CONSENSUS ALGORITHMS FOR A MULTIAGENT OPERATING SYSTEM

Evgeny Krokhalev†
Sirius University of Science and Technology
Sirius, Krasnodar region, Russia
krohalev.em@talantiuspeh.ru

Ivan Arkhipov*
St. Petersburg State University
St. Petersburg, Russia
ivan.arkhipov@spbu.ru

Oleg Granichin*
St. Petersburg State University
St. Petersburg, Russia
o.granichin@spbu.ru

Elizaveta Tarasova*
St. Petersburg State University
St. Petersburg, Russia
elizaveta.tarasova@spbu.ru

Article history:
Received 14.06.2026, Accepted 26.06.2026

Abstract

Edge multiagent systems call for operating-system-level agreement primitives that tolerate peer failures, lossy and unsynchronized messaging, and heterogeneous local state, rather than relying on a permanent timing master. Neighbor exchange is modeled as a time-delay multi-agent system under bounded communication delays. This article presents two complementary building blocks—monotone logical-time synchronization via max/SoftMax consensus with PI-style implementation, and consensus-based accelerated distributed simultaneous perturbation stochastic approximation (ADSPSA) for cooperative tracking under zeroth-order oracles—together with a multimodal composition principle: logical-time and spatial coordinates are stacked per agent, coupled through multimodal losses and a block preconditioner on the ADSPSA channel. Formal convergence analysis for the logical-time synchronization primitive is developed in companion works on monotonic logical time (an accepted ApPLIED 2026 contribution and a manuscript submitted to IEEE Access); fully coupled multimodal validation—joint logical-time and tracking experiments under delays and losses—remains outside the scope of this paper. For the Ψ-preconditioned recursion we develop a majorization framework that yields an asymptotic upper bound on the tracking-error covariance under explicit smoothness, centering, and spectral-contraction assumptions. Numerical experiments on a heterogeneous quadratic surrogate—isolating the tracking block but using a time/tracking block structure for Ψ—illustrate how an empirically tuned preconditioner can lower the asymptotic error level relative to the identity map, sometimes at the cost of slower transient decay.

Key words

Consensus, distributed algorithms, logical clock synchronization, multiagent operating systems, accelerated distributed simultaneous perturbation stochastic approximation (ADSPSA), time-delay systems.

1 Introduction

Computing platforms are becoming smaller, cheaper, and faster at the edge, which widens the design space for intelligent sensing, control, and data processing in the field [Baek et al., 2025]. Yet realizing this potential in practice is increasingly a *systems* problem rather than a raw-performance problem: loosely coupled collections of devices can often outperform a single centralized stack, but they are substantially harder to design, verify, and evolve. Monolithic applications tend to accumulate global modes, ad hoc feature flags, and brittle synchronization; as communication, timing, and partial failures enter the picture, the resulting coupling creates bottlenecks, latency, and operational drag that work against the very distribution that the hardware now invites [Vogels, 2009]. In other words, the pace at which applications can be composed often lags the pace

†Sections 1–3 were supported by the grant of the state program of the Sirius Federal Territory "Scientific and technological development of the Sirius Federal Territory" (Agreement No. 33-03 dated 01.08.2025).

*Sections 4–5 of this work were supported by St. Petersburg State University, project No. 121061000159-6.

at which hardware enables new deployments. A natural response is to structure computation as a *multiagent* system in which each agent runs autonomously and exchanges information with peers or with a coordinator only when needed, instead of constantly centralizing state. Such architectures are powerful, but programming them directly—as a zoo of bespoke protocols, schedulers, and retry policies—does not scale [Rice and Verhaegen, 2009; Shukla and Parker, 2025]. This motivates a unifying substrate: a common operating system layer that provides naming, messaging, timing, and resource isolation so that distributed agents can be developed as first-class citizens rather than as special cases bolted onto a single program.

A first basic service such an operating system must supply is *time*: agents often need a shared temporal reference so that actions or observations can be aligned. Establishing agreement purely by exchanging timestamps over the network is always entangled with propagation delay, jitter, and asymmetric paths, so "common time" is not a free primitive; neighbor exchanges are naturally modeled as a *time-delay system* in which messages carry stale information under bounded, possibly asymmetric communication delays (Section 2.1, (2)). A practical alternative is to anchor clocks to a coordinator when connectivity and trust allow it, which simplifies reasoning and scheduling. The difficulty is that many deployments are intermittently connected or physically constrained, so a design that *requires* a continuously available central time source becomes fragile: when that assumption breaks, distributed routines that implicitly depended on it can fail in subtle ways. Robust coordination therefore needs explicit time semantics, because many tasks genuinely require simultaneity—for example, multiple manipulators lifting opposite edges of a load in unison, or networked sensors taking measurements on a common schedule for fusion and calibration.

Many cooperative problems can, in principle, be viewed as solving a single joint objective: gather raw observations at a hub, run an optimizer or estimator, and broadcast instructions or state estimates back to the agents. That pattern is attractive when the model is stationary and the network is generous, but field systems increasingly operate at time scales where sensing, inference, and actuation must keep pace with the physical process. Links may be absent, rate-limited, or lossy, while safety and throughput still demand an answer *now*. The engineering implication is that decisions must often be formed *in place*, using local information and whatever timely messages are actually available, rather than after a full round-trip through a distant compute pool. The open question is how to realize such behavior reliably: what algorithms and interfaces let a network of agents approximate a coherent global computation without pretending that the center always sees everything first.

In realistic deployments, what must be kept consistent is rarely a single scalar: agents fuse *heterogeneous* quantities—delays and clock offsets alongside geometric or kinematic state—and those quantities interact in the same closed-loop controller. A tempting engineering shortcut is to run several specialized synchronization or estimation routines in parallel, each tuned to one modality, but on resource-constrained nodes that approach multiplies message traffic, bookkeeping, and worst-case timing, while making coupling between modalities an afterthought. A more scalable pattern is to expose one iterative procedure whose iterates encode the joint quantities of interest, so communication and computation amortize across modalities rather than competing for the same budget. A concrete instance is simultaneous alignment of a common time base *and* cooperative localization or tracking: the same interaction pattern that steers clocks should also be able to steer spatial estimates when both are safety-critical.

Scope of this article. *Physical context and area.* The motivating setting is a network of physical sensors that observe moving targets through noisy range-type measurements (Section 4.1, (22)) while exchanging data over channels with propagation delay and loss (Section 2.1). Such fleets arise in distributed sensing, cooperative localization, and synchronized sampling for fusion—problems at the interface of measurement physics, stochastic observation models, and cyber-physical coordination. The paper advances this interface in two directions: monotone logical-time synchronization tied to hardware clock drift and delayed timestamps (Section 3), and consensus-based cooperative tracking of spatial target state under zeroth-order measurement noise (Sections 4–5).

Physical content of the results. The manuscript does not propose a new fundamental physical law; rather, it supplies quantitative building blocks for *physically interpretable* distributed agreement: (i) a time-delay network model and delay-resilient logical-time updates that preserve a usable temporal ordering for physical events; (ii) an asymptotic upper bound on the tracking-error covariance matrix for preconditioned ADSPSA, i.e., a certifiable limit on spatial estimation error under stated smoothness and noise assumptions; and (iii) a multimodal composition rule (Section 3.7) in which time-like and spatial physical coordinates co-evolve under one communication budget. Fully coupled multimodal validation on moving targets with delays remains future work.

Organization. The manuscript is deliberately organized around two reusable blocks rather than one fully coupled deployment. Section 3 records the logical-time synchronization primitive at the level needed for OS-style composition—max/SoftMax consensus with PI-style realization and delay-resilient extensions—while formal convergence analysis of consensus-based tracking of logical time functions is developed in [Arkhipov et al., 2026a]; a related manuscript on monotonic logical-time synchronization is submitted to *IEEE Ac-*

cess [Arkhipov et al., 2026b]. The Ψ-preconditioned ADSPSA block (Sections 4–5) carries the main theoretical and numerical contribution here: an asymptotic covariance majorization framework together with preconditioner tuning on a synthetic quadratic benchmark. Section 3.7 formalizes the multimodal stack—stacked state, multimodal losses, block preconditioning, and a unified oracle—while convergence of the logical-time block is treated in [Arkhipov et al., 2026a; Arkhipov et al., 2026b]; validation of the fully coupled multimodal stack under realistic delays and losses remains future work (Section 6). Section 5 therefore isolates the tracking block but uses a heterogeneous block structure for Ψ as a surrogate for multimodal balancing.

1.1 Related work

Work on *multiagent operating systems* is still fragmented: most contributions target concrete application patterns and ship as middleware on a conventional OS, rather than revisiting scheduling, IPC, or time semantics for agent fleets. A representative system-level proposal is KAOS [Zhuo et al., 2025], which builds on Kylin, uses LLM-equipped agents with management roles and vertical collaboration, and studies shared-resource scheduling; it reports gains in scenario-based evaluations. Such work clarifies OS-level policy for agent interaction, but leaves aside the distributed algorithms needed for tightly coupled physical coordination—notably joint agreement on logical time and cooperative tracking over lossy, unsynchronized links—which we take up below.

Time synchronization is often engineered around a designated master, grandmaster, or fusion center, which is a single point of failure when *any* agent in a fleet may drop out or become unreachable. Decentralized alternatives instead propagate timing corrections through local interactions and graph-based consensus. Wang *et al.* [Wang et al., 2024] accelerate average-consensus clock updates in industrial wireless sensor networks by synthesizing *virtual multihop* links so that effective coupling resembles a denser topology. A companion line [Wang et al., 2025] couples event-triggered transmissions with average consensus to save bandwidth while explicitly compensating time-varying clock skew across nodes. Ullah *et al.* [Ullah et al., 2024] address uncertain higher-order nonlinear agent dynamics with a neuro-adaptive non-singular terminal sliding-mode law that guarantees distributed fixed-time agreement of timing-related states. He *et al.* [He et al., 2026] replace hand-crafted update rules by an unsupervised dual-attention network (DAUNet) that learns clock corrections from multi-agent wireless measurements. These contributions sharpen leaderless *hardware* clocking, but they do not subsume the semantics of lossy, unsynchronized messaging or joint synchronization of heterogeneous state such as logical time together with cooperative tracking. Arkhipov *et al.* [Arkhipov et al., 2026a] analyze consensus-based tracking of logical time functions in multi-agent networks; we refer the reader there for convergence guarantees and do not repeat that analysis here. A related manuscript on monotonic logical-time synchronization in decentralized networks is submitted to *IEEE Access* [Arkhipov et al., 2026b]. The present article builds on that logical-time perspective and couples it, at the modeling level, with ADSPSA-based cooperative tracking under a multiagent operating-system abstraction; fully coupled multimodal validation remains future work.

A foundational line for decentralized agreement under noise, switching topology, and measurement delays is the *local voting protocol* (LVP) of Amelina and Fradkov [Amelina and Fradkov, 2012; Amelina et al., 2015]. LVP achieves approximate consensus through neighbor-only updates with nonvanishing step sizes and is analyzed via averaged models; representative applications include load balancing in stochastic dynamic networks [Amelina et al., 2015; Amelina, 2013] and distributed node scheduling in multihop wireless networks [Vergados et al., 2018]. The consensus couplings used in this article inherit this LVP viewpoint—local exchanges, noisy and delayed measurements, and time-varying graphs—while extending it to monotone logical-time synchronization and derivative-free cooperative tracking.

The distributed *computational* core we build on is an accelerated consensus / accelerated distributed simultaneous perturbation stochastic approximation (ADSPSA) family tailored to multiagent networks. Erofeeva *et al.* [Erofeeva et al., 2024] introduce an accelerated decentralized load-balancing protocol in which agents exchange local mass estimates so that a conserved load spreads quickly over the communication graph. In the tracking setting, the same group proposes accelerated ADSPSA updates that follow time-varying targets under unknown-but-bounded disturbances using only two randomized function evaluations per step [Erofeeva et al., 2022]. Their subsequent accelerated consensus-based ADSPSA algorithm fuses neighborhood averaging with these derivative-free increments for multisensor multitarget tracking [Erofeeva et al., 2026]. Closely related work by Chernov *et al.* [Chernov et al., 2025a; Chernov et al., 2025b] develops accelerated SPSA-based consensus algorithms that can be viewed as accelerated variants of the basic consensus–SPSA methodology of [Erofeeva et al., 2026] for the multisensor multitarget problem; in particular, they treat mutual IoT device positioning as an application of the same derivative-free consensus framework. We view these mechanisms as reusable OS-level primitives—fast graph consensus for agreement, ADSPSA for gradient-free adaptation, and their combination for cooperative estimation.

Recent issues of *Cybernetics and Physics* address closely related problems at the cybernetics–physics interface pursued here. Erofeeva and Parsegov [Erofeeva and Parsegov, 2024] derive Laplacian spectral bounds for hierarchical graphs, emphasizing quantities that govern stability and convergence rates in decentralized multi-agent control and optimization—the same graph-Laplacian structure that enters our consensus regularizer (25). Imangazieva [Imangazieva, 2024] studies robust control of nonlinear physical plants with distributed state delay, complementing the time-delay neighbor-exchange model of Section 2.1. Tarasova and Moseiko [Tarasova and Moseiko, 2025] combine decentralized SPSA with consensus to adapt task-time predictions in multi-agent systems, illustrating within this journal the consensus-plus-derivative-free paradigm that underlies the ADSPSA building block developed below.

1.2 Contribution

Taken together, the prior art motivates multiagent operating systems and supplies useful consensus, clocking, and ADSPSA ingredients, yet three coupled aspects remain comparatively underdeveloped: *algorithmic* services aimed explicitly at an OS-style agent fleet (rather than a single application or a one-off sensor network), *monotone logical time* as a first-class synchronization target (as opposed to wall-clock skew alone), and *multimodal* distributed state in which time-like quantities can co-evolve with cooperative physical estimates under the same communication budget.

This paper contributes two reusable building blocks and a composition principle:

1. **Logical-time synchronization (Section 3).** We summarize distributed consensus–tracking procedures for monotone logical time: agents interact only through local channels that are lossy and unsynchronized (including delayed observations), without a permanently designated timing master. Max and SoftMax consensus operators, together with a PI-style realization and delay-resilient extensions, provide OS-ready synchronization primitives. Formal convergence guarantees for the logical-time tracking recursion appear in [Arkhipov et al., 2026a; Arkhipov et al., 2026b]; the present section focuses on operator choices and implementation details needed for composition.
2. **ADSPSA cooperative tracking and covariance majorization (Sections 4–5).** We specify a Ψ-preconditioned consensus-based ADSPSA recursion and develop a majorization framework that yields an asymptotic upper bound on the tracking-error covariance under explicit assumptions (smooth linearization, centered innovations, stationary majorants, and $\|A\|_2 < 1$). Numerical experiments on a synthetic quadratic benchmark study preconditioner tuning via empirical majorants.
3. **Multimodal composition (Section 3.7).** Logical-time and tracking coordinates are stacked per agent as $\theta_t^i = \mathrm{col}(\tilde{x}_t^i, r_t^{i,\mathrm{loc}})$, with multimodal losses $F_t^i = F_t^{i,\mathrm{time}} + F_t^{i,\mathrm{track}}$ and block preconditioning $\Psi = \mathrm{blkdiag}(\Psi_{\mathrm{time}}, \Psi_{\mathrm{track}})$. A single ADSPSA recursion updates all modalities under one communication budget. The present manuscript develops the two blocks separately, states the composition rule explicitly, and leaves fully coupled multimodal experiments and closed-loop analysis to future work.

The technical sections that follow instantiate these contributions: Section 3 presents monotonic logical-time synchronization and Section 3.7 formalizes multimodal composition; Section 4 develops the ADSPSA majorization analysis; Section 5 reports numerical evidence for preconditioner tuning.

2 Preliminaries

2.1 Communication graph, neighbors, and delays

We consider a network of n agents indexed by $\mathcal{N} \triangleq \{1, \dots, n\}$. Information exchange is described by a (possibly time-varying) directed graph $G_t = (\mathcal{N}, E_t)$, where an edge $(j, i) \in E_t$ means that agent i receives a message from agent j at time t. The corresponding in-neighbor set is

$$\mathcal{N}_t^i \triangleq \{j \in \mathcal{N} : (j, i) \in E_t\}. \tag{1}$$

Packet losses are modeled implicitly by missing edges (i.e., $(j, i) \notin E_t$ means that the message is not available at time t).

When messages are delayed, agent i's observation of a scalar/vector state broadcast by agent j corresponds to a past value. We write this generically as

$$y_t^{i,j} = x_j(t - \tau_t^{i,j}) + \nu_t^{i,j}, \qquad j \in \mathcal{N}_t^i, \tag{2}$$

where $\tau_t^{i,j} \in \mathbb{Z}_{\geq 0}$ is the (possibly asymmetric) communication delay and $\nu_t^{i,j}$ aggregates measurement/quantization noise. When needed, we assume a uniform delay bound $\tau_t^{i,j} \leq \tau_{\max}$.

Time-delay network model. Equations (1)–(2) describe a *time-delay multi-agent system*: agent i does not observe $x_j(t)$ directly, but only a delayed and noisy sample $x_j(t - \tau_t^{i,j})$. Delays may be asymmetric ($\tau_t^{i,j} \neq \tau_t^{j,i}$) and are assumed uniformly bounded when needed ($\tau_t^{i,j} \leq \tau_{\max}$). Packet losses are modeled by the time-varying graph G_t. This abstraction is central to logical-time synchronization (Section 3) and to the multimodal oracle (Section 3.7); the ADSPSA majorization study (Sections 4–5) uses a delay-free quadratic benchmark, while convergence under delays for the logical-time recursion is established in [Arkhipov et al., 2026a; Arkhipov et al., 2026b].

2.2 Laplacian coupling and stacking notation

For consensus-type couplings we use graph Laplacians. For an undirected weighted graph with weights $a_{ij} \geq 0$, the Laplacian is $L = D - A$, where $A = [a_{ij}]$ and $D = \mathrm{diag}(\sum_j a_{ij})$. When the graph varies with time we write L_t; for an averaged (time-invariant) Laplacian we write L_{av}.

For a vector $u^i \in \mathbb{R}^d$ held by each agent i, we use the stacked notation $\bar{u} \triangleq \mathrm{col}(u^1, \dots, u^n) \in \mathbb{R}^{nd}$. The Kronecker product $L \otimes I_d$ acts on stacked vectors in the standard way.

3 Synchronization of monotonic logical time

This section specifies the logical-time synchronization primitive at the algorithmic level; formal convergence analysis of the underlying consensus–tracking recursion is given in [Arkhipov et al., 2026a; Arkhipov et al., 2026b].

3.1 Problem statement

We consider a network of n nodes (agents) $\mathcal{N} = \{1, \dots, n\}$ that exchange messages over a time-varying communication graph $G_t = (\mathcal{N}, E_t)$, subject to losses and delays. Each node i has a local physical clock $x_i(t)$ (a hardware counter) whose rate may differ across nodes (frequency drift). Neighbor observations arrive with (possibly asymmetric) delay and measurement noise. The graph, neighbor sets, and the generic delay model are summarized in Section 2.1. Logical-time synchronization is thus a consensus–tracking problem on a time-delay network: each observation $y_t^{i,j}$ follows (2) with x_j replaced by the logical clock $\tilde{x}^j$.

The goal is to maintain at each node a *logical time* $\tilde{x}_i(t)$ that:

1. is **monotone** (never decreases):
$$\tilde{x}_i(t+1) \geq \tilde{x}_i(t), \quad \forall i \in \mathcal{N},\ \forall t \geq 0. \tag{3}$$
2. achieves **agreement** across nodes (small skew) despite delays and losses,
3. tracks a **virtual network time** $x^*(t)$ that acts as a global time reference.

We adopt the maximum-based virtual time:
$$x^*(t) \triangleq \max_{i \in \mathcal{N}} \tilde{x}_i(t). \tag{4}$$
This choice is natural for systems where logical time is not allowed to run backwards: lagging nodes should catch up, rather than forcing leaders to slow down.

3.2 Consensus-based tracking with monotonicity

Synchronization is cast as distributed *tracking* of $x^*(t)$ using only local neighbor exchanges. Let $\mathcal{N}_t^i$ be the set of neighbors from which node i successfully receives messages at time t. Let $y_t^{i,j}$ be node i's observation of neighbor j's logical time at time t (including delay and noise effects).

The generic local update (consensus operator) takes the form:
$$\tilde{x}_i(t+1) = \mathcal{C}_i\Big(\tilde{x}_i(t), \{y_t^{i,j}\}_{j \in \mathcal{N}_t^i}\Big). \tag{5}$$
Classic average-consensus converges to the mean of initial values. For logical clocks this is problematic because nodes above the mean must *decrease* their values, violating monotonicity.

3.3 Maximum consensus

The most direct way to guarantee monotonicity is to use the hard maximum:
$$\tilde{x}_i(t+1) = \max\Big\{\tilde{x}_i(t),\ \max_{j \in \mathcal{N}_t^i} y_t^{i,j}\Big\}. \tag{6}$$
By construction, $\tilde{x}_i(t+1) \geq \tilde{x}_i(t)$. The downside is **non-smoothness**: under noisy measurements, the max operator may switch abruptly between neighbors, degrading robustness and interacting poorly with control-loop implementations.

3.4 SoftMax consensus

To preserve monotonicity while making the update smooth, we use the log-sum-exp (SoftMax) approximation:
$$\mathrm{SoftMax}_T(z_1, \dots, z_m) = T \cdot \log\Big(\frac{1}{m}\sum_{k=1}^{m} \exp(z_k/T)\Big), \quad T > 0. \tag{7}$$
The corresponding update at node i is:
$$\tilde{x}_i(t+1) = \mathrm{SoftMax}_T\Big(\tilde{x}_i(t), \{y_t^{i,j}\}_{j \in \mathcal{N}_t^i}\Big). \tag{8}$$

Key properties for logical time synchronization include:

1. **Max approximation:**
$$\mathrm{SoftMax}_T(z_1, \dots, z_m) \leq \max_k z_k, \quad \max_k z_k \leq \mathrm{SoftMax}_T(z_1, \dots, z_m) + T \log m. \tag{9}$$
2. **ε-monotonicity:** Although $\mathrm{SoftMax}_T$ is increasing in each argument, a direct SoftMax update does not guarantee strict monotonicity $\tilde{x}_i(t+1) \geq \tilde{x}_i(t)$ when neighbor observations can fall below the current logical time. From (9), with M denoting the number of SoftMax inputs at node i (the local value plus available neighbor observations), one obtains the ε-monotonicity guarantee
$$\tilde{x}_i(t+1) \geq \tilde{x}_i(t) - \varepsilon, \qquad \varepsilon = T \log M. \tag{10}$$
3. **Tunable trade-off:** smaller T brings $\mathrm{SoftMax}_T$ closer to $\max$ and tightens the monotonicity margin $\varepsilon = T \log M$, while larger T yields smoother (often more noise-robust) updates at the cost of a larger admissible regression.

For numerical stability, one uses the standard log-sum-exp shift with $c = \max\{\tilde{x}_i(t), \max_j y_t^{i,j}\}$.

3.5 Implementation via PI control

In implementations, $\tilde{x}_i(t)$ is often realized as the physical clock plus controlled phase and rate corrections. Let $\Delta x_i^{\text{phys}}(t) \triangleq x_i(t+1)-x_i(t)$ be the physical increment. Each node maintains:

$\gamma_i(t)$: phase correction (proportional action),
$\beta_i(t)$: rate correction (integral action).

Let the consensus operator (max or SoftMax) define a correction term $\mathcal{C}_i(t)$, and define the tracking error as

$$e_i^{\text{ts}}(t) = \mathcal{C}_i(t) - \tilde{x}_i(t).$$

The logical clock is realized as a corrected physical counter,

$$\tilde{x}_i(t) = x_i(t) + \gamma_i(t), \tag{11}$$

while the physical tick is rate-corrected by the integral state,

$$x_i(t+1) = x_i(t) + \big(1 + \beta_i(t)\big)\, \Delta x_i^{\text{phys}}(t). \tag{12}$$

With $e_i^{\text{ts}}(t)$ in place of $e_i(t)$, a typical PI update is:

$$\gamma_i(t+1) = k_P\, e_i^{\text{ts}}(t), \tag{13}$$

$$\beta_i(t+1) = \beta_i(t) + k_I\, \frac{e_i^{\text{ts}}(t)}{\Delta x_i^{\text{phys}}(t)}. \tag{14}$$

Equations (8)–(14) (or their max-consensus counterpart (6)) define a closed-loop logical-time update: the consensus operator sets the target $\mathcal{C}_i(t)$, the PI states steer $\tilde{x}_i(t)$ toward that target without regressing below the SoftMax monotonicity margin (10), and (12) adjusts the hardware tick rate. The gains $k_P, k_I > 0$ tune convergence and robustness. In practice $\beta_i(t)$ is saturated to prevent runaway rate corrections.

3.6 Delay-resilient extension

In time-delay systems with asymmetric link delays, the tracking error $e_i^{\text{ts}}(t)$ may acquire a nonzero mean bias, causing the integral state $\beta_i(t)$ to drift and eventually hit hard limits. A delay-resilient extension addresses this via three mechanisms:

SNR-gated input correction. Each node estimates the bias over a sliding window of recent errors (e.g., a half-sample range estimator) and applies a correction to the consensus inputs only when the bias is statistically detectable (SNR gating), to avoid interfering with transients.

Beta consensus coupling. To suppress common-mode drift, the rate-correction states are coupled via a graph-Laplacian term:

$$\begin{aligned}\beta_i(t+1) = \beta_i(t) + k_I\, &\frac{e_i^{\text{ts}}(t)}{\Delta x_i^{\text{phys}}(t)} \\ &- \lambda_{\text{BC}} \sum_{j\in\mathcal{N}_t^i} \big(\beta_i(t) - \beta_j(t)\big), \\ &\lambda_{\text{BC}} > 0.\end{aligned} \tag{15}$$

Smooth stabilization outside a dead-zone. Instead of hard clipping, one can use a dead-zone leaky restoring force outside $|\beta_i| \le d$:

$$\begin{aligned}\beta_i(t+1) \leftarrow \beta_i(t+1) - \lambda_L\, \max(|\beta_i(t)| - d, 0)\, \beta_i(t),\\ \lambda_L > 0.\end{aligned} \tag{16}$$

This prevents integrator collapse over long horizons.

3.7 Multimodal composition

The logical-time primitive above is the *time modality* of a larger agreement service. In field deployments, agents must often keep consistent not only monotone logical clocks but also cooperative physical estimates—for example, target positions that enter the same closed-loop controller as scheduling and ordering decisions. We call such a joint state *multimodal*: its coordinates belong to qualitatively different modalities with different units, noise levels, and local dynamics, yet they are updated under a *single* communication budget.

Stacked per-agent state. Fix a time modality of dimension $d_{\text{time}} \ge 1$ and a tracking modality of dimension $d_{\text{track}} \ge 1$. Agent i maintains

$$\theta_t^i \triangleq \text{col}\big(\tilde{x}_t^i,\, r_t^{i,\text{loc}}\big) \in \mathbb{R}^d, \qquad d \triangleq d_{\text{time}} + d_{\text{track}}, \tag{17}$$

where $\tilde{x}_t^i \in \mathbb{R}^{d_{\text{time}}}$ is the logical-time block from Section 3 (typically $d_{\text{time}} = 1$) and $r_t^{i,\text{loc}} \in \mathbb{R}^{d_{\text{track}}}$ is agent i's local estimate of the cooperative tracking variables of Section 4.1 (with $d_{\text{track}} = mr$ from (21) when m targets in $\mathbb{R}^r$ are tracked). Stack all agents as usual: $\bar{\theta}_t = \text{col}(\theta_t^1, \dots, \theta_t^n) \in \mathbb{R}^{nd}$.

Shared neighbor exchanges. Both modalities reuse the same graph G_t, in-neighbor sets $\mathcal{N}_t^i$, and delay model (2) from Section 2.1. A single message round therefore delivers, for each $j \in \mathcal{N}_t^i$, delayed logical-time observations $y_t^{i,j}$ (as in Section 3.2) together with any co-observed tracking information available to sensor i. Running separate synchronization and estimation services in parallel would duplicate this traffic and decouple modalities at the protocol level; stacking (17) amortizes communication and computation across modalities.

Multimodal local losses. The time modality enters through mismatch to the virtual network time $x^*(t)$ and to delayed neighbor timestamps. A smooth surrogate is

$$F_t^{i,\mathrm{time}}(\bar{\theta}) \triangleq \tfrac{1}{2}\left\|\tilde{x}_t^i - x^*(t)\right\|^2 + \tfrac{\kappa_{\mathrm{ts}}}{2}\sum_{j\in\mathcal{N}_t^i}\left\|\tilde{x}_t^i - y_t^{i,j}\right\|^2, \qquad \kappa_{\mathrm{ts}}>0, \tag{18}$$

where $y_t^{i,j}$ follows (2) with x_j replaced by $\tilde{x}^j$. Section 3 gives an alternative *dedicated* realization of the time modality—SoftMax/PI updates that enforce monotone logical time directly—which can run alongside or inform the choice of $F_t^{i,\mathrm{time}}$. In the fully stacked ADSPSA formulation below, time coordinates are updated by the same recursion as the tracking block; the two presentations share the neighbor-exchange semantics but need not coincide step-for-step unless the oracle is engineered accordingly. The tracking modality contributes $F_t^{i,\mathrm{track}}(\bar{\theta})$, built from the relative measurements of Section 4.1 (e.g., squared distances (22) and co-observer exchanges). The multimodal objective aggregates both parts:

$$F_t^i(\bar{\theta}) \triangleq F_t^{i,\mathrm{time}}(\bar{\theta}) + F_t^{i,\mathrm{track}}(\bar{\theta}). \tag{19}$$

Unified zeroth-order oracle and ADSPSA recursion. Agents do not evaluate (19) explicitly. As in Section 4.1, each agent observes a single noisy zeroth-order oracle $y_t^i = f_{\xi_t^i}(\bar{\theta}) + v_t^i$ whose randomness ξ_t^i now encodes cooperative target triples *and* the choice of delayed timestamp observations. One outer iteration of the ADSPSA recursion of Section 4.2 then updates all coordinates of (17) using the same probing, consensus coupling, and graph Laplacian $L_{\mathrm{av}} \otimes I_d$ in (25).

Block preconditioning across modalities. The fixed matrix $\Psi \succ 0$ in (27) left-preconditions only the SPSA channel in (33). In multimodal operation it is natural to choose a block structure

$$\Psi = \mathrm{blkdiag}\big(\Psi_{\mathrm{time}},\, \Psi_{\mathrm{track}}\big), \qquad \Psi_{\mathrm{time}} \succ 0,\ \Psi_{\mathrm{track}} \succ 0, \tag{20}$$

so that time-like and spatial coordinates can be scaled differently when their Hessian curvatures or observation noise levels differ. Section 5 studies this balancing on a quadratic surrogate with heterogeneous blocks; validation of the fully coupled recursion (19) under delays and losses is deferred to Section 6.

Illustrative dimensions. Consider $n = 2$ agents, $m = 1$ target in $\mathbb{R}^2$, and scalar logical clocks ($d_{\mathrm{time}} = 1$, $d_{\mathrm{track}} = 2$, hence $d = 3$). Agent i then maintains $\theta_t^i = \mathrm{col}(\tilde{x}_t^i, r_{t,\mathrm{loc},1}^i, r_{t,\mathrm{loc},2}^i)$, where $(r_{t,\mathrm{loc},1}^i, r_{t,\mathrm{loc},2}^i)$ are the two components of $r_t^{i,\mathrm{loc}}$. Each ADSPSA step probes and updates all three coordinates jointly; $\Psi \in \mathbb{R}^{3\times 3}$ may take the block form (20) with a 1×1 time block and a 2×2 tracking block. Convergence analysis for the time modality under delays appears in [Arkhipov et al., 2026a; Arkhipov et al., 2026b]; the covariance majorization of Section 4 applies to the stacked recursion *once* the abstract assumptions of Section 4.4 are verified for the chosen multimodal F_t^i (this verification is not carried out here).

4 Asymptotic covariance bound for preconditioned consensus-based ADSPSA

4.1 Problem statement (multisensor–multitarget tracking)

We consider a network of n sensors that cooperatively track m moving targets in $\mathbb{R}^r$. Let $\mathcal{N} \triangleq \{1,\dots,n\}$ be the set of sensors and $\mathcal{M} \triangleq \{1,\dots,m\}$ the set of targets. At time t, sensor $i \in \mathcal{N}$ is located at $s_t^i \in \mathbb{R}^r$ and target $\ell \in \mathcal{M}$ has state $r_t^\ell \in \mathbb{R}^r$. Stack all target states into

$$r_t \triangleq \mathrm{col}(r_t^1,\dots,r_t^m) \in \mathbb{R}^{d_{\mathrm{track}}}, \qquad d_{\mathrm{track}} \triangleq mr. \tag{21}$$

In the *tracking-only* formulation of this subsection, each agent's local estimate has dimension $d = d_{\mathrm{track}}$; Section 3.7 stacks a time block on top, giving total per-agent dimension $d = d_{\mathrm{time}} + d_{\mathrm{track}}$.

Distance-type measurements. Each sensor i observes a subset $\mathcal{M}_t^i \subseteq \mathcal{M}$ via noisy squared distances

$$z_t^{i,\ell} = \|s_t^i - r_t^\ell\|^2 + w_t^{i,\ell}, \qquad \ell \in \mathcal{M}_t^i, \tag{22}$$

where $w_t^{i,\ell}$ is measurement noise. To respect communication constraints, sensor i uses at most $p \ge 1$ co-observing neighbors per pair (i,ℓ). Denote by $\mathcal{N}_t^i$ the in-neighbors of node i at time t, and by $\mathcal{C}_t^{i,\ell} \subseteq \{j \in \mathcal{N}_t^i : \ell \in \mathcal{M}_t^j\}$, $|\mathcal{C}_t^{i,\ell}| \le p$, the selected co-observers.

Nonstationary mean-risk optimization. Each sensor maintains a local estimate of r_t; stack all local estimates into a decision vector $\bar{\theta}_t \in \mathbb{R}^{nd}$. The tracking task is cast as minimizing a time-varying objective of the form

$$\bar{F}_t(\bar{\theta}) \triangleq \sum_{i\in\mathcal{N}} F_t^i(\bar{\theta}), \tag{23}$$

where each $F_t^i(\cdot)$ is an (implicit) loss induced by relative measurements available to sensor i at time t. Section 3.7 decomposes the multimodal case into time and tracking contributions $F_t^i = F_t^{i,\mathrm{time}} + F_t^{i,\mathrm{track}}$ with stacked per-agent state (17). In practice the exact function values are not accessible; instead, each agent observes a noisy zeroth-order oracle

$$y_t^i = f_{\xi_t^i}(\bar{\theta}) + v_t^i, \tag{24}$$

where ξ_t^i encodes the random choice of cooperative triples, delayed neighbor timestamps when the stack is multimodal, and the nonstationarity of targets, while v_t^i models unknown-but-bounded (or bounded-variance) disturbances.

Multimodal objective. When logical time is stacked with cooperative tracking as in Section 3.7, the per-agent dimension in (25) and Section 4.2 becomes $d = d_{\mathrm{time}} + d_{\mathrm{track}}$ (with $d_{\mathrm{track}} = mr$ from (21)), rather than $d = d_{\mathrm{track}}$ alone. The time modality enters through skew penalties (18); the tracking modality contributes $F_t^{i,\mathrm{track}}$ from distance-type and co-observer measurements such as (22). Their sum (19) defines the local term in $\bar{F}_t$ and (25). The consensus regularizer then couples *all* coordinates—logical time and spatial estimates alike—through the same $L_{\mathrm{av}} \otimes I_d$, while the preconditioner may be chosen in block form (20) to balance heterogeneous modalities in the SPSA channel. Monotone logical-time semantics (Section 3) are not automatic in this stacked ADSPSA form: preserving nondecreasing $\tilde{x}_t^i$ would require an explicit projection, constraint, or a hybrid update that keeps the PI/SoftMax channel for time and ADSPSA for spatial variables—an implementation choice left open here.

Consensus-regularized objective. Let $L_{\mathrm{av}} \in \mathbb{R}^{n\times n}$ be an averaged (undirected) graph Laplacian describing information exchange. The regularized objective is

$$\Phi_t(\bar{\theta}) \triangleq \sum_{i\in\mathcal{N}} F_t^i(\bar{\theta}) + \frac{\omega}{2}\,\bar{\theta}^\top \big(L_{\mathrm{av}} \otimes I_d\big)\bar{\theta}, \qquad \omega > 0. \tag{25}$$

In the sequel, Φ_t is the consensus-regularized objective that the recursion aims to minimize (up to stochastic zeroth-order noise). We use the neighbor/Laplacian and delay conventions summarized in Section 2.1–2.2.

4.2 Algorithm: accelerated consensus-based ADSPSA with Ψ matrix

We now specify the full accelerated consensus-based ADSPSA recursion in a vectorized form, augmented by a fixed positive definite matrix Ψ applied to the ADSPSA (zeroth-order) channel. The choice $\Psi = I$ recovers the accelerated consensus-based SPSA recursion for which convergence is proved in [Erofeeva et al., 2026]. For a fixed $\Psi \succ 0$, only the zeroth-order increment is left-preconditioned by $\bar{\Psi}$ in (33), while the consensus contribution is unchanged. Section 4.4 supplies an asymptotic covariance *upper bound* for the Ψ-modified scheme under Assumptions 1–4; verifying those assumptions for the full recursion below is not part of the present proof.

4.2.1 Random perturbations and stacked variables

For each sensor $i \in \mathcal{N}$ and outer iteration $k = 1, 2, \ldots$, draw a random perturbation $\Delta_k^i \in \mathbb{R}^d$ with independent components taking values $\pm 1/\sqrt{d}$ equiprobably. Stack

$$\bar{\Delta}_k \triangleq \mathrm{col}(\Delta_k^1, \ldots, \Delta_k^n) \in \mathbb{R}^{nd}. \tag{26}$$

Let $\bar{\theta}_{2k-1} \in \mathbb{R}^{nd}$ be the current stacked estimate and let $\bar{z}_{k-1} \in \mathbb{R}^{nd}$ be the current surrogate center.

4.2.2 Parameters and Ψ matrix

Following [Erofeeva et al., 2026], let $L > 0$ denote a Lipschitz constant of the (blockwise) gradient of the regularized objective (25) along the reference trajectory, and let $\mu > 0$ denote a strong-convexity parameter used in the accelerated schedule. Fix constants $h > 0$, $\omega > 0$, $\eta \in (0, \mu)$, $\gamma_0 > 0$, and a lower bound $\bar{\alpha} \in (0, 1)$. Let $\beta_k^+ \geq 0$, $\beta_k^- \geq 0$, and $\beta_k = \beta_k^+ + \beta_k^- > 0$.

Fix a symmetric positive definite matrix

$$\Psi \in \mathbb{R}^{d\times d}, \qquad \Psi \succ 0, \tag{27}$$

and define its block lift $\bar{\Psi} \triangleq I_n \otimes \Psi \in \mathbb{R}^{nd\times nd}$.

4.2.3 Recursion for α_k and γ_k

At each outer iteration $k \geq 1$, choose $\alpha_k \in [\bar{\alpha}, 1)$ as a solution of

$$\begin{aligned} \alpha_k^2 &= 2\Big(h - \frac{h^2 L}{2} - \bar{\alpha}\Big) \\ &\quad \times \Big((1-\alpha_k)\gamma_{k-1} + \alpha_k(\mu - \eta)\Big), \end{aligned} \tag{28}$$

and update

$$\gamma_k = (1-\alpha_k)\gamma_{k-1} + \alpha_k(\mu - \eta). \tag{29}$$

4.2.4 Two-point probing, direction estimation, and updates

Define the probing point

$$\begin{aligned} \tilde{x}_k &= \frac{\alpha_k \gamma_{k-1}}{\gamma_{k-1} + \alpha_k(\mu-\eta)}\,\bar{z}_{k-1} \\ &\quad + \frac{\gamma_k}{\gamma_{k-1} + \alpha_k(\mu-\eta)}\,\bar{\theta}_{2k-1}. \end{aligned} \tag{30}$$

Query the zeroth-order oracle at two points:

$$\bar{x}_{2k-1} = \tilde{x}_k - \beta_k^-\,\bar{\Delta}_k, \qquad \bar{x}_{2k} = \tilde{x}_k + \beta_k^+\,\bar{\Delta}_k. \tag{31}$$

Let $\bar{y}_{2k-1}$ and $\bar{y}_{2k}$ be the stacked noisy function values returned by (24) at $\bar{x}_{2k-1}$ and $\bar{x}_{2k}$, respectively (i.e., at probing steps $t = 2k-1$ and $t = 2k$).

Define the SPSA and consensus parts of the descent direction. Let L_{2k-1} be the (possibly time-varying) graph Laplacian used at outer iteration k; in a static graph one may take $L_{2k-1} = L_{\mathrm{av}}$ from (25). Then

$$\begin{aligned} \bar{g}_k^{\mathrm{spsa}} &\triangleq \Big(\mathrm{diag}\Big(\tfrac{\bar{y}_{2k}-\bar{y}_{2k-1}}{\beta_k}\Big) \otimes I_d\Big)\bar{\Delta}_k, \\ \bar{g}_k^{\mathrm{cons}} &\triangleq \omega\,(L_{2k-1} \otimes I_d)\,\tilde{x}_k, \end{aligned} \tag{32}$$

and set $\bar{g}_k = \bar{g}_k^{\mathrm{spsa}} + \bar{g}_k^{\mathrm{cons}}$.

Update of the main estimate. Apply $\bar{\Psi}$ to the SPSA part only:

$$\bar{\theta}_{2k} = \tilde{x}_k - h\,\bar{\Psi}\,\bar{g}_k^{\mathrm{spsa}} - h\,\bar{g}_k^{\mathrm{cons}}. \tag{33}$$

Define the copied iterate by

$$\bar{\theta}_{2k+1} = \bar{\theta}_{2k}. \tag{34}$$

Update of the surrogate center. Update $\bar{z}_k$ by

$$\bar{z}_k = \frac{(1-\alpha_k)\gamma_{k-1}\bar{z}_{k-1} + \alpha_k(\mu-\eta)\tilde{x}_k}{\gamma_k} - \frac{\alpha_k}{\gamma_k}\,\bar{g}_k. \tag{35}$$

4.3 Assumptions for covariance analysis

We state assumptions under which the linearized error dynamics admit an asymptotic *upper bound* on the second-moment matrix. They are stated abstractly for the full ADSPSA recursion of Section 4.2; a verifiable instance matching the synthetic benchmark in Section 5 is given in Proposition B (Appendix).

Assumption 1 (One-step smoothness and measurability). *Fix a filtration $(\mathcal{F}_k)_{k\ge 0}$ with $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$. The outer-iteration update $k \mapsto k+1$ admits a (local) smooth state-space representation whose first-order Taylor expansion around a reference trajectory is valid with a quadratic remainder (Appendix, Assumption A). Moreover, the tracking error e_k in* (41) *and the linearization coefficients $A_k, B_k(\Psi)$ can be chosen $\mathcal{F}_{k-1}$-measurable, while the aggregated innovation w_k at outer step k satisfies $\mathbb{E}[w_k \mid \mathcal{F}_{k-1}] = 0$.*

Assumption 2 (Centered noise and cross-term control). *In the setting of Assumption 1, the noise vector w_k is conditionally centered and its conditional covariance is majorized as in Appendix Assumption A(a). Cross-terms in the one-step second-moment expansion vanish or are absorbed into a PSD remainder; a sufficient condition is $\mathbb{E}[e_k w_k^\top \mid \mathcal{F}_{k-1}] = 0$ and $\mathbb{E}[b_k w_k^\top \mid \mathcal{F}_{k-1}] = 0$ almost surely (Appendix* (70)*).*

Assumption 3 (Stationary majorants). *There exist time-uniform (or asymptotically valid) majorants A, $B(\Psi)$, Ω, and $Q \succeq 0$ such that, with $\Sigma_k = \mathbb{E}[e_k e_k^\top]$,*

$$\mathbb{E}[A_k e_k e_k^\top A_k^\top] \preceq A\Sigma_k A^\top + \varepsilon_k^{(A)}, \tag{36}$$

$$\mathbb{E}\big[B_k(\Psi)\Omega_k B_k(\Psi)^\top\big] \preceq B(\Psi)\Omega B(\Psi)^\top + \varepsilon_k^{(B)}, \tag{37}$$

$$\mathbb{E}[\Gamma_k] \preceq Q + \varepsilon_k^{(\Gamma)}, \tag{38}$$

where Γ_k collects higher-order terms, drift cross-terms, and the contribution of a possibly moving reference θ_{2k} (via the aggregated drift b_k in Appendix Corollary A); the stationary bound Q in (39) *majorizes $\mathbb{E}[\Gamma_k]$ in the asymptotic regime, and $\varepsilon_k^{(A)}, \varepsilon_k^{(B)}, \varepsilon_k^{(\Gamma)} \succeq 0$ vanish along that regime. Consequently the covariance recursion is dominated by the* affine *map*

$$\mathcal{T}(\Sigma) \triangleq A\Sigma A^\top + B(\Psi)\Omega B(\Psi)^\top + Q, \tag{39}$$

up to a vanishing PSD remainder.

Assumption 4 (Spectral contraction of $\mathcal{T}$). *The affine map $\mathcal{T}$ in* (39) *preserves* PSD *order and is monotone in the Loewner sense. Moreover, with $\|\cdot\|_2$ denoting the spectral norm,*

$$\|A\|_2 < 1, \tag{40}$$

so that $\mathcal{T}$ is a contraction on symmetric matrices in $\|\cdot\|_2$ with constant $q = \|A\|_2^2$.

4.4 Theorem: Asymptotic covariance upper bound

Define the outer-iteration tracking error and its (second-moment) covariance:

$$e_k \triangleq \bar{\theta}_{2k} - \mathbf{1}_n \otimes \theta_{2k}, \qquad \Sigma_k \triangleq \mathbb{E}\big[e_k e_k^\top\big]\,. \tag{41}$$

Measurability convention. For the covariance analysis we evaluate e_k once $\bar{\theta}_{2k}$ is available and before the innovations that drive the transition $k \to k{+}1$. Under this convention $e_k, A_k, B_k(\Psi)$ are $\mathcal{F}_{k-1}$-measurable and the aggregated noise w_k is a martingale difference; see Assumption 1 and Corollary A.

Theorem 1 (Asymptotic covariance upper bound). *Suppose Assumptions 1–4 hold. Then there exists a unique $\Sigma^\star \succeq 0$ such that*

$$\Sigma^\star = \mathcal{T}(\Sigma^\star), \tag{42}$$

and there exists a sequence $E_k \succeq 0$ with $E_k \to 0$ (in spectral norm) such that

$$\Sigma_k \preceq \Sigma^\star + E_k, \qquad k \to \infty, \tag{43}$$

where $\preceq$ denotes the Loewner (positive-semidefinite) order.

Tuning Ψ matrix in practice. Since the fixed point $\Sigma^\star$ depends on the chosen preconditioner Ψ through the map $\mathcal{T}$, one can treat Ψ as a design parameter and search for a Ψ that makes the resulting upper bound as small as possible according to a chosen scalar surrogate (e.g., minimizing $\mathrm{Tr}(\Sigma^\star(\Psi))$ or $\|\Sigma^\star(\Psi)\|_2$ over an admissible family of positive definite Ψ).

Proof idea. The argument is a majorized fixed-point comparison for the second-moment matrix $\Sigma_k = \mathbb{E}[e_k e_k^\top]$, not a claim that the iterates $\bar{\theta}_{2k}$ converge to a single optimizer:

1. **Linearization.** Under Assumption 1, expand e_{k+1} as $A_k e_k + B_k(\Psi) w_k + b_k + r_k$ with $\mathcal{F}_{k-1}$-measurable coefficients and centered w_k (Appendix, Lemma A and Corollary A).
2. **Covariance majorization.** Expand $\Sigma_{k+1} = \mathbb{E}[e_{k+1}e_{k+1}^\top]$, control cross-terms using Assumption 2, and obtain

$$\Sigma_{k+1} \preceq \mathbb{E}[A_k e_k e_k^\top A_k^\top] + \mathbb{E}[B_k(\Psi)\Omega_k B_k(\Psi)^\top] + \mathbb{E}[\Gamma_k]$$

(Lemma A).

3. **Affine comparison map.** Replace time-varying coefficients by the stationary majorants in Assumption 3, yielding $\Sigma_{k+1} \preceq \mathcal{T}(\Sigma_k)+E_k^{(0)}$ with $E_k^{(0)} \to 0$ (Lemma A).
4. **Perturbed fixed point.** Because $\mathcal{T}$ is affine and $\|A\|_2 < 1$, the perturbed recursion in Lemma A gives $\Sigma_k \preceq \Sigma^\star + E_k$ with $E_k \to 0$.

The complete proof is given in Appendix A.

5 Simulations

We complement the preceding analysis with numerical experiments for the recursion specified in Section 4.2 (two-point zeroth-order queries, consensus coupling, and the Ψ-modified update (33) together with (30)–(35)).

For reproducibility we instantiate the scheme in a controlled *synthetic quadratic* benchmark: all agents share the same local model $f_i(x) = \frac{1}{2}x^\top Hx$ with a common $H \succ 0$, the reference trajectory is $\theta_{2k} \equiv 0$, and the communication graph is a static undirected ring (so $L_{2k-1} = L_{\mathrm{av}} = L$ in (32)). Proposition B (Appendix) verifies that this benchmark satisfies Assumptions 1–4, so the covariance upper bound of Theorem 1 applies to the tracking error reported below. We do not simulate the fully coupled logical-time plus tracking stack of Section 3.7; instead, we interpret the benchmark as a *multimodal surrogate* by partitioning the per-agent dimension as $d = d_{\mathrm{time}} + d_{\mathrm{track}}$ with $d_{\mathrm{time}} = 2$ and $d_{\mathrm{track}} = 4$ (so $d = 6$). The Hessian H and the tuned preconditioner $\Psi^\star$ are constrained to be block-diagonal with respect to this split, mimicking time-like versus spatial channels that differ in curvature and noise scale. Tuning Ψ then studies how block scaling balances heterogeneous modalities in the zeroth-order update—the same design question that arises for $\Psi = \mathrm{blkdiag}(\Psi_{\mathrm{time}}, \Psi_{\mathrm{track}})$ in (20). The code implements the same algebraic steps as Section 4.2 in per-agent coordinates, which is equivalent to the stacked notation used there.

Our focus is empirical: we illustrate how the choice of Ψ changes the *transient speed* of the Monte Carlo estimate of the error, while the averaged error still trends downward toward a stochastic steady state (up to small fluctuations from sampling).

5.1 Simulation setup

For each experiment index $j = 0, \ldots, 99$ we draw a random model instance with fixed dimension $d = 6$ ($d_{\mathrm{time}} = 2$, $d_{\mathrm{track}} = 4$) and $n = 10$ agents. The Hessian-like matrix $H \succ 0$ is generated randomly with eigenvalues in a wide range, with block structure aligned to the time/tracking split; the step h is uniform in $[0.045, 0.085]$; Q_1 is a scaled identity with diagonal entries uniform in $[5 \times 10^{-5}, 2 \times 10^{-4}]$; $\Omega = I$; and B_0 is a random matrix scaled by a factor uniform in $[0.15, 0.35]$. The algorithm parameters are fixed across all runs: $\omega = 0.2$, $\mu = 2.0$, $\eta = 0.1$, $\gamma_0 = 1.0$, $\beta_k^+ = \beta_k^- = 0.5$ (hence $\beta_k = 1.0$ in (32)), and observation noise standard deviation 0.05 in the ADSPSA channel.

We compare $\Psi = I$ with a matrix $\Psi^\star$ obtained as follows. We restrict Ψ to be symmetric positive definite and block-diagonal with 2×2 blocks matching the $d_{\mathrm{time}}/d_{\mathrm{track}}$ partition (three blocks along the diagonal of each agent's 6×6 slice, or equivalently block structure consistent with (20)), with eigenvalues of each block constrained to lie in $[0.2, 3]$. Using trajectories of the *mean* agent state $\bar{\theta}_k = \frac{1}{n}\sum_{i=1}^n \theta_{2k}^i$, we build an empirical linear-Gaussian majorant (A, S), inflate it to ensure stability, and minimize $\mathrm{Tr}(\Sigma^\star(\Psi))$ where $\Sigma^\star$ solves the discrete Lyapunov equation $\Sigma^\star = A\Sigma^\star A^\top + S$. This objective is the fixed-point covariance surrogate associated with the upper bounds in Section 4. Importantly, the horizon used for this identification is *decoupled* from the horizon used for reporting curves: we use `iters_tune` $= 100$ trajectories to fit the majorant and optimize $\Psi^\star$, and then evaluate the error curves for `iters_eval` $= 5000$ iterations. Thus $\Psi^\star$ does not change when the evaluation horizon is lengthened. The identification of $\Psi^\star$ is a design heuristic on top of the certified majorization template of Theorem 1; it is consistent with minimizing $\mathrm{Tr}(\Sigma^\star(\Psi))$ from (39) but does not constitute a separate theorem.

Error metrics and link to Theorem 1. With $\theta_{2k} \equiv 0$, the tracking error from (41) reduces to $e_k = \bar{\theta}_{2k}$. We therefore report the theorem-consistent surrogate

$$\mathbb{E}\left[\frac{1}{n}\|e_k\|_2^2\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \|\theta_{2k}^i\|_2^2\right], \qquad (44)$$

estimated by Monte Carlo averaging over 100 independent runs per model instance. In every run, the same random seed is used for $\Psi = I$ and $\Psi = \Psi^\star$ so that initial conditions and ADSPSA perturbations match. When $\theta_{2k} \not\equiv 0$, the same plots should be generated from $\|e_k\|_2^2$ rather than $\|\bar{\theta}_{2k}\|_2^2$ alone.

In all 100 experiments the curve (44) exhibits a clear decreasing trend toward a noise-dominated plateau (up to small fluctuations due to stochasticity). What varies across experiments is *how fast* the transient is and *how low* the plateau is for $\Psi = I$ versus $\Psi = \Psi^\star$.

5.2 Aggregate behavior across 100 instances

Fig. 1 summarizes the batch of 100 model instances: at each iteration k we plot the sample mean of (44) over instances, together with pointwise 95% confidence intervals (Student-t bands across the 100 experiments). On this aggregate view, the majorant-tuned preconditioner (orange) typically *descends more slowly* than the identity baseline (blue) during an extended early-to-mid horizon, and the relative transient speed varies noticeably from one random instance to another (as the representative plots below illustrate). Nevertheless, by $k = 5000$ the orange mean curve lies *below* the blue one, indicating a

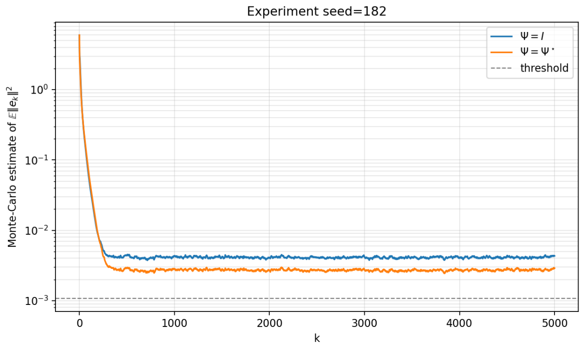

Figure 3. Monte Carlo estimate (44) versus iteration k for seed 182 ($k \leq 5000$): transient slopes are comparable, yet the majorant-tuned preconditioner (orange) reaches a lower plateau than the identity baseline (blue).

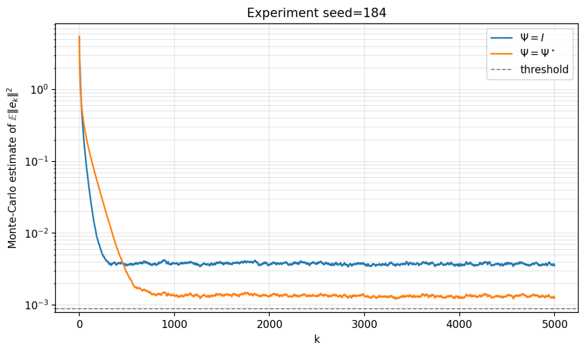

Figure 4. Monte Carlo estimate (44) versus iteration k for seed 184 ($k \leq 5000$): the majorant-tuned preconditioner (orange) attains a lower plateau but with noticeably slower decay in the mid-horizon compared to the identity baseline (blue).

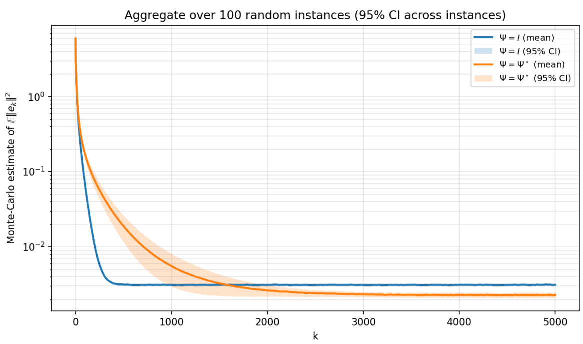

Figure 1. Batch summary over 100 random model instances: sample mean of the Monte Carlo estimate (44) versus iteration k, with pointwise 95% confidence intervals. The majorant-tuned preconditioner (orange) decays more slowly on average during much of the horizon yet reaches a lower level by $k = 5000$ than the identity baseline (blue).

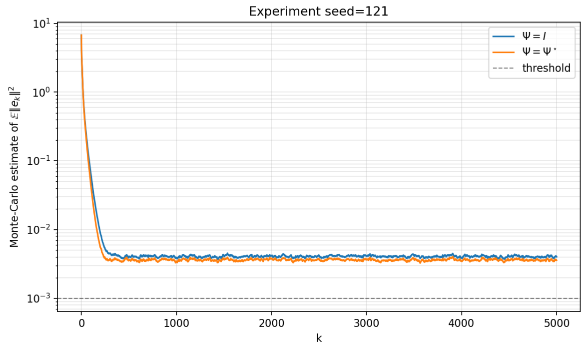

Figure 2. Monte Carlo estimate (44) versus iteration k for seed 121 ($k \leq 5000$): transient slopes are comparable, while the majorant-tuned preconditioner (orange) reaches a marginally lower steady-state level than the identity baseline (blue).

systematically lower noise floor when $\Psi = \Psi^\star$. In other words, optimizing the majorant surrogate trades faster initial decay for a smaller asymptotic error on average.

5.3 Representative behaviors

We highlight three representative trials from the batch of 100 experiments, indexed by $j \in \{21, 82, 84\}$ (random seeds 121, 182, and 184, respectively). Each panel uses the same Monte Carlo protocol as above but is evaluated for $k = 0, \ldots, 5000$.

Comparable transient, modestly lower tail (seed 121). Fig. 2 shows a case where the two curves largely track each other over the horizon; the majorant-tuned preconditioner (orange) still finishes with a slightly lower plateau than the identity baseline (blue).

Similar transient speed, lower tail (seed 182). Fig. 3 illustrates a case where the two curves track each other closely during much of the horizon, but the majorant-tuned preconditioner (orange) achieves a visibly lower noise floor by $k = 5000$. Here the benefit is mainly in the *quality* of the steady state rather than a dramatic change in early slope.

Lower tail but slower decay (seed 184). Finally, Fig. 4 shows a complementary regime: the majorant-tuned preconditioner (orange) eventually reaches a substantially smaller steady-state error than the identity baseline (blue), but its curve descends more slowly during a long middle phase—consistent with the slower average transient seen in Fig. 1.

6 Discussion

This work isolates a single design knob—a fixed positive definite matrix Ψ acting on the zeroth-order (ADSPSA) channel—and studies its interaction with consensus-based distributed optimization and covariance upper bounds. The simulations in Section 5 confirm that the scheme remains stable and that the averaged tracking error decreases toward a stochastic steady state, while the *transient* behavior can change substantially depending on how Ψ is chosen. We close with seven directions that appear natural for extending both the algorithm and the analysis.

SoftMax temperature in logical-time synchronization. Section 3.4 introduces the temperature T as a smooth surrogate for hard-max consensus. In practice, T controls not only update smoothness but also the monotonicity slack $\varepsilon = T \log M$ in (10). Smaller T keeps logical clocks closer to strict max-consensus behavior and reduces the worst-case regression allowed by SoftMax, which is attractive when causality or safety margins require tight monotonicity. Larger T dampens switching induced by noisy neighbor timestamps and can stabilize the PI loop in (13)–(14), at the price of permitting larger transient backward steps of order ε. A practical deployment therefore treats T as an application-level hyperparameter: it should be chosen jointly with the PI gains and the acceptable logical-time regression budget implied by downstream OS services (scheduling granularity, timeout semantics, and ordering buffers).

Preconditioning beyond the ADSPSA term. In Section 4.2, Ψ enters only through the ADSPSA contribution (cf. (33)), while the consensus term is left unscaled. A practical extension is to explore alternative splittings of the total direction g_k, e.g., applying a (possibly different) preconditioner to the consensus part, or using one map for the zeroth-order channel and another for consensus. Each choice changes the interaction between noise, coupling, and the contraction properties of the linearized dynamics, and would require revisiting both stability assumptions and the construction of majorants in the covariance bound.

Exploiting known cross-modal or cross-state structure. The present modeling stance treats cross-modal (or cross-component) dependence conservatively: we do not assume that an agent knows how different coordinates co-vary beyond what is implied by the local oracle and the graph. In applications such as robotics or navigation, however, kinematic or temporal relations are often explicit (e.g., how orientation is coupled to position, or how states evolve along a known trajectory). Incorporating such structure could take the form of a restricted family of Ψ (low-rank, sparsity patterns, Lie-group actions) or of tighter majorants that encode known couplings. The theoretical program would then connect identifiability of these relations to tighter covariance bounds rather than to a fully unstructured Ψ.

Methodology for constructing majorants. The general theorem provides an asymptotic upper bound once suitable majorants $(A, B(\Psi), \Omega, Q)$ are available, but it does not prescribe a constructive procedure for obtaining them from data or from first principles. The numerical study highlights that even in a simplified quadratic benchmark, the quality of an *empirical* majorant fit—and hence the induced choice of Ψ—can materially affect both the bound and the observed convergence profile. A pressing research line is therefore to develop *systematic* rules: when to estimate (A, S) from trajectories, how much inflation is needed for PSD/stability guarantees, how to align the state used in majorant identification with the error actually displayed in experiments, and how sensitive the tuned Ψ is to the tuning horizon. Bridging "data-driven majorants" with statistical guarantees (bias–variance, high-dimensional scaling) is an open problem.

Transient speed versus steady-state accuracy. Even when a choice of Ψ improves the stationary second-moment surrogate (or the empirical noise floor), the *rate* at which the averaged error decays can be faster, comparable, or slower than for the identity preconditioner. The simulations exhibit all three regimes. A complementary theory should therefore quantify *finite-horizon* performance—e.g., bounds on $\frac{1}{T}\sum_{k=0}^{T}\mathbb{E}\|e_k\|^2$ or mixing-time style estimates—and clarify which additional structural conditions make it possible to *simultaneously* improve tail level and transient slope. Without such conditions, optimizing a stationary majorant need not align with minimizing a finite-horizon cost, which partially explains the diversity of outcomes observed across random instances.

Verifying assumptions for the full ADSPSA recursion. The covariance bound is proved under abstract majorization assumptions; checking smoothness, centering, stationary majorants, and $\|A\|_2 < 1$ for the complete recursion of Section 4.2—especially with a moving reference θ_{2k} and a time-varying Laplacian L_{2k-1}—remains open. Closing this gap would turn the present framework from a proof template into a certifiable guarantee for deployed OS services.

Fully coupled multimodal deployments. Section 3.7 specifies the stacked state (17), multimodal losses (19), and block preconditioner (20), but the experiments in Section 5 study the ADSPSA block in isolation with a heterogeneous quadratic surrogate. A natural next step is to simulate the composed system under the SoftMax/PI logical-time dynamics of Section 3 together with moving-target tracking (22), using the same time-delay and loss model (2) throughout. Useful metrics would include logical-time skew $\max_i \tilde{x}_t^i - \min_j \tilde{x}_t^j$ alongside the tracking error (44), and the design question is whether a single $\Psi = \mathrm{blkdiag}(\Psi_{\mathrm{time}}, \Psi_{\mathrm{track}})$ can balance both modalities or whether separate preconditioners (or a consensus-side scaling) are needed when timestamp noise and spatial measurement noise differ by orders of magnitude. Extending the ADSPSA covariance analysis to the time-delayed multimodal oracle (24) remains open; convergence guarantees for the logical-time block in time-delay networks appear in [Arkhipov et al., 2026a; Arkhipov et al., 2026b].

7 Conclusion

Multiagent operating systems need agreement services that remain meaningful when any peer may fail, communication is lossy and unsynchronized, and time-like abstractions must advance in lockstep with cooperative physical state rather than being outsourced to a single grandmaster. This work took a step toward that goal by presenting two complementary OS-level building blocks—monotone logical-time synchronization and ADSPSA-based cooperative tracking—together with an explicit multimodal composition rule (Section 3.7). The scope was intentionally split across these blocks: Section 3 records the time-sync primitive, Section 4–5 develops the main ADSPSA theory and experiments, and fully coupled multimodal validation is left to future work (Section 6).

Section 3 formalized monotone logical-time updates under bounded delays; Section 3.7 stacked time and tracking with block preconditioning; Section 4 supplied an asymptotic covariance upper bound; and Section 5 showed that empirically tuned Ψ can lower the asymptotic error, sometimes at the cost of slower transients.

Several limitations remain: the benchmark is stylized, majorization assumptions are not verified for the full recursion, and the bounds are asymptotic rather than finite-horizon. Next steps include assumption verification, finite-time analysis, principled majorants, and deployment on real agent platforms.

APPENDIX

A Proof of Theorem 1

Lemma 1 (Shift of the external target) Let the reference trajectory $\{\theta_{2k}\}_{k\geq 0} \subset \mathbb{R}^d$ be arbitrary and define the outer-iteration tracking error as in (41),

$$e_k \triangleq \bar{\theta}_{2k} - \mathbf{1}_n \otimes \theta_{2k}.$$

Define the target shift

$$\delta_k^\theta \triangleq \mathbf{1}_n \otimes (\theta_{2(k+1)} - \theta_{2k}). \tag{45}$$

Then, for every $k \geq 0$,

$$e_{k+1} = \big(\bar{\theta}_{2(k+1)} - \mathbf{1}_n \otimes \theta_{2k}\big) - \delta_k^\theta. \tag{46}$$

Proof. By definition of e_{k+1},

$$e_{k+1} = \bar{\theta}_{2(k+1)} - \mathbf{1}_n \otimes \theta_{2(k+1)}.$$

Add and subtract the same vector $\mathbf{1}_n \otimes \theta_{2k}$ to obtain

$$e_{k+1} = \big(\bar{\theta}_{2(k+1)} - \mathbf{1}_n \otimes \theta_{2k}\big) - \mathbf{1}_n \otimes \big(\theta_{2(k+1)} - \theta_{2k}\big),$$

which is exactly (46) after using (45). □

Lemma 2 (Equivalent form of the Ψ-modified step) Consider the decomposition of the descent direction at outer iteration k into the SPSA and consensus parts $\bar{g}_k = \bar{g}_k^{\mathrm{spsa}} + \bar{g}_k^{\mathrm{cons}}$ as in (32), and the modified update (33)

$$\bar{\theta}_{2k} = \tilde{x}_k - h\,\bar{\Psi}\,\bar{g}_k^{\mathrm{spsa}} - h\,\bar{g}_k^{\mathrm{cons}}, \qquad \bar{\Psi} = I_n \otimes \Psi.$$

Define the corresponding "unmodified" (baseline) step by

$$\bar{\theta}_{2k}^{\mathrm{orig}} \triangleq \tilde{x}_k - h\big(\bar{g}_k^{\mathrm{spsa}} + \bar{g}_k^{\mathrm{cons}}\big). \tag{47}$$

Then the difference between the modified and unmodified steps satisfies

$$\bar{\theta}_{2k} - \bar{\theta}_{2k}^{\mathrm{orig}} = h\,(I_{nd} - \bar{\Psi})\,\bar{g}_k^{\mathrm{spsa}}. \tag{48}$$

Proof. Using $\bar{g}_k = \bar{g}_k^{\mathrm{spsa}} + \bar{g}_k^{\mathrm{cons}}$, expand the baseline step (47) as

$$\bar{\theta}_{2k}^{\mathrm{orig}} = \tilde{x}_k - h\,\bar{g}_k^{\mathrm{spsa}} - h\,\bar{g}_k^{\mathrm{cons}}.$$

Subtract this identity from the modified update (33):

$$\bar{\theta}_{2k} - \bar{\theta}_{2k}^{\mathrm{orig}} = -h\,\bar{\Psi}\,\bar{g}_k^{\mathrm{spsa}} + h\,\bar{g}_k^{\mathrm{spsa}} = h\,(I_{nd} - \bar{\Psi})\,\bar{g}_k^{\mathrm{spsa}},$$

which proves (48). □

Lemma 3 (Decomposition of the ADSPSA part) Let $(\mathcal{F}_k)_{k\geq 0}$ be a filtration such that $\bar{g}_k^{\mathrm{spsa}}$ is $\mathcal{F}_k$-measurable and $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$. Define

$$G_k \triangleq \mathbb{E}[\bar{g}_k^{\mathrm{spsa}} \mid \mathcal{F}_{k-1}], \qquad w_k^{\mathrm{spsa}} \triangleq \bar{g}_k^{\mathrm{spsa}} - G_k. \tag{49}$$

Then G_k is $\mathcal{F}_{k-1}$-measurable,

$$\bar{g}_k^{\mathrm{spsa}} = G_k + w_k^{\mathrm{spsa}}, \tag{50}$$

and

$$\mathbb{E}[w_k^{\mathrm{spsa}} \mid \mathcal{F}_{k-1}] = 0. \tag{51}$$

Proof. By definition in (49), G_k is a conditional expectation with respect to $\mathcal{F}_{k-1}$, hence it is $\mathcal{F}_{k-1}$-measurable. Identity (50) is just the rearrangement of the definition of w_k^{spsa}. Finally, using linearity of conditional expectation and $\mathcal{F}_{k-1}$-measurability of G_k,

$$\begin{aligned}\mathbb{E}[w_k^{\mathrm{spsa}} \mid \mathcal{F}_{k-1}] &= \mathbb{E}[\bar{g}_k^{\mathrm{spsa}} - G_k \mid \mathcal{F}_{k-1}] \\ &= \mathbb{E}[\bar{g}_k^{\mathrm{spsa}} \mid \mathcal{F}_{k-1}] - G_k = 0.\end{aligned} \tag{52}$$

□

Lemma 4 (Decomposition of the full direction) Let $\bar{g}_k$ be $\mathcal{F}_k$-measurable and let $\bar{G}_k$ be any $\mathcal{F}_{k-1}$-measurable random vector. Define

$$\tilde{w}_k \triangleq \bar{g}_k - \bar{G}_k. \tag{53}$$

Then the following statements are equivalent:

$$\begin{aligned} &\mathbb{E}[\tilde{w}_k \mid \mathcal{F}_{k-1}] = 0 \\ &\qquad \iff \quad \bar{G}_k = \mathbb{E}[\bar{g}_k \mid \mathcal{F}_{k-1}] \text{ a.s.} \end{aligned} \tag{54}$$

Proof. From (53) and linearity of conditional expectation,

$$\mathbb{E}[\tilde{w}_k \mid \mathcal{F}_{k-1}] = \mathbb{E}[\bar{g}_k \mid \mathcal{F}_{k-1}] - \mathbb{E}\big[\bar{G}_k \mid \mathcal{F}_{k-1}\big] .$$

Since $\bar{G}_k$ is $\mathcal{F}_{k-1}$-measurable, $\mathbb{E}[\bar{G}_k \mid \mathcal{F}_{k-1}] = \bar{G}_k$ almost surely. Therefore, $\mathbb{E}[\tilde{w}_k \mid \mathcal{F}_{k-1}] = 0$ holds if and only if $\bar{G}_k = \mathbb{E}[\bar{g}_k \mid \mathcal{F}_{k-1}]$ a.s., proving (54). □

Assumption 1 (One-step smoothness of the outer update) Consider the outer-iteration mapping induced by the algorithm in Section 4.2. There exists a measurable random vector U_k that collects all stochastic primitives revealed at outer iteration k (SPSA perturbations, oracle noise, and communication randomness). Fix the augmented state of one outer step as

$$s_k \triangleq \mathrm{col}(\bar{\theta}_{2k}, \bar{z}_k) \in \mathbb{R}^{2nd}. \tag{55}$$

There exists a measurable mapping

$$\begin{aligned} &\mathcal{H}_k : \mathbb{R}^{2nd} \times \mathsf{U} \to \mathbb{R}^{nd} \\ &\qquad \text{such that} \quad \bar{\theta}_{2(k+1)} = \mathcal{H}_k(s_k, U_k). \end{aligned} \tag{56}$$

Fix a (possibly time-varying) linearization point

$$s_k^\star \triangleq \mathrm{col}(\mathbf{1}_n \otimes \theta_{2k}, z_\star) \in \mathbb{R}^{2nd}, \tag{57}$$

where $z_\star \in \mathbb{R}^{nd}$ is an arbitrary reference point for the auxiliary variable (in the simplest choice, $z_\star = \mathbf{1}_n \otimes \theta_{2k}$). Assume that, with probability one, the map $s \mapsto \mathcal{H}_k(s, U_k)$ is continuously differentiable in a neighborhood of $s_k^\star$, and admits a first-order Taylor expansion with a quadratic remainder: there exist a Jacobian

$$J_k(U_k) \triangleq \nabla_s \mathcal{H}_k(s_k^\star, U_k) \in \mathbb{R}^{nd \times 2nd}, \tag{58}$$

and a remainder $\rho_k \in \mathbb{R}^{nd}$ such that

$$\mathcal{H}_k(s_k, U_k) = \mathcal{H}_k(s_k^\star, U_k) + J_k(U_k)\,(s_k - s_k^\star) + \rho_k, \tag{59}$$

where $\|\rho_k\| = O(\|s_k - s_k^\star\|^2)$ (locally, for almost all realizations).

Lemma 5 (Linearized one-step representation of the error) Assume Assumption A holds and define the outer error $e_k = \bar{\theta}_{2k} - \mathbf{1}_n \otimes \theta_{2k}$ as in (41). Then there exist (not necessarily unique) random matrices $A_k \in \mathbb{R}^{nd \times nd}$, $C_k \in \mathbb{R}^{nd \times nd}$, and random vectors $b_k, r_k \in \mathbb{R}^{nd}$ such that, for each $k \geq 0$,

$$e_{k+1} = A_k e_k + C_k(\bar{z}_k - z_\star) + b_k - \delta_k^\theta + r_k, \tag{60}$$

where δ_k^θ is the target shift in (45). Moreover, one can choose A_k, C_k, b_k, r_k to satisfy the following properties.

(i) b_k is the offset of the Taylor expansion at the linearization point: $b_k = \mathcal{H}_k(s_k^\star, U_k) - \mathbf{1}_n \otimes \theta_{2k}$. The external target shift $-\delta_k^\theta$ from Lemma A is kept explicit in (60); in Corollary A it is absorbed into the aggregated drift term (62).

(ii) A_k and C_k are taken from the Jacobian blocks $J_{k,\theta}$ and $J_{k,z}$ in (59). When the linearization point $s_k^\star$ is $\mathcal{F}_{k-1}$-measurable, A_k and C_k can be chosen $\mathcal{F}_{k-1}$-measurable as well.

(iii) The remainder $r_k = \rho_k$ is of quadratic order: $r_k = O(\|e_k\|^2 + \|\bar{z}_k - z_\star\|^2)$ locally.

(iv) The Ψ-dependence of the noise channel is made explicit in Corollary A via Lemma A.

Proof. Fix a realization of the stochastic primitives U_k. From the one-step map (56),

$$e_{k+1} = \mathcal{H}_k(s_k, U_k) - \mathbf{1}_n \otimes \theta_{2(k+1)}.$$

Use Lemma A to obtain

$$e_{k+1} = \big(\mathcal{H}_k(s_k, U_k) - \mathbf{1}_n \otimes \theta_{2k}\big) - \delta_k^\theta.$$

Apply the Taylor expansion (59) around $s_k^\star$ and write the Jacobian in block form $J_k(U_k) = [J_{k,\theta} \;\; J_{k,z}]$ conformably with $s_k = \mathrm{col}(\bar{\theta}_{2k}, \bar{z}_k)$. Then

$$J_k(U_k)(s_k - s_k^\star) = J_{k,\theta}(\bar{\theta}_{2k} - \mathbf{1}_n \otimes \theta_{2k}) + J_{k,z}(\bar{z}_k - z_\star).$$

Setting $A_k \triangleq J_{k,\theta}$, $C_k \triangleq J_{k,z}$, $r_k \triangleq \rho_k$, and $b_k \triangleq \mathcal{H}_k(s_k^\star, U_k) - \mathbf{1}_n \otimes \theta_{2k}$ yields (60). Items (i)–(iii) follow by construction; item (iv) is proved in Corollary A. □

Corollary 1 (Centered noise channel and explicit Ψ-dependence) In the setting of Lemma A, fix a filtration $(\mathcal{F}_k)$ such that all random quantities at outer iteration k are $\mathcal{F}_k$-measurable, $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$, and the coefficients A_k, C_k can be chosen $\mathcal{F}_{k-1}$-measurable. Then (60) can be written equivalently as

$$e_{k+1} = A_k e_k + B_k(\Psi)\, w_k + b_k + r_k, \tag{61}$$

where the drift term is

$$b_k \triangleq C_k(\bar{z}_k - z_\star) + \bar{b}_k - \delta_k^\theta, \tag{62}$$

with $\bar{b}_k \triangleq \mathcal{H}_k(s_k^\star, U_k) - \mathbf{1}_n \otimes \theta_{2k}$, w_k is an $\mathcal{F}_k$-measurable aggregated noise vector satisfying

$$\mathbb{E}[w_k \mid \mathcal{F}_{k-1}] = 0, \tag{63}$$

and $B_k(\Psi)$ depends on Ψ only through the SPSA channel. One may choose w_k so that its SPSA component is the martingale difference $w_k^{\rm spsa} = \bar{g}_k^{\rm spsa} - \mathbb{E}[\bar{g}_k^{\rm spsa} \mid \mathcal{F}_{k-1}]$ from Lemma A, with

$$B_k(\Psi)\, w_k = -h\, \bar{\Psi}\, w_k^{\rm spsa} + \tilde{w}_k, \tag{64}$$

where $\tilde{w}_k$ aggregates the remaining centered noise contributions (oracle noise, communication noise, and centered residuals) that are not premultiplied by $\bar{\Psi}$.

Proof. Start from (60) and absorb $C_k(\bar{z}_k - z_\star) - \delta_k^\theta$ into the redefined drift b_k in (62). The remaining stochastic terms in the Taylor expansion enter linearly through the SPSA and oracle channels; apply Lemma A to $\bar{g}_k^{\rm spsa}$ and Lemma A to collect all centered innovations into w_k, so that (63) holds. The factor $-h\bar{\Psi}$ on the SPSA martingale difference follows from the Ψ-modified update (33) and Lemma A. This yields the representation (61). □

Lemma 6 (One-step expansion of the second moment) Let $\Sigma_k \triangleq \mathbb{E}[e_k e_k^\top]$ as in (41). Assume that the second moments appearing below exist. If e_{k+1} satisfies (61), i.e., $e_{k+1} = A_k e_k + B_k(\Psi)\, w_k + b_k + r_k$, then

$$\Sigma_{k+1} = \mathbb{E}\Big[\big(A_k e_k + B_k(\Psi)\, w_k + b_k + r_k\big) \cdot \big(A_k e_k + B_k(\Psi)\, w_k + b_k + r_k\big)^\top\Big]. \tag{65}$$

Proof. By definition of the covariance (second-moment) matrix,

$$\Sigma_{k+1} \triangleq \mathbb{E}\big[e_{k+1} e_{k+1}^\top\big]\,. \tag{66}$$

From the one-step representation (61) we have

$$e_{k+1} = A_k e_k + B_k(\Psi)\, w_k + b_k + r_k. \tag{67}$$

Substituting (67) into (66) gives

$$\Sigma_{k+1} = \mathbb{E}\Big[\big(A_k e_k + B_k(\Psi)\, w_k + b_k + r_k\big) \cdot \big(A_k e_k + B_k(\Psi)\, w_k + b_k + r_k\big)^\top\Big], \tag{68}$$

which coincides with (65). □

Assumption 2 (Second-moment bounds and cross-term control) In the setting of Corollary A and Lemma A, fix a filtration $(\mathcal{F}_k)_{k\ge 0}$ such that all random quantities at outer iteration k are $\mathcal{F}_k$-measurable, $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$, and $e_k, A_k, B_k(\Psi)$ are $\mathcal{F}_{k-1}$-measurable. Assume:

(a) **Conditional covariance majorant for w_k.** There exists an $\mathcal{F}_{k-1}$-measurable matrix $\Omega_k \succeq 0$ such that

$$\mathbb{E}[w_k w_k^\top \mid \mathcal{F}_{k-1}] \preceq \Omega_k \qquad \text{a.s.} \tag{69}$$

(b) **Cross-term control (centering).** The aggregated noise w_k is conditionally centered as in (63). Because e_k and b_k are $\mathcal{F}_{k-1}$-measurable, the sufficient conditions

$$\begin{aligned} &\mathbb{E}[e_k w_k^\top \mid \mathcal{F}_{k-1}] = 0, \\ &\mathbb{E}[b_k w_k^\top \mid \mathcal{F}_{k-1}] = 0 \qquad \text{a.s.} \end{aligned} \tag{70}$$

hold automatically. Any remaining mixed terms involving r_k are absorbed into a PSD remainder Γ_k.

(c) **Finite second moments and higher-order remainder.** The vectors b_k and r_k have finite second moments. In addition, the higher-order remainder is dominated in second moment by a vanishing (or higher-order) term, e.g., there exists $R_k \succeq 0$ such that $\mathbb{E}[r_k r_k^\top] \preceq R_k$ and $R_k \to 0$.

Lemma 7 (Majorized recursion for the covariance matrix) Assume the one-step representation (61), Assumption A, and that A_k is $\mathcal{F}_{k-1}$-measurable. Let Ω_k satisfy (69). Then there exists a sequence of matrices $\Gamma_k \succeq 0$ such that

$$\Sigma_{k+1} \preceq \mathbb{E}[A_k e_k e_k^\top A_k^\top] + \mathbb{E}[B_k(\Psi)\, \Omega_k\, B_k(\Psi)^\top] + \mathbb{E}[\Gamma_k]. \tag{71}$$

One may choose Γ_k to aggregate:

quadratic and mixed terms involving r_k;
cross terms involving b_k (e.g., $\mathbb{E}[A_k e_k b_k^\top]$, $\mathbb{E}[b_k r_k^\top]$) whenever they are not zero;
the nonnegative gap between the true noise covariance $\mathbb{E}[w_k w_k^\top \mid \mathcal{F}_{k-1}]$ and its majorant Ω_k.

Proof. From Lemma A and (61), define

$$x_k \triangleq A_k e_k, \quad y_k \triangleq B_k(\Psi) w_k, \quad z_k \triangleq b_k + r_k, \tag{72}$$

so that $\Sigma_{k+1} = \mathbb{E}[(x_k + y_k + z_k)(x_k + y_k + z_k)^\top]$. Expanding the outer product and grouping terms gives the usual decomposition into quadratic blocks and cross-terms.

Main quadratic blocks. Because A_k is $\mathcal{F}_{k-1}$-measurable,

$$\mathbb{E}[x_k x_k^\top] = \mathbb{E}\big[\mathbb{E}[A_k e_k e_k^\top A_k^\top \mid \mathcal{F}_{k-1}]\big] = \mathbb{E}[A_k e_k e_k^\top A_k^\top]. \tag{73}$$

Second, by (69) and monotonicity of Loewner order under congruence,

$$\begin{aligned} \mathbb{E}[y_k y_k^\top] &= \mathbb{E}\big[B_k(\Psi)\, \mathbb{E}[w_k w_k^\top \mid \mathcal{F}_{k-1}]\, B_k(\Psi)^\top\big] \\ &\preceq \mathbb{E}[B_k(\Psi)\, \Omega_k\, B_k(\Psi)^\top]. \end{aligned} \tag{74}$$

Cross terms and remainder. Since e_k and b_k are $\mathcal{F}_{k-1}$-measurable and $\mathbb{E}[w_k \mid \mathcal{F}_{k-1}] = 0$, $\mathbb{E}[x_k y_k^\top] = \mathbb{E}[y_k x_k^\top] = 0$ and $\mathbb{E}[y_k z_k^\top] = \mathbb{E}[z_k y_k^\top] = 0$. The remaining mixed terms are bounded using $uv^\top + vu^\top \preceq uu^\top + vv^\top$ (since $(u-v)(u-v)^\top \succeq 0$) and collected into $\Gamma_k \succeq 0$ together with $z_k z_k^\top$ and the noise-covariance gap. Taking expectation yields (71). $\square$

Assumption 3 (Stationary majorants) Assume that there exist deterministic matrices $A \in \mathbb{R}^{nd\times nd}$, $B(\Psi) \in \mathbb{R}^{nd\times m}$, $\Omega \in \mathbb{R}^{m\times m}$, and $Q \succeq 0$ such that, for all sufficiently large k (or along a chosen asymptotic regime), the following Loewner majorants hold with $\Sigma_k = \mathbb{E}[e_k e_k^\top]$:

$$\mathbb{E}[A_k e_k e_k^\top A_k^\top] \preceq A\Sigma_k A^\top + \varepsilon_k^{(A)}, \tag{75}$$

$$\mathbb{E}\big[B_k(\Psi)\,\Omega_k\,B_k(\Psi)^\top\big] \preceq B(\Psi)\,\Omega\,B(\Psi)^\top + \varepsilon_k^{(B)}, \tag{76}$$

$$\mathbb{E}[\Gamma_k] \preceq Q + \varepsilon_k^{(\Gamma)}. \tag{77}$$

Here $\varepsilon_k^{(A)}, \varepsilon_k^{(B)}, \varepsilon_k^{(\Gamma)} \succeq 0$ vanish in the chosen asymptotic sense (e.g., $\|\varepsilon_k^{(\cdot)}\|_2 \to 0$). Define the affine comparison map

$$\mathcal{T}(\Sigma) \triangleq A\Sigma A^\top + B(\Psi)\,\Omega\,B(\Psi)^\top + Q. \tag{78}$$

Lemma 8 (Fixed map on PSD_{nd}) Assume Lemma A and Assumption A. With $\mathcal{T}$ as in (78), there exists a sequence $E_k^{(0)} \succeq 0$ with $\|E_k^{(0)}\|_2 \to 0$ such that

$$\Sigma_{k+1} \preceq \mathcal{T}(\Sigma_k) + E_k^{(0)}. \tag{79}$$

Proof. Start from the majorized recursion of Lemma A,

$$\begin{aligned}\Sigma_{k+1} \preceq\ & \mathbb{E}[A_k \Sigma_k A_k^\top] \\ & + \mathbb{E}[B_k(\Psi)\,\Omega_k\,B_k(\Psi)^\top] \\ & + \mathbb{E}[\Gamma_k].\end{aligned} \tag{80}$$

We majorize each term on the right-hand side using Assumption A.

Step 1: the drift block. By (75),

$$\mathbb{E}[A_k e_k e_k^\top A_k^\top] \preceq A\Sigma_k A^\top + \varepsilon_k^{(A)}. \tag{81}$$

Step 2: the noise block. By (76),

$$\begin{aligned}\mathbb{E}[B_k(\Psi)\,\Omega_k\,B_k(\Psi)^\top] \preceq\ & B(\Psi)\,\Omega\,B(\Psi)^\top \\ & + \varepsilon_k^{(B)}.\end{aligned} \tag{82}$$

Step 3: the remainder. By (77),

$$\mathbb{E}[\Gamma_k] \preceq Q + \varepsilon_k^{(\Gamma)}. \tag{83}$$

Step 4: collect terms. Summing (81), (82), and (83) and substituting into (80) yields

$$\begin{aligned}\Sigma_{k+1} \preceq\ & A\Sigma_k A^\top \\ & + B(\Psi)\,\Omega\,B(\Psi)^\top + Q \\ & + \varepsilon_k^{(A)} + \varepsilon_k^{(B)} + \varepsilon_k^{(\Gamma)}.\end{aligned} \tag{84}$$

Step 5: define the vanishing PSD error. Set

$$E_k^{(0)} \triangleq \varepsilon_k^{(A)} + \varepsilon_k^{(B)} + \varepsilon_k^{(\Gamma)}. \tag{85}$$

Since each summand is PSD, we have $E_k^{(0)} \succeq 0$. Moreover, $E_k^{(0)} \to 0$ in the chosen asymptotic sense because each $\varepsilon_k^{(\cdot)}$ vanishes by Assumption A. Finally, by the definition (78) of $\mathcal{T}$, the right-hand side of (84) equals $\mathcal{T}(\Sigma_k) + E_k^{(0)}$, which is precisely (79). $\square$

Assumption 4 (Spectral contraction of $\mathcal{T}$) Consider the affine map $\mathcal{T}$ in (78). Assume that $\mathcal{T}$ preserves PSD order, is monotone in the Loewner sense, and that

$$\|A\|_2 < 1, \tag{86}$$

where $\|\cdot\|_2$ is the spectral norm. Then, for all symmetric matrices Σ_1, Σ_2,

$$\|\mathcal{T}(\Sigma_1) - \mathcal{T}(\Sigma_2)\|_2 \le \|A\|_2^2\,\|\Sigma_1 - \Sigma_2\|_2. \tag{87}$$

Proof. Because $\mathcal{T}(\Sigma) = A\Sigma A^\top + K$ with fixed $K = B(\Psi)\Omega B(\Psi)^\top + Q$, $\mathcal{T}(\Sigma_1) - \mathcal{T}(\Sigma_2) = A(\Sigma_1 - \Sigma_2)A^\top$. Taking spectral norms gives (87). $\square$

Remark 1 (Loewner bound from a spectral norm) If X, Y are symmetric and $X \preceq Y$, then $\|X\|_2 \le \|Y\|_2$. Indeed, $Y - X \succeq 0$ implies $\lambda_{\min}(Y) \le \lambda_{\min}(X) \le \lambda_{\max}(X) \le \lambda_{\max}(Y)$, hence $\|X\|_2 = \max\{|\lambda_{\min}(X)|, |\lambda_{\max}(X)|\} \le \max\{|\lambda_{\min}(Y)|, |\lambda_{\max}(Y)|\} = \|Y\|_2$.

Lemma 9 (Fixed point and upper comparison) Under Assumption A, there exists a unique symmetric matrix $\Sigma^\star$ such that

$$\Sigma^\star = \mathcal{T}(\Sigma^\star). \tag{88}$$

Moreover, if the perturbed recursion

$$\Sigma_{k+1} \preceq \mathcal{T}(\Sigma_k) + E_k^{(0)}, \qquad E_k^{(0)} \succeq 0, \quad \|E_k^{(0)}\|_2 \to 0, \tag{89}$$

holds, then there exists a sequence $E_k \succeq 0$ with $\|E_k\|_2 \to 0$ such that

$$\Sigma_k \preceq \Sigma^\star + E_k. \tag{90}$$

Proof.

Step 1: existence and uniqueness of the fixed point. Because $\|A\|_2 < 1$, the map $\mathcal{T}$ is a contraction on the space of symmetric matrices equipped with $\|\cdot\|_2$, with constant $q = \|A\|_2^2 \in (0,1)$. By the Banach fixed-point theorem there exists a unique $\Sigma^\star$ satisfying (88).

Step 2: stability under a vanishing perturbation. Define $\Delta_k \triangleq \Sigma_k - \Sigma^\star$. Using $\Sigma^\star = \mathcal{T}(\Sigma^\star)$ and (89),

$$\begin{aligned}\Delta_{k+1} = \Sigma_{k+1} - \Sigma^\star \preceq \mathcal{T}(\Sigma_k) - \mathcal{T}(\Sigma^\star) + E_k^{(0)} \\ = A\Delta_k A^\top + E_k^{(0)}.\end{aligned} \tag{91}$$

Because $A\Delta_k A^\top \preceq \|A\|_2^2 \|\Delta_k\|_2 I_{nd}$ and $E_k^{(0)} \preceq \|E_k^{(0)}\|_2 I_{nd}$, transitivity of the Loewner order gives $\Delta_{k+1} \preceq q\|\Delta_k\|_2 I_{nd} + \|E_k^{(0)}\|_2 I_{nd}$ with $q = \|A\|_2^2$. Applying Remark A,

$$\|\Delta_{k+1}\|_2 \le q\,\|\Delta_k\|_2 + \|E_k^{(0)}\|_2. \tag{92}$$

Iterating (92) yields, for all $k \ge 1$,

$$\|\Delta_k\|_2 \le q^k \|\Delta_0\|_2 + \sum_{j=0}^{k-1} q^{k-1-j}\,\|E_j^{(0)}\|_2. \tag{93}$$

Since $q \in (0,1)$ and $\|E_k^{(0)}\|_2 \to 0$, the right-hand side of (93) tends to zero, hence $\|\Delta_k\|_2 \to 0$.

Step 3: from norm control to a Loewner upper comparison. Set $E_k \triangleq \|\Delta_k\|_2 I_{nd} \succeq 0$. Then $\Delta_k \preceq E_k$ and therefore $\Sigma_k = \Sigma^\star + \Delta_k \preceq \Sigma^\star + E_k$, which is (90). Moreover $\|E_k\|_2 = \|\Delta_k\|_2 \to 0$. □

Assumptions 1–4 of the main text correspond to Appendix Assumptions A–A together with the majorization relations in (75)–(77). By Corollary A and Lemma A,

$$\Sigma_{k+1} \preceq \mathbb{E}[A_k e_k e_k^\top A_k^\top] + \mathbb{E}[B_k(\Psi)\Omega_k B_k(\Psi)^\top] + \mathbb{E}[\Gamma_k].$$

Applying the stationary majorants and Lemma A gives $\Sigma_{k+1} \preceq \mathcal{T}(\Sigma_k) + E_k^{(0)}$ with $\|E_k^{(0)}\|_2 \to 0$. Lemma A then yields $\Sigma_k \preceq \Sigma^\star + E_k$ with $\|E_k\|_2 \to 0$, which is exactly the claim of Theorem 1. □

B Proof for the simulation benchmark

We verify that the synthetic quadratic experiment in Section 5 falls within the scope of Theorem 1. Throughout, $H \in \mathbb{R}^{d\times d}$ is symmetric positive definite, $L \in \mathbb{R}^{n\times n}$ is the (fixed) Laplacian of the undirected ring graph used in the experiments, $\bar{H} \triangleq I_n \otimes H$, and $\bar{L} \triangleq L \otimes I_d$.

Proposition 1 (Quadratic benchmark) Consider the recursion of Section 4.2 with the following specialization (matching Section 5.1):

(i) **Quadratic local losses.** Each agent observes $f_i(x) = \frac{1}{2}x^\top H x$, so the stacked objective is

$$\bar{F}(\bar\theta) = \tfrac{1}{2}\bar\theta^\top \bar{H}\,\bar\theta + \tfrac{\omega}{2}\bar\theta^\top \bar{L}\,\bar\theta.$$

(ii) **Static graph.** $L_{2k-1} = L$ for all k.

(iii) **Stationary reference.** $\theta_{2k} \equiv 0$, hence $e_k = \bar\theta_{2k}$ in (41).

(iv) **Noise model.** Independent mean-zero oracle noise with variance σ_v^2 on the ADSPSA channel (standard deviation 0.05 in the experiments) and independent Rademacher SPSA directions Δ_k^i as in (26).

(v) **Fixed schedule and step.** $\beta_k \equiv \beta > 0$ and the accelerated parameters (α_k, γ_k) are generated from fixed (μ, η, γ_0, h) as in (28)–(29); the step h and H are drawn as in Section 5.1 but satisfy

$$\|A_{\text{lin}}\|_2 < 1, \qquad A_{\text{lin}} \triangleq I_{nd} - h\,\bar\Psi\,\bar{H} - h\,\omega\,\bar{L}, \tag{94}$$

where $\bar\Psi = I_n \otimes \Psi$ (with $\Psi = I$ or the tuned $\Psi^\star$).

Then Assumptions 1–4 hold and Theorem 1 applies to the tracking-error covariance $\Sigma_k = \mathbb{E}[e_k e_k^\top]$.

Proof. We verify each assumption in order.

Step 1: Assumption 1 (smooth representation and measurability). Because $\bar{F}$ is quadratic, its gradient is the affine map $\nabla\bar{F}(\bar\theta) = \bar{H}\,\bar\theta + \omega\bar{L}\,\bar\theta = (\bar{H} + \omega\bar{L})\bar\theta$. Fix the filtration $\mathcal{F}_k = \sigma(\bar\theta_{2k}, \bar{z}_k, \text{past noise})$. At outer iteration k, the quantities $\bar\theta_{2k}$, $\tilde{x}_k$, and the coefficients assembled from (α_k, γ_k) are $\mathcal{F}_{k-1}$-measurable once the past iterates and past noise are in $\mathcal{F}_{k-1}$ (measurability convention before Section 4.4). The map $\bar\theta_{2k} \mapsto \bar\theta_{2(k+1)}$ induced by (33)–(35) is continuously differentiable; its Jacobian at $\bar\theta_{2k} = 0$ is A_{lin} in (94) up to the $\bar\Psi$-weighted SPSA channel (Lemma A). Because $\theta_{2k} \equiv 0$, the target shift δ_k^θ in Lemma A vanishes. Lemma A therefore applies with a quadratic remainder $r_k = O(\|e_k\|^2)$.

Step 2: Assumption 2 (centering and noise majorant). Let v_k denote the stacked oracle noise at iteration k and let ξ_k collect the SPSA directions $\bar\Delta_k$. By construction, $\mathbb{E}[v_k \mid \mathcal{F}_{k-1}] = 0$ and $\mathbb{E}[\bar\Delta_k \mid \mathcal{F}_{k-1}] = 0$. Aggregating the innovation at step k into w_k as in Corollary A yields $\mathbb{E}[w_k \mid \mathcal{F}_{k-1}] = 0$. Because e_k and the drift b_k are $\mathcal{F}_{k-1}$-measurable and $b_k = 0$ when $\theta_{2k} \equiv 0$, the sufficient cross-term conditions (70) hold. The conditional covariance $\mathbb{E}[w_k w_k^\top \mid \mathcal{F}_{k-1}]$ is bounded uniformly in k by a matrix $\Omega_k \preceq \Omega$ depending on σ_v^2, β, and d (SPSA scaling $1/\sqrt{d}$), so Appendix Assumption A(a) holds.

Step 3: Assumption 3 (stationary majorants). Near $\bar\theta = 0$, the linearized one-step error dynamics take the form (61) with $A_k \equiv A_{\text{lin}}$, $B_k(\Psi) \equiv B(\Psi)$ collecting the SPSA/oracle noise gain, and $b_k \equiv 0$. The Taylor remainder and higher-order terms are absorbed into Γ_k with $\mathbb{E}[\Gamma_k] \preceq Q$ for a fixed $Q \succeq 0$ that depends on h, $\|H\|$, and the neighborhood radius of the iterates (which remains bounded in the experiment because $\bar{F}$ is coercive and the step h is small). Thus (75)–(77) hold with

$\varepsilon_k^{(A)} = \varepsilon_k^{(B)} = \varepsilon_k^{(\Gamma)} = 0$ for the linearized model, and the comparison map (78) is

$$\mathcal{T}(\Sigma) = A_{\mathrm{lin}} \Sigma A_{\mathrm{lin}}^{\top} + B(\Psi)\, \Omega\, B(\Psi)^{\top} + Q.$$

This is exactly the surrogate Lyapunov majorant fitted in Section 5.1 when (A, S) is identified from trajectories.

Step 4: Assumption 4 (spectral contraction). The map (78) is affine and monotone on PSD_{nd}. By hypothesis (94), $\|A_{\mathrm{lin}}\|_2 < 1$, hence Assumption A holds with contraction constant $q = \|A_{\mathrm{lin}}\|_2^2$ (Appendix, proof after (87)). □

References

Amelina, N., Fradkov, A., Jiang, Y., and Vergados, D. J. (2015). Approximate consensus in stochastic networks with application to load balancing. *IEEE Trans. Inf. Theory*, **61** (4), pp. 1739–1752.

Amelina, N. O. (2013). Local voting protocol for decentralized load balancing of network with switched topology and noise in measurements. *Vestnik St. Petersbg. Univ., Math.*, **46** (3), pp. 118–125.

Amelina, N. O. and Fradkov, A. L. (2012). Approximate consensus in the dynamic stochastic network with incomplete information and measurement delays. *Autom. Remote Control*, **73** (11), pp. 1765–1783.

Arkhipov, I., Krokhalev, E., and Granichin, O. (2026a). Convergence analysis of consensus-based tracking algorithms for logical time functions in multi-agent networks. In *Proc. ApPLIED Workshop: Advanced Tools, Programming Languages, and PLatforms for Implementing and Evaluating algorithms for Distributed systems (ApPLIED 2026)*. Held in conjunction with PODC 2026, Egham, England.

Arkhipov, I., Krokhalev, E., Tarasova, E., and Granichin, O. (2026b). Monotonic logical time for cooperative tracking in decentralized multi-agent networks. *IEEE Access*. submitted.

Baek, Y. et al. (2025). Edge intelligence through in-sensor and near-sensor computing for the artificial intelligence of things. *npj Unconventional Computing*, **2** (1), pp. 25.

Chernov, A., Erofeeva, V., Granichin, O., Moseiko, E., and Sudomir, A. (2025a). SPSA-based consensus algorithm for mutual IoT device positioning. *IFAC-PapersOnLine*, **59** (14), pp. 133–138. Proc. 21st IFAC Int. Workshop Adaptation Learn. Control Signal Process. (ALCOSP).

Chernov, A., Erofeeva, V., Granichin, O., and Sudomir, A. (2025b). Accelerated SPSA-based consensus algorithm for mutual device positioning. In *Proc. 64th IEEE Conf. Decision Control (CDC)*, Rio de Janeiro, Brazil, pp. 8071–8076.

Erofeeva, V. et al. (2022). Accelerated simultaneous perturbation stochastic approximation for tracking under unknown-but-bounded disturbances. In *Proc. Amer. Control Conf. (ACC)*, pp. 1582–1587.

Erofeeva, V., Granichin, O., and Sergeenko, A. (2026). Accelerated consensus-based SPSA algorithm for multisensor multitarget tracking problem. *IEEE Trans. Autom. Control*. early access.

Erofeeva, V., Granichin, O., and Volodina, E. (2024). Accelerated decentralized load balancing in multi-agent networks. *IEEE Access*, **12**, pp. 161954–161967.

Erofeeva, V. and Parsegov, S. (2024). Properties of the Laplacian spectra of certain basic and hierarchical graphs. *Cybern. Phys.*, **13** (1), pp. 12–19.

He, H. et al. (2026). DAUNet: Unsupervised neural network based on dual attention for clock synchronization in multi-agent wireless. *Comput. Mater. Continua*, **86** (1), pp. 1–23.

Imangazieva, A. (2024). Robust control of one class of nonlinear systems with a distributed delay. *Cybern. Phys.*, **13** (1), pp. 42–48.

Rice, J. K. and Verhaegen, M. (2009). Distributed computations and control in multi-agent systems. In *Proc. 4th Int. Conf. Autonomous Robots and Agents (ICARA)*, pp. 44–49.

Shukla, V. and Parker, G. (2025). Multi-agent orchestration protocol for generative AI systems. In *Proc. 8th Int. Seminar Research of Information Technology and Intelligent Systems (ISRITI)*, pp. 54–59.

Tarasova, E. and Moseiko, E. (2025). Decentralized SPSA-based correction of task time predictions in adaptive multi-agent systems. *Cybern. Phys.*, **14** (1), pp. 67–73.

Ullah, S. et al. (2024). Neuro-adaptive non-singular terminal sliding mode control for distributed fixed-time synchronization of higher-order uncertain multi-agent nonlinear systems. *Inf. Sci.*, **659**, pp. 120087.

Vergados, D. J., Amelina, N., Granichin, O., Jiang, Y., and Kralevska, K. (2018). Toward optimal distributed node scheduling in a multihop wireless network through local voting. *IEEE Trans. Wireless Commun.*, **17** (1), pp. 400–414.

Vogels, W. (2009). Eventually consistent. *Commun. ACM*, **52** (1), pp. 40–44.

Wang, H. et al. (2024). A fast convergence scheme for distributed consensus time synchronization using multihop virtual links in industrial wireless sensor networks. *IEEE Sensors J.*, **24** (12), pp. 20009–20017.

Wang, H. et al. (2025). Event-triggered time synchronization with varying skews for wireless sensor networks based on average consensus. *IEEE Trans. Netw. Sci. Eng.*

Zhuo, Z. et al. (2025). KAOS: Large model multi-agent operating system. In *Proc. China Conf. Knowledge Graph Semantic Comput. Int. Joint Conf. Knowledge Graphs (CCKS-IJCKG)*, vol. 2229 of *Commun. Comput. Inf. Sci.*, Singapore, Springer, pp. 347–359.