: Feedback architectures of (a) target-centred AIF and (b) PF-DAIF, with $x_d^+ = x_d(t+\tau_f)$ and $u_d^+ = u_d(t+\tau_f)$.

4.1. Prediction of the current state

For the linear case with constant actuation delay, the current state is reconstructed from the delayed belief and the history of applied inputs,

$$\hat{x}_p(t) = e^{A\tau_u}\mu(t-\tau_u) + \int_{t-\tau_u}^{t} e^{A(t-s)} B\, u(s-\tau_u)\, ds. \tag{11}$$

Eq. (11) is the predictor of [7] evaluated at a different horizon. The Artstein reduction forecasts $x(t+\tau_u)$ and thereby removes the delay from the nominal loop. Eq. (11) instead reconstructs $x(t)$, because the prior must refer to the instant of the aligned evidence $\tilde{y}(t)$ with which it is fused. The actuation delay is compensated on the action side in Section 4.2. For time-varying delay, the predictor uses $\hat{\tau}_u(t)$, and the mismatch $\tilde{\tau}_u$ enters e_p. For the nonlinear plant, the predictor integrates

$$\dot{\hat{x}}_p(s) = \hat{f}(\hat{x}_p(s), u(s-\hat{\tau}_u), 0)\,, \qquad s \in [t-\hat{\tau}_u, t],$$

from $\mu(t-\hat{\tau}_u)$. Model mismatch, delay mismatch and the error of the initializing belief all enter e_p.

With $\hat{N}_k = \lfloor \hat{\tau}_{u,k}/T_s \rfloor$ and zero-order hold, the sampled predictor is

$$\begin{aligned} \hat{x}_{p,k} &= \Phi^{\hat{N}_k}\mu_{k-\hat{N}_k} + \sum_{j=1}^{\hat{N}_k} \Phi^{j-1}\Xi\, u_{k-\hat{N}_k-j}, \\ \Phi &= e^{AT_s}, \\ \Xi &= \int_0^{T_s} e^{As} B\, ds. \end{aligned} \tag{12}$$

It uses the inputs $u_{k-2\hat{N}_k}, \ldots, u_{k-\hat{N}_k-1}$, which have reached the plant. The inputs $u_{k-\hat{N}_k}, \ldots, u_{k-1}$ are still in flight, so the input buffer holds $2N_\tau$ entries with $N_\tau = \lceil \bar{\tau}_u/T_s \rceil$. A fractional delay is handled by evaluating Eq. (12) at the two neighbouring integers and interpolating.

4.2. Factorized generative model and free energy

With linear sensing, the three Gaussian factors of Eq. (10) are

$$\begin{aligned} p(\mu \mid \hat{x}_p) &= \mathcal{N}\big(\mu; \hat{x}_p, \Pi_\mu^{-1}\big), \\ p(\tilde{y} \mid \mu) &= \mathcal{N}\big(\tilde{y}; C\mu, \Pi_y^{-1}\big), \\ p(a \mid \hat{x}_p, x_d) &= \mathcal{N}\big(a; \hat{u}_p, \Pi_a^{-1}\big), \end{aligned}$$

and $h(\mu)$ replaces $C\mu$ for nonlinear sensing. The preferred action $\hat{u}_p$ is evaluated at a forecast of the predictor over a horizon $\tau_f \in \{0, \hat{\tau}_u\}$,

$$\begin{aligned} \hat{x}_p^+(t) &= \hat{x}_p(t + \tau_f \mid t), \\ \hat{u}_p(t) &= u_d(t + \tau_f) - K\big(\hat{x}_p^+(t) - x_d(t + \tau_f)\big), \end{aligned} \tag{13}$$

where u_d is a reference feedforward and K a state feedback gain. The forecast propagates $\hat{x}_p(t)$ through the nominal model with the commands already issued; for $\tau_f \le \hat{\tau}_u$, these commands determine the input over the whole forecast interval. With $\tau_f = \hat{\tau}_u$, the action is formed for its arrival time at the plant, up to one sampling interval; with $\tau_f = 0$, it is formed at the current state. Any locally Lipschitz map from $(\hat{x}_p^+, x_d)$ to $\hat{u}_p$ is admissible, including integral and derivative terms. With the horizon $N_f \in \{0, \hat{N}_k\}$ in samples, the sampled forecast is

$$\hat{x}_{p,k}^+ = \Phi^{N_f} \hat{x}_{p,k} + \sum_{j=1}^{N_f} \Phi^{j-1} \Xi \, u_{k-\hat{N}_k+N_f-j}. \tag{14}$$

The horizon is selected with the spectral test of Section 5.4: $\tau_f = 0$ if the spectral radius stays below one over the delay range of the application, and $\tau_f = \hat{\tau}_u$ otherwise. Since $\hat{x}_p$ is anchored at the belief $\hat{\tau}_u$ earlier, evidence reaches the action with this lag, and $\hat{u}_p$ is free of the current belief.

Only Π_y adapts, so the log-determinants of Π_μ and Π_a are constants and are discarded. The Laplace free energy is

$$\begin{aligned} F &= \frac{1}{2} e_\mu^\top \Pi_\mu e_\mu + \frac{1}{2} e_y^\top \Pi_y e_y + \frac{1}{2} e_a^\top \Pi_a e_a - \frac{1}{2} \ln |\Pi_y|, \\ e_\mu &= \mu - \hat{x}_p, \\ e_y &= \tilde{y} - C\mu, \\ e_a &= a - \hat{u}_p. \end{aligned} \tag{15}$$

The reference enters e_a through $\hat{u}_p$ alone and is absent from e_μ and e_y. Up to a constant, F is the negative log joint density of the three factors, so its minimizer in μ is the Gaussian fusion $\mu_{\mathrm{PFD}}^\star$ of Theorem 3.1. On the admissible precision set, $F \ge -\frac{1}{2} n_y \bar{\gamma}$.

4.3. Belief, action and precision dynamics

Algorithm 1 PF-DAIF sampled-data loop.

Require: nominal model $(\hat{f}, h)$; horizon τ_f; precisions Π_μ, Π_a; interval $[\underline{\gamma}, \bar{\gamma}]$; gains η_μ, η_a, Γ_γ, k_γ, K; sampling time T_s; box $\mathscr{U}$

Ensure: belief μ_k, applied input u_k

1: Initialize μ_0, a_0, $\gamma_{y,0}$, $\zeta_{\gamma,0}$ and set the input buffer of length $2N_\tau$ to zero.
2: **for** $k = 0, 1, \ldots$ **do**
3: Update $\hat{\tau}_{u,k}$, $\hat{\tau}_{y,k}$ and compute $\hat{x}_{p,k}$ by Eq. (12).
4: **if** a packet arrived **then**
5: Form $\tilde{y}_k$ by Eq. (3) and $e_{y,k} = \tilde{y}_k - h(\mu_k)$.
6: **else**
7: Omit the sensory term and hold $(\gamma_{y,k}, \zeta_{\gamma,k})$.
8: **end if**
9: Compute $\hat{x}^+_{p,k}$ by Eq. (14) and $\hat{u}_{p,k}$ by Eq. (13).
10: $\mu_{k+1} = \mu_k - \eta_\mu P_{\mu,k} \nabla_\mu F_k$, with $P_{\mu,k} = (\Pi_\mu + C^\top \Pi_{y,k} C)^{-1}$, or Π_μ^{-1} on a lost packet.
11: On a received packet, $\gamma_{y,k+1} = \mathrm{Proj}_{[\underline{\gamma}, \bar{\gamma}]}(\gamma_{y,k} + T_s \zeta_{\gamma,k})$ and $\zeta_{\gamma,k+1} = \zeta_{\gamma,k} - T_s(\Gamma_\gamma \nabla_{\gamma_y} F_k + k_\gamma \zeta_{\gamma,k})$, with outward components set to zero at the bounds.
12: $a_{k+1} = a_k - \eta_a \nabla_a F_k$, $u_{k+1} = \mathrm{Proj}_{\mathscr{U}}(a_{k+1})$; push u_{k+1} into the buffer.
13: **end for**

Belief and action follow the gradient flows $\dot{\mu} = -\Gamma_\mu \nabla_\mu F$ and $\dot{a} = -\Gamma_a \nabla_a F$ with $\Gamma_\mu, \Gamma_a \succ 0$, that is,

$$\dot{\mu} = -\Gamma_\mu \Big[\Pi_\mu (\mu - \hat{x}_p) - C^\top \Pi_y (\tilde{y} - C\mu) \Big], \tag{16}$$

$$\dot{a} = -\Gamma_a \Pi_a (a - \hat{u}_p), \tag{17}$$

with $H(\mu)^\top = (\partial h / \partial \mu)^\top$ in place of $C^\top$ for nonlinear sensing. Since $\hat{u}_p$ does not depend on μ, $\nabla_\mu F$ contains no term in x_d. Eq. (17) is a first-order filter that relaxes a toward $\hat{u}_p$.

Let $\gamma_y = \ln \mathrm{diag}(\Pi_y)$ element-wise. In the interior of $[\underline{\gamma}, \bar{\gamma}]$, the precision follows the damped gradient dynamics

$$\begin{aligned} \dot{\gamma}_y &= \zeta_\gamma, \\ \dot{\zeta}_\gamma &= -\Gamma_\gamma \nabla_{\gamma_y} F - k_\gamma \zeta_\gamma, \\ \frac{\partial F}{\partial \gamma_{y,i}} &= \frac{1}{2} e^{\gamma_{y,i}} e_{y,i}^2 - \frac{1}{2}, \end{aligned} \tag{18}$$

with $k_\gamma > 0$ and $\Gamma_\gamma \succ 0$ diagonal, and ζ_γ is projected onto the tangent cone of the interval at its bounds. A channel loses confidence while $|e_{y,i}| > \Pi_{y,ii}^{-1/2}$ and regains it below this threshold. The projection enforces Assumption 2.3 and excludes vanishing and unbounded precision.

Remark 4.1 (Precision damping). For fixed $e_{y,i} \neq 0$ and $\Gamma_\gamma = \Gamma I$, the unconstrained equilibrium is $\Pi^\star_{y,ii} = e_{y,i}^{-2}$, and the linearized precision mode is critically damped at $k_\gamma = \sqrt{2\Gamma}$. If $\ln \Pi^\star_{y,ii}$ lies outside $[\underline{\gamma}, \bar{\gamma}]$, the projection places the equilibrium at the nearest bound.

The applied input is $u = \mathrm{Proj}_{\mathscr{U}}(a)$ with $\mathscr{U}$ a closed convex box. A missing packet removes the sensory term and holds (γ_y, ζ_γ), so the belief evolves by prediction alone and packet loss needs no separate mode.

The belief step is preconditioned by the curvature $P_{\mu,k}$, so $\mu_{k+1} = \mu_k + \eta_\mu(\mu_k^\star - \mu_k)$ relaxes toward the instantaneous minimizer $\mu_k^\star$ with factor $\eta_\mu \in (0,1]$; on a lost packet, $\mu_k^\star = \hat{x}_{p,k}$. Each iteration requires one prediction, one alignment, one update of belief, action and precision, and one projection, without online optimization. The cost per sample is $O(n^3 + n_y n + n_u^2)$ for the updates and $O(N_\tau n n_u)$ for the buffer propagation. Algorithm 1 summarizes the sampled implementation.

5. Closed-loop analysis

This section establishes the closed-loop properties of PF-DAIF, proceeding from the inference dynamics to the physical loop. Section 5.1 bounds the predictor and projection residuals, Section 5.2 establishes contraction of the belief and admissibility of the precision, Section 5.3 establishes dissipation of the variational dynamics, and Section 5.4 decides delay-dependent stability of the sampled loop. Lemma 5.4 and Theorem 5.5 use only the projected precision dynamics and the constant precisions Π_μ, Π_a of Assumption 2.3, the residual bounds use Assumptions 2.2 and 2.4, and the delay results use the linear case with Assumption 2.1. Theorem 5.6 is necessary and sufficient for the linear, unsaturated, fixed-delay loop with frozen precision, whereas Theorem 5.13 and the cascade bounds of Section 6 are local sufficient conditions.

5.1. Residual bounds

The estimation error splits as $\mu - x = e_\mu + e_p$. Theorem 5.3 bounds e_μ, and Lemma 5.1 bounds e_p. A shift of the delay is assumed to perturb the nominal field by at most $\kappa_\tau|\tilde{\tau}_u|$, that is, $\left\|\hat{f}(z,u(s-\tau_u),0) - \hat{f}(z,u(s-\hat{\tau}_u),0)\right\| \le \kappa_\tau|\tilde{\tau}_u(s)|$ on $\mathscr{X}_c$.

Lemma 5.1 (Predictor error)**.** *Under Assumptions 2.1, 2.2 and 2.4, the predictor initialized at $\mu(t-\hat{\tau}_u)$ satisfies, on $\mathscr{X}_c$,*

$$\left\|e_p(t)\right\| \le e^{L_x\bar{\tau}_u}\left\|\mu(t-\hat{\tau}_u) - x(t-\hat{\tau}_u)\right\| + \frac{e^{L_x\bar{\tau}_u}-1}{L_x}\Big(\Delta_f + L_d\bar{d} + \bar{w} + \kappa_\tau \sup_{s\in[t-\bar{\tau}_u,t]}|\tilde{\tau}_u(s)|\Big). \quad (19)$$

Proof. See Appendix A.2 for details.

Eq. (19) trades delay against model quality. Let $\bar{e}_{\mathrm{est}}$ bound the delayed estimation error and $\bar{q}_p = \Delta_f + L_d\bar{d} + \bar{w} + \kappa_\tau\bar{\tilde{\tau}}_u > 0$. A prediction budget $\bar{e}_p^{\max} > \bar{e}_{\mathrm{est}}$ is met whenever

$$\bar{\tau}_u \le \frac{1}{L_x}\ln\left(\frac{\bar{e}_p^{\max} + \bar{q}_p/L_x}{\bar{e}_{\mathrm{est}} + \bar{q}_p/L_x}\right). \quad (20)$$

Model mismatch, unmodelled forces and delay errors therefore shrink the admissible delay.

The action error $e_a = a - \hat{u}_p$ obeys a linear estimate. Let $\frac{1}{2}(\Gamma_a\Pi_a + \Pi_a\Gamma_a^\top) \succeq \lambda_a I$ with $\lambda_a > 0$ and $\left\|\dot{\hat{u}}_p\right\| \le \bar{\upsilon}_a$. Then Eq. (17) gives

$$\|e_a(t)\| \le e^{-\lambda_a t}\|e_a(0)\| + \bar{\upsilon}_a/\lambda_a. \quad (21)$$

The sampled step of Algorithm 1 corresponds to $\Gamma_a = (\eta_a/T_s)I$, for which $\lambda_a = \eta_a\underline{\pi}_a/T_s$.

Lemma 5.2 (Projection residual)**.** *Let $\mathscr{U}$ be closed and convex and $u = \mathrm{Proj}_{\mathscr{U}}(a)$. Then $\|u-v\| \le \|a-v\|$ for every $v \in \mathscr{U}$ and $\|\Delta_a\| = \mathrm{dist}(a,\mathscr{U})$. If $\|a-\hat{u}_p\| \le \varepsilon_a$, then $\|\Delta_a\| \le \mathrm{dist}(\hat{u}_p,\mathscr{U}) + \varepsilon_a$, and $\|u-\hat{u}_p\| \le \varepsilon_a$ when $\hat{u}_p \in \mathscr{U}$.*

Proof. The projection onto a closed convex set is nonexpansive and fixes the set, and the triangle inequality gives the remaining bounds.

By Eq. (21), ε_a may be taken as $\|e_a(0)\| + \bar{v}_a/\lambda_a$.

5.2. Contraction of the belief and bounded precision

Theorem 5.3 (Contraction toward the predictor)**.** *Let $\mathscr{X}_c$ be convex and define the averaged Jacobian $\bar{H}(t) = \int_0^1 H\big(\hat{x}_p(t) + s\,e_\mu(t)\big)\,ds$, so that $h(\mu) - h(\hat{x}_p) = \bar{H} e_\mu$. Suppose that, on all of $\mathscr{X}_c$,*

$$\Gamma_\mu(\Pi_\mu + H(\mu)^\top \Pi_y \bar{H}) + (\cdot)^\top \succeq 2\alpha_\mu I, \qquad \alpha_\mu > 0. \tag{22}$$

Then

$$\begin{aligned} \big\|e_\mu(t)\big\| &\le e^{-\alpha_\mu t}\big\|e_\mu(0)\big\| + \frac{1}{\alpha_\mu}\sup_{0\le s\le t}\big\|\delta_p(s)\big\|, \\ \delta_p &= \Gamma_\mu H(\mu)^\top \Pi_y\big(\tilde{y} - h(\hat{x}_p)\big) - \dot{\hat{x}}_p. \end{aligned} \tag{23}$$

Proof. See Appendix A.3 for details.

For linear sensing, $\bar{H} = H = C$, and Eq. (22) holds with $\alpha_\mu = g_\mu \underline{\pi}_\mu$ for $\Gamma_\mu = g_\mu I$, $g_\mu > 0$. The preconditioned step of Algorithm 1 corresponds to $\Gamma_\mu = (\eta_\mu/T_s)(\Pi_\mu + C^\top \Pi_y C)^{-1}$, for which $\alpha_\mu = \eta_\mu/T_s$. For nonlinear sensing with $\Gamma_\mu = g_\mu I$, let $\bar{h}_1$ bound $\|H\|$, ℓ_H be the Lipschitz constant of H and D_c the diameter of $\mathscr{X}_c$. Since $\|\bar{H} - H(\mu)\| \le \frac{1}{2}\ell_H\big\|e_\mu\big\|$, Eq. (22) holds with $\alpha_\mu = g_\mu(\underline{\pi}_\mu - \frac{1}{2}\bar{h}_1\bar{\pi}_y\ell_H D_c)$ whenever this quantity is positive. A sufficiently stiff prior therefore contracts for any sensing map with Lipschitz Jacobian, whereas a weak prior requires a nearly affine one.

The forcing δ_p does not vanish at the stationary belief. For linear sensing and a frozen predictor, its first term equals $\Gamma_\mu(\Pi_\mu + C^\top \Pi_y C)(\mu^\star_{\mathrm{PFD}} - \hat{x}_p)$, so the belief converges to the fused point $\mu^\star_{\mathrm{PFD}}$ of Theorem 3.1, between the predictor and the aligned measurement.

Lemma 5.4 (Bounded adaptive precision)**.** *Under the projected dynamics of Eq.* (18)*, $\gamma_y \in [\underline{\gamma}, \bar{\gamma}]$ for all time, and ζ_γ is bounded whenever e_y is bounded. Hence Π_y satisfies Assumption 2.3 for all time.*

Proof. The projection keeps γ_y in the interval, and $k_\gamma > 0$ makes the ζ_γ dynamics input-to-state stable with respect to the bounded forcing $\Gamma_\gamma \nabla_{\gamma_y} F$.

5.3. Dissipation of the variational dynamics

For fixed predictor and precision, the belief subproblem is strongly convex with minimizer

$$\begin{aligned} \mu^\star &= (\Pi_\mu + C^\top \Pi_y C)^{-1}(\Pi_\mu \hat{x}_p + C^\top \Pi_y \tilde{y}) \\ &= \mu^\star_{\mathrm{PFD}}, \end{aligned} \tag{24}$$

and the action subproblem has minimizer $a^\star = \hat{u}_p$. Second-order precision adaptation requires a storage function that contains the precision rate.

Theorem 5.5 (Dissipation)**.** *Along Eqs.* (16)*,* (17) *and* (18)*, in the interior of the precision interval, the function* $\mathscr{V}_F = F + \frac{1}{2}\zeta_\gamma^\top \Gamma_\gamma^{-1}\zeta_\gamma$ *satisfies*

$$\begin{aligned} \dot{\mathscr{V}}_F &= -\left\|\nabla_\mu F\right\|_{\Gamma_\mu}^2 - \|\nabla_a F\|_{\Gamma_a}^2 - k_\gamma \left\|\zeta_\gamma\right\|_{\Gamma_\gamma^{-1}}^2 + \partial_t F, \\ \partial_t F &= -e_\mu^\top \Pi_\mu \dot{\hat{x}}_p + e_y^\top \Pi_y \dot{\tilde{y}} - e_a^\top \Pi_a \dot{\hat{u}}_p. \end{aligned} \tag{25}$$

Proof. See Appendix A.4 for details.

Since F is bounded below on the admissible precision set, $\mathscr{V}_F$ is a valid storage function. For constant $\hat{x}_p$, $\tilde{y}$ and $\hat{u}_p$ it is non-increasing, as required in Section 2.3. Along the closed loop, the free energy is raised only by the motion of the predictor, the aligned evidence and the preferred action, each weighted by its prediction error. In the sampled implementation at frozen precision, the preconditioned belief step gives $\mu_{k+1} - \mu^\star = (1-\eta_\mu)(\mu_k - \mu^\star)$. This step decreases the quadratic objective for every $\eta_\mu \in (0,2)$ and approaches $\mu^\star$ without overshoot for the implemented range $\eta_\mu \in (0,1]$; the action step decreases it for $\eta_a < 2/\bar{\pi}_a$.

5.4. Delay-dependent stability of the sampled loop

In the linear case with frozen precision and inactive projection, the loop of Algorithm 1 is a finite-dimensional linear difference equation. Its delay-dependent stability is therefore decided exactly by a spectral radius, as in semi-discretization methods for delay systems [35], without a Lyapunov–Krasovskii functional.

5.4.1. Sampled closed loop

Take $C = I$, $\mathscr{A}_y$ equal to the identity, as in both numerical studies, and a constant delay $\tau_u = NT_s$ with integer $N \geq 0$. With zero-order hold,

$$\begin{aligned} x_{k+1} &= \Phi x_k + \Xi u_{k-N} + \Xi_d d_k + w_k, \\ \Xi_d &= \int_0^{T_s} e^{As} B_d \, ds, \\ \tilde{y}_k &= x_k + \delta_{y,k}. \end{aligned} \tag{26}$$

With Π_y frozen, the controller of Algorithm 1 reads

$$\begin{aligned} \mu_{k+1} &= (1-\eta_\mu)\mu_k + \eta_\mu W(\Pi_\mu \hat{x}_{p,k} + \Pi_y \tilde{y}_k), \\ a_{k+1} &= (I - \Lambda_a)a_k + \Lambda_a \hat{u}_{p,k}, \\ u_k &= a_k, \end{aligned} \tag{27}$$

with $W = (\Pi_\mu + \Pi_y)^{-1}$, $\Lambda_a = \eta_a \Pi_a$ and $\hat{u}_{p,k} = r_k - K\hat{x}_{p,k}^+$, where $r_k = u_{d,k+N_f} + K x_{d,k+N_f}$. Define the input buffer $U_k = \mathrm{col}(u_{k-1}, \ldots, u_{k-2N})$, the belief history $\mathscr{H}_k = \mathrm{col}(\mu_k, \ldots, \mu_{k-N})$, the augmented state and the exogenous input

$$\begin{aligned} \xi_k &= \mathrm{col}(x_k, U_k, \mathscr{H}_k, a_k) \in \mathbb{R}^{n_\xi}, \\ \omega_k &= \mathrm{col}(d_k, w_k, \delta_{y,k}, r_k), \end{aligned} \tag{28}$$

with $n_\xi = (N+2)n + (2N+1)n_u$. Let S_j and E_j select $u_{k-j} = S_j U_k$ and $\mu_{k-j} = E_j \mathscr{H}_k$. The predictor of Eq. (12) and the forecast of Eq. (14) are linear in ξ_k,

$$\begin{aligned} \hat{x}_{p,k} &= \Upsilon_\mu \mathscr{H}_k + \Upsilon_u U_k, \\ \hat{x}^+_{p,k} &= \Upsilon^+_\mu \mathscr{H}_k + \Upsilon^+_u U_k, \end{aligned} \tag{29}$$

with $\Upsilon_\mu = \Phi^N E_N$ and $\Upsilon_u = \sum_{j=1}^N \Phi^{j-1}\Xi S_{N+j}$. For $\tau_f = \hat{\tau}_u$, $\Upsilon^+_\mu = \Phi^N \Upsilon_\mu$ and $\Upsilon^+_u = \Phi^N \Upsilon_u + \sum_{j=1}^N \Phi^{j-1}\Xi S_j$; for $\tau_f = 0$, $\Upsilon^+_\mu = \Upsilon_\mu$ and $\Upsilon^+_u = \Upsilon_u$. Let Z_U and $Z_{\mathscr{H}}$ shift every block down by one place and discard the last, and let D_U and $D_{\mathscr{H}}$ insert into the top block. The closed loop is

$$\xi_{k+1} = \mathbb{A}_N \xi_k + \mathbb{G}_N \omega_k, \tag{30}$$

with

$$\begin{aligned} \mathbb{A}_N &= \begin{bmatrix} \Phi & \Xi S_N & 0 & 0 \\ 0 & Z_U & 0 & D_U \\ D_{\mathscr{H}} B_x & D_{\mathscr{H}} B_U & Z_{\mathscr{H}} + D_{\mathscr{H}} B_{\mathscr{H}} & 0 \\ 0 & -\Lambda_a K \Upsilon^+_u & -\Lambda_a K \Upsilon^+_\mu & I - \Lambda_a \end{bmatrix}, \\ \mathbb{G}_N &= \begin{bmatrix} \Xi_d & I & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & D_{\mathscr{H}} B_x & 0 \\ 0 & 0 & 0 & \Lambda_a \end{bmatrix}, \end{aligned} \tag{31}$$

where $B_{\mathscr{H}} = (1-\eta_\mu)E_0 + \eta_\mu W \Pi_\mu \Upsilon_\mu$, $B_U = \eta_\mu W \Pi_\mu \Upsilon_u$ and $B_x = \eta_\mu W \Pi_y$. For $N = 0$, the buffer is empty, $\mathscr{H}_k = \mu_k = \hat{x}_{p,k}$, the plant row reads $x_{k+1} = \Phi x_k + \Xi a_k + \Xi_d d_k + w_k$, and $\xi_k = \mathrm{col}(x_k, \mu_k, a_k)$. Integral and derivative terms in $\hat{u}_p$ add one state block each.

The baselines of Section 7 take the same form. Target-centred AIF has $B_{\mathscr{H}} = (1-\eta_\mu)E_0$, $B_U = 0$, $B_x = \eta_\mu(\Pi_x + \Pi_y)^{-1}\Pi_y$, an additional input column $\eta_\mu(\Pi_x + \Pi_y)^{-1}\Pi_x$ on $x_{d,k}$, and $\hat{u}_p$ evaluated at μ_k. The Smith-predictor PID controller has the observer $\mu_{k+1} = \mu_k + \kappa_o(\tilde{y}_k - \mu_k)$, forecasts μ_k over N steps with its internal model and applies its law without action filter.

5.4.2. Fixed delay

Theorem 5.6 (Exact delay-dependent stability)**.** *For the linear plant with $C = I$, frozen admissible Π_y, inactive projection and constant delay $\tau_u = NT_s$, the residual-free loop ($\omega \equiv 0$) of Eq.* (30) *is exponentially stable if and only if $\rho(\mathbb{A}_N) < 1$. In that case, let $P \succ 0$ solve $\mathbb{A}_N^\top P \mathbb{A}_N - P = -Q$ for some $Q \succ 0$ and set $c_0 = \lambda_{\min}(Q)/\lambda_{\max}(P)$. If $\sup_k \|\omega_k\| \le \bar{\omega}$, then*

$$\limsup_{k\to\infty} \|\xi_k\|^2 \le \frac{2(2-c_0)}{c_0^2} \frac{\lambda_{\max}(\mathbb{G}_N^\top P \mathbb{G}_N)}{\lambda_{\min}(P)} \bar{\omega}^2. \tag{32}$$

Proof. See Appendix A.5 for details.

The tracking error $e_k = x_k - x_{d,k}$ is a linear function of ξ_k and the bounded reference, so Eq. (32) is the sampled form of Eq. (4) without e_p and Δ_a. When the plant row of $\mathbb{A}_N$ carries the true plant and the predictor rows carry the nominal model, the model residual is contained in $\mathbb{A}_N$. Delay mismatch and active projection enter as perturbations bounded by Lemmas 5.1 and 5.2.

Lemma 5.7 (Delay margin and critical mode)**.** *On a family of integer delays, the certified stable grid points are* $\{NT_s : \rho(\mathbb{A}_N) < 1\}$*. Let* N_* *be the first tested integer with* $\rho(\mathbb{A}_{N_*}) \geq 1$*; the loop at* N_* *is not exponentially stable. A dominant eigenvalue* λ_c *of* $\mathbb{A}_{N_*}$ *with* $|\lambda_c| \geq 1$ *and* $\arg\lambda_c \notin \{0, \pi\}$ *is a non-decaying oscillatory sampled mode, and* $\arg\lambda_c/T_s$ *is its angular frequency up to aliasing. Since the dimension of* $\mathbb{A}_N$ *changes with N, the test classifies each integer delay separately.*

Proof. See Appendix A.5 for details.

5.4.3. Switching delay and packet loss

Random transmission delays and packet dropouts are naturally analysed in the mean-square sense [36].

Definition 5.8 (Exponential mean-square stability)**.** A residual-free sampled random linear system is exponentially mean-square stable if there exist $\kappa_0 \geq 1$ and $\lambda_0 \in (0,1)$ with $\mathbb{E}\,\|\xi_k\|^2 \leq \kappa_0\lambda_0^k\mathbb{E}\,\|\xi_0\|^2$ for every square-integrable initial state and every $k \geq 0$.

Theorem 5.9 (Mean-square stability)**.** *Let the delay be* $N_k \in \{N_1, \ldots, N_{\bar{\ell}}\}$ *with* $\Pr(N_k = N_\ell) = p_\ell$*, and let a packet be lost with probability* $p_{\mathrm{loss},\ell}$ *in mode* ℓ*, independently across steps and of the state. Embed every mode in the state of* $\bar{N} = \max_\ell N_\ell$*, with the shifts acting on the full buffer and history and* S_{N_ℓ}*,* E_{N_ℓ} *selecting within it. Let* $\mathbb{A}_\ell^{\mathrm{rx}} = \mathbb{A}_{N_\ell}$*, and let* $\mathbb{A}_\ell^{\mathrm{loss}}$ *be* $\mathbb{A}_{N_\ell}$ *with* $\Pi_y = 0$ *in* $B_{\mathscr{H}}$*,* B_U *and* B_x*. If there exist* $P \succ 0$ *and* $\alpha > 0$ *with*

$$\sum_{\ell=1}^{\bar{\ell}} p_\ell \Big[(1 - p_{\mathrm{loss},\ell})\,(\mathbb{A}_\ell^{\mathrm{rx}})^\top P \mathbb{A}_\ell^{\mathrm{rx}} + p_{\mathrm{loss},\ell}\,(\mathbb{A}_\ell^{\mathrm{loss}})^\top P \mathbb{A}_\ell^{\mathrm{loss}}\Big] - P \preceq -\alpha I, \tag{33}$$

then the residual-free loop is exponentially mean-square stable, and $\limsup_k \mathbb{E}\,\|\xi_k\|^2 \leq c_\omega \bar{\omega}_2^2$ *for a finite* c_ω *whenever* $\sup_k \mathbb{E}\,\|\omega_k\|^2 \leq \bar{\omega}_2^2$*.*

Proof. See Appendix A.6 for details.

Collect the $n_{\mathrm{m}} = 2\bar{\ell}$ combinations of delay mode and packet status into mode matrices $\mathbb{A}_j$ with probabilities p_j, namely $p_\ell(1 - p_{\mathrm{loss},\ell})$ for $\mathbb{A}_\ell^{\mathrm{rx}}$ and $p_\ell p_{\mathrm{loss},\ell}$ for $\mathbb{A}_\ell^{\mathrm{loss}}$. The second-moment map and its adjoint are

$$\begin{aligned} \mathscr{T}(X) &= \sum_{j=1}^{n_{\mathrm{m}}} p_j \mathbb{A}_j X \mathbb{A}_j^\top, \\ \mathscr{T}^*(Y) &= \sum_{j=1}^{n_{\mathrm{m}}} p_j \mathbb{A}_j^\top Y \mathbb{A}_j, \end{aligned} \tag{34}$$

so that $X_k = \mathbb{E}[\xi_k\xi_k^\top]$ obeys $X_{k+1} = \mathscr{T}(X_k)$ in the residual-free loop, and Eq. (33) reads $\mathscr{T}^*(P) - P \preceq -\alpha I$. Since $\mathscr{T}$ is a positive map, Eq. (33) is feasible if and only if $\rho(\mathscr{T}) < 1$ [37], which is an exact test for independent switching and Bernoulli loss. In matrix form, $\rho(\mathscr{T}) = \rho(\mathscr{M})$ with $\mathscr{M} = \sum_j p_j \mathbb{A}_j \otimes \mathbb{A}_j$.

Mean-square stability bounds the asymptotic decay of a weighted quadratic energy and does not require the Euclidean norm to contract at each step. The short-lived peaks of the reduced loop in Section 7.4 follow from this property.

Lemma 5.10 (Finite-time mean-square envelope)**.** *Consider the residual-free loop of Theorem 5.9. Suppose that $P \succ 0$ and $Q \succ 0$ satisfy $\mathscr{T}^*(P) - P = -Q$, and let $c_Q = \lambda_{\min}(Q)/\lambda_{\max}(P)$. Then $0 < c_Q \leq 1$ and, for every deterministic initial state ξ_0,*

$$\mathbb{E}\,\|\xi_k\|^2 \leq \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}(1-c_Q)^k\,\|\xi_0\|^2\,. \tag{35}$$

In particular, $\rho(\mathscr{T}) < 1$ ensures asymptotic mean-square decay, while the condition number of P admits transient amplification of $\mathbb{E}\,\|\xi_k\|^2$ relative to $\|\xi_0\|^2$.

Proof. With $V_k = \xi_k^\top P \xi_k$, independence of the mode gives $\mathbb{E}[V_{k+1} \mid \xi_k] = V_k - \xi_k^\top Q \xi_k \leq (1-c_Q)V_k$. Iteration and the eigenvalue bounds of P give Eq. (35), and $Q \preceq P$ gives $c_Q \leq 1$.

Section 8.4 evaluates the envelope for the reduced quadrotor loop. A mode matrix with unit spectral radius, such as a lost-packet mode, is compatible with mean-square decay of the independent mixture. The exact worst-case amplification over initial directions follows from the adjoint map.

Lemma 5.11 (Exact finite-horizon mean-square amplification)**.** *Consider the residual-free loop of Theorem 5.9, let $Y_0 = I$ and $Y_{k+1} = \mathscr{T}^*(Y_k)$. For every deterministic initial state ξ_0,*

$$\begin{aligned} \mathbb{E}\,\|\xi_k\|^2 &= \xi_0^\top Y_k \xi_0, \\ g_k &:= \sup_{\|\xi_0\|=1} \mathbb{E}\,\|\xi_k\|^2 \\ &= \lambda_{\max}(Y_k). \end{aligned} \tag{36}$$

Consequently, over a finite window $0 \leq k \leq \bar{k}$,

$$\begin{aligned} G_{\bar{k}} &= \max_{0\leq k\leq \bar{k}} g_k \\ &= \max_{0\leq k\leq \bar{k}} \lambda_{\max}\left[(\mathscr{T}^*)^k(I)\right] \end{aligned} \tag{37}$$

is the worst-case amplification of the squared norm over initial directions, and $\mathbb{E}\,\|\xi_k\|^2 \leq g_k\,\|\xi_0\|^2$ for every prescribed ξ_0.

Proof. The adjoint relation gives $\mathbb{E}\,\|\xi_k\|^2 = \mathrm{tr}[(\mathscr{T}^*)^k(I)\,\xi_0\xi_0^\top] = \xi_0^\top Y_k \xi_0$, and maximizing the Rayleigh quotient gives g_k.

Hence $\rho(\mathscr{T}) < 1$ and $G_{\bar{k}} > 1$ are compatible: non-normal modes and mixing between state directions amplify the transient while the second moment still decays asymptotically.

5.4.4. Adaptive precision

Theorem 5.12 (Common Lyapunov function over the precision box)**.** *Let Π_μ be diagonal and $\Theta = W\Pi_\mu = \mathrm{diag}(\theta_i)$ as in Lemma 3.2. Then $B_x = \eta_\mu(I-\Theta)$, $B_{\mathscr{H}} = (1-\eta_\mu)E_0 + \eta_\mu\Theta\Upsilon_\mu$ and $B_U = \eta_\mu\Theta\Upsilon_u$, so $\mathbb{A}_N$ is affine in Θ. Let $\Theta^{(v)}$, $v = 1,\ldots,2^n$, be the vertices of the box $\theta_i \in [\pi_{\mu,i}/(\pi_{\mu,i}+e^{\bar{\gamma}}), \bar{\theta}_i]$, with $\bar{\theta}_i = \pi_{\mu,i}/(\pi_{\mu,i}+e^{\underline{\gamma}})$, or $\bar{\theta}_i = 1$ if packets may be lost. If some $P \succ 0$ satisfies $\mathbb{A}_N(\Theta^{(v)})^\top P \mathbb{A}_N(\Theta^{(v)}) - P \prec 0$ at every vertex, then $V = \xi^\top P \xi$ is a common Lyapunov function for every precision trajectory in the box.*

Proof. The blocks are affine in Θ, and $\mathbb{A} \mapsto \mathbb{A}^\top P \mathbb{A} - P$ is matrix-convex, so negativity at the vertices extends to the whole box.

Lemma 5.4 keeps Π_y in the precision interval, and a lost packet corresponds to $\Theta = I$. For each fixed delay, Theorem 5.12 therefore covers the adaptive precision under arbitrary switching. With $\bar{\theta}_i = 1$, it also covers arbitrary packet loss, which requires $\rho(\Phi) < 1$ because Φ is a diagonal block of $\mathbb{A}_N(I)$; for a marginally stable plant such as the double integrator, random loss is covered by Theorem 5.9 instead.

5.4.5. Nonlinear plant

For the nonlinear plant, the delay enters a local sufficient condition. Let $u^\circ(x,t)$ be the preferred input at the true current state for $\tau_f = 0$, and write the tracking-error dynamics as $\dot{e} = F_0(e,t) + \Delta_\tau(t)$ with $F_0(e,t) = f(x,u^\circ(x,t),0) - \dot{x}_d(t)$ and

$$\Delta_\tau(t) = f\big(x(t), u(t-\tau_u(t)), d(t)\big) - f\big(x(t), u^\circ(x(t),t), 0\big) + w(t).$$

Theorem 5.13 (Local delay-dependent bound)**.** *Suppose that $\dot{e} = F_0(e,t)$ admits a Lyapunov function V_0 on $\mathscr{X}_c$ with $c_1 \|e\|^2 \le V_0 \le c_2 \|e\|^2$, $\nabla V_0^\top F_0 \le -c_3 \|e\|^2$ and $\|\nabla V_0\| \le c_4 \|e\|$, that solutions remain in $\mathscr{X}_c$, and that*

$$\|\Delta_\tau(t)\| \le L_\tau \sup_{s \in [t-\bar{\tau}_u, t]} \|e(s)\| + \bar{\omega}_{\mathrm{nl}},$$

with $\bar{\omega}_{\mathrm{nl}}$ a class-$\mathscr{K}$ function of $\bar{d} + \bar{w} + \bar{\delta}_y + \bar{e}_p + \bar{\Delta}_a + \bar{v}_d$. If, for some $q > 1$ and $q_R^2 = q c_2 / c_1$,

$$2 c_4^2 L_\tau^2 q_R^2 < c_3^2, \tag{38}$$

then the loop is locally uniformly ultimately bounded on $\mathscr{X}_c$ and satisfies Eq. (4)*; if all residuals vanish, the reference is locally asymptotically stable.*

Proof. See Appendix A.7 for details.

Forward invariance of $\mathscr{X}_c$ under nonzero disturbances follows from a strict inward-pointing Lyapunov inequality on its boundary, and a moving reference requires the nominal dissipation estimates to hold uniformly in time. Condition (38) reads

$$\begin{aligned} L_\tau &< L_{\tau,\mathrm{crit}} \\ &:= \frac{c_3}{\sqrt{2}\, c_4 q_R} \\ &= \frac{c_3}{c_4} \sqrt{\frac{c_1}{2 q c_2}}. \end{aligned} \tag{39}$$

This is a threshold on the gain from delayed state to disturbance. It becomes a bound on $\bar{\tau}_u$ through an upper bound on $L_\tau(\bar{\tau}_u)$ valid on the whole invariant set; the gain L_τ depends on the delay and on the Lipschitz constants of f, of the command history and of u°. Eq. (38) is therefore a

Razumikhin-type delay bound [34], and it yields the bound of Eq. (4). Section 6 adds the attitude and cascade bounds required by the quadrotor.

5.5. Invariance conditions for the nonlinear bound

Theorem 5.13 requires the comparison inequality on a compact sublevel set $\mathscr{E}_c = \{e : V_0(e) \le \bar{V}\}$ of the tracking error, uniformly for the errors in $\mathscr{E}_c$ and the delayed histories compatible with it. At admissible comparison times, the Razumikhin estimate reads $\dot{V}_0 \le -\alpha_e \|e\|^2 + c_5 \bar{\omega}_{\mathrm{nl}}^2$ with $\alpha_e = c_3/2 - c_4^2 L_\tau^2 q_R^2 / c_3 > 0$ and $c_5 = c_4^2/c_3$. On the boundary $V_0(e) = \bar{V}$, $\|e\|^2 \ge \bar{V}/c_2$, so

$$\bar{\omega}_{\mathrm{nl}}^2 < \frac{\alpha_e \bar{V}}{c_2 c_5} \tag{40}$$

gives an inward-pointing derivative at comparison times. Forward invariance additionally requires the initial histories to lie in $\mathscr{E}_c$ and the inequality to hold across switching of the active saturation and projection sets. For the quadrotor, the position loop receives the thrust-direction error as a disturbance and the attitude loop receives commands that depend on the requested acceleration, so the two ISS gains must be evaluated on compatible sets and satisfy the small-gain inequality of Section 6.

6. Quadrotor model and cascade analysis

This section specializes the analysis to an underactuated quadrotor. It introduces the rigid-body model and the preferred action of the translational loop, bounds the quantities required by Section 5, and closes the attitude and translation loops through a cascade argument.

6.1. Model and preferred action

Let $p \in \mathbb{R}^3$ be the inertial position, $v_I \in \mathbb{R}^3$ the inertial velocity, $R \in \mathrm{SO}(3)$ the attitude and $\Omega \in \mathbb{R}^3$ the body angular velocity. With mass m, inertia J, gravitational acceleration g, $e_3 = \mathrm{col}(0,0,1)$, aerodynamic force D_a, thrust T, torque M and process noise w_v, w_Ω,

$$\begin{aligned} \dot{p} &= v_I, \\ m\dot{v}_I &= -mge_3 + TRe_3 + D_a + w_v, \\ \dot{R} &= R\Omega^\wedge, \\ J\dot{\Omega} &= -\Omega^\wedge J\Omega + M + w_\Omega, \end{aligned} \tag{41}$$

where $z^\wedge b = z \times b$. The configuration evolves on $\mathbb{R}^3 \times \mathrm{SO}(3)$, and the physical state (p, v_I, R, Ω) has twelve coordinates with three local coordinates for the attitude. The outer loop acts on the translational state $\mathrm{col}(p, v_I) \in \mathbb{R}^6$, whereas the predictor, action filter, command buffer, adaptive precision and attitude error enlarge the closed-loop state.

The vehicle is underactuated, with four inputs for six degrees of freedom [33]. The off-board controller computes a commanded acceleration $u \in \mathbb{R}^3$, which reaches the vehicle after the actuation delay $\tau_u(t)$. The onboard loop converts the received command into thrust and torque,

$$\begin{aligned} T &= m\big(u(t-\tau_u) + ge_3\big)^\top Re_3, \\ M &= -k_R e_R - k_\Omega e_\Omega + \Omega \times J\Omega, \end{aligned} \tag{42}$$

where $e_R = \frac{1}{2}(R_d^\top R - R^\top R_d)^\vee$ and $e_\Omega = \Omega$ are the geometric errors of [38], and R_d aligns the body axis with $u(t-\tau_u) + ge_3$ at zero yaw. The onboard loop uses the measured attitude and rate at 200 Hz, and thrust and torque are clipped and slew-limited. These effects, together with the omitted angular-rate feedforward, enter the analysis through the torque residual $\bar{\delta}_M$ of Section 6.2.

Position and velocity are measured, so Eq. (2) holds with $h(x) = \mathrm{col}(p, v_I)$, and the belief is the translational state $\mu = \mathrm{col}(\hat{p}, \hat{v}_I) \in \mathbb{R}^6$ with $C = I_6$; the nominal model is a double integrator driven by the delayed command, so the attitude dynamics and D_a enter Δ_f. The latent action is the commanded acceleration, with $\mathscr{U} = [-3,3]^3$ m/s^2. The preferred action of Eq. (13) is

$$\hat{u}_p = \ddot{p}_d - K_{\mathrm{pos}}(\hat{p}_p - p_d) - K_{\mathrm{vel}}(\hat{v}_{I,p} - \dot{p}_d), \tag{43}$$

that is, $u_d = \ddot{p}_d$ and $K = [K_{\mathrm{pos}}\ K_{\mathrm{vel}}]$. For this outer loop, a double integrator per axis, the spectral radius of $\mathbb{A}_N$ with $\tau_f = 0$ crosses one at 0.425 s, above the largest delay of the nominal scenario, so the rule of Section 4.2 selects $\tau_f = 0$.

6.2. Quadrotor-specific bounds

The results of Section 5 apply to the translational loop once its Lipschitz constant, attitude loop and cascade are bounded. With reference scales for position, velocity and angular rate, $\|\Omega\| \le \bar{\Omega}$ and $0 \le T \le \bar{T}$ on $\mathscr{X}_c$, the rigid-body field of Eq. (41) is Lipschitz in the scaled norm with

$$L_x \le c_{\mathrm{sc}} \max\left\{1, \frac{\bar{T}}{m}, \bar{\Omega}, \frac{2\|J\|\bar{\Omega}}{\lambda_{\min}(J)}\right\} + L_{D_a}, \tag{44}$$

where c_{sc} depends on the scaling and L_{D_a} is the Lipschitz constant of the aerodynamic force. Lemma 5.1 and Eq. (20) apply with this constant, so a larger angular-rate bound or a stronger unmodelled force reduces the admissible delay at a fixed prediction budget.

On $\mathrm{SO}(3)$, define the attitude potential $\Psi_R(R, R_d) = \frac{1}{2}\mathrm{tr}(I - R_d^\top R)$. The implementation applies the zero-rate damping law $-k_R e_R - k_\Omega \Omega$ of Eq. (42), whereas the desired rotation constructed from delayed acceleration commands varies in time. A frozen target has the standard local Lyapunov estimate, a moving target adds forcing by the reference rate, and the diagonal-gain, saturated and slew-limited implementation enters through a residual bound. Let $\bar{e}_{\mathrm{att}}$ bound the error of the measured attitude and rate, and let $\bar{\delta}_M$ bound the realized torque mismatch, including clipping and actuator rate limits.

Theorem 6.1 (Attitude ISS for a frozen attitude target)**.** *Let $R_d \in \mathrm{SO}(3)$ be constant, $\Omega_d = 0$, and $k_R, k_\Omega > 0$ be scalar gains. Consider the ideal compensated attitude law with an additive torque and sensing residual of magnitude at most $\bar{d}_R$. On a compact, positively invariant set of (R, Ω) on which $\Psi_R \le \psi_{\max} < 2$, there exist $C_R, \alpha_R, c_R > 0$ such that*

$$\Psi_R(t) + \|\Omega(t)\|^2 \le C_R e^{-\alpha_R t}\left(\Psi_R(0) + \|\Omega(0)\|^2\right) + c_R \bar{d}_R^2.$$

For $\bar{d}_R = 0$, the equilibrium $(R_d, 0)$ is locally exponentially stable on sufficiently small invariant energy sublevels.

Proof. See Appendix A.8 for details.

Lemma 6.2 (Forcing by a moving attitude command)**.** *Suppose $R_d(t)$ is differentiable between command switching instants with $\dot{R}_d = R_d\,\Omega_d^{\wedge}$. With $\tilde{R} = R_d^{\top} R$, so that $e_R = \frac{1}{2}(\tilde{R} - \tilde{R}^{\top})^{\vee}$, and $\mathscr{C}(\tilde{R}) = \frac{1}{2}(\mathrm{tr}(\tilde{R})I - \tilde{R}^{\top})$, the implemented zero-rate damping law satisfies*

$$\begin{aligned} \dot{e}_R &= \mathscr{C}(\tilde{R})\big(\Omega - \tilde{R}^{\top}\Omega_d\big), \\ \dot{\Psi}_R &= e_R^{\top}\big(\Omega - \tilde{R}^{\top}\Omega_d\big). \end{aligned} \tag{45}$$

Consequently, for $\mathscr{V}_0 = \frac{1}{2}\Omega^{\top} J\Omega + k_R\Psi_R$ and ideal gyroscopic cancellation,

$$\dot{\mathscr{V}}_0 = -k_{\Omega}\,\|\Omega\|^2 - k_R e_R^{\top}\tilde{R}^{\top}\Omega_d + \Omega^{\top} d_M, \tag{46}$$

where d_M collects the realized torque and measurement residuals. At a command jump t_j with $R_d^{-} = R_d(t_j^{-})$ and $R_d^{+} = R_d(t_j^{+})$, the physical R and Ω are continuous, and

$$\begin{aligned} |\mathscr{V}_0(t_j^{+}) - \mathscr{V}_0(t_j^{-})| &= k_R\,|\Psi_R(R(t_j), R_d^{+}) - \Psi_R(R(t_j), R_d^{-})| \\ &\le \frac{\sqrt{3}}{2} k_R \left\| R_d^{+} - R_d^{-} \right\|_F . \end{aligned} \tag{47}$$

Proof. See Appendix A.9 for details.

By Eq. (46), a moving target adds the supply term $-k_R e_R^{\top}\tilde{R}^{\top}\Omega_d$, so the frozen-target estimate of Theorem 6.1 extends to $\Omega_d \neq 0$ under bounds on the reference rate and on the switching. Eq. (47) bounds a single command jump, and a dwell time or a summable sequence of jumps extends the estimate across switching. A moving-reference ISS margin thus follows from bounds on Ω_d, on the jumps of R_d and on the actuator mismatch over a compatible invariant set.

Define the position error $e_{\mathrm{pos}} = p - p_d$, the thrust-axis error $e_b = Re_3 - b_{3,d}$ with commanded axis $b_{3,d} = R_d e_3$, and the cascade error $z_{\mathrm{cas}} = \mathrm{col}(e_{\mathrm{pos}}, e_b, e_R, e_{\Omega})$. Assume that (C1) the translational loop driven by the exact thrust axis is locally ISS from e_b to e_{pos} with linear gain $\chi_{\mathrm{pos}} \ge 0$, (C2) the attitude loop is locally ISS from the position coupling to e_b with linear gain $\chi_{\mathrm{att}} \ge 0$, and (C3) $\chi_{\mathrm{pos}}\chi_{\mathrm{att}} < 1$. Theorem 5.13, with $\|e_b\|$ added to the arguments of $\bar{\omega}_{\mathrm{hl}}$, supplies (C1) for the delayed translational loop. Theorem 6.1 gives ISS for a frozen target, and (C2) follows with Lemma 6.2 once Ω_d and the jumps of R_d are bounded by the position error.

Theorem 6.3 (Cascade practical stability)**.** *Under Assumptions 2.1 to 2.4 and (C1)–(C3), there exist $\beta \in \mathscr{KL}$ and $\chi \in \mathscr{K}$ with*

$$\left\|e_{\mathrm{pos}}(t)\right\| \le \beta\left(\left\|z_{\mathrm{cas},[-\bar{\tau}_u,0]}\right\|_{\infty}, t\right) + \chi(\bar{d} + \bar{w} + \bar{\delta}_y + \bar{e}_p + \bar{\Delta}_a + \bar{\delta}_M + \bar{e}_{\mathrm{att}} + \bar{\upsilon}_d).$$

Proof. By (C1) and (C2) the two loops are ISS subsystems, and the small-gain theorem [39] under (C3) gives the bound.

7. Simulation results

This section compares PF-DAIF with target-centred AIF and a Smith-predictor PID controller in four examples: a scalar estimator for the stationary belief offset of Theorem 3.1, a delayed linear

benchmark for the spectral test of Theorem 5.6, a reduced sampled model for the mean-square results of Theorem 5.9 under packet loss, and an underactuated quadrotor under time-varying delay, packet loss, gusts and sensor bias. In all examples, $\mathscr{A}_y = I$ and the delay estimate is exact, $\tilde{\tau}_u = 0$, whereas the model and sensing residuals are nonzero.

7.1. Comparison methods and performance indices

Two baselines are compared with PF-DAIF. Conventional AIF uses the target-centred prior with fixed precisions, evaluates its preferred action at the current belief and filters it with the action dynamics of Eq. (17). On a lost packet, its belief relaxes toward the reference. The Smith-predictor PID controller (Smith-PID) updates a fixed-gain observer, forecasts it over the delay with the in-flight inputs of its internal model and applies a PID law without action filter. On a lost packet, its observer is propagated by the model. All controllers use the same references, noise and packet-loss realizations and input limits. For the jth Monte Carlo run with N_t samples, the tracking and estimation errors are

$$\mathrm{RMSE}_{x,j} = \Big(\frac{1}{N_t}\sum_{k=1}^{N_t} \big\|\Sigma_x(x_{j,k} - x_{d,k})\big\|^2\Big)^{1/2},$$
$$\mathrm{RMSE}_{\mu,j} = \Big(\frac{1}{N_t}\sum_{k=1}^{N_t} \big\|\Sigma_\mu(\mu_{j,k} - x_{j,k})\big\|^2\Big)^{1/2},$$

and RMSE_x and RMSE_μ denote their averages over N_{MC} runs. The matrices Σ_x and Σ_μ select the evaluated coordinates. In the linear study, Σ_x selects the positions and $\Sigma_\mu = I$; in the quadrotor study, both select the position, and errors are given in metres.

7.2. Stationary belief offset

Theorem 3.1 is first evaluated on a scalar estimator without feedback, and its closed-loop counterpart follows in Section 7.5.3. The scalar channel is fully observed, with $x = 0$, $\delta_y = 0.12$ m, $e_p = -0.08$ m, $\pi_x = \pi_\mu = 2\ \mathrm{m}^{-2}$ and $\pi_y = 3\ \mathrm{m}^{-2}$, and the reference offset $x_d - x$ varies from -2 to 2 m. The AIF error changes linearly from -0.728 to 0.872 m with slope $\theta^x = 0.4$, whereas the PF-DAIF stationary error stays at 0.040 m (Fig. 2(b)). The AIF belief follows $\dot{\mu} = -g_\mu[(\pi_x + \pi_y)\mu - \pi_x x_d - \pi_y \tilde{y}]$, and the PF-DAIF belief follows the same flow with x_d replaced by $\hat{x}_p$, for $g_\mu = 0.4$ and an initial error of 0.7 m. The AIF beliefs converge to offsets that depend on the reference, whereas the PF-DAIF belief converges to 0.040 m (Fig. 2(a)).

7.3. Linear benchmark with input delay

The first example consists of two coupled oscillators,

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -2 & -0.6 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \\ 0.4 & 0 & -1.5 & -0.5 \end{bmatrix} x(t) + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} u(t-\tau_u) + d(t), \tag{48}$$
$$y_\tau(t) = x(t) + \nu_y(t),$$

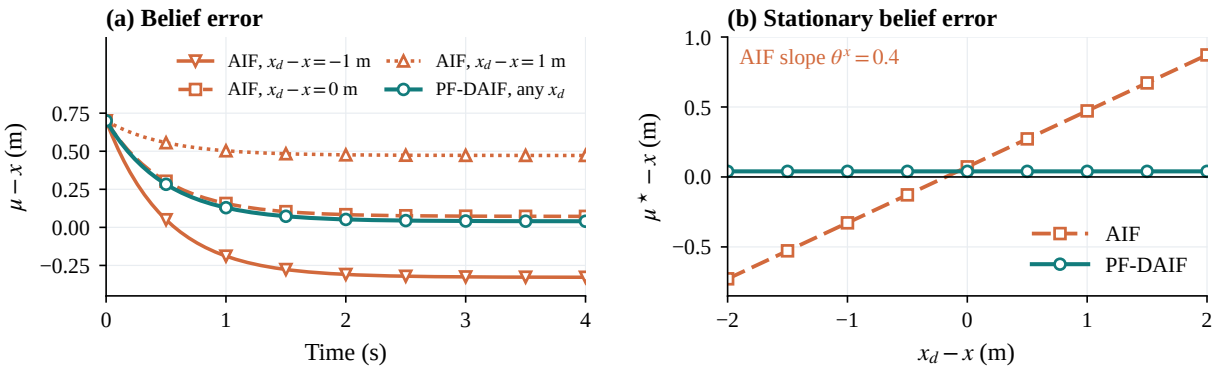


Figure 2: Scalar estimator. (a) Belief error over time. (b) Stationary belief error versus reference offset.

with positions x_1, x_3 and velocities x_2, x_4. The controllers use this nominal model, whereas the plant scales A_{21}, A_{22}, A_{43} and A_{44} by 1.1, 1.15, 0.9 and 0.85, so $\Delta_f \neq 0$. The disturbance is $d(t) = \mathrm{col}(0, 0.13\sin 0.9t + 0.07\cos 0.13t, 0, 0.11\cos 0.7t)$, the reference positions are $0.65\sin 0.37t + 0.2$ and $0.55\cos 0.31t - 0.25$ with matching velocities, and the measurement noise is Gaussian. PF-DAIF uses the forecast law $\tau_f = \hat{\tau}_u$ with $u_d = 0$, so $\xi_k = \mathrm{col}(x_k, U_k, \mathscr{H}_k, a_k)$ as in Eq. (28). AIF uses lower gains, because the PF-DAIF gains applied at the current belief lose stability at 0.525 s (Section 7.3.1). Packets are lost independently with a fixed probability, all quantities are dimensionless, and Table 2 lists the parameters.

7.3.1. Delay-dependent stability

The modulus of the dominant complex eigenvalue pair of $\mathbb{A}_N$ in Eq. (31) is evaluated for $N = 0, \ldots, 42$ (Table 3 and Fig. 3(a)), with the precision fixed at its initial value, the true plant in the plant row and the nominal model in the predictor rows. The integral terms of the baselines add a real eigenvalue at 0.9993, independent of the delay.

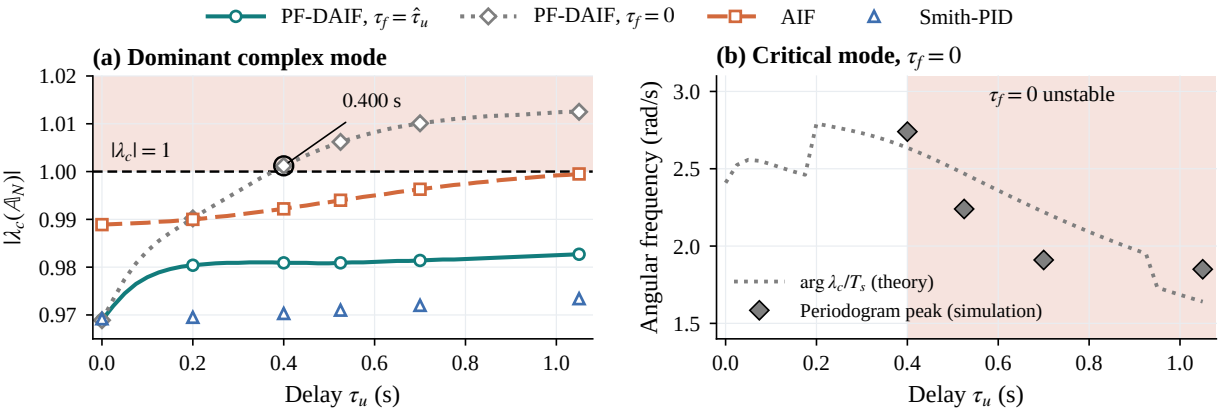


Figure 3: Linear benchmark. (a) Dominant complex eigenvalue modulus of $\mathbb{A}_N$. (b) Critical-mode frequency of PF-DAIF with $\tau_f = 0$.

For fixed gains, the forecast horizon determines stability. Under the present-state law $\tau_f = 0$, the modulus increases monotonically with the delay and crosses one at $N = 16$ (0.400 s), where the

Table 2: Parameters of the linear benchmark.

Symbol	Meaning	Value
PF-DAIF		
η_μ, η_a	belief relaxation factor, action step	0.50, 0.05
Π_μ, Π_a	prior and action precisions	$\mathrm{diag}(5,2,5,2)$, $5I_2$
$\Pi_y(0)$	initial sensory precision	$\mathrm{diag}(6,2,6,2)$
K	position and velocity gains per channel	2.88, 1.98
Γ_γ, k_γ	precision rate and damping	$0.6I_4$, 2 s^{-1}
$\underline{\gamma}$, $\bar{\gamma}$	log-precision interval	-6, 6
τ_f	forecast horizon	$\hat{\tau}_u$
Baselines		
AIF	position, velocity, integral gains; Π_x; fixed Π_y	1, 1, 0.15; $\mathrm{diag}(5,2,5,2)$; $\mathrm{diag}(6,2,6,2)$
Smith-PID	position, velocity, integral gains; observer gain κ_o	2.88, 1.98, 0.15; 0.5
Scenario		
T_s, τ_u	sampling time, nominal delay	0.025 s, 0.70 s ($N = 28$)
$\mathscr{U}$	input box per channel	$[-4,4]$
$x(0) = \mu(0)$	initial state and belief	$\mathrm{col}(0.9,0,-0.7,0)$
σ_ν	noise standard deviations	$\mathrm{col}(0.025,0.045,0.025,0.045)$
T_f, N_{MC}	duration, Monte Carlo runs	8 s, 3

Table 3: Dominant complex eigenvalue modulus of $\mathbb{A}_N$ and critical-mode frequency of PF-DAIF with $\tau_f = 0$.

		Dominant complex modulus					Critical mode (rad/s)	
N	τ_u (s)	PF-DAIF $\tau_f = \hat{\tau}_u$	PF-DAIF $\tau_f = 0$	AIF	AIF PF-DAIF gains	Smith-PID	$\arg\lambda_c/T_s$	Peak
0	0.000	0.9689	0.9689	0.9889	0.9835	0.9692	2.41	–
8	0.200	0.9804	0.9902	0.9900	0.9888	0.9695	2.79	–
16	0.400	0.9809	**1.0012**	0.9922	0.9968	0.9703	2.64	2.74
21	0.525	0.9809	1.0062	0.9940	**1.0006**	0.9710	2.47	2.24
28	0.700	0.9814	1.0101	0.9963	1.0040	0.9720	2.22	1.91
42	1.050	0.9827	1.0125	0.9995	1.0069	0.9734	1.64	1.85

critical mode has angular frequency 2.64 rad/s. A predictor-centred prior alone therefore leaves the actuation delay in the loop, as stated in Section 4.1. Under the forecast law, the modulus stays between 0.969 and 0.983 up to 1.05 s, so the delay is compensated on the action side over

the whole range. With the PF-DAIF gains, AIF crosses one at 0.525 s, and with its own lower gains it approaches one, which reflects the trade-off between gain and delay of a law without forecast. Smith-PID stays between 0.969 and 0.974. At the crossing delay, the simulated tracking error of the present-state law oscillates at 2.74 rad/s, against 2.64 rad/s predicted by Lemma 5.7 (Fig. 3(b)). Beyond the crossing, the input saturates and the measured frequency falls below the linear prediction at 0.525 and 0.70 s; at 1.05 s, it exceeds the linear prediction.

Simulations of 40 s without packet loss locate the same boundary, with the delay increased from 0.30 to 1.05 s in steps of one sample. The present-state law tracks within 7% of the forecast law up to 0.375 s and loses stability at 0.400 s, the delay of the spectral crossing. The steady-state RMSE rises from 0.27 to 0.61 and then to 5.0, with the input saturated on more than 90% of the samples. The forecast law remains stable over the whole tested range.

7.3.2. Tracking and estimation performance

At the nominal delay of 0.70 s (Table 4), PF-DAIF tracks within 5% of Smith-PID at a comparable input energy $\mathscr{E}_u = \sum_k T_s \|u_k\|^2$, and its estimation error is an order of magnitude below that of AIF. The remaining gap to Smith-PID reflects the information lag of Section 4.2: Smith-PID forecasts from its current observer state, PF-DAIF from a belief N samples old. The present-state law saturates the input on 73% of the samples, consistent with $\rho(\mathbb{A}_{28}) > 1$.

Table 4: Performance of the linear benchmark at $\tau_u = 0.70$ s without packet loss.

Method	RMSE_x	RMSE_μ	$\mathscr{E}_u$
AIF	0.717	0.497	2.4
Smith-PID	**0.441**	**0.048**	10.0
PF-DAIF, $\tau_f = \hat{\tau}_u$	0.459	0.055	9.4
PF-DAIF, $\tau_f = 0$	1.812	0.284	172.9

The delay is then varied from 0 to 1.05 s and the loss rate from 0 to 70%, giving 30 combinations. The tracking RMSE lies between 0.41 and 0.49 for PF-DAIF, between 0.40 and 0.46 for Smith-PID and between 0.64 and 0.81 for AIF. Loss raises the tracking error of PF-DAIF by at most 1.3%, because a lost packet leaves the belief on the predictor and the action on the forecast. This insensitivity agrees with Theorem 5.9: at the nominal delay, the second-moment map of the forecast loop has spectral radius 0.966 at 30% loss, 0.979 at 70% and 0.988 at 99%, so Eq. (33) is feasible for every tested loss rate. The estimation errors separate the methods more sharply. The AIF error lies between 0.40 and 0.92 and increases with loss, since each lost packet relaxes its belief toward the reference. PF-DAIF stays between 0.05 and 0.16, and Smith-PID between 0.05 and 0.11. Over all combinations, PF-DAIF reduces the estimation error of AIF by 80.8% to 89.4%. This ordering follows Lemma 3.2: the target-centred belief carries $\theta_i^x(x_{d,i} - x_i)$, whereas the predictor-centred belief carries $\theta_i e_{p,i}$, which is bounded by Lemma 5.1.

7.4. Mean-square stability under packet loss

The mean-square results are tested on the single-axis reduced sampled model of Section 8.2, which has two modes for received and lost packets. The evolution of $\mathbb{E}\|\xi_k\|^2$ over 80 samples

at loss rates of 0, 30 and 60% follows from the second-moment recursion and from 2000 Monte Carlo runs (Fig. 4). At 30% loss, the spectral radius of the second-moment map stays below one at every tested delay, so the reduced loop is mean-square stable (Fig. 5(a)). The initial state has a unit position offset, an exact initial belief and a zero input history, and is normalized to unit norm. The second moment is not monotone: it rises to eight times its initial value within ten samples, as the commands build up, and then decays, the transient growth admitted by Lemma 5.10. Packet loss leaves this peak unchanged and slows the late decay, and at $k = 80$ the exact expected squared norm is 0.287, 0.415 and 0.840 for loss probabilities 0, 0.30 and 0.60. The Monte Carlo means agree with the exact recursion to within 1.1% (Section 8.3).

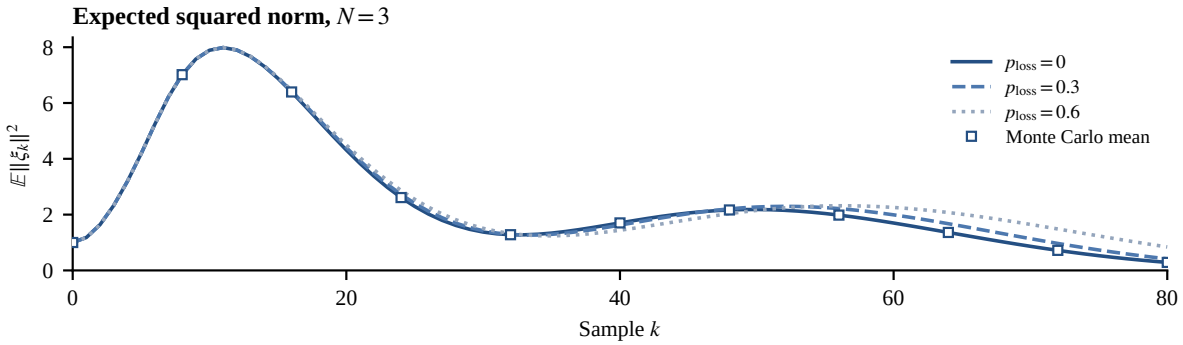


Figure 4: $\mathbb{E}\,\|\xi_k\|^2$ of the reduced sampled model at $N = 3$ under packet loss.

7.4.1. Effect of the delay

The input delay of the same reduced model is varied over $N = 1,\dots,5$, with $T_s = 0.025$ s, gains $K_{\text{pos}} = K_{\text{vel}} = 2.4$ and independent packet loss $p_{\text{loss}} = 0.30$. Each N changes the dimension of the delay history, and the matrix $\mathscr{M}$ of Section 5.4.3, denoted $\mathscr{M}_N$, is formed from the mode matrices of that dimension. Transient growth is measured by G_{80} of Lemma 5.11, the largest expected squared norm over unit initial states within 80 steps, computed with the adjoint recursion $Y_{k+1} = \mathscr{T}^*(Y_k)$. All spectral radii remain below one (Fig. 5(a)), while G_{80} grows from 23.6 at $N = 1$ to 70.8 at $N = 5$ (Fig. 5(b)). A longer delay thus amplifies the transient while the loop remains mean-square stable, and Section 7.6 quantifies the same effect on the quadrotor.

7.5. Quadrotor example

7.5.1. Simulation scenario and results

The quadrotor of Section 6.1 is simulated with the parameters of Tables 5 and 6. The delay, the external force and the sensor bias are, in seconds, newtons and metres,

$$\tau_u(t) = 0.08 + 0.04\sin 0.7t + 0.08\,\mathbf{1}_{[6.5,9.5]}(t), \tag{49}$$

$$D_a(t) = \operatorname{col}(1.2,-0.8,0.45)\,\mathbf{1}_{[6,9.5]}(t) + 0.15\operatorname{col}(\sin 0.7t,\cos 0.5t,\sin 0.3t) - 0.08\,v_I(t), \tag{50}$$

$$\varphi_y(t) = \operatorname{col}\big(\operatorname{col}(0.40,0,-0.20)\,\mathbf{1}_{[5.8,9.8]}(t),\ \operatorname{col}(0,0,0.10\sin 3t)\big), \tag{51}$$

where $\mathbf{1}_{\mathscr{I}}$ is the indicator of the interval $\mathscr{I}$, and the two blocks of φ_y act on the position and velocity measurements. The position bias acts on the x and z channels between 5.8 and 9.8 s and

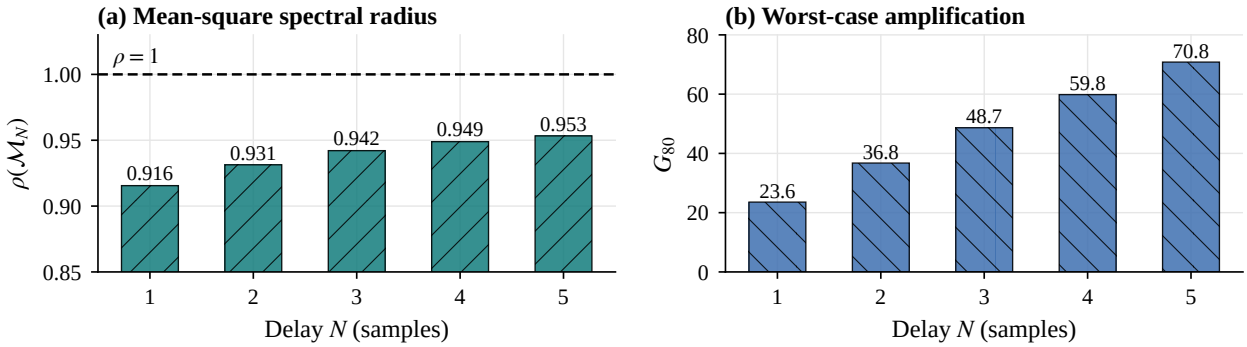


Figure 5: Reduced sampled model at $p_{\text{loss}} = 0.3$ versus delay. (a) Spectral radius of $\mathscr{M}_N$. (b) Worst-case amplification G_{80}.

overlaps with the gust from 6 to 9.5 s and the increased delay from 6.5 to 9.5 s. The y channel remains unbiased, and the vertical velocity measurement carries an oscillating bias. The delay is smooth between its two switching instants, where Assumption 2.1 holds.

Table 5: Plant and scenario parameters of the quadrotor example.

Symbol	Meaning	Value
m, J	mass, inertia	1.5 kg, $\mathrm{diag}(0.028, 0.030, 0.055)$ kg m^2
$T_{\min}$, $T_{\max}$	thrust limits	1.47, 29.43 N
$M_{\max}$	torque limits	$(0.28, 0.28, 0.14)$ N m
k_R, k_Ω	attitude and rate gains	$\mathrm{diag}(1.1, 1.1, 0.5)$, $\mathrm{diag}(0.25, 0.25, 0.15)$
T_s	outer-loop sampling time	0.025 s
$\tau_u(t)$	actuation delay	Eq. (49), range $[0.04, 0.175]$ s
p_{loss}	Bernoulli packet-loss probability	0.3
$p_d(t)$	reference	$\mathrm{col}(2.5\cos 0.22t,\ 2.5\sin 0.22t,\ 1.2 + 0.06t)$ m
σ_p, σ_v, σ_w	position, velocity and process noise	0.055 m, 0.035 m/s, 0.025 m/s^2
$\varphi_y(t)$	sensor bias	Eq. (51)
T_f, N_{MC}	duration, Monte Carlo runs	10 s, 5

The mean tracking RMSE is 0.417 m for PF-DAIF, 0.411 m for Smith-PID and 0.679 m for AIF (Fig. 6), and the mean estimation RMSE is 0.293, 0.263 and 0.443 m. Before the perturbed interval, the estimation RMSE is 0.055 m for PF-DAIF, 0.028 m for Smith-PID and 0.174 m for AIF. The AIF belief is pulled toward the reference whenever a packet is lost, and this offset grows with its tracking error, as Eq. (7) predicts. Within the perturbed interval, the estimation RMSE rises to 0.454 m for PF-DAIF, 0.410 m for Smith-PID and 0.649 m for AIF, since none of the three estimators can distinguish a persistent bias from a state offset. The tracking RMSE in this interval is 0.578 m for PF-DAIF, 0.593 m for Smith-PID and 0.987 m for AIF. The sensory precision remains in its admissible interval, as Lemma 5.4 guarantees, and it rises during the bias window, because the belief follows the biased measurement and the innovation stays below its threshold.

Table 6: Controller parameters of the quadrotor example.

Symbol	Meaning	Value
PF-DAIF		
Π_μ, Π_a	prior and action precisions	$\mathrm{diag}(5,5,6,2,2,2)$, $5I_3$
$\Pi_y(0)$	initial sensory precision	$\mathrm{diag}(6,6,6,2,2,2)$
η_μ, η_a	belief relaxation factor, action step	0.50, 0.05
Γ_γ, k_γ	precision rate and damping	$0.8I_6$, 5 s^{-1}
$\underline{\gamma}$, $\bar{\gamma}$	log-precision interval	-5, 4
K_{pos}, K_{vel}, τ_f	gains of Eq. (43), forecast horizon	$2.40I_3$, $2.40I_3$, 0
Baselines		
AIF	Π_x; fixed Π_y; η_μ, η_a, Π_a, K_{pos}, K_{vel}	$\mathrm{diag}(5,5,6,2,2,2)$; $\Pi_y(0)$; as PF-DAIF
Smith-PID	observer gain κ_o; integral gain; K_{pos}, K_{vel}	0.15; $0.15I_3$; as PF-DAIF

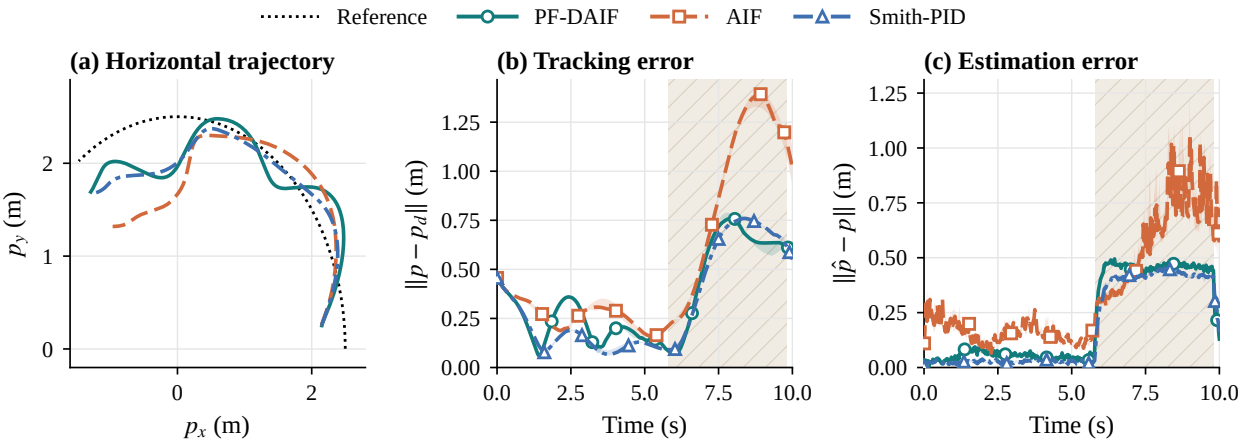


Figure 6: Quadrotor response. (a) Horizontal trajectory of one run. (b) Tracking error. (c) Estimation error. Median and interquartile range over five runs; shading marks the disturbance interval.

7.5.2. State responses

For the representative run of Fig. 6(a), the six translational coordinates $(p_x, p_y, p_z, v_{I,x}, v_{I,y}, v_{I,z})$ and the six rotational coordinates (roll, pitch, yaw, Ω_x, Ω_y, Ω_z) are plotted in Figs. 7 and 8. The Euler angles are computed from the rotation matrix, Ω is expressed in the body frame and v_I in the inertial frame. Before the perturbed interval, all controllers follow the reference. Within it, the vertical bias drives all three vehicles upward, and AIF departs furthest from the horizontal reference. PF-DAIF has larger roll and pitch excursions, which remain below 10°.

The inner loop tracks the attitude $R^{\mathrm{in}}_{d,k} = R_d(u(t_k - \tau_u(t_k)))$ constructed from the delayed acceleration command, whereas the outer loop generates $u(t_k)$. The attitude tracking error is

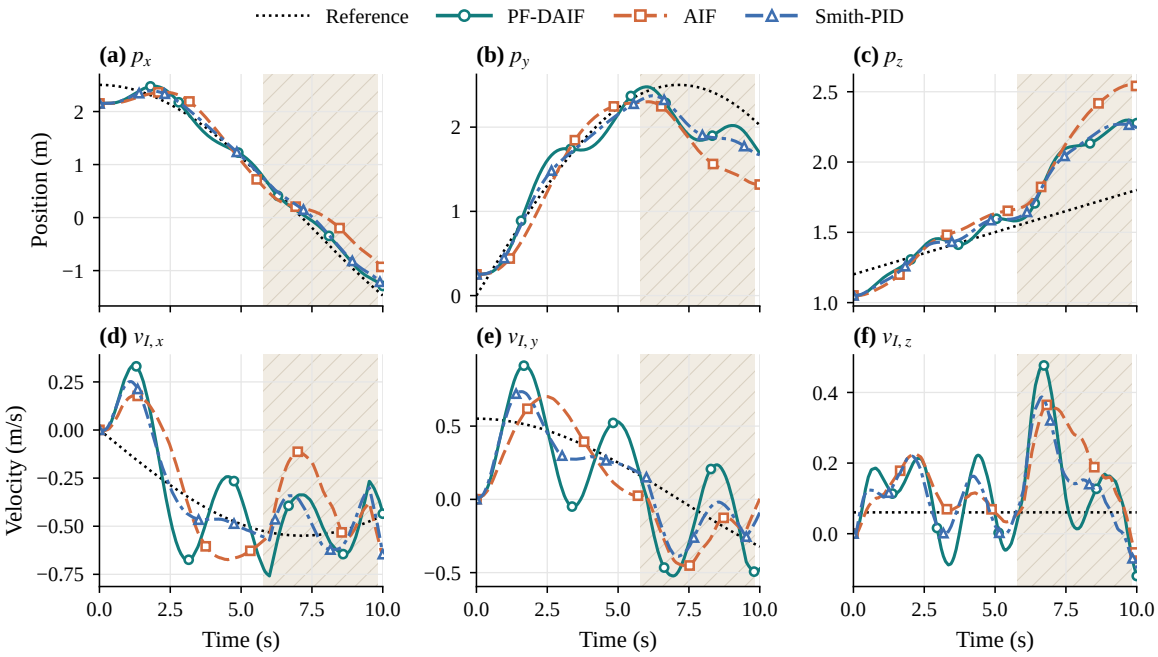


Figure 7: Translational states of the quadrotor. (a)–(c) Position. (d)–(f) Inertial velocity.

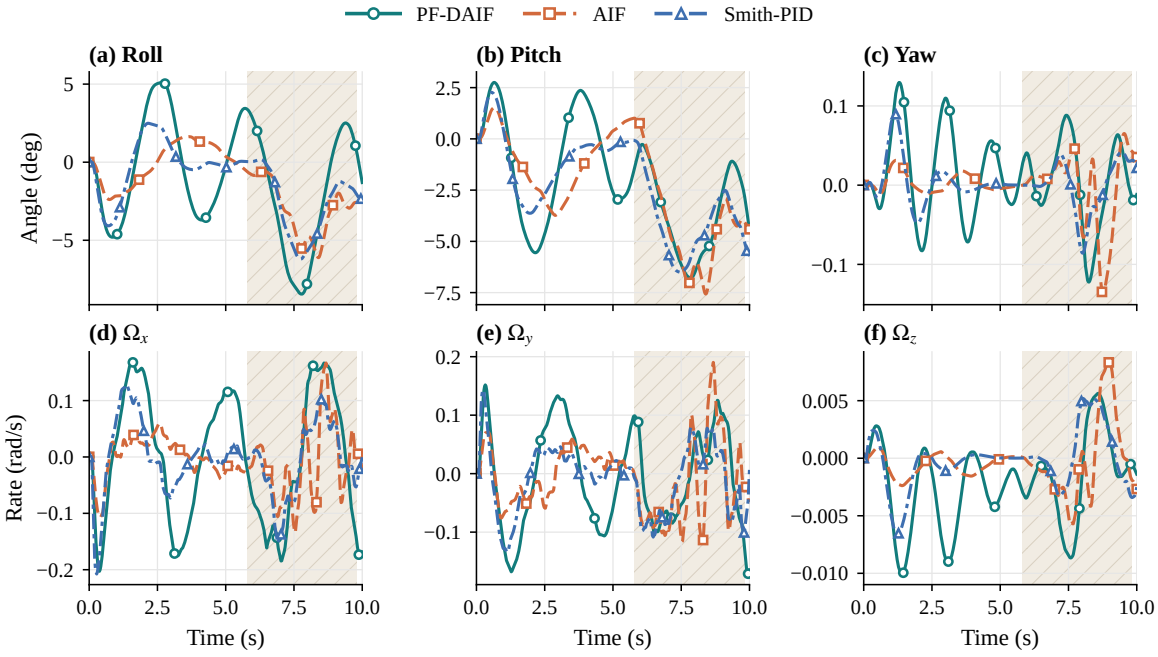


Figure 8: Rotational states of the quadrotor. (a)–(c) Euler angles. (d)–(f) Body rates.

therefore measured by

$$\begin{aligned} R_{e,k} &= (R_{d,k}^{\text{in}})^{\top} R_k, \\ \theta_{R,k} &= \left\| \text{Log}(R_{e,k})^{\vee} \right\|, \\ \Psi_{R,k} &= \tfrac{1}{2} \text{tr}(I - R_{e,k}), \end{aligned} \tag{52}$$

where $\theta_{R,k}$ is the principal geodesic rotation angle in radians and $\Psi_{R,k} = 1 - \cos\theta_{R,k}$. Since the inner loop damps Ω to zero without angular-rate feedforward, the attitude error is driven by the moving command, as described by Lemma 6.2.

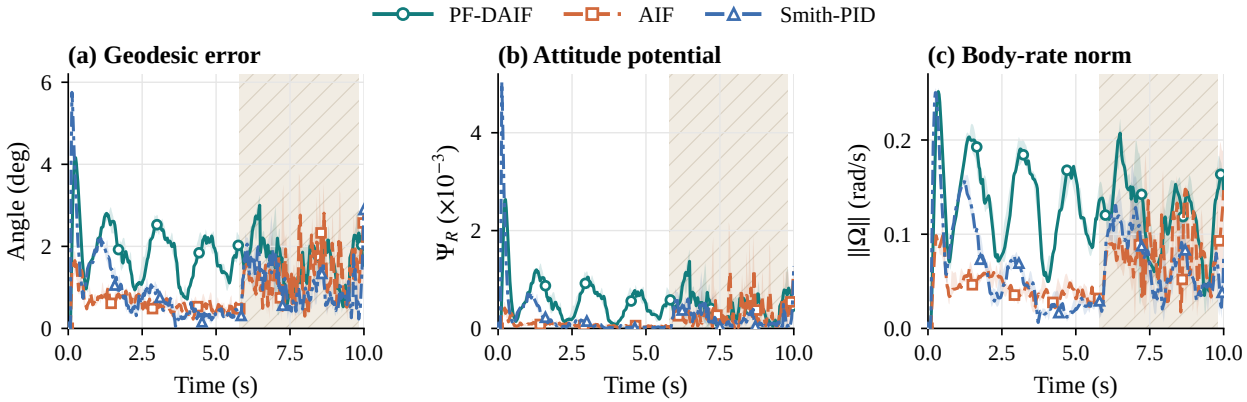


Figure 9: Inner-loop attitude tracking. (a) Geodesic error. (b) Attitude potential Ψ_R. (c) Body-rate norm. Median and interquartile range over five runs.

The RMS geodesic attitude error, averaged over five runs, is 1.913° for PF-DAIF, 1.253° for AIF and 1.264° for Smith-PID (Fig. 9). Relative to AIF, the smaller translational tracking error of PF-DAIF thus comes with a larger inner-loop attitude error. Each error is measured with respect to the delayed attitude command of the respective controller.

The forcing term of Lemma 6.2 depends on the variation of the attitude command, which is measured by the rotation between successive commands and its sum over the run,

$$\begin{aligned} \delta\theta_{d,k} &= \left\| \mathrm{Log}[(R^{\mathrm{in}}_{d,k-1})^{\top} R^{\mathrm{in}}_{d,k}]^{\vee} \right\|, \\ \mathscr{L}_d &= \sum_{k=1}^{N_t} \delta\theta_{d,k}. \end{aligned} \tag{53}$$

Averaged over five runs, the maximum of $\delta\theta_{d,k}$ is 1.223° for PF-DAIF, 1.854° for AIF and 4.178° for Smith-PID, and $\mathscr{L}_d$ is 91.23°, 112.63° and 94.37°. PF-DAIF therefore commands the smoothest attitude target of the three controllers. The actuation delay changes discontinuously at 6.5 and 9.5 s, so $\delta\theta_{d,k}/T_s$ approximates both $\|\Omega_d\|$ and the command jumps.

7.5.3. Robustness to reference shifts, disturbances and gains

Horizontal reference offsets of -0.8, 0 and $+0.8$ m give estimation RMSEs of 0.291, 0.293 and 0.300 m for PF-DAIF and 0.438, 0.443 and 0.525 m for AIF. This insensitivity is the closed-loop counterpart of the reference independence in Theorem 3.1. When the sensor bias is doubled, all estimators degrade, and when the gust is doubled the tracking RMSE is 0.756 m for PF-DAIF, 0.779 m for Smith-PID and 1.355 m for AIF. Across nine feedback-gain settings, the estimation RMSE of PF-DAIF ranges from 0.293 to 0.307 m, against 0.379 to 0.548 m for AIF. Removing the predictor improves the nominal tracking RMSE (0.369 against 0.417 m) but worsens it at three times the nominal delay (1.678 against 0.915 m).

7.6. Effect of the actuation delay on the quadrotor

The effect of the delay on the full nonlinear loop, including the inner attitude loop, actuator saturation, measurement noise and packet loss, is quantified by scaling the delay profile as $\tau_u(t) = s_\tau[0.08 + 0.04\sin(0.7t) + 0.08\,\mathbf{1}_{[6.5,9.5]}(t)]$ s, with $s_\tau \in \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2\}$ and five runs for each controller and scale. Each scale is labelled by its maximum delay over the run.

The tracking energy of the vehicle is measured by

$$V_{\mathrm{diag},k} = \tfrac{1}{2}\left(\left\|p_k - p_{d,k}\right\|^2 + 0.5\left\|v_{I,k} - \dot{p}_{d,k}\right\|^2 + 0.5\left\|e_{R,k}\right\|^2 + 0.1\left\|\Omega_k\right\|^2\right), \tag{54}$$

where $e_{R,k} = \frac{1}{2}(R_{d,k}^\top R_k - R_k^\top R_{d,k})^\vee$ and $R_{d,k}$ is reconstructed from the current commanded acceleration. The fraction of samples with $V_{\mathrm{diag},k+1} > V_{\mathrm{diag},k}$ measures the frequency of tracking-energy gains of the disturbed vehicle.

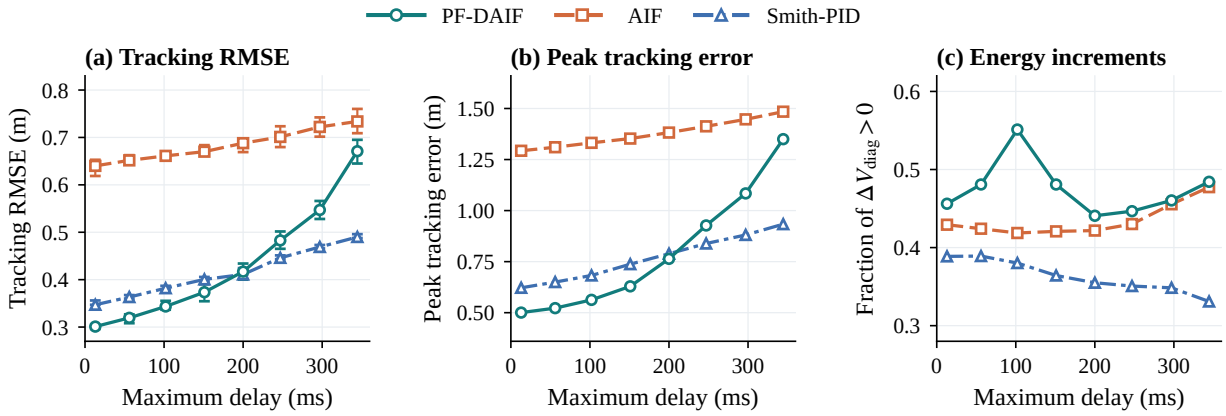


Figure 10: Quadrotor under scaled actuation delay. (a) Tracking RMSE, mean and standard deviation over five runs. (b) Peak tracking error. (c) Fraction of positive increments of V_{diag}.

The mean tracking RMSE of PF-DAIF rises from 0.301 m without scheduled delay (Fig. 10(a)) to 0.417 m at the nominal profile and 0.671 m at twice the nominal profile, whose largest delay is about 0.35 s. The corresponding Smith-PID values are 0.347, 0.411 and 0.490 m, and AIF stays above 0.64 m. The ranking therefore crosses as the delay grows: PF-DAIF tracks best for short delays, and Smith-PID tracks better from about the nominal profile onward. Two effects analysed above account for the crossing: the quadrotor runs the present-state law $\tau_f = 0$, which leaves the actuation delay in the loop (Section 7.3.1), and the predictor of PF-DAIF is anchored at a belief one delay old, so the residual bound of Lemma 5.1 grows exponentially with the delay, whereas Smith-PID forecasts from its current observer state. The peak tracking error has the same crossing (Fig. 10(b)), and positive increments of V_{diag} persist under continued forcing for all controllers (Fig. 10(c)).

At $s_\tau = 1.5$, nine gain pairs $(K_{\mathrm{pos}}, K_{\mathrm{vel}})$ are tested for each controller with five runs each (Fig. 11). The tracking error of PF-DAIF depends strongly on the gains and is smallest at the largest velocity gain, in line with the trade-off between gain and delay of Section 7.3.1.

7.7. Operating region of the quadrotor

The operating region is evaluated on a 3×3 grid of delay and gust scales and on a 3×3 grid of initial pitch angles and horizontal position offsets, with sensor bias and packet loss as in the

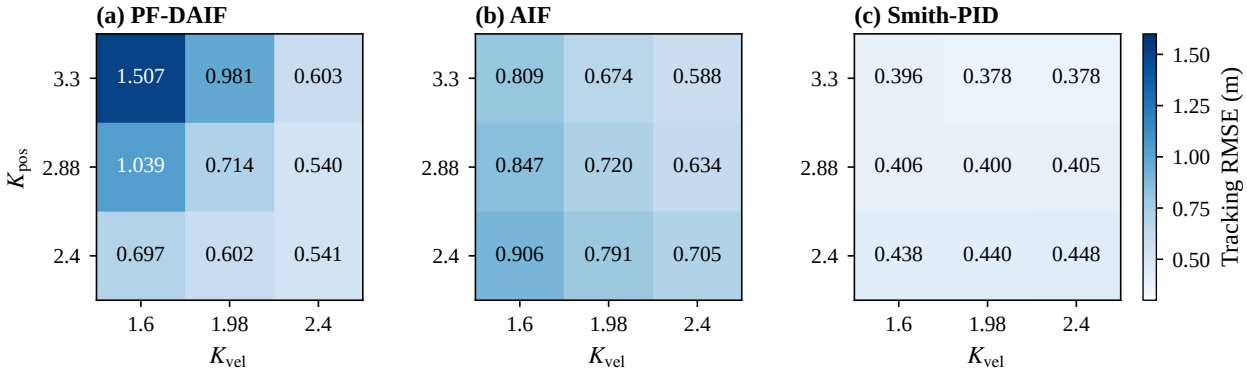


Figure 11: Mean tracking RMSE at $s_\tau = 1.5$ over the gain grid. (a) PF-DAIF. (b) AIF. (c) Smith-PID.

nominal scenario and two runs for each controller and condition. The delay scale multiplies the whole delay profile and the gust scale the gust amplitude. The initial pitch angles are 0°, 15° and 30°, and the horizontal offsets are -0.4, 0 and 0.4 m.

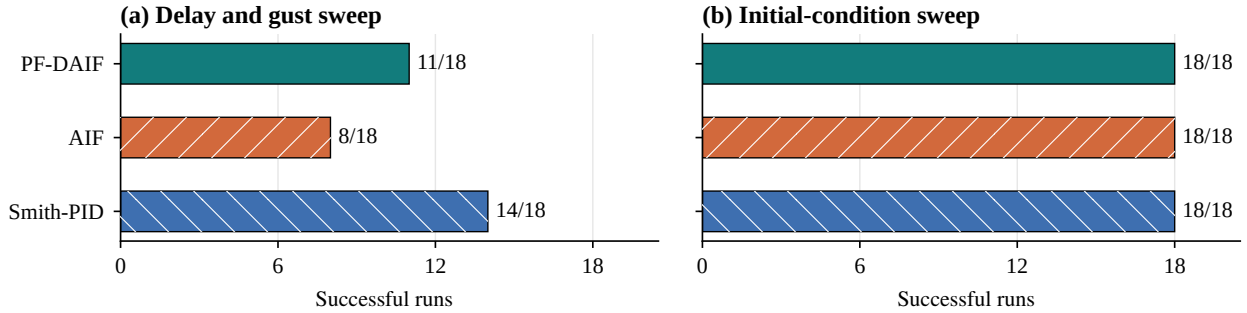


Figure 12: Successful runs out of 18. (a) Delay and gust scales. (b) Initial conditions.

A run is counted as successful if, over 10 s, the altitude stays positive, the tilt angle stays below 60° and the peak tracking error stays below 3 m. For the grid of delay and gust (Fig. 12(a)), 11 of 18 runs are successful for PF-DAIF, 8 for AIF and 14 for Smith-PID; at the most severe combination, the mean tracking RMSE of PF-DAIF reaches 3.20 m. For the grid of initial conditions (Fig. 12(b)), all 18 runs are successful for all three methods. The lower count of PF-DAIF relative to Smith-PID at large delays agrees with the ranking crossover of Section 7.6.

8. Supplementary analysis

This section gives the linearization of the quadrotor subsystems at hover, the reduced sampled model used in Section 7.4, its second-moment evolution and its mean-square Lyapunov function.

8.1. Linearization at hover

The attitude and translational subsystems are linearized at hover with the gains of Tables 5 and 6. For the fixed desired orientation $R_d = I$ and zero angular rate, the first-order attitude coordinates ε_R

obey

$$\frac{d}{dt}\begin{bmatrix}\varepsilon_R\\ \Omega\end{bmatrix} = A_{\mathrm{att}}\begin{bmatrix}\varepsilon_R\\ \Omega\end{bmatrix},\qquad A_{\mathrm{att}} = \begin{bmatrix}0 & I\\ -J^{-1}k_R & -J^{-1}k_\Omega\end{bmatrix}. \tag{55}$$

With the inertia and attitude gains of Table 5, the largest real part of the six eigenvalues is $-1.3636\ \mathrm{s}^{-1}$. For an ideal thrust axis, the reduced translational model is $\dot{p} = v$, $\dot{v} = a$, with the discrete action filter replaced by its continuous equivalent $\dot{a} = -\lambda_a(a + K_{\mathrm{pos}}p + K_{\mathrm{vel}}v)$, $\lambda_a = \eta_a\pi_a/T_s = 10\ \mathrm{s}^{-1}$. Its Jacobian is

$$A_{\mathrm{tr}} = \begin{bmatrix}0 & I & 0\\ 0 & 0 & I\\ -\lambda_a K_{\mathrm{pos}} I & -\lambda_a K_{\mathrm{vel}} I & -\lambda_a I\end{bmatrix}, \tag{56}$$

with spectral abscissa $-1.4532\ \mathrm{s}^{-1}$ at $K_{\mathrm{pos}} = K_{\mathrm{vel}} = 2.40$. Solving $A^\top P + PA = -I$ for $A \in \{A_{\mathrm{att}}, A_{\mathrm{tr}}\}$ gives the values of Table 7, so both linearized subsystems are exponentially stable.

Table 7: Stability margins of the subsystems linearized at hover.

Subsystem	$\max \mathrm{Re}\,\lambda(A)\ (\mathrm{s}^{-1})$	$\lambda_{\min}(P)$	$\lambda_{\max}(P)$	$\left\|A^\top P + PA + I\right\|_2$
Attitude, fixed target	-1.3636	0.05736	2.37372	2.3×10^{-15}
Translation with action filter	-1.4532	0.05394	4.07848	8.2×10^{-15}

8.2. Reduced sampled model

The reduced model describes one translational axis of the quadrotor with ideal attitude tracking, an integer delay, fixed precisions and no input saturation. The parameters are $T_s = 0.025$ s, $K_{\mathrm{pos}} = K_{\mathrm{vel}} = 2.4$, $\eta_\mu = 0.5$, $\eta_a = 0.05$, $\Pi_a = 5$, fixed prior and sensory precisions $(5,2)$ and $(6,2)$ for position and velocity, $\Phi = \left(\begin{smallmatrix}1 & T_s\\ 0 & 1\end{smallmatrix}\right)$ and $\Xi = \mathrm{col}(T_s^2/2, T_s)$. In the quadrotor implementation of Algorithm 1, the action update is applied in the sample of the measurement: at sample k, the belief μ_k fuses the current measurement with the predictor anchored at μ_{k-N}, and the updated action is applied in the same sample, $u_k = a_{k+1}$. For an integer delay $N \geq 1$, the augmented state $\mathrm{col}(x_k, \mu_{k-1}, \ldots, \mu_{k-N}, u_{k-1}, \ldots, u_{k-2N}, a_k)$ therefore has dimension $4N+3$, one belief block fewer than Eq. (28), in which the current measurement enters μ_{k+1} and $u_k = a_k$. The plant is advanced with u_{k-N}, and the action filter has factor $1 - \eta_a\Pi_a = 0.75$. The received and lost packets give the linear maps $\xi_{k+1} = \mathbb{A}_N^{\mathrm{rx}}\xi_k$ and $\xi_{k+1} = \mathbb{A}_N^{\mathrm{loss}}\xi_k$. Table 8 lists the spectral radius of the received mode and of $\mathscr{M}_N = 0.7\,\mathbb{A}_N^{\mathrm{rx}} \otimes \mathbb{A}_N^{\mathrm{rx}} + 0.3\,\mathbb{A}_N^{\mathrm{loss}} \otimes \mathbb{A}_N^{\mathrm{loss}}$ for Bernoulli reception with probability 0.7.

The lost-packet mode alone has $\rho(\mathbb{A}_N^{\mathrm{loss}}) = 1$ on this grid, so Schur stability of each mode is not required for mean-square decay under independent reception. With the update order of Eq. (27), the same parameters give $\rho(\mathscr{M}_N)$ between 0.952 and 0.974 for $N = 1, \ldots, 5$, so both orders are mean-square stable on this grid.

Table 8: Spectral radii of the reduced sampled model.

N	τ_u (s)	Dimension	$\rho(\mathbb{A}_N^{\mathrm{rx}})$	$\rho(\mathscr{M}_N)$
1	0.025	7	0.95357	0.91552
2	0.050	11	0.95527	0.93135
3	0.075	15	0.96201	0.94209
4	0.100	19	0.96759	0.94892
5	0.125	23	0.97187	0.95329

8.3. Second-moment evolution at $N = 3$

For $N = 3$ and the initial state ξ_0 of Section 7.4, the recursion $X_{k+1} = \mathscr{T}(X_k)$ with $X_0 = \xi_0\xi_0^\top$ gives $\mathbb{E}\,\|\xi_k\|^2 = \operatorname{tr} X_k$. Table 9 compares it with the mean over 2000 Monte Carlo runs.

Table 9: Second moment of the reduced sampled model at $N = 3$: recursion and Monte Carlo mean.

Loss p_{loss}	$\operatorname{tr} X_{40}$	Monte Carlo, $k = 40$	$\operatorname{tr} X_{80}$	Monte Carlo, $k = 80$
0.0	1.70220	1.70220	0.28693	0.28693
0.3	1.62301	1.62315	0.41476	0.41579
0.6	1.43954	1.43700	0.84041	0.84883

8.4. Mean-square Lyapunov function at $N = 3$

For the same reduced switched model with $N = 3$ and $p_{\mathrm{loss}} = 0.30$, the discrete stochastic Lyapunov equation $P - 0.7(\mathbb{A}_3^{\mathrm{rx}})^\top P \mathbb{A}_3^{\mathrm{rx}} - 0.3(\mathbb{A}_3^{\mathrm{loss}})^\top P \mathbb{A}_3^{\mathrm{loss}} = I_{15}$ has the solution $P \succ 0$ with $\lambda_{\min}(P) = 1.0000$ and $\lambda_{\max}(P) = 1473.65$. In the notation of Lemma 5.10, $c_Q = 6.7859 \times 10^{-4}$ and $\lambda_{\max}(P)/\lambda_{\min}(P) = 1473.65$, which gives the bound

$$\mathbb{E}\,\|\xi_k\|^2 \le 1473.65\,(1 - 6.7859 \times 10^{-4})^k\,\|\xi_0\|^2 . \tag{57}$$

For a unit initial state, this bound equals 1434.17 at $k = 40$ and 1395.76 at $k = 80$. It certifies mean-square decay, while the exact second moments are those of Table 9.

9. Conclusion

PF-DAIF separates state reconstruction from control preference in active inference for delayed networked feedback. An input-history predictor centres the state prior, aligned measurements with bounded adaptive precision supply the evidence, and the reference enters only a projected action evaluated at a forecast of the arrival state. The fixed-point analysis locates the estimation bias of target-centred AIF in the reference offset of its prior, and the augmented model of the sampled loop yields an exact spectral test for fixed delay together with mean-square conditions under switching delay and packet loss. One variational objective thus provides a reference-independent estimator with adaptive channel weights and a delay-compensating forecast action. Between packets, the

belief remains on the predictor instead of relaxing toward the reference, so the estimation error stays far below that of target-centred AIF, the tracking error of the linear benchmark is insensitive to packet loss, and the quadrotor estimate is insensitive to reference shifts. The forecast action removes the oscillatory instability of present-state feedback over the tested delay range, and the tracking accuracy matches Smith-predictor control at moderate delays. The study has several limitations. At long delays, the present-state action of the quadrotor and a predictor anchored at a belief one delay old accumulate delay and model error, and Smith-predictor control, which forecasts from its current observer state, tracks more accurately. A persistent sensor bias remains indistinguishable from a state offset for all estimators considered, and PF-DAIF incurs a larger inner-loop attitude error under moving attitude references. The mean-square analysis assumes independent switching and loss, the simulations use undelayed sensing, and the nonlinear and cascade conditions are local, with Lipschitz, invariance and small-gain constants still to be evaluated for the full quadrotor loop. Future work will address bias-aware sensing models, predictors anchored at fresher evidence, computable certificates for the full nonlinear loop and flight experiments.

Acknowledgements

This research was supported by the National Natural Science Foundation of China under Grants 62203358 and 62233014.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix A. Proofs

Appendix A.1. Proof of Theorem 3.1 and Lemma 3.2

Setting the gradient of Eq. (5) to zero gives the two stationary beliefs. Substituting $\hat{x}_p = x + e_p$ and $\tilde{y} = x + \delta_y$ and subtracting x gives Eqs. (6) and (7). For Eq. (8), write

$$W\Pi_\mu e_p = W^{1/2}(W^{1/2}\Pi_\mu W^{1/2})W^{-1/2}e_p.$$

The symmetric matrices $W^{1/2}\Pi_\mu W^{1/2}$ and $W^{1/2}\Pi_y W^{1/2}$ are positive definite and sum to I, so each has its spectrum in $(0,1)$. Hence $\left\|W\Pi_\mu e_p\right\|_{W^{-1}} \le \left\|e_p\right\|_{W^{-1}}$, and likewise for the sensing term. Lemma 3.2 is the scalar case.

Appendix A.2. Proof of Lemma 5.1

Subtracting the nominal predictor from the true flow on $[t-\hat{\tau}_u, t]$, the initial mismatch is $\mu(t-\hat{\tau}_u) - x(t-\hat{\tau}_u)$, and model, disturbance and noise mismatches enter additively under Assumption 2.4. A

delay shift perturbs the input seen by the nominal flow by at most $\kappa_\tau \sup_s |\tilde{\tau}_u(s)|$ over the window. Grönwall's inequality over a horizon of at most $\bar{\tau}_u$ gives Eq. (19), and solving $e^{L_x \bar{\tau}_u} \bar{e}_{\mathrm{est}} + (e^{L_x \bar{\tau}_u} - 1)\bar{q}_p / L_x \le \bar{e}_p^{\max}$ for $\bar{\tau}_u$ gives Eq. (20).

Appendix A.3. Proof of Theorem 5.3

Along Eq. (16) with $H(\mu)^\top$ in place of $C^\top$,

$$\dot{e}_\mu = -\Gamma_\mu \Pi_\mu e_\mu + \Gamma_\mu H(\mu)^\top \Pi_y \big(\tilde{y} - h(\mu)\big) - \dot{\hat{x}}_p.$$

Since $\tilde{y} - h(\mu) = \tilde{y} - h(\hat{x}_p) - \bar{H} e_\mu$, one obtains $\dot{e}_\mu = -\Gamma_\mu(\Pi_\mu + H(\mu)^\top \Pi_y \bar{H}) e_\mu + \delta_p$. Eq. (22) gives $\frac{d}{dt}\frac{1}{2}\left\|e_\mu\right\|^2 \le -\alpha_\mu \left\|e_\mu\right\|^2 + \left\|e_\mu\right\| \left\|\delta_p\right\|$, and the comparison lemma [40] yields Eq. (23).

Appendix A.4. Proof of Theorem 5.5

In the interior of the precision interval,

$$\dot{\mathscr{V}}_F = \nabla_\mu F^\top \dot{\mu} + \nabla_a F^\top \dot{a} + \nabla_{\gamma_y} F^\top \zeta_\gamma + \zeta_\gamma^\top \Gamma_\gamma^{-1} \dot{\zeta}_\gamma + \partial_t F.$$

The first two terms give $-\left\|\nabla_\mu F\right\|_{\Gamma_\mu}^2 - \left\|\nabla_a F\right\|_{\Gamma_a}^2$. Substituting Eq. (18), the terms $\nabla_{\gamma_y} F^\top \zeta_\gamma$ and $-\zeta_\gamma^\top \nabla_{\gamma_y} F$ cancel, and $-k_\gamma \zeta_\gamma^\top \Gamma_\gamma^{-1} \zeta_\gamma$ remains. Differentiating Eq. (15) with respect to $\hat{x}_p$, $\tilde{y}$ and $\hat{u}_p$ at fixed (μ, a, γ_y) gives $\partial_t F$. The sampled claims follow from $\mu_{k+1} - \mu^\star = (1 - \eta_\mu)(\mu_k - \mu^\star)$ and from the descent lemma for a quadratic of curvature Π_a.

Appendix A.5. Proof of Theorem 5.6 and Lemma 5.7

For a linear time-invariant difference equation, $\rho(\mathbb{A}_N) < 1$ is necessary and sufficient for exponential stability. With $V_k = \xi_k^\top P \xi_k$, Young's inequality with $\varepsilon > 0$ gives

$$\begin{aligned} V_{k+1} &\le (1+\varepsilon)\, \xi_k^\top (P - Q) \xi_k + (1 + \varepsilon^{-1}) \lambda_{\max}(\mathbb{G}_N^\top P \mathbb{G}_N) \bar{\omega}^2 \\ &\le (1+\varepsilon)(1 - c_0) V_k + (1 + \varepsilon^{-1}) \lambda_{\max}(\mathbb{G}_N^\top P \mathbb{G}_N) \bar{\omega}^2. \end{aligned}$$

The choice $\varepsilon = c_0/(2(1-c_0))$, or any ε if $c_0 = 1$, makes the first factor $1 - c_0/2$ and $1 + \varepsilon^{-1} = (2 - c_0)/c_0$. Iterating and using $\|\xi_k\|^2 \le V_k / \lambda_{\min}(P)$ gives Eq. (32). Lemma 5.7 applies the spectral condition to each integer delay separately.

Appendix A.6. Proof of Theorem 5.9

Let $V_k = \xi_k^\top P \xi_k$. Since the mode is independent of ξ_k, Eq. (33) gives $\mathbb{E}[V_{k+1} \mid \xi_k] \le V_k - \alpha \|\xi_k\|^2 \le (1 - c_V) V_k$ in the residual-free case, with $c_V = \alpha / \lambda_{\max}(P) \in (0, 1]$. Taking expectations gives $\mathbb{E} V_k \le (1 - c_V)^k \mathbb{E}[\xi_0^\top P \xi_0]$. With exogenous input, the cross and quadratic terms are bounded as in the proof of Theorem 5.6, which gives $\mathbb{E} V_{k+1} \le (1 - c_V/2) \mathbb{E} V_k + \sigma \bar{\omega}_2^2$ for some $\sigma > 0$. Under independent switching, $X_k = \mathbb{E}[\xi_k \xi_k^\top]$ evolves by the second-moment map $\mathscr{T}$ of Eq. (34), whose spectral radius is below one if and only if a $P \succ 0$ satisfying Eq. (33) exists [37].

Appendix A.7. Proof of Theorem 5.13

Let ω_{nl} collect the residuals not contained in the delayed state term, with $\|\omega_{\mathrm{nl}}\| \le \bar{\omega}_{\mathrm{nl}}$. Along the delayed loop, Young's inequality gives

$$\dot{V}_0 \le -\frac{c_3}{2}\|e\|^2 + \frac{c_4^2 L_\tau^2}{c_3}\sup_{s\in[t-\bar{\tau}_u,t]}\|e(s)\|^2 + c_5\|\omega_{\mathrm{nl}}\|^2$$

with $c_5 = c_4^2/c_3$. Under the Razumikhin condition $V_0(e(s)) \le qV_0(e(t))$ on $[t-\bar{\tau}_u, t]$, the quadratic bounds give $\|e(s)\| \le q_R\|e(t)\|$, and Eq. (38) yields $\dot{V}_0 \le -\alpha_e\|e\|^2 + c_5\|\omega_{\mathrm{nl}}\|^2$ with $\alpha_e = c_3/2 - c_4^2 L_\tau^2 q_R^2/c_3 > 0$. The Razumikhin theorem [34] gives local uniform ultimate boundedness with a class-$\mathscr{K}$ bound in $\bar{\omega}_{\mathrm{nl}}$, and local asymptotic stability when all residuals vanish.

Appendix A.8. Proof of Theorem 6.1

On the sublevel, $\|e_R\|^2 = \Psi_R(2-\Psi_R)$ [38], so $\frac{1}{2}\|e_R\|^2 \le \Psi_R \le \|e_R\|^2/(2-\psi_{\max})$. Consider $\mathscr{V}_R = \frac{1}{2}e_\Omega^\top J e_\Omega + k_R\Psi_R + \varepsilon_V e_R^\top J e_\Omega$, which is equivalent to $\|e_R\|^2 + \|e_\Omega\|^2$ for small $\varepsilon_V > 0$. Without residuals, $J\dot{e}_\Omega = -k_R e_R - k_\Omega e_\Omega$ and $\dot{e}_R = \mathscr{C}(\tilde{R})e_\Omega$ with $\left\|\mathscr{C}(\tilde{R})\right\| \le 1$, so

$$\dot{\mathscr{V}}_R = -k_\Omega\|e_\Omega\|^2 - \varepsilon_V k_R\|e_R\|^2 - \varepsilon_V k_\Omega e_R^\top e_\Omega + \varepsilon_V e_\Omega^\top \mathscr{C}(\tilde{R})^\top J e_\Omega,$$

and a small ε_V gives $\dot{\mathscr{V}}_R \le -\lambda_{\min}(Q_R)(\|e_R\|^2 + \|e_\Omega\|^2)$ with $Q_R \succ 0$. The residuals enter the torque additively, with a norm bounded by a class-$\mathscr{K}$ function of $\bar{e}_{\mathrm{att}} + \bar{\delta}_M$ by Lipschitz continuity of Eq. (42) on the sublevel, and Young's inequality absorbs them. On the positively invariant sublevel, the comparison lemma applied to this dissipation inequality gives the stated bound.

Appendix A.9. Proof of Lemma 6.2

From $\tilde{R} = R_d^\top R$, the identities $\dot{R} = R\Omega^\wedge$ and $\dot{R}_d = R_d\Omega_d^\wedge$ yield $\dot{\tilde{R}} = \tilde{R}\Omega^\wedge - \Omega_d^\wedge\tilde{R} = \tilde{R}(\Omega - \tilde{R}^\top\Omega_d)^\wedge$. Taking the skew-symmetric part and using the standard differential identity on $\mathrm{SO}(3)$ yields $\dot{e}_R = \mathscr{C}(\tilde{R})(\Omega - \tilde{R}^\top\Omega_d)$, and differentiating $\Psi_R = (3-\mathrm{tr}\,\tilde{R})/2$ gives the second formula of Eq. (45). With $J\dot{\Omega} = -k_R e_R - k_\Omega\Omega + d_M$ after gyroscopic cancellation, the derivative of $\mathscr{V}_0$ is Eq. (46). At a jump of R_d, the physical R and Ω are continuous, so the kinetic term is unchanged. Since $\Psi_R(R,R_d) = \frac{1}{2}(3-\mathrm{tr}(R_d^\top R))$, the Cauchy–Schwarz inequality in the Frobenius inner product gives $|\Delta\Psi_R| = \frac{1}{2}|\mathrm{tr}((R_d^+ - R_d^-)^\top R)| \le \frac{1}{2}\left\|R_d^+ - R_d^-\right\|_F\|R\|_F = \frac{\sqrt{3}}{2}\left\|R_d^+ - R_d^-\right\|_F$, which proves Eq. (47).

References

[1] Richard JP. Time-delay systems: an overview of some recent advances and open problems. Automatica 2003;39:1667–94.

[2] Yan XP, Liu FB, Zhang CH. Multiple stability switches and Hopf bifurcation in a damped harmonic oscillator with delayed feedback. Nonlinear Dyn 2020;99:2011–30.

[3] Zhu XL, Zhang X, Wei J, Lin H. Output-based dynamic event-triggered control for networked control systems with delays and packet losses without acknowledgements. IEEE Trans Autom Control 2023;68:7120–35.

[4] Zhang XM, Han QL, Ge X. A novel approach to H_∞ performance analysis of discrete-time networked systems subject to network-induced delays and malicious packet dropouts. Automatica 2022;136:110010.

[5] Feng Y, Tan Y, Gu G, Chen X. Remote state estimation with enhanced robustness in the presence of data packet dropouts. IEEE Trans Autom Control 2022;67:6552–66.

[6] Smith OJM. Closer control of loops with dead time. Chem Eng Prog 1957;53:217–9.
[7] Artstein Z. Linear systems with delayed controls: a reduction. IEEE Trans Autom Control 1982;27:869–79.
[8] Krstic M. Delay compensation for nonlinear, adaptive, and PDE systems. Boston: Birkhäuser; 2009.
[9] González A, García P. Output-feedback anti-disturbance predictor-based control for discrete-time systems with time-varying input delays. Automatica 2021;129:109627.
[10] Bresch-Pietri D, Krstic M. Delay-adaptive predictor feedback for systems with unknown long actuator delay. IEEE Trans Autom Control 2010;55:2106–12.
[11] Karafyllis I. Stabilization by means of approximate predictors for systems with delayed input. SIAM J Control Optim 2011;49:1100–23.
[12] Espitia N, Perruquetti W. Predictor-feedback prescribed-time stabilization of LTI systems with input delay. IEEE Trans Autom Control 2022;67:2784–99.
[13] Fang Q, Zhang Z. Inexact predictor feedback for multi-input nonlinear systems with distinct input delays. Automatica 2024;159:111399.
[14] Kong S, Bresch-Pietri D. Prediction-based controller for linear systems with stochastic input delay. Automatica 2022;138:110149.
[15] Cui L, Pang B, Krstic M, Jiang ZP. Learning-based adaptive optimal control of linear time-delay systems: a value iteration approach. Automatica 2025;171:111944.
[16] Sharma M, Kar I. Control of a quadrotor with network induced time delay. ISA Trans 2021;111:132–43.
[17] Liu K, Wang R, Zheng S, Dong S, Sun G. Fixed-time disturbance observer-based robust fault-tolerant tracking control for uncertain quadrotor UAV subject to input delay. Nonlinear Dyn 2022;107:2363–90.
[18] Kang Y, Luo D, Xin B, Cheng J, Yang T, Zhou S. Robust leaderless time-varying formation control for nonlinear unmanned aerial vehicle swarm system with communication delays. IEEE Trans Cybern 2023;53:5692–705.
[19] Fridman E. Introduction to time-delay systems: analysis and control. Systems & Control: Foundations & Applications. Cham: Birkhäuser; 2014.
[20] Lee SH, Park M, Kwon OM. Improvements on stability criteria for linear systems with a time-varying delay via novel delay-dependent Lyapunov functionals. ISA Trans 2024;152:269–76.
[21] Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci 2010;11:127–38.
[22] Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active inference: a process theory. Neural Comput 2017;29:1–49.
[23] Friston KJ. Variational filtering. NeuroImage 2008;41:747–66.
[24] Buckley CL, Kim CS, McGregor S, Seth AK. The free energy principle for action and perception: a mathematical review. J Math Psychol 2017;81:55–79.
[25] Baltieri M, Buckley CL. PID control as a process of active inference with linear generative models. Entropy 2019;21:257.
[26] Meera AA, Wisse M. Free energy principle based state and input observer design for linear systems with colored noise. In: Proc Am Control Conf; 2020, p. 5052–8.
[27] Pezzato C, Ferrari R, Hernández Corbato C. A novel adaptive controller for robot manipulators based on active inference. IEEE Robot Autom Lett 2020;5:2973–80.
[28] Meo C, Franzese G, Pezzato C, Spahn M, Lanillos P. Adaptation through prediction: multisensory active inference torque control. IEEE Trans Cogn Dev Syst 2023;15:32–41.
[29] Bos F, Meera AA, Benders D, Wisse M. Free energy principle for state and input estimation of a quadcopter flying in wind. In: Proc IEEE Int Conf Robot Autom; 2022, p. 5389–95.
[30] Baioumy M, Pezzato C, Ferrari R, Hawes N. Unbiased active inference for classical control. In: Proc IEEE/RSJ Int Conf Intell Robots Syst; 2022, p. 12787–94.
[31] Ji M, Lyu Y, Pan K, Zhang X, Pan Q, Li Y, et al. Delay feedback active inference based on predicted states for uncertain systems with input delay. In: Proc 62nd IEEE Conf Decis Control; 2023, p. 926–31.
[32] Ji M, Pan K, Zhang X, Pan Q, Dai X, Lyu Y. Integration of sense and control for uncertain systems based on delayed feedback active inference. Entropy 2024;26:990.
[33] Gajbhiye S, Cabecinhas D, Silvestre C, Cunha R. Geometric finite-time inner-outer loop trajectory tracking control strategy for quadrotor slung-load transportation. Nonlinear Dyn 2022;107:2291–308.
[34] Hale JK, Verduyn Lunel SM. Introduction to functional differential equations. New York: Springer; 1993.

[35] Insperger T, Stépán G. Semi-discretization for time-delay systems: stability and engineering applications. New York: Springer; 2011.
[36] Hu Z, Deng F, Su Y, Luo S, Liu Y, Chen K. Stability analysis and synthesis for networked systems with noisy sampling intervals, random transmission delays, and packet dropouts. IEEE Trans Autom Control 2024;69:1273–80.
[37] Costa OLV, Fragoso MD, Marques RP. Discrete-time Markov jump linear systems. London: Springer; 2005.
[38] Lee T, Leok M, McClamroch NH. Geometric tracking control of a quadrotor UAV on SE(3). In: Proc 49th IEEE Conf Decis Control; 2010, p. 5420–5.
[39] Jiang ZP, Teel AR, Praly L. Small-gain theorem for ISS systems and applications. Math Control Signals Syst 1994;7:95–120.
[40] Khalil HK. Nonlinear systems. 3rd ed. Upper Saddle River: Prentice Hall; 2002.