ACCELERATED CONSENSUS AND DECENTRALIZED OPTIMIZATION OVER SLOWLY VARYING NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study average consensus over slowly time-varying networks and its application to decentralized optimization. We introduce WAVE, a windowed Chebyshev method that limits the accumulated effect of network variation by restarting the recurrence after finite windows. If χ bounds the network condition number and successive communication operators satisfy $\|L_{k+1} - L_k\| \le \beta$, WAVE reaches ϵ-consensus in $O((\sqrt{\chi} + \min\{\beta\chi^2, \chi\}) \ln(e/\epsilon))$ communication rounds. This rate matches the fixed-network $\sqrt{\chi}$ dependence when $\beta = O(\chi^{-3/2})$ and smoothly interpolates across the full range of network variation. For piecewise-constant networks where each change is detected when it occurs and consecutive changes are at least τ rounds apart, WAVE reaches ϵ-consensus in $O((\sqrt{\chi} + \chi/\tau) \ln(e/\epsilon))$ communication rounds. Finally, we use WAVE as the consensus step in an accelerated decentralized optimization method for α-smooth convex local objectives with a μ-strongly convex average. For sufficiently slow network variation, every agent obtains an ϵ-solution to the global optimization problem after $\widetilde{O}(\sqrt{\alpha/\mu}\,\sqrt{\chi})$ communication rounds, matching the fixed-network dependence.

1 INTRODUCTION

The average consensus problem requires a network of agents to compute the mean of their private initial vectors using only neighbor-to-neighbor communication. It is a basic component of distributed computation, multi-agent control, distributed estimation, and decentralized optimization (DeGroot, 1974; Tsitsiklis, 1984; Jadbabaie et al., 2003; Olfati-Saber & Murray, 2004; Moreau, 2005; Nedić et al., 2018; Gorbunov et al., 2022; Rogozin et al., 2023). Related distributed optimization problems arise in estimation, control, resource allocation, and machine learning (Rabbat & Nowak, 2004; Forero et al., 2010; Beck et al., 2014; Giselsson et al., 2013; Konečný et al., 2016; McMahan et al., 2017). In both settings, communication is often the dominant cost, making the number of communication rounds a central complexity measure.

For a fixed normalized gossip operator with condition number χ on the disagreement subspace, standard gossip requires $O(\chi \ln(e/\epsilon))$ communication rounds. Classical Chebyshev acceleration improves this bound to $O(\sqrt{\chi} \ln(e/\epsilon))$ (Xiao & Boyd, 2004; Golub & Varga, 1961; Gutknecht & Röllin, 2002; Nguyen et al., 2024). For fixed networks, the same $\sqrt{\chi}$ network dependence appears in the optimal communication complexity of smooth strongly convex decentralized optimization (Scaman et al., 2017). For arbitrarily time-varying networks, however, the worst-case network dependence is linear in χ (Kovalev et al., 2021b;a). This gap motivates a fundamental question: how does communication complexity depend on the rate of network variation, and when can the fixed-network square-root acceleration be retained?

Network models and approach. We consider two complementary models of network variation. In the *metric-drift* model, the normalized gossip operator may change at every communication round, while consecutive normalized gossip matrices satisfy $\|L_{k+1} - L_k\| \le \beta$. In the *piecewise-constant* model, the operator may change abruptly but remains fixed between change points, with consecutive changes at least τ communication rounds apart. For both models, we introduce WAVE (**W**indowed **A**cceleration for **V**arying n**E**tworks). It applies the classical Chebyshev recurrence on the full spectral interval $[\chi^{-1}, 1]$ and restarts after finite windows. The metric-drift schedule chooses

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

window lengths using β and χ, while the piecewise-constant schedule restarts at changes reported by an ideal external detector.

Contributions. Our main contribution is a consensus method (WAVE) whose communication complexity adapts to the magnitude and frequency of network changes. We consider two complementary settings: small changes may occur at every communication round, or the network may change arbitrarily at detected change points and remain fixed between them. In both settings, our bounds interpolate between the fixed-network $\sqrt{\chi}$ dependence and the worst-case χ dependence. Below, $\widetilde{O}$ suppresses logarithmic factors.

- *Consensus.* We prove that WAVE reaches ϵ-consensus in

$$\widetilde{O}\big(\sqrt{\chi}+\min\{\beta\chi^2,\chi\}\big) \quad \text{and} \quad \widetilde{O}\Big(\sqrt{\chi}+\frac{\chi}{\tau}\Big)$$

  communication rounds under metric drift and detected piecewise-constant changes, respectively. The bounds interpolate between the accelerated $\sqrt{\chi}$ and robust χ dependences, recovering the former when $\beta = O(\chi^{-3/2})$ or $\tau = \Omega(\sqrt{\chi})$. Thus acceleration persists under small changes at every round or arbitrary detected jumps that are sufficiently separated, without requiring a persistent connected subgraph. An exact variation-of-constants argument controls the perturbations caused by noncommuting communication operators within a Chebyshev window and determines its length. The piecewise-constant schedule adapts without knowing τ. When $\tau \geq \sqrt{\chi}\ln(2n)+2$, a flooding-and-rollback protocol implements change detection inside the network with the same communication order plus $O(\sqrt{\chi}\ln n)$ rounds.
- *Decentralized optimization.* For α-smooth, μ-strongly convex local objectives, let $\kappa := \alpha/\mu$. Using WAVE for mixing in accelerated gradient tracking (Li & Lin, 2024) gives $O(\sqrt{\kappa}\ln(1/\epsilon))$ gradient evaluations and multiplies each consensus network factor above by $O(\sqrt{\kappa}\ln(1/\epsilon))$. In the accelerated regimes this yields $O(\sqrt{\kappa\chi}\ln(1/\epsilon))$, with the optimal fixed-network $\sqrt{\kappa\chi}$ dependence and a single accuracy logarithm (Scaman et al., 2017). When only the average objective is strongly convex, WAVE-STM uses restarted Similar Triangles iterations (Nesterov, 2004) and gives the same network factors with $\widetilde{O}(\sqrt{\kappa})$ gradient evaluations and a further accuracy logarithm in communication.

Related work and comparison. For arbitrary time variation, ADOM and ADOM+ attain the optimal worst-case network dependence (Kovalev et al., 2021b;a). Subgradient-push, gradient-tracking, and primal-dual methods provide convergence under broad connectivity assumptions (Nedić & Olshevsky, 2015; Nedić et al., 2017; Maros & Jaldén, 2018; Li & Lin, 2024; Rogozin et al., 2021). These guarantees are designed for worst-case variation and do not improve continuously as the network changes more slowly. Existing acceleration results use different restrictions on network variation. Rogozin et al. (2019) allow arbitrarily large changes and retain the fixed-network rate when the number of changing rounds among the first N iterations is $O(N/[\sqrt{\kappa\chi}\log(\kappa\chi)])$, so the admissible fraction of changing rounds decreases with $\kappa\chi$. Our metric-drift model allows changes at every round, while our detected piecewise-constant model recovers the accelerated network factor when $\tau = \Omega(\sqrt{\chi})$, independently of the objective condition number κ. Metelev et al. (2023) require all communication graphs to share a connected spanning subgraph, and their network factor is controlled by the spectral gap of this persistent subgraph. Montijano et al. (2013) give sufficient conditions for convergence of a Chebyshev iteration over time-varying networks, but do not derive a communication bound that adapts to the rate of variation. Metelev et al. (2024) consider a stationary, uniformly geometrically mixing Markov chain of networks and obtain bounds with additional mixing-dependent terms. They also show that, for first-order decentralized optimization, removing one edge and adding another at every round can retain a linear dependence on χ. Li & Lin (2024) use repeated gossip for their time-varying accelerated method; WAVE replaces this mixing step when variation is slow. Our deterministic bounds interpolate between the fixed-network and arbitrary-variation regimes through the operator drift or the time between detected changes. To the best of our knowledge, our metric-drift result is the first deterministic average-consensus complexity bound that gives this interpolation as an explicit function of the one-step operator-norm change. For time-varying quadratics, Pasechnyuk-Vilensky & Takáč (2026) study the first-order sensitivity of Chebyshev realizations to Hessian perturbations. Here we bound the full perturbation accumulated in a window and use it to obtain consensus communication guarantees.

Organization. Section 2 formulates the consensus problem and introduces WAVE. Section 3 states the main consensus guarantees. Sections 4 and 5 analyze the metric-drift and piecewise-constant models, respectively. Section 6 presents the application to decentralized optimization. Section 7 reports the numerical results, and Section 8 concludes the paper. The supplementary material contains complete proofs, implementation details, the extension to memoryless changes, and additional experiments.

2 Problem Statement and Algorithm

This section formulates the consensus problem, specifies the two models of network variation, and introduces WAVE. We first recall the classical Chebyshev recurrence for a fixed network, then explain how finite-window restarts extend it to changing communication operators.

Notation. The norm $\|\cdot\|$ is Euclidean for vectors and spectral for matrices, and $\mathrm{spec}(A)$ denotes the set of eigenvalues of a square matrix A. For any positive integer p, $\mathbf{1}_p$ is the vector of p ones and I_p is the $p \times p$ identity matrix; $\otimes$ denotes the Kronecker product. The complete notation is collected in Table 1 (Appendix A).

2.1 Consensus and Network Model

We consider a network of agents represented at each communication round k by an undirected connected graph $\mathcal{G}_k = (V, \mathcal{E}_k)$, where $V = \{1, \dots, n\}$ is the set of agents and $\mathcal{E}_k$ is the set of edges. Agent i holds $u_k^{(i)} \in \mathbb{R}^d$ and may exchange vectors only with its current neighbors. The goal is for every agent to compute the average $\bar{u}_0 := n^{-1} \sum_{i=1}^n u_0^{(i)}$ of the initial vectors.

Standard decentralized optimization formulations model local communication by a symmetric positive semidefinite gossip matrix whose sparsity agrees with the communication graph (Scaman et al., 2017; Nguyen et al., 2024). We measure network variation by the distance between successive communication operators, which changes under rescaling. We therefore place all communication operators on the same scale.

Definition 2.1 (Normalized gossip matrix)**.** A matrix $L \in \mathbb{R}^{n \times n}$ associated with a connected graph $G = (V, \mathcal{E})$ is a normalized gossip matrix if $L = L^\top \succeq 0$, $\ker L = \mathrm{span}\{\mathbf{1}_n\}$, $L_{ij} = 0$ whenever $i \neq j$ and $\{i, j\} \notin \mathcal{E}$, and $\|L\| \leq 1$.

Throughout the paper, L_k denotes a normalized gossip matrix associated with $\mathcal{G}_k$, with the same normalization used at every round. Concretely, the unnormalized operators are divided by one common upper bound on their norms rather than rescaled independently at each round. Consequently, $\|L_{k+1} - L_k\|$ compares operators expressed on the same scale. The kernel condition identifies the consensus subspace, while the sparsity condition makes multiplication by $L_k \otimes I_d$ implementable using one neighbor exchange. We therefore count one such multiplication as one communication round.

The Chebyshev coefficients must be valid for every operator encountered by the algorithm. We therefore impose one uniform spectral bound on the whole operator sequence.

Assumption 2.2 (Uniform spectral gap)**.** There is a known $\chi \geq 1$ such that

$$\mathrm{spec}(L_k) \setminus \{0\} = \mathrm{spec}\big(L_k|_{\mathbf{1}_n^\perp}\big) \subseteq [\chi^{-1}, 1] \qquad \text{for every } k \geq 0.$$

Thus χ is a uniform upper bound on the condition number of the communication operators on $\mathbf{1}_n^\perp$.

Since χ is a conservative upper bound, any valid bound $\chi < 4$ may be replaced by 4. We therefore assume $\chi \geq 4$ throughout. For $1 \leq \chi < 4$, standard gossip already reaches ϵ-consensus in $O(\ln(e/\epsilon))$ communication rounds under arbitrary network variation.

2.2 Network Variation

We consider two complementary restrictions on network variation. Metric drift bounds the size of each change but allows the network to change at every round. The piecewise-constant model allows arbitrarily large changes but requires consecutive changes to be separated in time.

Assumption 2.3 (Metric drift)**.** For a known $\beta \in [0, 1]$, $\|L_{k+1} - L_k\| \le \beta$ for every $k \ge 0$.

The parameter β is available to the metric-drift scheduler. No commutativity is assumed between consecutive operators.

Assumption 2.4 (Piecewise-constant network with minimum spacing)**.** There are change points $0 = t_0 < t_1 < t_2 < \cdots$ such that $L_k = L_{t_i}$ for $t_i \le k < t_{i+1}$ and $t_{i+1} - t_i \ge \tau$, where $\tau \ge 1$. A new matrix may depend on the previous network history, and its distance from the preceding matrix is unrestricted.

For piecewise-constant networks, an ideal external detector reports a change at the corresponding round boundary, before the new operator is used. The algorithm does not know τ. The basic analysis treats reports as available at no communication cost; Proposition 5.1 implements them by messages inside the network under a longer spacing condition.

2.3 Chebyshev acceleration and WAVE

For the analysis, stack the local states as $\boldsymbol{u}_k := ((u_k^{(1)})^\top, \dots, (u_k^{(n)})^\top)^\top \in \mathbb{R}^{nd}$ and define $\boldsymbol{e}_k := \boldsymbol{u}_k - \mathbf{1}_n \otimes \bar{u}_0$. To describe communication between these stacked states, we use $\boldsymbol{L}_k = L_k \otimes I_d$.

Definition 2.5 (ϵ-consensus)**.** For $\epsilon \in (0, 1)$, a stacked state $\boldsymbol{u} \in \mathbb{R}^{nd}$ is an ϵ-consensus solution relative to $\boldsymbol{u}_0$ if $\|\boldsymbol{u} - \mathbf{1}_n \otimes \bar{u}_0\| \le \epsilon \|\boldsymbol{u}_0 - \mathbf{1}_n \otimes \bar{u}_0\|$.

Fixed network. If $\boldsymbol{L}_k \equiv \boldsymbol{L}$, every mean-preserving method in the linear operator model (Appendix B) that uses m communication rounds returns $\boldsymbol{e}_m = p(\boldsymbol{L})\boldsymbol{e}_0$ for a polynomial p with $\deg p \le m$ and $p(0) = 1$. By Assumption 2.2, it contracts the disagreement by at most $\max_{\lambda \in [\chi^{-1}, 1]} |p(\lambda)|$. The best such polynomial is built from the Chebyshev polynomials $T_0(x) = 1$, $T_1(x) = x$, $T_{t+1}(x) = 2xT_t(x) - T_{t-1}(x)$, which are bounded by one on $[-1, 1]$ and grow rapidly outside it. Let

$$z_\chi(\lambda) := \frac{1 + \chi^{-1} - 2\lambda}{1 - \chi^{-1}}, \qquad z_0 := z_\chi(0), \qquad P_t^\chi(\lambda) := \frac{T_t(z_\chi(\lambda))}{T_t(z_0)}. \tag{1}$$

The affine map z_χ sends the admissible disagreement spectrum $[\chi^{-1}, 1]$ to $[-1, 1]$, while $z_0 > 1$. Hence $P_t^\chi(0) = 1$, and the residual P_t^χ is uniformly small on $[\chi^{-1}, 1]$.

Proposition 2.6 (Classical Chebyshev residual)**.** *For $m \ge 1$ and $\chi > 1$, P_m^χ is the unique minimizer of*

$$\rho_m^\chi := \min_{\deg p \le m,\ p(0)=1} \max_{\lambda \in [\chi^{-1}, 1]} |p(\lambda)|, \qquad \rho_m^\chi = \frac{1}{T_m(z_0)} = \operatorname{sech}(m\theta_\chi) \le 2e^{-2m/\sqrt{\chi}}, \tag{2}$$

where $\theta_\chi = 2\,\operatorname{artanh}(\chi^{-1/2})$.

Thus a degree-$\Theta(\sqrt{\chi})$ Chebyshev polynomial contracts the disagreement by a constant factor, and its repetition yields the fixed-network $O(\sqrt{\chi}\ln(1/\epsilon))$ complexity (Golub & Varga, 1961; Gutknecht & Röllin, 2002; Scaman et al., 2017). WAVE also uses windows with $m \ll \sqrt{\chi}$, for which $2e^{-2m/\sqrt{\chi}}$ may exceed one; the exact sech value still gives progress.

Lemma 2.7 (Short-window progress)**.** *If $\chi \ge 4$, $m \ge 1$, and $m^2 \le 2\chi$, then*

$$\rho_m^\chi \le \exp\left(-\frac{4m^2}{5\chi}\right). \tag{3}$$

A window of length $m \le \sqrt{\chi}$ thus earns order m/χ of logarithmic contraction per round, between the gossip rate $1/\chi$ ($m = 1$) and the accelerated rate $1/\sqrt{\chi}$ ($m \asymp \sqrt{\chi}$). The complexity of WAVE is therefore governed by the window length that network variation admits. Appendix C contains the proofs and an illustration of the residuals.

Chebyshev window. Dividing the Chebyshev recurrence by $T_{t+1}(z_0)$ gives $P_{t+1}^\chi(\lambda) = a_t z_\chi(\lambda) P_t^\chi(\lambda) - c_t P_{t-1}^\chi(\lambda)$, where $a_t := 2T_t(z_0)/T_{t+1}(z_0)$ and $c_t := T_{t-1}(z_0)/T_{t+1}(z_0)$ are computed without forming the growing values $T_t(z_0)$:

$$a_0 = \frac{2}{z_0}, \qquad a_t = \frac{4}{4z_0 - a_{t-1}}, \qquad c_t = \frac{a_{t-1}a_t}{4} \quad (t \ge 1). \tag{4}$$

The coefficients do not depend on the operator, and each step uses one multiplication by it, that is, one communication round. The recurrence can therefore use the operator of the current round.

Definition 2.8 (Chebyshev window)**.** For $\boldsymbol{L} = L \otimes I_d$, let $z_\chi(\boldsymbol{L}) := ((1+\chi^{-1})I_{nd} - 2\boldsymbol{L})/(1-\chi^{-1})$. For given window length $s \geq 1$ and starting round k, initialize $\boldsymbol{v}_0 = \boldsymbol{u} \in \mathbb{R}^{nd}$ and compute

$$\boldsymbol{v}_1 = z_0^{-1} z_\chi(\boldsymbol{L}_k)\boldsymbol{v}_0, \qquad \boldsymbol{v}_{t+1} = a_t z_\chi(\boldsymbol{L}_{k+t})\boldsymbol{v}_t - c_t \boldsymbol{v}_{t-1}, \quad t = 1, \ldots, s-1. \tag{5}$$

We call the resulting map $\mathcal{C}_{k,s} : \mathbb{R}^{nd} \to \mathbb{R}^{nd}$, $\mathcal{C}_{k,s}(\boldsymbol{u}) := \boldsymbol{v}_s$, a Chebyshev window, and set $\mathcal{C}_{k,0}(\boldsymbol{u}) := \boldsymbol{u}$.

Proposition 2.9 (Chebyshev representation on an unchanged window)**.** *Suppose that $\boldsymbol{L}_{k+t} = \boldsymbol{L}$ for $t = 0, \ldots, s-1$. Then $\boldsymbol{v}_t = P_t^\chi(\boldsymbol{L})\boldsymbol{u}$ for every $t = 0, \ldots, s$. In particular, $\mathcal{C}_{k,s}(\boldsymbol{u}) = P_s^\chi(\boldsymbol{L})\boldsymbol{u}$.*

Hence a window of h rounds on a single operator contracts the disagreement by $\rho_h^\chi = 1/T_h(z_0)$, whether or not it is interrupted. When the operator changes inside a window, the recurrence is no longer a polynomial of one matrix because consecutive operators need not commute, and its deviation from P_h^χ grows with h. WAVE limits the window length and restarts: the next window starts at $t = 0$ from the current state alone, discarding $\boldsymbol{v}_{s-1}$.

Window lengths. Under metric drift, every window has length

$$\widetilde{m}_\beta := \min\left\{\sqrt{\chi}, (3\beta\chi)^{-1}\right\}, \qquad m_\beta := \max\{1, \lfloor \widetilde{m}_\beta \rfloor\}, \tag{6}$$

where $(3\beta\chi)^{-1} = +\infty$ for $\beta = 0$. The second term keeps the drift error of a window, $O(\beta m^3)$, below its progress, of order m^2/χ (Section 4). For a piecewise-constant network, τ is unknown, and the length adapts by doubling: it starts at one, doubles after every unchanged window up to $m_{\max} := \lfloor\sqrt{\chi}\rfloor$, and resets to one after a detected change. Long unchanged intervals thus reach the accelerated length, while frequent changes keep the windows short. The detector reports a change before round r when $L_r \neq L_{r-1}$, once and before L_r is used. The counter q below records the certified logarithmic contraction: a drift window of length m_β earns $m_\beta^2/(5\chi)$ (Theorem 4.2), and a piecewise window of h executed rounds earns $\ln T_h(z_0)$.

Algorithm 1: WAVE$(\boldsymbol{u}, k; \delta, \mathcal{M})$

Input: $\boldsymbol{u} \in \mathbb{R}^{nd}$, next round k, target $\delta \in (0,1)$, model $\mathcal{M} \in \{\text{drift}, \text{piecewise}\}$; known χ and, for metric drift, β

$q \leftarrow 0$; $m \leftarrow m_\beta$ for metric drift, and $m \leftarrow 1$ for a piecewise-constant network;
while $q < \ln(1/\delta)$ **do**
  if $\mathcal{M} = \text{drift}$ **then**
    $(\boldsymbol{u}, k) \leftarrow (\mathcal{C}_{k,m}(\boldsymbol{u}), k+m)$; $q \leftarrow q + m^2/(5\chi)$;
  else
    $k_0 \leftarrow k$; execute the window until either m rounds are completed or a change is reported before the next round;
    $h \leftarrow$ number of executed rounds;
    $(\boldsymbol{u}, k) \leftarrow (\mathcal{C}_{k_0,h}(\boldsymbol{u}), k_0 + h)$; $q \leftarrow q + \ln T_h(z_0)$;
    if *a change was reported before round k* **then**
      $m \leftarrow 1$;
    else
      $m \leftarrow \min\{2m, m_{\max}\}$;
return $(\boldsymbol{u}, k)$;

To state the output guarantee, let $\Pi := n^{-1}\mathbf{1}_n\mathbf{1}_n^\top$, $\boldsymbol{\Pi} := \Pi \otimes I_d$, and $\boldsymbol{\Pi}_\perp := (I_n - \Pi) \otimes I_d$. The matrices $\boldsymbol{\Pi}$ and $\boldsymbol{\Pi}_\perp$ are the orthogonal projections onto the consensus and disagreement subspaces of $\mathbb{R}^{nd}$, respectively. Every window preserves the consensus component, and the credit of every window bounds the logarithm of its contraction on the disagreement subspace. The analysis below shows that, when the algorithm returns $(\widehat{\boldsymbol{u}}, k^+) = \text{WAVE}(\boldsymbol{u}, k; \delta)$, it satisfies

$$\boldsymbol{\Pi}\widehat{\boldsymbol{u}} = \boldsymbol{\Pi}\boldsymbol{u}, \qquad \|\boldsymbol{\Pi}_\perp \widehat{\boldsymbol{u}}\| \leq \delta \|\boldsymbol{\Pi}_\perp \boldsymbol{u}\|. \tag{7}$$

Here and below, once the scheduler is fixed, we suppress $\mathcal{M}$ in the notation for a WAVE call.

3 MAIN RESULTS

The first theorem covers the full range of metric drift with one window-length rule and requires no commutativity between successive operators.

Theorem 3.1 (Consensus under metric drift)**.** *Let Assumptions 2.2 and 2.3 hold with $\chi \geq 4$. For $\epsilon \in (0,1)$, run Algorithm 1 on $(\boldsymbol{u}_0, 0; \epsilon)$ with $\mathcal{M} = \text{drift}$ and m_β from (6). Then its output $(\widehat{\boldsymbol{u}}, K)$ is an ϵ-consensus solution and*

$$K = O\left(\left(\sqrt{\chi} + \min\{\beta\chi^2, \chi\}\right) \ln \frac{e}{\epsilon}\right). \tag{8}$$

Moreover, every completed window reduces the disagreement norm by a factor at most $\exp(-m_\beta^2/(5\chi))$.

For small drift, long windows retain Chebyshev acceleration. As β increases, the window shortens and the additional term grows as $\beta\chi^2$ until it reaches the robust $O(\chi)$ scale, where $m_\beta = 1$ and the recurrence becomes its degree-one endpoint.

Corollary 3.2 (Fixed-network rate under small drift)**.** *Under the assumptions of Theorem 3.1, if $\beta \leq (3\chi^{3/2})^{-1}$, then $m_\beta = \lfloor\sqrt{\chi}\rfloor$ and $K = O\big(\sqrt{\chi}\ln\frac{e}{\epsilon}\big)$. This includes the fixed-network case $\beta = 0$.*

The corollary still allows the operator to change at every round; only the norm of each change is restricted. The constant 3 comes from the explicit stability estimate of Lemma 4.1, and the proof gives $K \leq 10\sqrt{\chi}\ln(1/\epsilon) + \sqrt{\chi}$ in this regime. The second theorem replaces control of the jump size by control of the time between jumps. The detector restarts the recurrence at a change, and the doubling rule increases the attempted window during each unchanged interval without using τ.

Theorem 3.3 (Consensus under piecewise-constant changes)**.** *Let Assumptions 2.2 and 2.4 hold with $\chi \geq 4$, and assume access to the ideal change detector. For $\epsilon \in (0,1)$, run Algorithm 1 on $(\boldsymbol{u}_0, 0; \epsilon)$ with $\mathcal{M} = \text{piecewise}$. Then its output $(\widehat{\boldsymbol{u}}, K)$ is an ϵ-consensus solution and*

$$K = O\left(\left(\sqrt{\chi} + \frac{\chi}{\tau}\right) \ln \frac{e}{\epsilon}\right). \tag{9}$$

The term χ/τ is the amortized cost of restarts: for $\tau \geq \sqrt{\chi}$ the bound is $O(\sqrt{\chi}\ln(e/\epsilon))$, and $\tau = 1$ gives the robust $O(\chi\ln(e/\epsilon))$ dependence. The same bound holds in expectation when the change times are memoryless with mean spacing $\bar{\tau}$ (Theorem E.3). Together, the results give two routes to accelerated consensus on changing networks: changes may be frequent but small, or abrupt but separated. Neither requires a connected subgraph that persists throughout the run. The bounds also quantify partial acceleration between the fully accelerated and worst-case endpoints. For example, $\beta = \Theta(\chi^{-5/4})$ in the metric-drift model or $\tau = \Theta(\chi^{1/4})$ in the piecewise-constant model gives $O(\chi^{3/4}\ln(e/\epsilon))$ rounds.

4 METRIC-DRIFT NETWORK

We compare each moving WAVE window with a fixed reference operator. For a window beginning at round k, denote the sum drift of Laplacian matrix as $\boldsymbol{E}_t := \boldsymbol{L}_{k+t} - \boldsymbol{L}_k$, so that $\|\boldsymbol{E}_t\| \leq \beta t$ under metric drift (Assumption 2.3). The recurrence uses the actual operators:

$$\boldsymbol{e}_1 = \frac{z_\chi(\boldsymbol{L}_k)}{z_0}\boldsymbol{e}_0, \qquad \boldsymbol{e}_{t+1} = a_t z_\chi(\boldsymbol{L}_{k+t})\boldsymbol{e}_t - c_t\boldsymbol{e}_{t-1}, \quad t = 1, \ldots, m-1. \tag{10}$$

If every $\boldsymbol{E}_t$ vanishes, then $\boldsymbol{e}_m = P_m^\chi(\boldsymbol{L}_k)\boldsymbol{e}_0$ by Proposition 2.9. To quantify the effect of changing operators, let U_r denote the Chebyshev polynomial of the second kind, defined by $U_0(x) = 1$, $U_1(x) = 2x$, and $U_{r+1}(x) = 2xU_r(x) - U_{r-1}(x)$. For $1 \leq t \leq m$, the exact variation-of-constants identity is

$$\boldsymbol{e}_t = P_t^\chi(\boldsymbol{L}_k)\boldsymbol{e}_0 - \frac{4}{1-\chi^{-1}} \sum_{j=2}^{t} \frac{T_{j-1}(z_0)}{T_t(z_0)} U_{t-j}(z_\chi(\boldsymbol{L}_k))\, \boldsymbol{E}_{j-1}\boldsymbol{e}_{j-1}.$$

Each perturbation acts on the actual iterate e_{j-1} and is then propagated by a polynomial of the reference operator. The matrix order is preserved, so no commutativity is needed. On $(\mathbf{1}_n \otimes \boldsymbol{I}_d)^\perp$,

$\|U_{t-j}(z_\chi(\boldsymbol{L}_k))\| \le t-j+1$ and $\|\boldsymbol{E}_{j-1}\| \le \beta(j-1)$. Summing these responses produces the cubic short-window term below. The normalization ratios give the second, geometrically damped bound. Appendix D proves the identity and both estimates.

Lemma 4.1 (Stability of a moving full-spectrum window)**.** *Let Assumptions 2.2 and 2.3 hold. Let a window satisfy (10) and set* $M_t := \max_{0\le s\le t}\|\boldsymbol{e}_s\|, D_t := \min\left\{\frac{2}{3}\beta(t^3-t), 2\beta t\chi\right\}$. *Then* $\|\boldsymbol{e}_t - P_t^\chi(\boldsymbol{L}_k)\boldsymbol{e}_0\| \le D_t M_t$ *for every* $0 \le t \le m$. *Consequently, if* $D_m < 1$, *then*

$$\|\boldsymbol{e}_m\| \le \left(\rho_m^\chi + \frac{D_m}{1-D_m}\right)\|\boldsymbol{e}_0\|. \tag{11}$$

The two terms in D_t are the short-window and geometrically damped response bounds, and $D_1 = 0$ because a single step uses a single operator. For $m \le \sqrt{\chi}$, the fixed-reference progress is of order m^2/χ, while metric drift contributes $O(\beta m^3)$. Requiring the perturbation to remain below the reference progress gives $\beta m\chi = O(1)$ and explains the choice $m_\beta \asymp \min\{\sqrt{\chi}, (\beta\chi)^{-1}\}$ in (6).

Theorem 4.2 (Uniform certificate for a full-spectrum window)**.** *Let Assumptions 2.2 and 2.3 hold with* $\chi \ge 4$. *If* $m = 1$, *no restriction on* β *is required. If* $m \ge 2$, *suppose* $m^2/\chi \le 1$ *and* $\beta m\chi \le 1/3$. *Then every complete Chebyshev window satisfies*

$$\|\boldsymbol{e}_m\| \le \exp\left(-\frac{m^2}{5\chi}\right)\|\boldsymbol{e}_0\|. \tag{12}$$

Hence, for every $\epsilon \in (0,1)$, *repeated complete windows reach* ϵ*-consensus within* $K(\epsilon) \le (5\chi/m)\ln(1/\epsilon) + m$ *communication rounds.*

The main idea behind the proof of Theorem 4.2 is to compare the contraction of an unchanged Chebyshev window with the perturbation accumulated while the operator moves. For $m \le \sqrt{\chi}$, these quantities scale respectively as m^2/χ and βm^3, so the condition $\beta m\chi \le 1/3$ keeps the perturbation below the useful progress and yields (12). When $m = 1$, the window uses only one operator and therefore needs no drift condition. The complete argument, including the constants, is given in Appendix D. Any shorter integer window, for example one obtained by replacing the constant 3 in (6) by a larger one, satisfies the same hypotheses.

Choosing the window length. If $m_\beta \ge 2$, then $m_\beta \le \sqrt{\chi}$ and $\beta m_\beta\chi \le 1/3$, so Theorem 4.2 applies; if $m_\beta = 1$, its degree-one part applies without a drift condition. When $\widetilde{m}_\beta \ge 2$, the rounding inequality $m_\beta \ge \widetilde{m}_\beta/2$ gives $\chi/m_\beta \le 2\max\{\sqrt{\chi}, 3\beta\chi^2\}$. When $\widetilde{m}_\beta < 2$, the degree-one cost is $O(\chi\ln(e/\epsilon))$ and $\beta\chi^2 > \chi/6$. This proves Theorem 3.1 with one recurrence and one length formula; the transition from acceleration to the robust endpoint is caused only by the integer value of m.

5 Piecewise-constant networks

At a detected change, Algorithm 1 keeps the state reached by the current window and starts a new length-one attempt before using the new operator. The jump magnitude therefore does not enter the contraction estimate: successive graphs may have entirely different edge sets, provided each operator satisfies the common spectral bound. Every window runs on a single operator and, by Propositions 2.9 and 2.6, a window of $h \le m_{\max}$ rounds, complete or interrupted, earns $\ln T_h(z_0) \ge 4h^2/(5\chi)$ by Lemma 2.7. A change restarts the build-up of acceleration, but preserves the progress already made. The term χ/τ accounts for this restart cost over unchanged intervals. In particular, attaining the fixed-network rate only requires each interval to support a constant-factor reduction, rather than an entire run to the final accuracy ϵ.

5.1 Minimum spacing between changes

Consider r consecutive rounds with a fixed operator following a length-one start. If the largest completed window has length $m < m_{\max}$, then $r < 4m$: completed windows use fewer than $2m$ rounds, and the next attempt has length at most $2m$. If $r \ge 4m_{\max}$, the shorter windows use fewer than $2m_{\max}$ rounds, leaving time for at least $r/(5m_{\max})$ completed windows of length $m_{\max}$. These

two cases give at least $r \min\{r, m_{\max}\}/(20\chi)$ of certified logarithmic reduction in the disagreement norm. On an interval between changes, $r \geq \tau$, so the certified reduction is at least $1/(40(\sqrt{\chi}+\chi/\tau))$ per round; calls that begin or end inside an interval are handled in Appendix E.

5.2 Detecting changes inside the network

The ideal detector can be replaced by local observation and flooding. For this extension, assume synchronous neighbor exchanges, that every agent knows $n \geq 2$ and χ, and that it observes its current row of L_k before round k. A message may carry a round index alongside the usual vector. Set $D_n := \lfloor (\sqrt{\chi}/2)\ln(2n) \rfloor + 1$. Agents flood each observed change index and retain $D_n + 1$ local checkpoints. Exactly D_n rounds after a change, they synchronously restore its checkpoint and restart the recurrence. A prospective output is returned only after D_n further rounds have ruled out an unprocessed earlier change; numerical computation and flooding continue during this validation. Appendix E.5 specifies the protocol.

Proposition 5.1 (Flooding detector)**.** *Under Assumptions 2.2 and 2.4, suppose $\tau \geq 2D_n$ and the preceding local-observation and message assumptions hold. The flooding-and-rollback version of Algorithm 1 reaches ϵ-consensus, for $\epsilon \in (0,1)$, in*

$$K = O\Big(\Big(\sqrt{\chi} + \frac{\chi}{\tau}\Big)\ln\frac{e}{\epsilon} + \sqrt{\chi}\ln n\Big)$$

communication rounds, without knowing τ. It uses $O(D_n d)$ local vector storage and one change index per neighbor message.

The nonzero-support graph of each operator has diameter at most D_n. Indeed, the spectral bound gives $\big\|P^{\chi}_{D_n}(L) - \Pi\big\| \leq 2e^{-2D_n/\sqrt{\chi}} < 1/n$, so every entry of $P^{\chi}_{D_n}(L)$ is positive. If two agents were more than D_n hops apart, the corresponding entry of any polynomial of degree at most D_n in L would be zero, a contradiction. Thus the same spectral information used for acceleration also bounds the reporting delay. Consequently, every rollback occurs before the next change; after deleting the discarded rounds, the retained computation is an ideal-detector run with segment lengths at least $\tau/2$. The condition $\tau \geq \sqrt{\chi}\ln(2n) + 2$ is a sufficient, slightly stronger spacing condition.

6 Application to decentralized optimization

We consider

$$\min_{x\in\mathbb{R}^d} f(x) := \frac{1}{n}\sum_{i=1}^{n} f_i(x),$$

where each f_i is convex and α-smooth, while f is μ-strongly convex. Let $x^\star := \arg\min_x f(x)$ and $\kappa := \alpha/\mu$. WAVE-STM combines restarted Similar Triangles iterations (Nesterov, 2004; Gorbunov et al., 2020) with WAVE communication. The agents start from a common point x_0 and are given $G_0 > 0$ such that $n^{-1}\sum_i \|\nabla f_i(x_0)\|^2 \leq G_0^2$.

Theorem 6.1 (WAVE-STM)**.** *Let Assumption 2.2 and either Assumption 2.3 or 2.4 hold with $\chi \geq 4$. For every $\epsilon > 0$,* WAVE-STM *returns an ϵ-solution at every agent:*

$$\max_{1\leq i\leq n}\big\{f(y^{(i)}_{\text{out}}) - f(x^\star)\big\} \leq \epsilon. \tag{13}$$

Each agent evaluates $\widetilde{O}(\sqrt{\kappa})$ local gradients, and the number of communication rounds satisfies

$$K = \begin{cases} \widetilde{O}\big(\sqrt{\kappa}\,[\sqrt{\chi} + \min\{\beta\chi^2, \chi\}]\big), & \textit{metric drift},\\ \widetilde{O}\big(\sqrt{\kappa}\,[\sqrt{\chi} + \chi/\tau]\big), & \textit{piecewise-constant changes}. \end{cases} \tag{14}$$

Here $\widetilde{O}$ hides logarithmic factors in the target accuracy, initialization, κ, and n. The piecewise-constant guarantee uses the ideal external change detector.

Consequently, when $\beta = O(\chi^{-3/2})$ or $\tau = \Omega(\sqrt{\chi})$, the leading communication dependence is $\widetilde{O}(\sqrt{\kappa\chi})$, recovering the fixed-network scaling up to logarithmic factors.

Remark 6.2. Guarantee (13) is stronger than the usual objective bound for the network average $\bar{y}_{\text{out}} := n^{-1}\sum_{i=1}^{n} y^{(i)}_{\text{out}}$. Indeed, convexity of f gives $f(\bar{y}_{\text{out}}) - f(x^\star) \leq \frac{1}{n}\sum_{i=1}^{n}\big(f(y^{(i)}_{\text{out}}) - f(x^\star)\big) \leq \epsilon$. Moreover, (13) guarantees an ϵ-solution directly at every agent, whereas the averaged iterate need not be locally available.

7 NUMERICAL EXPERIMENTS

We compare WAVE with gossip, minimax Richardson iteration, and the time-varying method of Montijano et al. (2013). We also run the classical Chebyshev semi-iteration (Golub & Varga, 1961) continuously, without restarts, to test whether restarts are needed when the network changes. Every method processes the same operators and spectral bounds. We report the exact worst-case gap $r_k := \max_{e_0 \neq 0} \|e_k\| / \|e_0\|$; the level 10^{-6} is used only to summarize first-passage times.

Metric drift. Figure 1(a) uses the controlled family of Appendix H.4, with $\chi = 1.6 \cdot 10^5$ and $\beta\chi^{3/2} \approx 30$. This drift is 90 times the sufficient threshold for the fixed-network rate, and the WAVE rule selects $m = 4$. WAVE reaches $r_k \leq 10^{-6}$ in about $2.8 \cdot 10^5$ rounds, versus $1.1 \cdot 10^6$ for Montijano and Richardson; gossip remains above the target after $1.25 \cdot 10^6$ rounds. Chebyshev without restarts is unstable and leaves the plotted range within $5 \cdot 10^4$ rounds.

Piecewise-constant changes. Figure 1(b) uses 32 independent pairs of connected sparse random geometric graphs with $n = 100$ and $\chi \in [339, 2873]$ (median 905). The operator changes every $\mathrm{round}(\sqrt{\chi})$ rounds, with each change reported to WAVE. Median first-passage counts are 180 rounds for WAVE, 344 for Montijano and Richardson, and 422 for gossip. WAVE is first on all 32 pairs. Chebyshev reaches the target on only 12 pairs within the 1000-round budget.

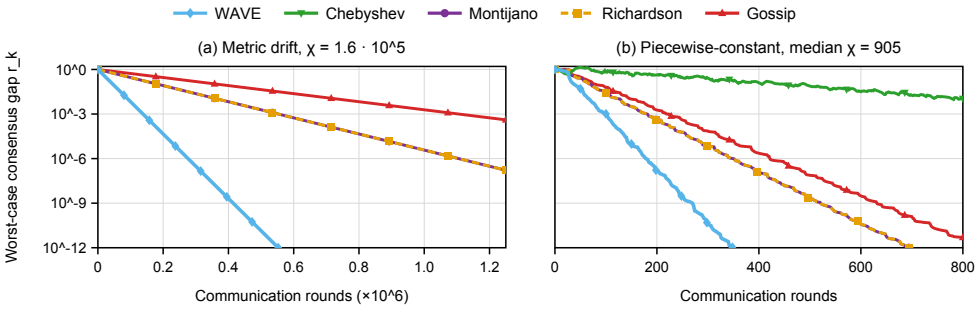

Figure 1: Worst-case consensus gap against communication rounds. Classical Chebyshev is run without restarts. (a) Controlled metric drift with $\chi = 1.6 \cdot 10^5$ and $m = 4$. (b) Pointwise median over 32 sparse random-geometric graph pairs, with changes every $\mathrm{round}(\sqrt{\chi})$ rounds.

8 CONCLUSION

WAVE shows that Chebyshev acceleration of consensus survives network variation if the recurrence is restarted before perturbations from changing, noncommuting operators outweigh its progress. To our knowledge, its bounds are the first consensus guarantees that interpolate between the fixed-network $\sqrt{\chi}$ and the arbitrary-variation χ rates as explicit functions of the one-step drift β, $O((\sqrt{\chi} + \min\{\beta\chi^2, \chi\}) \ln(e/\epsilon))$, and of the spacing τ between changes, $O((\sqrt{\chi} + \chi/\tau) \ln(e/\epsilon))$; the fixed-network rate is retained whenever $\beta \leq (3\chi^{3/2})^{-1}$ or $\tau \geq \sqrt{\chi}$. These guarantees require no persistent connected subgraph and no knowledge of τ, hold in expectation for memoryless change times, and remain valid, up to an additive $O(\sqrt{\chi} \ln n)$ term, when sufficiently separated changes are detected by flooding inside the network. For convex local objectives with a strongly convex average, WAVE-STM multiplies these network factors by $\widetilde{O}(\sqrt{\kappa})$ and thus retains the fixed-network $\widetilde{O}(\sqrt{\kappa\chi})$ dependence in the same two regimes. In the experiments, WAVE outperforms all compared methods in both regimes.

AI USE STATEMENT

Generative AI tools were used for language editing and to improve the clarity of the manuscript. We reviewed all resulting changes and take full responsibility for the final content.

REFERENCES

Amir Beck, Angelia Nedić, Asuman Ozdaglar, and Marc Teboulle. An $O(1/k)$ gradient method for network resource allocation problems. *IEEE Transactions on Control of Network Systems*, 1(1): 64–73, 2014.

Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69 (345):118–121, 1974.

Olivier Devolder. *Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization*. PhD thesis, Université catholique de Louvain, Louvain-la-Neuve, 2013.

Olivier Devolder, François Glineur, and Yurii Nesterov. First-order methods of smooth convex optimization with inexact oracle. *Mathematical Programming*, 146(1–2):37–75, 2014.

Pedro A. Forero, Alfonso Cano, and Georgios B. Giannakis. Consensus-based distributed support vector machines. *Journal of Machine Learning Research*, 11:1663–1707, 2010.

Pontus Giselsson, Minh Dang Doan, Tamas Keviczky, Bart De Schutter, and Anders Rantzer. Accelerated gradient methods and dual decomposition in distributed model predictive control. *Automatica*, 49(3):829–833, 2013.

Gene H. Golub and Richard S. Varga. Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order richardson iterative methods. *Numerische Mathematik*, 3(1): 147–168, 1961.

Eduard Gorbunov, Marina Danilova, and Alexander Gasnikov. Stochastic optimization with heavy-tailed noise via accelerated gradient clipping. In *Advances in Neural Information Processing Systems*, volume 33, pp. 15042–15053, 2020.

Eduard Gorbunov, Alexander Rogozin, Aleksandr Beznosikov, Darina Dvinskikh, and Alexander Gasnikov. Recent theoretical advances in decentralized distributed convex optimization. In *High-Dimensional Optimization and Probability*, volume 191 of *Springer Optimization and Its Applications*, pp. 253–325. Springer, Cham, 2022.

Martin H. Gutknecht and Stefan Röllin. The chebyshev iteration revisited. *Parallel Computing*, 28 (2):263–283, 2002.

Ali Jadbabaie, Jie Lin, and A. Stephen Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency, 2016. NIPS Workshop on Private Multi-Party Machine Learning.

Nikita Kornilov, Eduard Gorbunov, Mohammad Alkousa, Fedor Stonyakin, Pavel Dvurechensky, and Alexander Gasnikov. Intermediate gradient methods with relative inexactness. *Journal of Optimization Theory and Applications*, 207(3):62, 2025. doi: 10.1007/s10957-025-02809-y.

Dmitry Kovalev, Elnur Gasanov, Alexander Gasnikov, and Peter Richtárik. Lower bounds and optimal algorithms for smooth and strongly convex decentralized optimization over time-varying networks. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22325–22335, 2021a.

Dmitry Kovalev, Egor Shulgin, Peter Richtárik, Alexander Rogozin, and Alexander Gasnikov. ADOM: Accelerated decentralized optimization method for time-varying networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5784–5793, 2021b.

Huan Li and Zhouchen Lin. Accelerated gradient tracking over time-varying graphs for decentralized optimization. *Journal of Machine Learning Research*, 25(274):1–52, 2024.

Mihaly Maros and Joakim Jaldén. PANDA: A dual linearly converging method for distributed optimization over time-varying undirected graphs. In *Proceedings of the 57th IEEE Conference on Decision and Control*, pp. 6520–6525, 2018.

H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282, 2017.

Dmitry Metelev, Alexander Rogozin, Dmitry Kovalev, and Alexander Gasnikov. Is consensus acceleration possible in decentralized optimization over slowly time-varying networks? In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 24532–24554, 2023.

Dmitry Metelev, Aleksandr Beznosikov, Alexander Rogozin, Alexander Gasnikov, and Anton Proskurnikov. Decentralized optimization over slowly time-varying graphs: Algorithms and lower bounds. *Computational Management Science*, 21(1):8, 2024. Article 8.

Eduardo Montijano, Juan I. Montijano, and Carlos Sagüés. Chebyshev polynomials in distributed consensus applications. *IEEE Transactions on Signal Processing*, 61(3):693–706, 2013.

Luc Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005.

Angelia Nedić and Alex Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 60(3):601–615, 2015.

Angelia Nedić, Alex Olshevsky, and Wei Shi. Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM Journal on Optimization*, 27(4):2597–2633, 2017.

Angelia Nedić, Alex Olshevsky, and Michael G. Rabbat. Network topology and communication-computation tradeoffs in decentralized optimization. *Proceedings of the IEEE*, 106(5):953–976, 2018.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Boston, 2004.

N. T. Nguyen, Alexander V. Rogozin, and Alexander V. Gasnikov. Average-case optimization analysis for distributed consensus algorithms on regular graphs. *Russian Journal of Nonlinear Dynamics*, 20(5):907–931, 2024.

Reza Olfati-Saber and Richard M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004.

Dmitry Pasechnyuk-Vilensky and Martin Takáč. Chebyshev-exact acceleration under hessian variation, i: Sine–jacobi method. arXiv:2606.16671, 2026.

Michael Rabbat and Robert Nowak. Distributed optimization in sensor networks. In *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks*, pp. 20–27, 2004.

Alexander Rogozin, C. A. Uribe, Alexander Gasnikov, Nikolay Malkovsky, and Angelia Nedić. Optimal distributed convex optimization on slowly time-varying graphs. *IEEE Transactions on Control of Network Systems*, 7(2):829–841, 2019.

Alexander Rogozin, Vladimir Lukoshkin, Alexander Gasnikov, Dmitry Kovalev, and Egor Shulgin. Towards accelerated rates for distributed optimization over time-varying networks. In *Optimization and Applications (OPTIMA 2021)*, volume 13078 of *Lecture Notes in Computer Science*, pp. 258–272. Springer, Cham, 2021.

Alexander Rogozin, Alexander Gasnikov, Aleksandr Beznosikov, and Dmitry Kovalev. Decentralized convex optimization over time-varying graphs. In *Encyclopedia of Optimization*, pp. 1–17. Springer, Cham, 2023.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal algorithms for smooth and strongly convex distributed optimization in networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3027–3036, 2017.

John N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1984.

Alexander Vasin, Alexander Gasnikov, Pavel Dvurechensky, and Vladimir Spokoiny. Accelerated gradient methods with absolute and relative noise in the gradient. *Optimization Methods and Software*, 38(6):1180–1229, 2023.

Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.

SUPPLEMENTARY MATERIAL: CONTENTS

A. Notation and asymptotic conventions . . . 13
B. Model and implementation details . . . 14
C. Classical Chebyshev facts . . . 15
D. Proofs for metric drift . . . 17
E. Piecewise-constant networks . . . 20
In-network change detection and rollback . . . 24
F. Proofs for the optimization results . . . 25
Accelerated gradient tracking with WAVE . . . 32
G. Examples and interpretation . . . 33
H. Supplementary numerical experiments . . . 37

A NOTATION AND ASYMPTOTIC CONVENTIONS

Table 1 summarizes the notation and defines the comparison symbol $\asymp$ used in the main scaling statements. Symbols used only within individual proofs are defined locally.

Table 1: Consolidated notation and asymptotic conventions used throughout the paper.

Symbol	Definition / convention	Meaning
Linear-algebra and consensus conventions		
n, d, i	n agents, local dimension d, agent index $i \in \{1, \dots, n\}$	Network size, local vector dimension, and agent index. A parenthesized superscript, as in $u_k^{(i)}$, is reserved for the agent index.
$\mathbf{1}_n, \mathbf{1}_n^\perp$	$\mathbf{1}_n = (1, \dots, 1)^\top$, $\mathbf{1}_n^\perp = \{v \in \mathbb{R}^n : \mathbf{1}_n^\top v = 0\}$	All-ones vector and disagreement subspace.
$\Pi, \Pi_\perp$	$\Pi = \mathbf{1}_n \mathbf{1}_n^\top / n$, $\Pi_\perp = I_n - \Pi$	Orthogonal projectors onto the consensus and disagreement subspaces.
$\mathbf{\Pi}, \mathbf{\Pi}_\perp$	$\mathbf{\Pi} = \Pi \otimes I_d$, $\mathbf{\Pi}_\perp = \Pi_\perp \otimes I_d$	Block projectors on stacked vectors in $\mathbb{R}^{nd}$.
$u_k^{(i)}, \boldsymbol{u}_k$	$\boldsymbol{u}_k = ((u_k^{(1)})^\top, \dots, (u_k^{(n)})^\top)^\top$	Local and stacked consensus states at communication round k.
$\bar{u}_0$	$\bar{u}_0 = n^{-1} \sum_i u_0^{(i)}$	Initial average.
$\boldsymbol{e}_k, \boldsymbol{e}_{(w)}$	$\boldsymbol{e}_k = \boldsymbol{u}_k - \mathbf{1}_n \otimes \bar{u}_0$, $\boldsymbol{e}_{(w)} := \boldsymbol{e}_{k_0+wm}$ for equal-length windows starting at k_0	Global consensus error and error at a window boundary.
$\lambda, \lambda_{\max}(A)$, $\lambda_{\min}^+(A)$	λ is reserved for eigenvalues or scalar spectral arguments; $\lambda_{\max}(A)$ and $\lambda_{\min}^+(A)$ are the largest and smallest positive eigenvalues	Spectral variables and extremal positive eigenvalues.
$\mathrm{spec}(A)$, $A\|_S$	$\mathrm{spec}(A)$ is the set of eigenvalues of A; $A\|_S$ is the restriction of A to an invariant subspace S	Spectrum and restricted operator.
$\chi(A)$	$\chi(A) = \lambda_{\max}(A)/\lambda_{\min}^+(A)$ and $\chi(cA) = \chi(A)$ for $c > 0$	Condition number on $(\ker A)^\perp$; invariant under positive scalar normalization.
$\|\cdot\|$, $\langle \cdot, \cdot \rangle$	Euclidean norm for vectors, spectral norm for matrices, and the Euclidean inner product	Norm and inner-product conventions.
$\lesssim, \gtrsim, \asymp$	For nonnegative a, b, $a \lesssim b$ means $a \le Cb$ for an absolute constant $C > 0$ independent of the problem parameters; $a \gtrsim b$ is the reverse inequality and $a \asymp b$ means both	Comparison up to absolute multiplicative constants. Parameter-dependent constants are stated explicitly when used.
Network and variation quantities		
V	$V = \{1, \dots, n\}$	Fixed set of agent indices.
$\mathcal{G}_k = (V, \mathcal{E}_k)$	Connected undirected communication graph at round k	Time-varying communication network.
χ	$\mathrm{spec}(L_k\|_{\mathbf{1}_n^\perp}) \subseteq [\chi^{-1}, 1]$ for every k	Known conservative spectral-envelope condition bound for the normalized communication operators.
$L_k, \boldsymbol{L}_k$	$\boldsymbol{L}_k = L_k \otimes I_d$, and $\mathrm{spec}(L_k\|_{\mathbf{1}_n^\perp}) \subseteq [\chi^{-1}, 1]$	Normalized agent-level gossip matrix and its lifted operator on stacked states.
$\mathbb{L}(\chi)$	$\{L :$ L is a normalized gossip matrix and $\mathrm{spec}(L\|_{\mathbf{1}_n^\perp}) \subseteq [\chi^{-1}, 1]\}$	Communication-operator class used in the model.
β	$\|L_{k+1} - L_k\| \le \beta$	One-step operator drift in the metric-drift model.
t_i, τ_i, τ	$\tau_i = t_{i+1} - t_i$ and $\tau_i \ge \tau$ under the minimum-spacing assumption	Change times, constant-segment lengths, and their minimum spacing.
$\bar{\tau}$	$\mathbb{P}(B_k = 1) = 1/\bar{\tau}$; equivalently, $\mathbb{E}T_i = \bar{\tau}$ for the geometric segment lengths T_i	Mean segment length in the memoryless change-time model.
Communication and WAVE indices		
k, K	$k = 0, \dots, K-1$ indexes communication updates; states are indexed by $k = 0, \dots, K$	Global communication index and total number of communication rounds.

Continued on the next page.

Table 1 continued.

Symbol	Definition / convention	Meaning
w, J	$w = 0, \dots, J-1$	Chebyshev-window index and number of complete windows.
t	$t = 0, \dots, m$ inside a single Chebyshev window	Local step index within a window.
$\widetilde{m}_\beta, m_\beta$	Defined by (6)	Continuous drift target and selected integer window length.
$m, m_{\max}, h$	m is the current attempted length, $m_{\max} = \lfloor\sqrt{\chi}\rfloor$, and h is the number of rounds executed in an attempted window	Attempted length, dyadic cap, and executed length.
q	Increment $m_\beta^2/(5\chi)$ per drift window and $\ln T_h(z_0)$ per piecewise window of h rounds	Certified logarithmic contraction accumulated by Algorithm 1.
D_n	$D_n = \lfloor \frac{\sqrt{\chi}}{2} \ln(2n) \rfloor + 1$	Delay of the flooding detector (Proposition 5.1).
Chebyshev quantities		
T_m, U_m	Chebyshev polynomials of the first and second kind	Classical polynomial basis used in the accelerated recurrence.
$z_\chi(\lambda), z_0$	$z_\chi(\lambda) = (1 + \chi^{-1} - 2\lambda)/(1 - \chi^{-1})$, $z_0 = z_\chi(0)$	Affine map sending the full interval $[\chi^{-1}, 1]$ to $[-1, 1]$ and the image of the origin.
$P_m^\chi(\lambda)$	$P_m^\chi(\lambda) = T_m(z_\chi(\lambda))/T_m(z_0)$	Full-spectrum shifted and normalized Chebyshev residual executed by WAVE.
ρ_m^χ	$\rho_m^\chi = \max_{[\chi^{-1},1]} \lvert P_m^\chi \rvert$	Full-spectrum minimax residual value.
a_t, c_t	$a_t = 2T_t(z_0)/T_{t+1}(z_0)$ for $t \geq 0$ and $c_t = T_{t-1}(z_0)/T_{t+1}(z_0)$ for $t \geq 1$	Coefficients of the Chebyshev three-term recurrence.
Optimization quantities and indices		
f_i, f	$f(x) = n^{-1} \sum_{i=1}^n f_i(x)$	Local objectives and global average objective.
$x^\star$	$x^\star = \arg\min f(x)$	Unique global optimizer.
α, μ, κ	Each f_i is convex and α-smooth; the average f is μ-strongly convex; $\kappa = \alpha/\mu$	Local smoothness, strong convexity of the average, and their ratio.
G_0, R, H	$n^{-1} \sum_i \lVert\nabla f_i(x_0)\rVert^2 \leq G_0^2$; $R = G_0/\mu$, $H = (1+\kappa)G_0$	Single supplied initialization bound and its derived radius and heterogeneity bounds.
WAVE$(\boldsymbol{u}, k; \delta)$	Defined in Algorithm 1; returns an updated state and the next unused global communication round	Direct WAVE call with requested relative disagreement factor δ.
$\ln_+(t)$	$\ln_+(t) := \max\{0, \ln t\}$	Nonnegative part of the natural logarithm.
ℓ	$\ell = 0, \dots, N-1$ within an STM stage	Inner optimization-iteration index.
$A_\ell, \omega_{\ell+1}$	$A_{\ell+1} = A_\ell + \omega_{\ell+1}$, $\omega_{\ell+1} = (\ell+2)/(2\alpha)$	Weights of the Similar Triangles recurrence.
s, S	$s = 0, \dots, S-1$	Optimization-stage index and number of restart stages.
$N_0, r_s, r_\epsilon, \delta_\star$	$N_0 = \lceil 8\sqrt{\kappa} \rceil$, $r_s = 2^{-s}R$; r_ϵ and $\delta_\star$ are defined in (31) and (32)	Stage length, deterministic radius schedule, and the single inner-call WAVE tolerance.

B Additional model and implementation details

B.1 Normalization and the linear operator model

For a symmetric positive-semidefinite matrix A with a positive eigenvalue, its condition number on $(\ker A)^\perp$ is

$$\chi(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}^+(A)}.$$

It is invariant under positive scalar normalization. If the system supplies an unnormalized weighted Laplacian $\widetilde{L}_k$, a fixed certified scale $\Lambda \geq \sup_k \lambda_{\max}(\widetilde{L}_k)$ may therefore be absorbed into the communication weights by setting $L_k = \widetilde{L}_k/\Lambda$. The same scale must be used at all rounds, since round-dependent rescaling changes metric drift.

In the linear-operator model used here, coefficients may depend on the round number, the stated a priori parameters, and past external change notifications, but not on a matrix-specific eigendecomposition, the exact spectrum, or a topology-specific finite-time aggregation scheme. Generated vectors may be stored without restriction. The algorithms use no internal randomization: in the memoryless change-time model, randomness belongs only to the network change process.

B.2 Finite-window implementation and progress accounting

Algorithm 2 evaluates the coefficient ratios in (4) without forming the potentially large values $T_t(z_0)$. In exact arithmetic it is the same same recurrence and reproduces every fixed-operator Chebyshev iterate.

Algorithm 2: CHEBYSHEV-WINDOW$(\boldsymbol{u}; k, m)$: finite run of the classical Chebyshev semi-iteration

Input: current state $\boldsymbol{u} \in \mathbb{R}^{nd}$; starting communication round k; common spectral-envelope parameter χ; window length $m \geq 1$

$z_0 \leftarrow (1+\chi^{-1})/(1-\chi^{-1})$;
$a_0 \leftarrow 2/z_0$;
$\boldsymbol{v}_{\mathrm{prev}} \leftarrow \boldsymbol{u}$;
$\boldsymbol{v}_{\mathrm{curr}} \leftarrow (a_0/2) z_\chi(\boldsymbol{L}_k)\boldsymbol{u}$;
for $t = 1, \ldots, m-1$ **do**
  $a_t \leftarrow 4/(4z_0 - a_{t-1})$; // $= 2T_t(z_0)/T_{t+1}(z_0)$
  $c_t \leftarrow a_{t-1}a_t/4$; // $= T_{t-1}(z_0)/T_{t+1}(z_0)$
  $\boldsymbol{v}_{\mathrm{new}} \leftarrow a_t z_\chi(\boldsymbol{L}_{k+t})\boldsymbol{v}_{\mathrm{curr}} - c_t \boldsymbol{v}_{\mathrm{prev}}$;
  $\boldsymbol{v}_{\mathrm{prev}} \leftarrow \boldsymbol{v}_{\mathrm{curr}}$;
  $\boldsymbol{v}_{\mathrm{curr}} \leftarrow \boldsymbol{v}_{\mathrm{new}}$;
return $\boldsymbol{v}_{\mathrm{curr}}$;

Lemma B.1 (Mean preservation and accumulation of certified progress)**.** *Every call of Algorithm 1 preserves the block average. Moreover, if $\mathbf{u}^{(r)}$ and q_r denote, respectively, the state and the value of the progress counter after the rth update of the outer loop, then*

$$\boldsymbol{\Pi}\boldsymbol{u}^{(r)} = \boldsymbol{\Pi}\boldsymbol{u}^{(0)}, \qquad \left\|\boldsymbol{\Pi}_\perp \boldsymbol{u}^{(r)}\right\| \leq e^{-q_r} \left\|\boldsymbol{\Pi}_\perp \boldsymbol{u}^{(0)}\right\|.$$

Consequently, termination with $q_r \geq \ln(1/\delta)$ implies

$$\left\|\boldsymbol{\Pi}_\perp \boldsymbol{u}^{(r)}\right\| \leq \delta \left\|\boldsymbol{\Pi}_\perp \boldsymbol{u}^{(0)}\right\|.$$

Proof. Since $\boldsymbol{L}_k\boldsymbol{\Pi} = \boldsymbol{\Pi}\boldsymbol{L}_k = 0$, $\boldsymbol{\Pi} z_\chi(\boldsymbol{L}_k) = z_0\boldsymbol{\Pi}$. The first iterate of a Chebyshev window therefore preserves the mean because $a_0/2 = 1/z_0$. For every subsequent local step, the Chebyshev recurrence gives

$$a_t z_0 - c_t = \frac{2z_0 T_t(z_0) - T_{t-1}(z_0)}{T_{t+1}(z_0)} = 1.$$

Induction on the local step thus shows that every complete or interrupted Chebyshev window preserves the input mean.

Under metric drift, every window has length m_β and norm at most $e^{-m_\beta^2/(5\chi)}$ on the disagreement subspace by Theorem 4.2, which is its credit. Under the piecewise-constant scheduler, every window, complete or interrupted by a report, runs on one operator; by Propositions 2.9 and 2.6 a window of h rounds has norm at most $1/T_h(z_0) = e^{-\ln T_h(z_0)}$ on the disagreement subspace, which is again its credit. Multiplication of these successive bounds proves $\left\|\boldsymbol{\Pi}_\perp \boldsymbol{u}^{(r)}\right\| \leq e^{-q_r} \left\|\boldsymbol{\Pi}_\perp \boldsymbol{u}^{(0)}\right\|$. No commutativity between different windows or operators is used. □

C Classical Chebyshev facts

Here we collect the classical polynomial background, an illustrative residual plot, and the proofs used in Sections 2, 4, and 5. These details are separated from the main argument because the polynomial construction is standard; the moving-operator stability analysis is not.

C.1 Residual-polynomial viewpoint and illustration

For a fixed operator, any mean-preserving method in this linear-operator model gives

$$\boldsymbol{e}_m = p_m(\boldsymbol{L})\boldsymbol{e}_0, \qquad \deg p_m \leq m, \qquad p_m(0) = 1.$$

The spectral theorem therefore yields

$$\|\boldsymbol{e}_m\| \leq \max_{\lambda \in [\chi^{-1}, 1]} |p_m(\lambda)| \, \|\boldsymbol{e}_0\|.$$

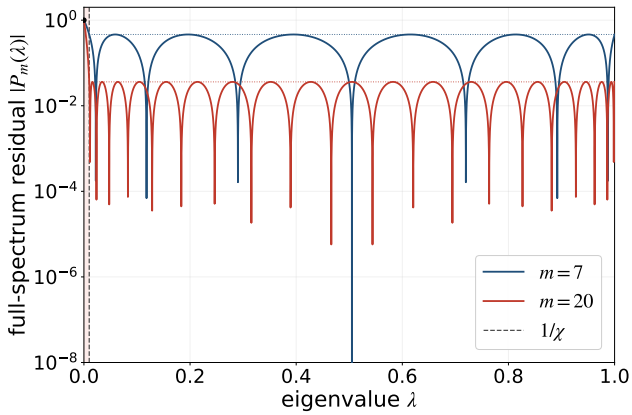

Figure 2: Full-spectrum residuals $|P_m^\chi(\lambda)|$ for $m = 7$ and $m = 20$ with $\chi = 100$. Both polynomials use the same interval $[\chi^{-1}, 1]$; only the finite window length changes.

The affine map z_χ sends the admissible interval to $[-1, 1]$, so the classical alternation theorem gives the shifted, normalized Chebyshev residual in Proposition 2.6. In particular,

$$\max_{\lambda\in[\chi^{-1},1]} |P_m^\chi(\lambda)| \le 2e^{-2m/\sqrt{\chi}}.$$

C.2 Proof of Proposition 2.6

The affine map $x = z_\chi(\lambda)$ sends $[\chi^{-1}, 1]$ onto $[-1, 1]$ and sends the normalization point $\lambda = 0$ to $z_0 > 1$. Hence (2) is equivalent to

$$\min_{\substack{\deg q\le m\\ q(z_0)=1}} \max_{x\in[-1,1]} |q(x)|.$$

The polynomial $q_\star(x) = T_m(x)/T_m(z_0)$ is feasible and has uniform norm $1/T_m(z_0)$ on $[-1, 1]$.

Let

$$x_j := \cos(j\pi/m), \qquad j = 0, \dots, m,$$

so that $x_0 > x_1 > \cdots > x_m$ and $T_m(x_j) = (-1)^j$. Let ℓ_j be the Lagrange cardinal polynomial associated with these nodes. Since $z_0 > x_0$, $\operatorname{sign} \ell_j(z_0) = (-1)^j$. Interpolation of T_m at the $m + 1$ nodes therefore gives

$$T_m(z_0) = \sum_{j=0}^{m} \ell_j(z_0)(-1)^j = \sum_{j=0}^{m} |\ell_j(z_0)|.$$

For every polynomial q of degree at most m with $q(z_0) = 1$, Lagrange interpolation yields

$$1 = \left|\sum_{j=0}^{m} \ell_j(z_0)q(x_j)\right| \le \|q\|_{L^\infty[-1,1]} \sum_{j=0}^{m} |\ell_j(z_0)| = T_m(z_0)\,\|q\|_{L^\infty[-1,1]}\,.$$

Hence $\|q\|_{L^\infty[-1,1]} \ge 1/T_m(z_0)$, and $q_\star = T_m/T_m(z_0)$ attains equality.

If equality holds, equality must hold both in the triangle inequality and in all $m+1$ pointwise bounds above. Consequently,

$$q(x_j) = \frac{(-1)^j}{T_m(z_0)} = q_\star(x_j), \qquad j = 0, \dots, m.$$

Two polynomials of degree at most m agreeing at $m + 1$ distinct points are identical. Thus $q = q_\star$, proving uniqueness.

Finally,

$$\theta_\chi := \operatorname{arccosh} z_0 = 2\operatorname{artanh}(\chi^{-1/2}) \ge 2\chi^{-1/2},$$

so

$$T_m(z_0) = \cosh(m\theta_\chi) \geq \tfrac{1}{2}e^{m\theta_\chi} \geq \tfrac{1}{2}e^{2m/\sqrt{\chi}},$$

which proves the final inequality in (2). □

C.3 Proof of Proposition 2.9

The identity holds for $t = 0$. Since $P_1^\chi(\lambda) = z_\chi(\lambda)/z_0$, it also holds for $t = 1$. Assume it holds for t and $t-1$. The Chebyshev recurrence and the closed forms of a_t, c_t in Table 1 give

$$\begin{aligned} P_{t+1}^\chi(\lambda) &= \frac{2T_t(z_0)}{T_{t+1}(z_0)} z_\chi(\lambda) P_t^\chi(\lambda) - \frac{T_{t-1}(z_0)}{T_{t+1}(z_0)} P_{t-1}^\chi(\lambda) \\ &= a_t z_\chi(\lambda) P_t^\chi(\lambda) - c_t P_{t-1}^\chi(\lambda). \end{aligned}$$

Substitution of the induction hypothesis into (5) with $\boldsymbol{L}_{k+t} \equiv \boldsymbol{L}$ yields $\boldsymbol{v}_{t+1} = P_{t+1}^\chi(\boldsymbol{L})\boldsymbol{u}$.

□

C.4 Proof of Lemma 2.7

Set

$$s := \chi^{-1/2}, \qquad x := m\theta_\chi = 2m \operatorname{artanh} s, \qquad h := \frac{m^2}{\chi}.$$

Proposition 2.6 gives

$$\max_{\lambda \in [\chi^{-1}, 1]} |P_m^\chi(\lambda)| = \operatorname{sech} x.$$

Since $\operatorname{artanh} s \geq s$, one has $x \geq 2\sqrt{h}$. For the upper range, $s \leq 1/2$, $ms \leq \sqrt{2}$, and the function $\operatorname{artanh}(s)/s$ is increasing on $(0, 1)$. Therefore

$$x = 2(ms)\frac{\operatorname{artanh} s}{s} \leq 2\sqrt{2}\ln 3 < \frac{25}{8}.$$

The last strict inequality follows, for example, from $\sqrt{2} < 99/70$ and $\ln 3 < 11/10$.

We next record a scalar estimate valid for $0 \leq x \leq 25/8$:

$$\ln\cosh x \geq \frac{x^2}{5}. \tag{15}$$

Indeed, let $F(x) = \ln\cosh x - x^2/5$ and $G(x) = F'(x) = \tanh x - 2x/5$. Since $G'(x) = \operatorname{sech}^2 x - 2/5$ is strictly decreasing on $(0, \infty)$, G can change sign from positive to negative at most once. Thus the minimum of F on $[0, 25/8]$ occurs at an endpoint. Now $F(0) = 0$, while

$$F(25/8) \geq \frac{25}{8} - \ln 2 - \frac{125}{64} > \frac{25}{8} - \frac{7}{10} - \frac{125}{64} > 0.$$

This proves (15). The rational logarithm bounds used above follow directly from the positive exponential series.

Combining $x^2 \geq 4h$ with (15) yields

$$\operatorname{sech} x = \exp(-\ln\cosh x) \leq \exp\left(-\frac{x^2}{5}\right) \leq \exp\left(-\frac{4m^2}{5\chi}\right) \leq \exp\left(-\frac{m^2}{2\chi}\right).$$

This is (3). □

D Proofs for the metric-drift results

We first establish the exact variation-of-constants decomposition and the cumulative response bounds, then prove the direct metric-drift window estimate. All operators below act on the disagreement subspace of $\mathbb{R}^{nd}$.

Let U_r denote the Chebyshev polynomial of the second kind. For a horizon m, define

$$\boldsymbol{\xi}_0 := -\frac{2}{(1-\chi^{-1})z_0}\boldsymbol{E}_0\boldsymbol{e}_0, \qquad \boldsymbol{\xi}_t := -\frac{2a_t}{1-\chi^{-1}}\boldsymbol{E}_t\boldsymbol{e}_t \quad (t \geq 1),$$

and

$$\mathcal{R}^{\chi}_{m,j}(\boldsymbol{L}_k) := \frac{T_j(z_0)}{T_m(z_0)}U_{m-j}(z_\chi(\boldsymbol{L}_k)).$$

Also set $\theta_\chi := \operatorname{arccosh} z_0 = 2\operatorname{artanh}(\chi^{-1/2})$. For scalar spectral arguments write

$$R^{\chi}_{m,j}(\lambda) := \frac{T_j(z_0)}{T_m(z_0)}U_{m-j}(z_\chi(\lambda)).$$

Lemma D.1 (Variation of constants)**.** *Under the hypotheses of Lemma 4.1,*

$$\boldsymbol{e}_m = P^{\chi}_m(\boldsymbol{L}_k)\boldsymbol{e}_0 + \sum_{j=1}^{m}\mathcal{R}^{\chi}_{m,j}(\boldsymbol{L}_k)\boldsymbol{\xi}_{j-1}, \tag{16}$$

and $\|\boldsymbol{\xi}_t\| \leq 8\,\|\boldsymbol{E}_t\|\,\|\boldsymbol{e}_t\|$.

Lemma D.2 (Response sums)**.** *For $j = 1, \ldots, m$ and $\lambda \in [\chi^{-1}, 1]$,*

$$\left|\frac{T_j(z_0)}{T_m(z_0)}U_{m-j}(z_\chi(\lambda))\right| \leq 2(m-j+1)e^{-(m-j)\theta_\chi}. \tag{17}$$

Moreover,

$$\sum_{j=1}^{m}\sup_{[\chi^{-1},1]}|R^{\chi}_{m,j}| \leq 2\min\left\{\frac{m(m+1)}{2}, \frac{1}{(1-e^{-\theta_\chi})^2}\right\}, \tag{18}$$

and

$$\sum_{j=1}^{m}(j-1)\sup_{[\chi^{-1},1]}|R^{\chi}_{m,j}| \leq 2\min\left\{\frac{m^3}{6}, \frac{m}{(1-e^{-\theta_\chi})^2}\right\}. \tag{19}$$

D.1 Proofs of Lemmas D.1, D.2, and 4.1

Identity. To remove the time-dependent normalization in the recurrence, define the auxiliary sequence $\boldsymbol{b}_t := T_t(z_0)\,\boldsymbol{e}_t$ and set $\widehat{\boldsymbol{Z}} := z_\chi(\boldsymbol{L}_k)$. From (10) and (4), for $t \geq 1$,

$$\begin{aligned}\boldsymbol{b}_{t+1} &= T_{t+1}(z_0)\big[a_t z_\chi(\boldsymbol{L}_{k+t})\boldsymbol{e}_t - c_t\boldsymbol{e}_{t-1}\big] = 2T_t(z_0)\,z_\chi(\boldsymbol{L}_{k+t})\boldsymbol{e}_t - T_{t-1}(z_0)\boldsymbol{e}_{t-1}\\ &= 2\widehat{\boldsymbol{Z}}\boldsymbol{b}_t - \boldsymbol{b}_{t-1} + T_{t+1}(z_0)\,\boldsymbol{\xi}_t.\end{aligned}$$

because $z_\chi(\boldsymbol{L}_{k+t}) - z_\chi(\boldsymbol{L}_k) = -\frac{2}{1-\chi^{-1}}\boldsymbol{E}_t$. Moreover,

$$T_{t+1}(z_0)\boldsymbol{\xi}_t = -\frac{2a_tT_{t+1}(z_0)}{1-\chi^{-1}}\boldsymbol{E}_t\boldsymbol{e}_t = -\frac{4T_t(z_0)}{1-\chi^{-1}}\boldsymbol{E}_t\boldsymbol{e}_t = 2T_t(z_0)\big(z_\chi(\boldsymbol{L}_{k+t}) - \widehat{\boldsymbol{Z}}\big)\boldsymbol{e}_t.$$

Similarly, $\boldsymbol{b}_1 = z_0\boldsymbol{e}_1 = z_\chi(\boldsymbol{L}_k)\boldsymbol{e}_0 = \widehat{\boldsymbol{Z}}\boldsymbol{b}_0 + T_1(z_0)\boldsymbol{\xi}_0$. Thus $(\boldsymbol{b}_t)$ solves the constant-coefficient Chebyshev recurrence $\boldsymbol{b}_{t+1} = 2\widehat{\boldsymbol{Z}}\boldsymbol{b}_t - \boldsymbol{b}_{t-1}$ driven by the impulses $T_j(z_0)\boldsymbol{\xi}_{j-1}$ added at times $j = 1, \ldots, m$. The homogeneous solution with data $\boldsymbol{b}_0 = \boldsymbol{e}_0$, $\boldsymbol{b}_1 = \widehat{\boldsymbol{Z}}\boldsymbol{b}_0$ is $\boldsymbol{b}^{\mathrm{h}}_t = T_t(\widehat{\boldsymbol{Z}})\boldsymbol{e}_0$ (the initial conditions of T_t). The response at time m to a unit impulse $\boldsymbol{w}$ added at time j is $U_{m-j}(\widehat{\boldsymbol{Z}})\boldsymbol{w}$, since the sequence $U_{t-j}(\widehat{\boldsymbol{Z}})\boldsymbol{w}$ satisfies the homogeneous recurrence for $t > j$ and takes the values 0 and $\boldsymbol{w}$ at times $j-1$ and j, respectively ($U_0 = 1$, $U_1(z) = 2z$). Superposition gives $\boldsymbol{b}_m = T_m(\widehat{\boldsymbol{Z}})\boldsymbol{e}_0 + \sum_{j=1}^{m}U_{m-j}(\widehat{\boldsymbol{Z}})T_j(z_0)\boldsymbol{\xi}_{j-1}$; dividing by $T_m(z_0)$ yields (16). The same derivation with horizon $t \leq m$ in place of m gives the partial-horizon identity used in the bootstrap below.

Norm of the impulses. For $t \geq 1$, $\|\boldsymbol{\xi}_t\| \leq \frac{2a_t}{1-\chi^{-1}}\,\|\boldsymbol{E}_t\|\,\|\boldsymbol{e}_t\| \leq \frac{4}{1-\chi^{-1}}\,\|\boldsymbol{E}_t\|\,\|\boldsymbol{e}_t\| \leq 8\,\|\boldsymbol{E}_t\|\,\|\boldsymbol{e}_t\|$ ($a_t \leq 2$, $\chi^{-1} \leq \frac{1}{4}$); for $t = 0$, $\frac{2}{(1-\chi^{-1})z_0} \leq \frac{2}{1-\chi^{-1}} \leq 4 \leq 8$.

Bounds on the response operators. For every $\lambda \in [\chi^{-1}, 1]$, $z_\chi(\lambda) \in [-1, 1]$ and, for every integer $r \geq 0$,

$$|U_r(\cos\vartheta)| = \left|\frac{\sin((r+1)\vartheta)}{\sin\vartheta}\right| \leq r+1.$$

At $\vartheta = 0$ or π, the quotient is understood by continuity. Taking $r = m - j$ and using

$$\frac{T_j(z_0)}{T_m(z_0)} = \frac{\cosh(j\theta_\chi)}{\cosh(m\theta_\chi)} \leq \frac{e^{j\theta_\chi}}{\frac{1}{2}e^{m\theta_\chi}} = 2e^{-(m-j)\theta_\chi}$$

gives (17). Summing with $r = m - j$,

$$\sum_{j=1}^{m} \sup_{[\chi^{-1},1]} |R^\chi_{m,j}| \leq 2\sum_{r=0}^{m-1}(r+1)e^{-r\theta_\chi}$$
$$\leq 2\min\Big\{\frac{m(m+1)}{2}, \sum_{r\geq 0}(r+1)x^r\Big|_{x=e^{-\theta_\chi}}\Big\} = 2\min\Big\{\frac{m(m+1)}{2}, \frac{1}{(1-e^{-\theta_\chi})^2}\Big\},$$

which is (18). For (19), bounding $e^{-(m-j)\theta_\chi} \leq 1$ and using $\sum_{j=1}^{m}(j-1)(m-j+1) = \sum_{i=1}^{m-1} i(m-i) = \frac{m(m-1)(m+1)}{6} \leq \frac{m^3}{6}$ gives the first branch; bounding $(j-1) \leq m$ and reusing the geometric sum gives the second.

□

Proof of Lemma 4.1. Apply the partial-horizon form of (16) with horizon t. The term with $j = 1$ vanishes because $\boldsymbol{E}_0 = 0$. For $j \geq 2$, the identity $T_j(z_0)a_{j-1} = 2T_{j-1}(z_0)$ gives

$$\mathcal{R}^\chi_{t,j}(\boldsymbol{L}_k)\boldsymbol{\xi}_{j-1} = -\frac{4}{1-\chi^{-1}}\frac{T_{j-1}(z_0)}{T_t(z_0)}U_{t-j}\big(z_\chi(\boldsymbol{L}_k)\big)\boldsymbol{E}_{j-1}\boldsymbol{e}_{j-1}.$$

On the disagreement subspace one has $\|U_{t-j}(z_\chi(\boldsymbol{L}_k))\| \leq t-j+1$, and $\|\boldsymbol{E}_{j-1}\| \leq \beta(j-1)$, $\|\boldsymbol{e}_{j-1}\| \leq M_t$. Hence

$$\|\boldsymbol{e}_t - P^\chi_t(\boldsymbol{L}_k)\boldsymbol{e}_0\| \leq \frac{4\beta M_t}{1-\chi^{-1}}\sum_{j=2}^{t}\frac{T_{j-1}(z_0)}{T_t(z_0)}(j-1)(t-j+1). \tag{20}$$

Short-window branch. Since $T_{i+1}(z_0) = 2z_0T_i(z_0) - T_{i-1}(z_0) \geq z_0T_i(z_0)$ for $i \geq 1$ and $T_1(z_0) = z_0T_0(z_0)$, the ratio in (20) is at most z_0^{-1} for $j \leq t$. Moreover $(1-\chi^{-1})z_0 = 1+\chi^{-1}$ and $\sum_{j=2}^{t}(j-1)(t-j+1) = (t^3-t)/6$, so the right-hand side of (20) is at most

$$\frac{4}{1+\chi^{-1}}\cdot\frac{t^3-t}{6}\beta M_t \leq \frac{2}{3}\beta(t^3-t)M_t.$$

Damped branch. Since $T_{j-1}(z_0)/T_t(z_0) = \cosh((j-1)\theta_\chi)/\cosh(t\theta_\chi) \leq 2e^{-(t-j+1)\theta_\chi}$ and $j-1 \leq t$, the sum in (20) is at most

$$2t\sum_{r\geq 1} re^{-r\theta_\chi} = \frac{t}{2\sinh^2(\theta_\chi/2)} = \frac{t(\chi-1)}{2},$$

where $\sinh^2(\theta_\chi/2) = 1/(\chi-1)$. Multiplying by $4\beta M_t/(1-\chi^{-1})$ gives $2\beta t\chi M_t$. Taking the better of the two estimates gives the asserted perturbation bound.

Because $\|P^\chi_t(\boldsymbol{L}_k)\| \leq 1$ and $D_t \leq D_m$ for $t \leq m$, this bound implies $M_m \leq \|\boldsymbol{e}_0\| + D_mM_m$. If $D_m < 1$, then $M_m \leq \|\boldsymbol{e}_0\|/(1-D_m)$. Applying the perturbation bound once more at $t = m$ and using $\|P^\chi_m(\boldsymbol{L}_k)\| \leq \rho^\chi_m$ proves (11). □

D.2 Proof of Theorem 4.2

If $m = 1$, then

$$P^\chi_1(\lambda) = 1 - \frac{2\lambda}{1+\chi^{-1}}, \qquad \max_{[\chi^{-1},1]}|P^\chi_1| = \frac{\chi-1}{\chi+1} \leq e^{-2/\chi} \leq e^{-1/(5\chi)},$$

since $\ln((\chi+1)/(\chi-1)) = 2\,\mathrm{artanh}(\chi^{-1}) \geq 2/\chi$. This calculation uses only the current operator and is valid for arbitrary drift.

Now let $m \geq 2$ and let $h = m^2/\chi \in (0, 1]$. The conditions of the theorem and Lemma 4.1 give

$$D_m \leq \frac{2}{3}\beta m^3 = \frac{2}{3}(\beta m\chi)h \leq \frac{2h}{9} \leq \frac{2}{9}.$$

Therefore (11) and Lemma 2.7 imply

$$\|\boldsymbol{e}_m\| \leq \left(e^{-4h/5} + \frac{2h/9}{1-2h/9}\right)\|\boldsymbol{e}_0\| \leq \left(e^{-4h/5} + \frac{2h}{7}\right)\|\boldsymbol{e}_0\|.$$

For $0 < h \leq 1$, write $A = e^{-h/5}$ and $B = e^{-4h/5}$. The inequality $e^x \geq 1 + x$ gives $A \geq 4/5$ and $A \geq B(1 + 3h/5)$. Consequently,

$$A - B \geq \frac{3hA}{5+3h} \geq \frac{3h}{10} > \frac{2h}{7}.$$

Hence (12) holds in both cases.

Complexity. Each complete window earns at least $m^2/(5\chi)$ nats. Hence

$$J = \left\lceil \frac{5\chi}{m^2} \ln \frac{1}{\epsilon} \right\rceil$$

windows suffice. Since $K = Jm$ and $\lceil a \rceil m \leq am + m$, this gives the communication bound stated in Theorem 4.2. □

D.3 Proof of Theorem 3.1 and Corollary 3.2

Let $\widetilde{m}_\beta, m_\beta$ be given by (6). If $m_\beta \geq 2$, then $m_\beta \leq \widetilde{m}_\beta \leq \sqrt{\chi}$ and $\beta m_\beta \chi \leq 1/3$, so Theorem 4.2 applies. If $m_\beta = 1$, its degree-one case applies. Thus every complete window earns $m_\beta^2/(5\chi)$ nats, and the communication bound in Theorem 4.2 holds with $m = m_\beta$.

If $\widetilde{m}_\beta \geq 2$, then $m_\beta = \lfloor \widetilde{m}_\beta \rfloor \geq \widetilde{m}_\beta/2$ and

$$\frac{\chi}{m_\beta} \leq 2\frac{\chi}{\widetilde{m}_\beta} = 2\max\{\sqrt{\chi}, 3\beta\chi^2\}.$$

Moreover, $\widetilde{m}_\beta \geq 2$ implies $\beta\chi^2 \leq \chi/6 < \chi$, so the second term is exactly the unsaturated quantity in $\min\{\beta\chi^2, \chi\}$. Together with $m_\beta \leq \sqrt{\chi}$, this bounds that communication estimate by an absolute multiple of the right-hand side of (8).

If $\widetilde{m}_\beta < 2$, then $m_\beta = 1$. Because $\chi \geq 4$, this can occur only when $(3\beta\chi)^{-1} < 2$, hence $\beta\chi^2 > \chi/6$. If $\beta \leq 1/\chi$, then $\min\{\beta\chi^2, \chi\} = \beta\chi^2 > \chi/6$; if $\beta > 1/\chi$, the minimum equals χ. In either case the degree-one cost $5\chi\ln(1/\epsilon) + 1$ is bounded by the claimed right-hand side after enlarging one absolute constant.

Finally, if $\beta \leq (3\chi^{3/2})^{-1}$, then the first argument in the minimum defining $\widetilde{m}_\beta$ is active, so $m_\beta = \lfloor\sqrt{\chi}\rfloor \geq \sqrt{\chi}/2$. The communication estimate in Theorem 4.2 then gives $K \leq (5\chi/m_\beta)\ln(1/\epsilon) + m_\beta \leq 10\sqrt{\chi}\ln(1/\epsilon) + \sqrt{\chi}$, proving Corollary 3.2. □

E Piecewise-constant networks

E.1 Detector timing and scheduler convention

Before round $k \geq 1$, the external detector reports whether $L_k \neq L_{k-1}$. If it reports a change, the current window attempt ends before round k, and round k is the first round of a new length-one attempt. A change notification before round k is processed only once: after it resets the attempted length, the new attempt executes round k using L_k and checks for further changes only before subsequent rounds. Thus no communication round is skipped or counted twice.

An attempt of length m starting at round k completes if no change is reported before rounds $k+1, \ldots, k+m-1$. A change reported before round $k+m$ does not erase a window that has already

completed; it resets the next attempted length to one. If no change is reported, the next attempted length is $\min\{2m, m_{\max}\}$, where

$$m_{\max} = \lfloor \sqrt{\chi} \rfloor, \qquad \frac{\sqrt{\chi}}{2} \le m_{\max} \le \sqrt{\chi}.$$

Every window, complete or interrupted, earns the credit $\ln T_h(z_0)$, where h is its number of executed rounds. By Lemma 2.7, every window of length $1 \le h \le m_{\max}$ earns at least $4h^2/(5\chi)$; the bounds below use only completed windows and discard the nonnegative credit of interrupted ones.

The first t iterations of a window, $t \ge 1$, use a single operator L_k. By Proposition 2.9, they apply $P_t^\chi(\boldsymbol{L}_k)$ to the stacked state. The map z_χ in (1) sends $[\chi^{-1}, 1]$ onto $[-1, 1]$, while $z_0 > 1$. For $s = \cos\theta \in [-1, 1]$ one has $T_t(s) = \cos(t\theta)$, whereas $T_t(z_0) = \cosh(t \operatorname{arccosh} z_0) \ge 1$. Thus the definition (1) gives

$$\begin{aligned} \max_{\lambda \in [\chi^{-1}, 1]} |P_t^\chi(\lambda)| &= \frac{\max_{s \in [-1,1]} |T_t(s)|}{T_t(z_0)} \\ &= \frac{1}{\cosh(t \operatorname{arccosh} z_0)} \le 1. \end{aligned}$$

By the spectral theorem and the uniform spectral bound, $\|\boldsymbol{\Pi}_\perp P_t^\chi(\boldsymbol{L}_k)\boldsymbol{u}\| \le \|\boldsymbol{\Pi}_\perp \boldsymbol{u}\|$ for every stacked state $\boldsymbol{u}$. The same argument applies to a window interrupted by a change and to the state observed after any prescribed number of completed rounds; zero completed rounds leave the state unchanged. Since the norm after t iterations is at most $1/T_t(z_0) = e^{-\ln T_t(z_0)}$, if q is the progress counter in Algorithm 1, the disagreement norm at every window boundary is at most e^{-q} times its value at the beginning of the call.

E.2 Accumulated progress in a stable segment

Consider $r \ge 1$ consecutive rounds with a fixed operator, beginning with a length-one attempt. Completed window lengths double until they reach $m_{\max}$; subsequent completed windows have length $m_{\max}$. We bound the progress credited within these r rounds.

Let m be the largest window length completed by round r. If $m < m_{\max}$, the completed lengths through m sum to less than $2m$, and the next attempted length is at most $2m$. Since that attempt has not completed by round r, we have $r < 4m$, and therefore $m \ge (r+1)/4$. If $m = m_{\max}$, the bound is immediate. Thus $m \ge \min\{(r+1)/4, m_{\max}\}$. This single completed window earns at least $4m^2/(5\chi)$, so the total credited progress is at least

$$\frac{4m^2}{5\chi} \ge \frac{4}{5\chi} \min\left\{\frac{r+1}{4}, m_{\max}\right\}^2. \tag{21}$$

For $r \ge 4m_{\max}$, we also count completed windows of length $m_{\max}$. The shorter windows preceding the first such window use $2^{\lceil \log_2 m_{\max} \rceil} - 1 < 2m_{\max}$ rounds. Consequently, at least $\lfloor (r - 2m_{\max})/m_{\max} \rfloor \ge r/(5m_{\max})$ windows of length $m_{\max}$ complete by round r. Each earns at least $4m_{\max}^2/(5\chi)$, giving total credited progress of at least

$$\frac{r}{5m_{\max}} \cdot \frac{4m_{\max}^2}{5\chi} = \frac{4rm_{\max}}{25\chi}. \tag{22}$$

For $r < 4m_{\max}$, (21) is at least $r \min\{r, m_{\max}\}/(20\chi)$: if $(r+1)/4 \le m_{\max}$, it is at least $r^2/(20\chi)$; otherwise it is at least $4m_{\max}^2/(5\chi)$, which is at least $rm_{\max}/(5\chi)$ because $r < 4m_{\max}$. For $r \ge 4m_{\max}$, (22) gives a stronger bound. Hence every such segment of r rounds earns at least

$$\frac{r \min\{r, m_{\max}\}}{20\chi}. \tag{23}$$

E.3 Minimum spacing and arbitrary call starts

Lemma E.1 (Deterministic doubling from an arbitrary start)**.** *Under the hypotheses of Theorem 3.3, let a fresh* WAVE *call start at an arbitrary global communication round, reset its attempted length*

to one, and thereafter reset at every detected change. There is an absolute constant C_{call} such that, for every $\delta \in (0,1)$, its finishing round k^+ satisfies

$$k^+ - k \le C_{\mathrm{call}} \left(\sqrt{\chi} + \frac{\chi}{\tau}\right) \ln \frac{e}{\delta}.$$

The scheduler does not require prior knowledge of τ.

Proof. Continue the call hypothetically without its stopping test. Let q_K be the value of the certified progress counter q after its first K rounds, counting only windows completed by then. Set

$$a_\tau := \frac{1}{80} \min\left\{\frac{\tau}{\chi}, \frac{1}{m_{\max}}\right\}.$$

On any complete interval between changes of length $r \ge \tau$, (23) gives an average credited progress per round of at least

$$\frac{\min\{r, m_{\max}\}}{20\chi} \ge \frac{\min\{\tau, m_{\max}\}}{20\chi} \ge a_\tau.$$

The last inequality uses $m_{\max}^2 \ge \chi/4$.

A piece of length r at either boundary of the call may be shorter than τ. If $r < 4m_{\max}$, its credited progress is at least $0 \ge a_\tau(r - 4m_{\max})$. If $r \ge 4m_{\max}$, (22) gives credited progress of at least $4rm_{\max}/(25\chi) \ge a_\tau r$. Thus every boundary piece contributes at least $a_\tau(r - 4m_{\max})$.

Partition the first K call rounds at their change boundaries. All interior pieces are complete intervals between changes, and at most the first and last pieces are boundary pieces. If these coincide, count their overhead only once. Summing the preceding bounds gives

$$q_K \ge a_\tau \max\{K - 8m_{\max}, 0\}.$$

Consequently, the stopping threshold $L = \ln(1/\delta)$ has been reached by $K_0 = \lceil 8m_{\max} + L/a_\tau \rceil$ rounds. Since $a_\tau^{-1} = 80 \max\{\chi/\tau, m_{\max}\}$ and $m_{\max} \le \sqrt{\chi}$, the claimed estimate follows from $L + 1 = \ln(e/\delta)$. □

Proof of Theorem 3.3. Apply Lemma E.1 at $k = 0$ with $\delta = \epsilon$. This gives (9) after renaming an absolute constant. □

E.4 Memoryless change times

This is a stochastic variant of the piecewise-constant model. The operator is fixed between reported change points, but the segment lengths are geometric rather than bounded below by a deterministic τ.

Assumption E.2 (Memoryless change times)**.** Let $(B_k)_{k\ge 0}$ be independent Bernoulli variables, independent of the past, with $\mathbb{P}(B_k = 1) = 1/\bar{\tau}$, where $\bar{\tau} \ge 1$. If $B_k = 0$, then $L_{k+1} = L_k$. If $B_k = 1$, the detector reports a new segment boundary and a different normalized gossip matrix satisfying Assumption 2.2 may be selected as a function of the past. Thus only the change times are memoryless.

Theorem E.3 (Consensus under memoryless change times)**.** *Let Assumptions 2.2 and E.2 hold with $\chi \ge 4$, and assume access to the ideal change detector. For $\epsilon \in (0,1)$, run Algorithm 1 on $(\boldsymbol{u}_0, 0; \epsilon)$ with $\mathcal{M} =$ piecewise, and let $(\widehat{\boldsymbol{u}}, K)$ be its output. Then $K < \infty$ almost surely, $\widehat{\boldsymbol{u}}$ is an ϵ-consensus solution almost surely, and*

$$\mathbb{E}K = O\left(\left(\sqrt{\chi} + \frac{\chi}{\bar{\tau}}\right) \ln \frac{e}{\epsilon}\right). \tag{24}$$

In particular, $\bar{\tau} \ge \sqrt{\chi}$ recovers the fixed-network $O(\sqrt{\chi}\ln(e/\epsilon))$ dependence in expectation.

Proof of Theorem E.3. We bound the duration D_δ of a fresh certified call from any starting round. Its disagreement hitting time is no larger than D_δ. For the analysis, continue the scheduler hypothetically after the call terminates, and let q_K be the value of its certified progress counter after K rounds of this continuation. All probabilities and expectations below may be conditioned on the

history at an almost surely finite stopping-time start. Future change indicators remain independent Bernoulli variables with probability $1/\bar{\tau}$; the new call starts with a length-one attempt.

First, a deterministic bound holds for every switching sequence. Partition any K call rounds into constant-operator pieces of lengths r_i. Each piece starts with a length-one attempt, so (23) and $\min\{r_i, m_{\max}\} \geq 1$ give

$$q_K \geq \sum_i \frac{r_i \min\{r_i, m_{\max}\}}{20\chi} \geq \frac{K}{20\chi}. \tag{25}$$

Hence $D_\delta \leq \lceil 20\chi \ln(1/\delta) \rceil \leq C_0 \chi \ln(e/\delta)$, which gives the required order when $1 \leq \bar{\tau} < 2$.

Suppose $2 \leq \bar{\tau} \leq 4m_{\max}$. Let T_i be successive constant-operator segment lengths from the call start. By memorylessness, the T_i are i.i.d. geometric random variables with mean $\bar{\tau}$, even conditional on the call-start history. Set

$$Y_i := \frac{4}{5\chi} \min\left\{ \frac{T_i + 1}{4}, m_{\max} \right\}^2, \qquad \mu_Y := \mathbb{E}Y_i, \qquad L := \ln \frac{1}{\delta}.$$

By (21), Y_i is a lower bound on the progress earned in segment i, and $0 \leq Y_i \leq 4m_{\max}^2/(5\chi) \leq 4/5$. Moreover,

$$\mathbb{P}(T_i \geq \lceil \bar{\tau} \rceil) = (1 - 1/\bar{\tau})^{\lceil \bar{\tau} \rceil - 1} \geq (1 - 1/\bar{\tau})^{\bar{\tau}} \geq \tfrac{1}{4}.$$

On this event, $\bar{\tau} \leq 4m_{\max}$ implies $Y_i \geq \bar{\tau}^2/(20\chi)$. Therefore

$$\mu_Y \geq \frac{\bar{\tau}^2}{80\chi}. \tag{26}$$

Let $N := \inf\{n \geq 1 : \sum_{i=1}^n Y_i \geq L\}$. For each integer $a \geq 1$, the event $\{N \geq i\}$ depends only on $Y_1, \ldots, Y_{i-1}$. Independence and the overshoot bound give

$$\mu_Y \mathbb{E}(N \wedge a) = \mathbb{E} \sum_{i=1}^{N \wedge a} Y_i \leq L + \tfrac{4}{5}.$$

Monotone convergence and (26) imply $\mathbb{E}N \leq 80\chi(L+1)/\bar{\tau}^2 < \infty$. The certified call ends no later than the end of segment N. Because $\{N \geq i\}$ is independent of T_i, Tonelli's theorem gives

$$\mathbb{E}D_\delta \leq \mathbb{E} \sum_{i=1}^{N} T_i = \bar{\tau}\mathbb{E}N \leq \frac{80\chi}{\bar{\tau}}(L+1).$$

Finally, suppose $\bar{\tau} > 4m_{\max}$. Partition future call time into disjoint blocks of $4m_{\max}$ rounds. Call a block clean if no change is reported before its second through last rounds. A change before its first round merely starts a new length-one attempt there. The clean-block events involve disjoint change indicators and are i.i.d., with probability

$$p_{\mathrm{cl}} = (1 - 1/\bar{\tau})^{4m_{\max} - 1} \geq (1 - 1/\bar{\tau})^{\bar{\tau}} \geq \tfrac{1}{4}.$$

Within a clean block, the attempt in progress at its start finishes after at most its attempted length. From that attempt through the next completed window of length $m_{\max}$, the total number of rounds is at most $A + m_{\max} < 3m_{\max} < 4m_{\max}$, where $A = 2^{\lceil \log_2 m_{\max} \rceil} - 1$. Thus a window of length $m_{\max}$ completes within the block, regardless of the attempt state at its start. It earns at least $b := 4m_{\max}^2/(5\chi) \geq 1/5$.

For block j, put $Z_j := b\mathbf{1}\{\text{block } j \text{ is clean}\}$. The variables Z_j are i.i.d., $0 \leq Z_j \leq 4/5$, and $\mathbb{E}Z_j = bp_{\mathrm{cl}} \geq 1/20$. The progress earned by the end of block n is at least $\sum_{j=1}^n Z_j$. If $M := \inf\{n \geq 1 : \sum_{j=1}^n Z_j \geq L\}$, the same bounded stopping-time calculation gives $\mathbb{E}M \leq 20(L+1)$. Consequently,

$$\mathbb{E}D_\delta \leq 4m_{\max}\mathbb{E}M \leq 80m_{\max}\left(\ln \frac{1}{\delta} + 1\right).$$

Together with (25) and $m_{\max} \leq \sqrt{\chi}$, these cases give the conditional $O((\sqrt{\chi} + \chi/\bar{\tau}) \ln(e/\delta))$ bound. In particular, $D_\delta < \infty$ almost surely. The matrices selected at change points may depend on the past: the argument uses only the change indicators and progress bounds uniform over admissible operators. Taking the call start at zero and $\delta = \epsilon$ proves (24). □

E.5 FLOODING, ROLLBACK, AND COMMITTED OUTPUTS

Throughout this subsection $n \geq 2$, $\chi \geq 4$, and $D := D_n = \lfloor (\sqrt{\chi}/2) \ln(2n) \rfloor + 1$. The additional local-observation and message assumptions are those stated before Proposition 5.1. Global round indices and window schedules are synchronized. Changes to rows, including changes in edge weights or the disappearance of an edge, are observed before the new row is used. No agent needs to know which other rows have changed.

Lemma E.4 (Diameter of the operator support)**.** *Let L be a normalized gossip matrix satisfying $\operatorname{spec}(L|_{\mathbf{1}_n^\perp}) \subseteq [\chi^{-1}, 1]$. Its undirected nonzero-support graph, with edges $\{i, j\}$ whenever $i \neq j$ and $L_{ij} \neq 0$, is connected and has diameter at most*

$$\left\lfloor \frac{\operatorname{arccosh}(n)}{\theta_\chi} \right\rfloor + 1 \leq D, \qquad \theta_\chi := 2 \operatorname{artanh}(\chi^{-1/2}).$$

Proof. We use the polynomial diameter argument of **?**. If the support graph were disconnected, its component indicators would give linearly independent kernel vectors, contradicting $\ker L = \operatorname{span}\{\mathbf{1}_n\}$. Put $r := \lfloor \operatorname{arccosh}(n)/\theta_\chi \rfloor + 1$ and $M := P_r^\chi(L)$. Since $P_r^\chi(0) = 1$, the spectral theorem and the Chebyshev residual bound give

$$M = \Pi + M(I_n - \Pi), \qquad \|M(I_n - \Pi)\| \leq \frac{1}{\cosh(r\theta_\chi)} < \frac{1}{n}.$$

Every entry of M is therefore strictly positive. On the other hand, $(L^j)_{ab} = 0$ whenever the support-graph distance from a to b exceeds j. As M is a polynomial of degree r in L, positivity forces every such distance to be at most r. Finally, $\theta_\chi \geq 2/\sqrt{\chi}$ and $\operatorname{arccosh}(n) \leq \ln(2n)$ imply $r \leq D$. □

Change propagation. At a change time t, every agent whose row changes attaches t to its messages starting in round t. Agents forward the largest change index they have learned. They do not alter the numerical schedule merely upon learning an index. All agents process that index only at the common boundary $t + D$ and only once. Since $\tau \geq 2D$, the operator is fixed through rounds $t, \ldots, t + D - 1$. Its nonzero-support graph is consequently fixed and is a subgraph of every available communication graph during these rounds, even if additional unused links vary. Information traverses one support edge per round, so Lemma E.4 ensures that every agent knows t before boundary $t + D$. No later change can overtake this propagation. A single timestamp per message therefore suffices.

Checkpoints and synchronized rollback. For the numerical computation, the ordinary stopping test of Algorithm 1 is temporarily disabled: window attempts continue even when their progress counter exceeds $L := \ln(1/\epsilon)$. At each boundary s, after crediting any just-completed window and processing any rollback at that boundary, each agent records its local state $u_s^{(i)}$ and the common scalar data (q_s, h_s). Here q_s is the credit of completed windows and h_s is the number of already executed rounds of the current, unfinished attempt; $h_s = 0$ if a window has just ended or restarted. The most recent $D+1$ checkpoints are retained. For a change at t, the checkpoint at t is taken before any use of L_t.

At boundary $t + D$, every agent restores $u_t^{(i)}$, replaces its counter by

$$q \leftarrow q_t + \ln T_{h_t}(z_0),$$

and resets the attempted length to one. The physical round index is not rolled back. If $h_t = 0$, the added credit is zero; in particular, the credit of a window completed exactly at t is not counted twice. The restored state is the output after h_t iterations using the old operator only. The new length-one attempt begins in physical round $t + D$, using the new operator. Local notifications never cause an earlier or asynchronous numerical restart. Apart from these common rollbacks, the ordinary doubling schedule is used.

Output validation. All candidate handling occurs at common boundaries, after any scheduled rollback and its counter correction. At a window-completion or rollback boundary b with $q \geq L$, if no candidate is pending, the agents save the local output candidate $v^{(i)} := u_b^{(i)}$ and its index b.

They continue numerical computation, row observation, and timestamp forwarding during the next D rounds. If a rollback subsequently occurs for a change time $t < b$, they cancel this candidate; after correcting the counter, a new candidate may be created at that same boundary if its counter meets the threshold. A rollback for $t \geq b$ does not cancel the saved candidate, because no operator from round t or later contributed to it. At boundary $b + D$, after processing any rollback due there, an uncancelled candidate is returned by every agent. The saved candidate is retained separately until cancellation or return. Thus neither an unconfirmed threshold crossing nor a locally received flag can halt the communication needed to validate or cancel an output.

Proof of Proposition 5.1. First continue the numerical protocol hypothetically forever, ignoring the return instruction. Delete the intervals of physical rounds

$$[t_i, t_i + D), \qquad i \geq 1,$$

from the operator sequence. They are disjoint because $\tau \geq 2D$. The first remaining constant segment has length at least τ, and each subsequent complete segment has length at least $\tau - D \geq \tau/2$. At boundary $t_i + D$, the synchronized rollback restores the state reached with the old operator at t_i, credits the corresponding completed iterations, and restarts on the new operator. Therefore the states, counters, and schedules on the retained rounds agree exactly with the ideal-detector scheduler on the compressed sequence. The temporary states and credits inside a deleted interval play no role in this retained computation. The argument is pathwise and does not require the new operators to be independent of the past.

By Lemma E.1, the ideal-detector run reaches its stopping threshold after at most

$$K' \leq C_{\mathrm{call}} \left(\sqrt{\chi} + \frac{2\chi}{\tau} \right) \ln \frac{e}{\epsilon}$$

retained rounds. Let B be the corresponding physical boundary, taking the boundary after a rollback when a deleted interval is compressed to the terminal boundary. If N_B intervals have been deleted before B, then $B = K' + DN_B$ (or the same inequality with the displayed upper bound for K'). Minimum spacing gives $N_B \leq B/\tau + 1$, hence

$$B \leq K' + \frac{D}{\tau} B + D \leq K' + \frac{B}{2} + D, \qquad B \leq 2K' + 2D.$$

Every false candidate is cancelled: if its state or credit was affected by an as-yet unprocessed change at $t < b$, the common rollback at $t + D$ occurs after its creation and no later than $b + D$. Conversely, a candidate not cancelled by $b + D$ contains no such uncorrected change, and coincides with a threshold-reaching state of the retained ideal-detector computation. It thus satisfies the claimed consensus bound. A change at or after b cannot invalidate that saved state. At the retained boundary B, all earlier deleted intervals have already been corrected. Any candidate still pending there is valid; if none is pending, the threshold-reaching state at B creates one. Consequently the actual protocol returns no later than $B + D$, so

$$K \leq 2K' + 3D.$$

Finally $D \leq \sqrt{\chi} \ln n + 1$ for $n \geq 2$. This proves the stated round complexity. Checkpoints use $O(Dd)$ real coordinates per agent plus $O(D)$ scalar scheduling data, with one extra local candidate vector. Each neighbor message carries one timestamp, requiring $O(\log(K + 1))$ bits up to termination; this is bandwidth overhead, not an additional exchange round. The D validation exchanges are included in the bound on K. □

F Proofs for the optimization results

For completeness, we make the logarithmic factors hidden in Theorem 6.1 explicit. Define

$$\Lambda_\epsilon := 1 + \ln_+ \frac{G_0^2}{\mu\epsilon}, \qquad \Gamma_{\epsilon,n} := \Lambda_\epsilon + \ln \kappa + \ln_+ \frac{\sqrt{n}}{\kappa}.$$

The proof below gives

$$N_{\mathrm{grad}} = O(\sqrt{\kappa}\, \Lambda_\epsilon)$$

local-gradient evaluations per agent and replaces $\widetilde{O}(\sqrt{\kappa})$ in (14) by $O(\sqrt{\kappa}\, \Lambda_\epsilon \Gamma_{\epsilon,n})$.

F.1 Complete schedule

Algorithm 3 combines restarted Similar Triangles iterations Nesterov (2004); Gorbunov et al. (2020) with WAVE communication. For $G_0 > 0$, set

$$N_0 := \lceil 8\sqrt{\kappa} \rceil, \qquad S := 1 + \left\lceil \max\left\{0, \log_2 \frac{G_0}{2\sqrt{\mu\epsilon}}\right\}\right\rceil,$$
$$\delta_\star := \left[60000\,\kappa\left(\kappa + (1+\kappa)\max\left\{1, \frac{G_0}{\sqrt{\mu\epsilon}}\right\}\right)\right]^{-1}. \tag{27}$$

Each WAVE call is Algorithm 1: the global round counter k continues across calls, while the piecewise-constant doubling schedule starts anew at length one.

The optimization proof uses the mean-preservation and contraction contract in (7), together with the following call-cost bounds. Let $\mathcal{F}_k$ denote all information available before round k, including the observed change indicators and operators and the current algorithm state. Under Assumption E.2, future change indicators are independent of $\mathcal{F}_k$.

Corollary F.1 (Certified WAVE call)**.** *Under Assumption 2.2 and the corresponding variation assumption (Assumption 2.3, 2.4, or E.2), a fresh call* $(\widehat{\boldsymbol{u}}, k^+) = \mathrm{WAVE}(\boldsymbol{u}, k; \delta)$, $\delta \in (0,1)$, *run with the corresponding scheduler satisfies the respective bound below for absolute constants* $C_{\mathrm{m}}, C_{\mathrm{d}}, C_{\mathrm{r}} > 0$*:*

$$k^+ - k \le C_{\mathrm{m}}\big(\sqrt{\chi} + \min\{\beta\chi^2, \chi\}\big)\ln\frac{e}{\delta}, \quad \textit{metric drift}, \tag{28}$$
$$k^+ - k \le C_{\mathrm{d}}\big(\sqrt{\chi} + \chi/\tau\big)\ln\frac{e}{\delta}, \quad \textit{piecewise-constant, minimum spacing}, \tag{29}$$
$$\mathbb{E}[k^+ - k \mid \mathcal{F}_k] \le C_{\mathrm{r}}\left(\sqrt{\chi} + \frac{\chi}{\bar{\tau}}\right)\ln\frac{e}{\delta}, \quad \textit{memoryless change times}. \tag{30}$$

The last line holds almost surely for every stopping-time start k; both detector-assisted bounds use the ideal external change detector, and neither τ nor $\bar{\tau}$ is supplied to the scheduler.

Algorithm 3: WAVE-STM: restarted STM with WAVE communication

Input: common point x_0; G_0 satisfying $n^{-1}\sum_i \|\nabla f_i(x_0)\|^2 \le G_0^2$; α, μ; target $\epsilon > 0$; a fixed WAVE scheduler

$\boldsymbol{v} \leftarrow \mathbf{1}_n \otimes x_0$; $k \leftarrow 0$;
if $G_0 = 0$ **then return** $(\boldsymbol{v}, k)$;
Set $N_0, S, \delta_\star$ by (27);
for $s = 0, \dots, S-1$ **do**
  if $s \ge 1$ **then** $(\boldsymbol{v}, k) \leftarrow \mathrm{WAVE}(\boldsymbol{v}, k; 1/4)$;
  $\boldsymbol{x}_0 = \boldsymbol{y}_0 = \boldsymbol{z}_0 \leftarrow \boldsymbol{v}$; $A_0 \leftarrow 0$;
  for $\ell = 0, \dots, N_0 - 1$ **do**
    $\omega_{\ell+1} \leftarrow (\ell+2)/(2\alpha)$; $A_{\ell+1} \leftarrow A_\ell + \omega_{\ell+1}$;
    $\boldsymbol{x}_{\ell+1} \leftarrow (A_\ell \boldsymbol{y}_\ell + \omega_{\ell+1}\boldsymbol{z}_\ell)/A_{\ell+1}$;
    $g_{\ell+1}^{(i)} \leftarrow \nabla f_i(x_{\ell+1}^{(i)})$ for all i; stack into $\boldsymbol{g}_{\ell+1}$;
    $(\boldsymbol{z}_{\ell+1}, k) \leftarrow \mathrm{WAVE}(\boldsymbol{z}_\ell - \omega_{\ell+1}\boldsymbol{g}_{\ell+1}, k; \delta_\star)$;
    $\boldsymbol{y}_{\ell+1} \leftarrow (A_\ell \boldsymbol{y}_\ell + \omega_{\ell+1}\boldsymbol{z}_{\ell+1})/A_{\ell+1}$;
  $\boldsymbol{v} \leftarrow \boldsymbol{y}_{N_0}$;
if $\kappa < \sqrt{n}$ **then** $(\boldsymbol{v}, k) \leftarrow \mathrm{WAVE}(\boldsymbol{v}, k; \kappa/\sqrt{n})$;
$\boldsymbol{y}_{\mathrm{out}} \leftarrow \boldsymbol{v}$; **return** $(\boldsymbol{y}_{\mathrm{out}}, k)$;

Corollary F.2 (WAVE-STM under memoryless change times)**.** *Let Assumptions 2.2 and E.2 hold with $\chi \ge 4$, and assume access to the ideal change detector. With Λ_ϵ and $\Gamma_{\epsilon,n}$ defined above, run Algorithm 3 with the piecewise-constant* WAVE *scheduler. It returns an ϵ-solution at every agent almost surely, uses the same gradient bound, and satisfies*

$$\mathbb{E}K = O\big(\sqrt{\kappa}\,\Lambda_\epsilon\Gamma_{\epsilon,n}[\sqrt{\chi} + \chi/\bar{\tau}]\big).$$

F.2 Proof scales and the WAVE interface

A stage means one execution of the inner STM loop in Algorithm 3. Its parameters $N_0, S, \delta_\star$ are specified in (27). Throughout the proof, $\kappa = \alpha/\mu \geq 1$. For $G_0 > 0$, define

$$R := \frac{G_0}{\mu}, \qquad H := (1+\kappa)G_0, \qquad r_\epsilon := \min\left\{R, \sqrt{\frac{\epsilon}{\mu}}\right\}. \tag{31}$$

The initialization argument below proves $\|x_0 - x^\star\| \leq R$ and $(n^{-1}\sum_i \|\nabla f_i(x^\star)\|^2)^{1/2} \leq H$. Set $r_s := 2^{-s}R$ for $s = 0, 1, \ldots$. The main-text parameter formulas give the exact identities

$$\begin{aligned} S &= 1 + \left\lceil \max\left\{0, \log_2\left(R\sqrt{\frac{\mu}{4\epsilon}}\right)\right\}\right\rceil, \\ \delta_\star &= \left[60000\,\kappa\left(\kappa + \frac{H}{\mu r_\epsilon}\right)\right]^{-1}, \end{aligned} \tag{32}$$

because $H/(\mu r_\epsilon) = (1+\kappa)\max\{1, G_0/\sqrt{\mu\epsilon}\}$. The quantities R, H, r_ϵ, r_s are used only in the analysis.

Communication interface. The optimization proof uses exact mean preservation and the contraction in (7), followed by the call-cost bounds of Corollary F.1. Each call is Algorithm 1: the global communication counter continues across calls, and each piecewise-constant call starts a fresh doubling schedule at length one. Use a common spectral envelope $\chi \geq 4$ for the scheduler. A smaller valid envelope may be enlarged to 4, changing the stated complexity bounds only by absolute factors.

Proof of Corollary F.1. The metric estimate follows from Theorem 3.1 after shifting the starting round; the spectral and drift assumptions are preserved by this shift. The minimum-spacing estimate for a call starting anywhere within an interval between changes is Lemma E.1.

For memoryless change times, use the filtration $\mathcal{F}_t$ defined above. For a fixed $\delta \in (0,1)$, start a call at an almost surely finite stopping time K, with an $\mathcal{F}_K$-measurable input. Conditional on $\mathcal{F}_K$, future boundary indicators retain their original i.i.d. law. The certified-progress argument in the proof of Theorem E.3 bounds the duration until the accumulated certificate reaches $\ln(1/\delta)$, uniformly over the initial operator and admissible history-dependent choices of subsequent operators. Applied conditionally, it gives

$$\mathbb{E}[K^+ - K \mid \mathcal{F}_K] \leq C_{\mathrm{r}}\left(\sqrt{\chi} + \frac{\chi}{\bar{\tau}}\right)\ln\frac{e}{\delta} \quad \text{almost surely.}$$

This is a bound on the actual duration of Algorithm 1, and proves (30). Both detector-assisted estimates use the ideal external change detector. □

F.3 STM with additive gradient errors

We use the similar-triangles form of the fast gradient method (STM; see Nesterov (2004); Gorbunov et al. (2020) and, for the inexact-oracle context, Devolder et al. (2014); Devolder (2013); Kornilov et al. (2025)): given a convex α-smooth f on $\mathbb{R}^d$, use estimates $\tilde{g}_{\ell+1} = \nabla f(x_{\ell+1}) + \theta_{\ell+1}$ with $\|\theta_{\ell+1}\| \leq \eta$. The errors may depend on the entire history; no unbiasedness or independence is assumed. The recursion is given below; within this subsection, (z_ℓ) denotes the STM iterate sequence and is unrelated to the Chebyshev scalar $z_0 = z_\chi(0)$.

$$\begin{aligned} &\omega_{\ell+1} = \frac{\ell+2}{2\alpha}, \qquad A_{\ell+1} = A_\ell + \omega_{\ell+1}, \qquad A_0 = 0, \qquad x_0 = y_0 = z_0, \\ &x_{\ell+1} = \frac{A_\ell y_\ell + \omega_{\ell+1} z_\ell}{A_{\ell+1}}, \qquad z_{\ell+1} = z_\ell - \omega_{\ell+1}\tilde{g}_{\ell+1}, \qquad y_{\ell+1} = \frac{A_\ell y_\ell + \omega_{\ell+1} z_{\ell+1}}{A_{\ell+1}}. \end{aligned} \tag{33}$$

Note $A_N = \sum_{\ell=0}^{N-1}\frac{\ell+2}{2\alpha} = \frac{N(N+3)}{4\alpha} \geq \frac{N^2}{4\alpha}$, $\alpha\omega_{\ell+1}^2 \leq A_{\ell+1}$ (equivalent to $(\ell+2)^2 \leq (\ell+1)(\ell+4)$), and $W_N := \sum_{\ell=0}^{N-1}\omega_{\ell+1}^2 \leq \frac{(N+2)^3}{12\alpha^2}$.

Lemma F.3 (STM with additive noise)**.** *Let f be convex and α-smooth with a minimizer $x^\star$, write $f^\star := f(x^\star)$ and $R_0 := \|x_0 - x^\star\|$, and let (33) be run with $\|\theta_{\ell+1}\| \le \eta$ for all $\ell < N$, $N \ge 4$. Then, with $C_N := A_N + \alpha W_N \le \frac{(N+2)^3}{6\alpha}$,*

(i) (*trajectory*) $\max_{\ell \le N} \|z_\ell - x^\star\| \le \bar\rho := R_0 + 2\eta C_N + \eta\sqrt{2W_N}$, *and all points x_ℓ, y_ℓ lie in the ball of radius $\bar\rho$ around $x^\star$. Moreover, this estimate is causal: for every $T \le N$, if the error bounds are known only for $\ell = 0, \dots, T-1$, then*

$$\max_{\ell \le T} \|z_\ell - x^\star\| \le R_0 + 2\eta C_T + \eta\sqrt{2W_T} \le R_0 + 2\eta C_N + \eta\sqrt{2W_N}, \tag{34}$$

where $C_T := A_T + \alpha W_T$.

(ii) (*accuracy*) $f(y_N) - f^\star \le \dfrac{2\alpha R_0^2}{N^2} + 3\,\eta R_0 N + \dfrac{6\,\eta^2 N^4}{\alpha}$.

Proof. Master inequality. Write $\omega = \omega_{\ell+1}$, $A = A_{\ell+1}$. By α-smoothness and $y_{\ell+1} - x_{\ell+1} = -\frac{\omega^2}{A}\tilde g_{\ell+1}$,

$$\begin{aligned} f(y_{\ell+1}) &\le f(x_{\ell+1}) - \frac{\omega^2}{A}\langle \nabla f(x_{\ell+1}),\, \tilde g_{\ell+1}\rangle + \frac{\alpha\omega^4}{2A^2}\|\tilde g_{\ell+1}\|^2 \\ &\le f(x_{\ell+1}) - \frac{\omega^2}{2A}\|\tilde g_{\ell+1}\|^2 + \frac{\omega^2}{A}\langle \theta_{\ell+1},\, \tilde g_{\ell+1}\rangle . \end{aligned}$$

using $\langle \nabla f, \tilde g\rangle = \|\tilde g\|^2 - \langle \theta, \tilde g\rangle$ and $\alpha\omega^2 \le A$. Multiplying by $A = A_\ell + \omega$ and applying convexity twice, $A_\ell f(x_{\ell+1}) \le A_\ell f(y_\ell) + A_\ell\langle \nabla f(x_{\ell+1}),\, x_{\ell+1} - y_\ell\rangle$ and $\omega f(x_{\ell+1}) \le \omega f(u) + \omega\langle \nabla f(x_{\ell+1}),\, x_{\ell+1} - u\rangle$ for any u, and noting $A_\ell(x_{\ell+1} - y_\ell) + \omega(x_{\ell+1} - u) = \omega(z_\ell - u)$ by the definition of $x_{\ell+1}$, we get

$$\begin{aligned} A_{\ell+1} f(y_{\ell+1}) \le A_\ell f(y_\ell) + \omega f(u) + \omega\langle \tilde g_{\ell+1},\, z_\ell - u\rangle \\ - \omega\langle \theta_{\ell+1},\, z_\ell - u\rangle - \frac{\omega^2}{2}\|\tilde g_{\ell+1}\|^2 + \omega^2\langle \theta_{\ell+1},\, \tilde g_{\ell+1}\rangle . \end{aligned}$$

Finally, from $z_{\ell+1} = z_\ell - \omega\tilde g_{\ell+1}$, $\omega\langle \tilde g_{\ell+1},\, z_\ell - u\rangle = \frac12\|z_\ell - u\|^2 - \frac12\|z_{\ell+1} - u\|^2 + \frac{\omega^2}{2}\|\tilde g_{\ell+1}\|^2$; the quadratic terms cancel, and telescoping over $\ell = 0, \dots, N-1$ (with $A_0 = 0$) gives, for every u,

$$\begin{aligned} A_N\big(f(y_N) - f(u)\big) \le \tfrac12\|z_0 - u\|^2 - \tfrac12\|z_N - u\|^2 \\ - \sum_{\ell=0}^{N-1} \omega_{\ell+1}\langle \theta_{\ell+1},\, z_\ell - u\rangle \\ + \sum_{\ell=0}^{N-1} \omega_{\ell+1}^2\langle \theta_{\ell+1},\, \tilde g_{\ell+1}\rangle . \end{aligned} \tag{35}$$

Trajectory bound. An easy induction shows

$$y_\ell \in \operatorname{conv}\{z_0, \dots, z_\ell\}, \qquad x_{\ell+1} \in \operatorname{conv}\{z_0, \dots, z_\ell\}.$$

Hence, with $\rho_T := \max_{\ell \le T}\|z_\ell - x^\star\|$, all queried points lie within ρ_T of $x^\star$. Moreover,

$$\|\nabla f(x_{\ell+1})\| = \|\nabla f(x_{\ell+1}) - \nabla f(x^\star)\| \le \alpha\rho_\ell.$$

Take $u = x^\star$ in (35), drop $f(y_N) - f^\star \ge 0$, and bound $\langle \theta, \tilde g\rangle \le \eta(\alpha\rho + \eta)$ and $|\langle \theta, z_\ell - x^\star\rangle| \le \eta\rho$: for every $T \le N$,

$$\tfrac12\|z_T - x^\star\|^2 \le \tfrac12 R_0^2 + \eta\,(A_T + \alpha W_T)\,\rho_T + \eta^2 W_T \le \tfrac12 R_0^2 + \eta\, C_N\,\rho_N + \eta^2 W_N.$$

Taking T to realize ρ_N (note $\rho_0 = R_0$ is included), we get $\rho_N^2 \le R_0^2 + 2\eta C_N\rho_N + 2\eta^2 W_N$, whence $\rho_N \le \eta C_N + \sqrt{\eta^2 C_N^2 + R_0^2 + 2\eta^2 W_N} \le R_0 + 2\eta C_N + \eta\sqrt{2W_N} = \bar\rho$, which is (i). The same argument stopped at any $T \le N$ uses only the errors $\theta_1, \dots, \theta_T$ and gives (34); no bound on future oracle errors is needed.

Accuracy. Using (35) with $u = x^\star$ once more, now keeping the left side and inserting $\rho_N \le \bar{\rho}$,

$$f(y_N) - f^\star \;\le\; \frac{\frac{1}{2}R_0^2 + \eta C_N \bar{\rho} + \eta^2 W_N}{A_N} \;\le\; \frac{\frac{1}{2}R_0^2 + \eta C_N R_0 + 2\eta^2 C_N^2 + \eta^2 C_N \sqrt{2W_N} + \eta^2 W_N}{A_N}.$$

With $A_N \ge N^2/(4\alpha)$, $C_N \le (N+2)^3/(6\alpha)$, $W_N \le (N+2)^3/(12\alpha^2)$ and $N \ge 4$ (so $N+2 \le \frac{3}{2}N$): $\frac{1}{2}R_0^2/A_N \le 2\alpha R_0^2/N^2$; $\eta C_N R_0/A_N \le \frac{2}{3}\eta R_0 (N+2)^3/N^2 \le \frac{2}{3}(\frac{3}{2})^3 \eta R_0 N \le 3\eta R_0 N$ (indeed $2.25 \le 3$); and the three η^2 terms are bounded by $\big(\frac{2}{9}(\frac{3}{2})^6 + \frac{4}{6\sqrt{6}}(\frac{3}{2})^{9/2} + \frac{1}{3}(\frac{3}{2})^3\big)\eta^2 N^4/\alpha \le (2.6 + 1.7 + 1.2)\,\eta^2 N^4/\alpha \le 6\eta^2 N^4/\alpha$, using $N^{2.5}, N \le N^4$.

□

F.4 Initialization from a single gradient bound

Because f is μ-strongly convex and differentiable,

$$\mu \left\|x_0 - x^\star\right\| \le \left\|\nabla f(x_0)\right\| = \left\|\frac{1}{n}\sum_i \nabla f_i(x_0)\right\| \le \left(\frac{1}{n}\sum_i \left\|\nabla f_i(x_0)\right\|^2\right)^{1/2} \le G_0.$$

Hence $\|x_0 - x^\star\| \le R = G_0/\mu$. Smoothness of the individual objectives further gives

$$\begin{aligned}\left(\frac{1}{n}\sum_i \left\|\nabla f_i(x^\star)\right\|^2\right)^{1/2} &\le \left(\frac{1}{n}\sum_i \left\|\nabla f_i(x_0)\right\|^2\right)^{1/2} + \left(\frac{1}{n}\sum_i \left\|\nabla f_i(x^\star) - \nabla f_i(x_0)\right\|^2\right)^{1/2} \\ &\le G_0 + \alpha \left\|x^\star - x_0\right\| \le (1+\kappa) G_0 = H.\end{aligned}$$

This proves the two derived bounds asserted after (31). If $G_0 = 0$, then $\nabla f(x_0) = 0$, and strong convexity implies $x_0 = x^\star$.

F.5 Mean and disagreement decomposition

For this proof only, define

$$F(\boldsymbol{x}) := \sum_{i=1}^n f_i(x^{(i)}), \qquad \boldsymbol{x}^\star := \mathbf{1}_n \otimes x^\star.$$

For a stacked vector $\boldsymbol{v}$, write

$$\bar{v} := \frac{1}{n}(\mathbf{1}_n^\top \otimes I_d)\boldsymbol{v}.$$

For the inner loop of Algorithm 3, exact mean preservation in (7) gives

$$\bar{z}_{\ell+1} = \bar{z}_\ell - \omega_{\ell+1}\frac{1}{n}\sum_i \nabla f_i(x_{\ell+1}^{(i)}).$$

The averaged $\boldsymbol{x}$- and $\boldsymbol{y}$-updates are the same convex combinations as in (33). Thus the averaged iterates follow STM for f, with gradient estimate

$$\begin{aligned}\widetilde{g}_{\ell+1} &:= \frac{1}{n}\sum_i \nabla f_i(x_{\ell+1}^{(i)}), \\ \left\|\widetilde{g}_{\ell+1} - \nabla f(\bar{x}_{\ell+1})\right\| &\le \frac{\alpha}{\sqrt{n}}\left\|\boldsymbol{\Pi}_\perp \boldsymbol{x}_{\ell+1}\right\|.\end{aligned} \tag{36}$$

The inequality follows by local smoothness, the triangle inequality, and Cauchy–Schwarz applied to $n^{-1}\sum_i \alpha \left\|x_{\ell+1}^{(i)} - \bar{x}_{\ell+1}\right\|$.

Lemma F.4 (Disagreement within one stage)**.** *Fix a stage radius $r > 0$ and an error level $\eta \ge 0$, let $\hat{\rho} := r + 2\eta C_{N_0} + \eta\sqrt{2W_{N_0}}$, and set*

$$M_\ell := \max\left\{\left\|\boldsymbol{\Pi}_\perp \boldsymbol{x}_\ell\right\|, \left\|\boldsymbol{\Pi}_\perp \boldsymbol{y}_\ell\right\|, \left\|\boldsymbol{\Pi}_\perp \boldsymbol{z}_\ell\right\|\right\}, \qquad \Delta := \delta_\star \omega_{\max}\sqrt{n}(\alpha\hat{\rho} + H),$$

where $\omega_{\max} := \omega_{N_0} \le 5\sqrt{\kappa}/\alpha$. Fix $T \le N_0$. Suppose $\|\bar{x}_{\ell+1} - x^\star\| \le \hat{\rho}$ for $0 \le \ell < T$, $M_0 \le \Delta$, and $\delta_\star \le 1/(12\sqrt{\kappa})$. Then $M_\ell \le 2\Delta$ for $0 \le \ell \le T$. In particular, the one-step bound uses only the current averaged query point.

Proof. The $\boldsymbol{x}$- and $\boldsymbol{y}$-updates are convex combinations, while (7) gives

$$\|\mathbf{\Pi}_\perp \boldsymbol{z}_{\ell+1}\| \le \delta_\star \left(M_\ell + \omega_{\ell+1} \left\|\mathbf{\Pi}_\perp \boldsymbol{g}_{\ell+1}\right\|\right).$$

The bound derived above at $x^\star$ and α-smoothness imply

$$\begin{aligned}\left\|\mathbf{\Pi}_\perp \boldsymbol{g}_{\ell+1}\right\| &\le \|\nabla F(\boldsymbol{x}_{\ell+1}) - \nabla F(\boldsymbol{x}^\star)\| + \|\mathbf{\Pi}_\perp \nabla F(\boldsymbol{x}^\star)\| \\ &\le \alpha\left(M_\ell + \sqrt{n}\,\|\bar{x}_{\ell+1} - x^\star\|\right) + \sqrt{n}H.\end{aligned}$$

Consequently,

$$\|\mathbf{\Pi}_\perp \boldsymbol{z}_{\ell+1}\| \le \delta_\star(1 + \omega_{\max}\alpha)M_\ell + \Delta \le \frac{1}{2}M_\ell + \Delta.$$

Together with the convex-combination updates this yields $M_{\ell+1} \le \max\{M_\ell, \frac{1}{2}M_\ell + \Delta\}$. Induction from $M_0 \le \Delta$ proves $M_\ell \le 2\Delta$. □

F.6 Stage induction with one inner accuracy

Proof of Theorem 6.1 and Corollary F.2. Assume $G_0 > 0$; the zero case was handled above. First consider any realization on which all calls terminate and satisfy (7). The argument through the local-output step is deterministic; almost-sure termination in the memoryless model is established in the cost analysis below.

At the start of stage s, we inductively require that the mean of its input lie within r_s of $x^\star$. This holds at $s = 0$ by the initialization bound; the stage reset preserves the mean. For stage s, put $r = r_s$ and

$$\eta_s := \frac{\mu r}{1200\sqrt{\kappa}}, \qquad N_0 = \lceil 8\sqrt{\kappa}\rceil \le 9\sqrt{\kappa}.$$

The trajectory radius in Lemma F.3 satisfies

$$\hat{\rho}_s := r + 2\eta_s C_{N_0} + \eta_s\sqrt{2W_{N_0}} \le 1.4r. \tag{37}$$

Indeed, using $N_0 + 2 \le 11\sqrt{\kappa}$ and $\alpha = \mu\kappa$, the two additional terms are bounded by

$$2\eta_s C_{N_0} \le \frac{1331}{3600}r < 0.37r, \qquad \eta_s\sqrt{2W_{N_0}} \le \frac{11^{3/2}}{1200\sqrt{6}}\kappa^{-3/4}r < 0.02r.$$

Define

$$\omega_{\max} := \omega_{N_0} \le \frac{5\sqrt{\kappa}}{\alpha}, \qquad \Delta_s := \delta_\star\omega_{\max}\sqrt{n}(\alpha\hat{\rho}_s + H).$$

At stage zero the input is consensual, so $M_0 = 0$. If $s \ge 1$, let $M_{\text{in}} := \|\mathbf{\Pi}_\perp \boldsymbol{v}\|$ before the reinitialization call. The preceding stage gives $M_{\text{in}} \le 2\Delta_{s-1}$. Here $r_{s-1} = 2r_s$, $\eta_{s-1} = 2\eta_s$, and N_0 is fixed; hence the definition in (37) gives the exact identity $\hat{\rho}_{s-1} = 2\hat{\rho}_s$. Consequently, $\alpha\hat{\rho}_{s-1} + H \le 2(\alpha\hat{\rho}_s + H)$, and therefore

$$\Delta_{s-1} \le 2\Delta_s.$$

The fixed reinitialization call therefore gives $M_0 \le \frac{1}{4}M_{\text{in}} \le \Delta_s$. No stage-zero reset is needed.

We next close the mean and disagreement estimates simultaneously. The definition (32) gives

$$\delta_\star \le \frac{1}{60000\kappa^2} \le \frac{1}{12\sqrt{\kappa}}.$$

Suppose causally that the mean iterates through inner iteration t lie within $\hat{\rho}_s$ and that $M_j \le 2\Delta_s$ for $j \le t$. The formula for $\boldsymbol{x}_{t+1}$ is a convex combination of $\boldsymbol{y}_t$ and $\boldsymbol{z}_t$. It therefore first gives $\|\bar{x}_{t+1} - x^\star\| \le \hat{\rho}_s$ and $\|\mathbf{\Pi}_\perp \boldsymbol{x}_{t+1}\| \le M_t$, using only already established iterates. Equation (36) then gives

$$\begin{aligned}\|\widetilde{g}_{t+1} - \nabla f(\bar{x}_{t+1})\| &\le 2\delta_\star\omega_{\max}\alpha(\alpha\hat{\rho}_s + H) \\ &\le 10\delta_\star\sqrt{\kappa}(1.4\alpha r + H) \\ &\le 14\delta_\star\sqrt{\kappa}\,\mu r\left(\kappa + \frac{H}{\mu r}\right).\end{aligned}$$

Every executed stage satisfies $r_s \geq r_\epsilon$. Indeed, the final radius is larger than $\sqrt{\epsilon/\mu}$ when more than one stage is used, and the assertion is immediate in the one-stage case. Hence the last display and (32) imply

$$\|\widetilde{g}_{t+1} - \nabla f(\bar{x}_{t+1})\| \leq \frac{14\mu r}{60000\sqrt{\kappa}} < \eta_s.$$

The one-step estimate in Lemma F.4 now bounds $\boldsymbol{z}_{t+1}$, and the convex-combination update bounds $\boldsymbol{y}_{t+1}$; hence it advances the disagreement induction to time $t+1$. Only after that step do we apply the causal trajectory estimate (34) through iteration $t+1$, whose noise bounds have just been established. This advances the mean induction in a causal order. Thus both estimates hold throughout the stage without assuming future error bounds.

Lemma F.3(ii), $N_0 \geq 8\sqrt{\kappa}$, and $N_0 \leq 9\sqrt{\kappa}$ now give

$$f(\bar{y}_{N_0}) - f(x^\star) \leq \left(\frac{1}{32} + \frac{27}{1200} + \frac{6 \cdot 6561}{1200^2}\right)\mu r^2 \leq \frac{\mu r^2}{8}.$$

Strong convexity implies $\|\bar{y}_{N_0} - x^\star\| \leq r/2$, so the next stage starts within $r_{s+1} = r_s/2$.

Let $\boldsymbol{v}_{\mathrm{fin}}$ be the value of $\boldsymbol{v}$ after the last stage and before the optional final WAVE call. Let $t_\epsilon := \sqrt{\epsilon/\mu}$. The definition of S ensures $r_{S-1} \leq 2t_\epsilon$, and the preceding stage estimate therefore gives

$$f(\bar{v}_{\mathrm{fin}}) - f(x^\star) \leq \frac{\mu r_{S-1}^2}{8} \leq \frac{\epsilon}{2}. \tag{38}$$

It also gives $r_{S-1} \leq 2r_\epsilon$, including the one-stage case.

Final local-output step. Before the final call,

$$\|\boldsymbol{\Pi}_\perp \boldsymbol{v}_{\mathrm{fin}}\| \leq B_{\mathrm{out}} := 2\delta_\star \omega_{\max}\sqrt{n}(1.4\alpha r_{S-1} + H).$$

Since $r_{S-1} \leq 2r_\epsilon$,

$$B_{\mathrm{out}} \leq \frac{28\sqrt{n}}{60000\,\kappa^{3/2}}\, r_\epsilon.$$

Write $\delta_{\mathrm{loc}} := \min\{1, \kappa/\sqrt{n}\}$ for the effective contraction of the final step, including the case where it is skipped. Then the returned vector $\boldsymbol{y}_{\mathrm{out}}$ satisfies

$$\|\boldsymbol{\Pi}_\perp \boldsymbol{y}_{\mathrm{out}}\| \leq \frac{28}{60000\sqrt{\kappa}} r_\epsilon < \rho_{\mathrm{loc}} := \frac{1}{3}\sqrt{\frac{\epsilon}{\alpha}}.$$

When $\delta_{\mathrm{loc}} = 1$, the inequality already holds and the call is skipped. Otherwise the final WAVE call establishes it while preserving the mean. Thus $\bar{y}_{\mathrm{out}} = \bar{v}_{\mathrm{fin}}$ and $\left\|y_{\mathrm{out}}^{(i)} - \bar{y}_{\mathrm{out}}\right\| \leq \rho_{\mathrm{loc}}$ for every agent. From (38) and smooth convexity,

$$\|\nabla f(\bar{y}_{\mathrm{out}})\| \leq \sqrt{2\alpha(f(\bar{y}_{\mathrm{out}}) - f(x^\star))} \leq \sqrt{\alpha\epsilon}.$$

Therefore

$$\begin{aligned} f(y_{\mathrm{out}}^{(i)}) - f(x^\star) &\leq \frac{\epsilon}{2} + \sqrt{\alpha\epsilon}\,\rho_{\mathrm{loc}} + \frac{\alpha}{2}\rho_{\mathrm{loc}}^2 \\ &\leq \left(\frac{1}{2} + \frac{1}{3} + \frac{1}{18}\right)\epsilon < \epsilon. \end{aligned}$$

This proves the local-output guarantees in Theorem 6.1 and Corollary F.2.

Gradient and communication counts. With $\Lambda_\epsilon, \Gamma_{\epsilon,n}$ as in Theorem 6.1, the stage schedule gives $S = O(\Lambda_\epsilon)$. The exact gradient count per agent is $N_0 S$, proving the stated gradient bound. From (31),

$$\frac{H}{\mu r_\epsilon} = (1+\kappa)\max\left\{1, \frac{G_0}{\sqrt{\mu\epsilon}}\right\}.$$

Writing $b = \max\{1, G_0/\sqrt{\mu\epsilon}\}$, we have $\kappa + (1+\kappa)b \leq 3\kappa b$. Hence

$$\ln\frac{e}{\delta_\star} \leq 1 + \ln(180000) + 2\ln\kappa + \frac{1}{2}\ln_+\frac{G_0^2}{\mu\epsilon} = O(\Lambda_\epsilon + \ln\kappa).$$

There are $N_0 S$ calls with factor $\delta_\star$, $S-1$ resets with factor $1/4$, and one final call only if $\kappa < \sqrt{n}$. Thus the sum of the call-cost logarithms is bounded by

$$O\left(\sqrt{\kappa}\,\Lambda_\epsilon(\Lambda_\epsilon + \ln\kappa) + 1 + \ln_+ \frac{\sqrt{n}}{\kappa}\right) = O(\sqrt{\kappa}\,\Lambda_\epsilon\Gamma_{\epsilon,n}).$$

Corollary F.1 gives the two deterministic communication bounds in (14).

For the memoryless model, fix the objectives and the supplied inputs. The number of scheduled calls is the deterministic integer $N_{\mathrm{call}} = N_0 S + (S-1) + \mathbf{1}_{\{\kappa<\sqrt{n}\}}$; their requested factors δ_j are also deterministic. Let $K_0 = 0$ and let K_j be the next unused round at the start of call j, so $K_{N_{\mathrm{call}}}$ is the total communication count. The scheduler makes its decisions from the observed history. Consequently, each call endpoint is a stopping time. Inductively, the conditional finite-duration bound (30) ensures that every K_j is almost surely finite. The input vector of call j is $\mathcal{F}_{K_j}$-measurable, though it need not be deterministic. For $j = 0, \dots, N_{\mathrm{call}} - 1$, the tower property gives

$$\mathbb{E}(K_{j+1} - K_j) = \mathbb{E}\big[\mathbb{E}(K_{j+1} - K_j \mid \mathcal{F}_{K_j})\big] \le C_{\mathrm{r}} \left(\sqrt{\chi} + \frac{\chi}{\bar{\tau}}\right) \ln \frac{e}{\delta_j}.$$

Summing proves the expected communication bound in Corollary F.2. All N_{call} calls terminate and satisfy their contraction guarantees simultaneously almost surely. The preceding pathwise proof therefore establishes the local-output guarantee almost surely as well. □

F.7 Accelerated gradient tracking with WAVE

When every local objective is strongly convex, WAVE can also replace the mixing operation in accelerated gradient tracking (Li & Lin, 2024). We first verify the properties required of the resulting mixing maps.

Lemma F.5 (A WAVE call is an average-preserving contraction)**.** *Fix the operator sequence, scheduler, starting round, and, in the piecewise-constant model, change reports. A* WAVE *call is a linear map $A \otimes I_d$ with*

$$A\mathbf{1}_n = \mathbf{1}_n, \qquad \mathbf{1}_n^\top A = \mathbf{1}_n^\top, \qquad A = \Pi + \Pi_\perp A \Pi_\perp, \qquad \|A - \Pi\| \le \delta, \qquad \|A\| \le 1.$$

Proof. The scheduler and the stopping rule depend on the network sequence and change reports, but not on the input vector. Every recurrence step is linear and acts identically on all vector coordinates, so the full call has the form $A \otimes I_d$. Every window fixes consensus vectors and preserves their average; hence $A\Pi = \Pi A = \Pi$ and therefore $A = \Pi + \Pi_\perp A \Pi_\perp$. The call contract (7) gives $\|A - \Pi\| = \|\Pi_\perp A \Pi_\perp\| \le \delta$. The two displayed blocks act on orthogonal subspaces, and thus $\|A\| = \max\{1, \|A - \Pi\|\} = 1$. □

Exact packed implementation. Let A_k denote the map of the kth WAVE call, with target $\delta = e^{-1}$. Use the initialization of Li & Lin (2024, Algorithm 1), replacing W^0 by A_0; the common initialization makes this first mixing operation immaterial. For $k \ge 1$, given $\boldsymbol{x}^k, \boldsymbol{z}^k, \boldsymbol{s}^{k-1}$ and $\boldsymbol{y}^{k-1}$, compute

$$\boldsymbol{y}^k = \theta \boldsymbol{z}^k + (1-\theta)\boldsymbol{x}^k, \qquad \boldsymbol{g}^k = \big(\nabla f_i(y_i^k)\big)_{i=1}^n.$$

Each agent packs its three local d-vectors

$$\left(s_i^{k-1},\ \frac{\mu\eta}{\theta} y_i^k + z_i^k,\ x_i^k\right)$$

into one $3d$-vector and runs one WAVE call. After unpacking, write

$$\widetilde{\boldsymbol{s}}^k = A_k \boldsymbol{s}^{k-1}, \qquad \widetilde{\boldsymbol{z}}^k = A_k\left(\frac{\mu\eta}{\theta}\boldsymbol{y}^k + \boldsymbol{z}^k\right), \qquad \widetilde{\boldsymbol{x}}^k = A_k \boldsymbol{x}^k,$$

where A_k abbreviates $A_k \otimes I_d$ on stacked d-vectors. Then set

$$\begin{aligned} \boldsymbol{s}^k &= \widetilde{\boldsymbol{s}}^k + \boldsymbol{g}^k - \boldsymbol{g}^{k-1}, \\ \boldsymbol{z}^{k+1} &= \frac{\widetilde{\boldsymbol{z}}^k - (\eta/\theta)\boldsymbol{s}^k}{1 + \mu\eta/\theta}, \\ \boldsymbol{x}^{k+1} &= \theta \boldsymbol{z}^{k+1} + (1-\theta)\widetilde{\boldsymbol{x}}^k. \end{aligned}$$

These are exactly equations (13a)–(13d) of Li & Lin (2024) with the same virtual matrix A_k in all three mixing locations. The three vectors are communicated in parallel, so this is one WAVE call per outer iteration, with a constant-factor increase in message size.

Corollary F.6 (Accelerated gradient tracking with WAVE)**.** *Let Assumption 2.2 and either Assumption 2.3 or 2.4 hold with $\chi \geq 4$, and let every f_i be μ-strongly convex and α-smooth. Run accelerated gradient tracking (Li & Lin, 2024, eqs. (13a)–(13d), Theorem 3) from a common point, with $\eta = (1-e^{-1})^3/(4244\alpha)$ and $\theta = \sqrt{\mu\eta}/2$. Replace its mixing matrix by one* WAVE *call with target e^{-1}, applied jointly to the three vectors specified above. There is a constant $C_0 > 0$, depending on the initialization and first virtual iterate, such that, writing $H_\epsilon := 1 + \ln_+(nC_0/\epsilon)$, every agent obtains an ϵ-solution using $O(\sqrt{\kappa}H_\epsilon)$ gradient evaluations. With the ideal detector in the piecewise-constant model, the number of communication rounds satisfies*

$$K = \begin{cases} O\big(\sqrt{\kappa}[\sqrt{\chi} + \min\{\beta\chi^2, \chi\}]H_\epsilon\big)\,, & \textit{metric drift}, \\ O\big(\sqrt{\kappa}[\sqrt{\chi} + \chi/\tau]H_\epsilon\big)\,, & \textit{piecewise-constant changes}, \end{cases}$$

Fixed networks belong to both models. Thus the fixed-network lower bound of Scaman et al. (2017) applies to their gossip oracle model, and Corollary F.6 matches its $\sqrt{\kappa\chi}$ dependence and single accuracy logarithm whenever $\beta \leq (3\chi^{3/2})^{-1}$ or $\tau \geq \sqrt{\chi}$.

Proof of Corollary F.6. The strong-convexity result used here is Theorem 3 of Li & Lin (2024). Although a virtual WAVE map may be dense and have signed entries, their proof remains valid for the sequence $\{A_k\}$. Indeed, column stochasticity is used to obtain their averaged recursion (15), while double stochasticity and their estimate (9) are used to control disagreement. Lemma F.5 gives the corresponding identities and the stronger one-step estimate

$$\|((A_k - \Pi) \otimes I_d)\boldsymbol{v}\| \leq e^{-1}\,\|\boldsymbol{\Pi}_\perp \boldsymbol{v}\|\,.$$

The only further mixing estimate in their general γ-step proof is their (10) for products of fewer than γ matrices. Here $\gamma = 1$, so it reduces to the identity map. Entrywise positivity and one-hop sparsity are therefore not used in the convergence inequalities; they only describe how one multiplication by W^k is implemented. In our construction, A_k is instead implemented by the legal neighbor-to-neighbor rounds of one WAVE call. Thus their Theorem 3 applies algebraically with $\gamma = 1$, $\sigma_1 = e^{-1}$, the displayed η, and $\theta = \sqrt{\mu\eta}/2$.

Let C_0 be the constant C in that theorem; for $\gamma = 1$ it depends only on the initialization and the first virtual iterate. For every integer $T \geq 1$, the theorem gives

$$f(\bar{x}^{T+1}) - f(x^\star) \leq (1-\theta)^{T+1} C_0, \tag{39}$$

$$\frac{1}{n}\left\|\boldsymbol{\Pi}_\perp \boldsymbol{x}^T\right\|^2 \leq (1-\theta)^{T+1}\frac{4C_0}{\alpha}. \tag{40}$$

The two estimates refer to different time indices. Apply (39) with T and (40) with $T+1$, run one additional outer iteration, and return the stored local states x_i^{T+1}. Since f is convex and α-smooth,

$$f(x_i) - f(x^\star) \leq 2\big(f(\bar{x}) - f(x^\star)\big) + \alpha\,\|x_i - \bar{x}\|^2\,.$$

Together with $\left\|x_i^{T+1} - \bar{x}^{T+1}\right\|^2 \leq \left\|\boldsymbol{\Pi}_\perp \boldsymbol{x}^{T+1}\right\|^2$, this yields

$$f(x_i^{T+1}) - f(x^\star) \leq (2+4n)C_0(1-\theta)^{T+1} \qquad (i = 1, \ldots, n).$$

Hence $T = O(\theta^{-1}[1 + \ln_+(nC_0/\epsilon)]) = O(\sqrt{\kappa}[1 + \ln_+(nC_0/\epsilon)])$ outer iterations suffice. Each iteration evaluates one local gradient and makes one WAVE call. For $\delta = e^{-1}$, Corollary F.1 bounds the call cost by $2C_{\mathrm{m}}(\sqrt{\chi} + \min\{\beta\chi^2, \chi\})$ or $2C_{\mathrm{d}}(\sqrt{\chi} + \chi/\tau)$, respectively. Multiplication by the number of outer iterations proves the two communication bounds. □

G Supplementary examples and interpretation

We give a stationary-momentum instability example and relate its perturbation to relative inexactness. These observations are independent of the positive complexity proofs.

G.1 STATIONARY-MOMENTUM RESONANCE EXAMPLE

A periodic modulation of a quadratic coefficient can amplify the auxiliary state in a momentum recurrence, much as in a parametrically excited oscillator. The example below uses the stationary heavy-ball recurrence obtained from a generic family of limiting Chebyshev coefficients. This is not the recurrence executed by WAVE, whose coefficients are fixed by $[\chi^{-1}, 1]$ and whose momentum is restarted after every finite window. For this subsection, define

$$z_a(\lambda) := \frac{1 + a - 2\lambda}{1 - a}, \qquad q_a := \frac{1 - \sqrt{a}}{1 + \sqrt{a}}.$$

Then $a_t \to 2q_a$, $c_t \to q_a^2$, and

$$2q_a z_a(L) = (1 + q_a^2)I - \frac{4}{(1 + \sqrt{a})^2} L,$$

so the stationary limit is

$$e_{t+1} = \big((1 + \gamma)I - hL_t\big)e_t - \gamma e_{t-1}, \qquad h = \frac{4}{(1 + \sqrt{a})^2}, \quad \gamma = q_a^2. \tag{41}$$

For a static matrix with spectrum in $[a, 1]$, this recurrence has the usual asymptotic factor q_a.

Proposition G.1 (Resonant divergence of the stationary tuning)**.** *Let $n \geq 3$, $a \in (0, 1/16]$, and $\chi^{-1} \leq a/8$. Put*

$$\lambda_c := \frac{1 + \gamma}{h} = \frac{1 + a}{2}, \qquad A := 3\sqrt{a}.$$

Let $Q \in \mathbb{R}^{n \times (n-1)}$ have orthonormal columns spanning $\mathbf{1}^\perp$, and denote its second column by v_{osc}. For each phase $s \in \{-1, +1\}$, define the commuting family

$$L_t = Q\ \mathrm{diag}\big(\chi^{-1}, \lambda_t, \lambda_3, \ldots, \lambda_{n-1}\big)Q^\top, \qquad \lambda_t = \lambda_c\big(1 + s(-1)^t A\big),$$

where the remaining eigenvalues $\lambda_j \in [\chi^{-1}, 1]$ are fixed.

For either phase, the family is admissible and satisfies

$$\|L_{t+1} - L_t\| = 2A\lambda_c = 3(1 + a)\sqrt{a} \leq 4\sqrt{a}.$$

Consider the initialization $e_{-1} = 0$ and let

$$e_0^{\mathrm{osc}} := \langle v_{\mathrm{osc}}, e_0 \rangle.$$

If $e_0^{\mathrm{osc}} \neq 0$, then the choice $s = +1$ makes (41) divergent. More precisely, there exists a constant $C_a > 0$, depending only on a, such that

$$\|e_{2t}\| \geq C_a |e_0^{\mathrm{osc}}| (1 + \sqrt{a})^t \qquad \textit{for all sufficiently large } t.$$

If $A > 0$ is regarded as the amplitude parameter of the corresponding scalar period-two recurrence, its exact instability threshold is

$$A > \frac{1 - \gamma}{1 + \gamma} = \frac{2\sqrt{a}}{1 + a}.$$

Equivalently, in terms of the operator-drift amplitude $\beta := 2A\lambda_c$, the threshold is

$$\beta > 2\sqrt{a}.$$

The parameter a belongs only to this stationary comparator; the condition $\chi^{-1} \leq a/8$ separates its tuning from the full-spectrum coefficients on $[\chi^{-1}, 1]$ used by WAVE. The proposition motivates finite restarts. Its proof follows the comparison below.

Connection with inexact first-order methods. On a disagreement coordinate, replacing the reference operator $\boldsymbol{L}_k$ by $\boldsymbol{L}_k + E_t$ changes the gradient from $\boldsymbol{L}_k e$ to $(\boldsymbol{L}_k + E_t)e$. The resulting structured error is $E_t e$, with

$$\|E_t e\| \leq \|E_t\|\,\|e\|\,.$$

Equation (16) keeps this operator structure explicit. A black-box inexact-oracle reduction instead treats the same term as arbitrary gradient noise and can therefore give a weaker robustness certificate. This relative-error interpretation connects the analysis to Devolder et al. (2014); Devolder (2013); Vasin et al. (2023); Kornilov et al. (2025).

G.2 Proof of Proposition G.1

The columns of Q form a fixed orthonormal basis of $\mathbf{1}^\perp$, so (41) decouples into independent scalar recurrences. Let

$$e_t^{\mathrm{osc}} := \langle v_{\mathrm{osc}}, e_t \rangle$$

denote the coordinate in the oscillating eigendirection. With

$$\lambda_t = \lambda_c\big(1 + s(-1)^t A\big), \qquad h\lambda_c = 1 + \gamma,$$

the corresponding scalar recurrence is

$$e_{t+1}^{\mathrm{osc}} = -s(-1)^t(1+\gamma)A\, e_t^{\mathrm{osc}} - \gamma e_{t-1}^{\mathrm{osc}}. \tag{42}$$

For admissibility,

$$\lambda_c = \frac{1+a}{2} \in \left[\frac{1}{2}, \frac{17}{32}\right], \qquad A = 3\sqrt{a} \le \frac{3}{4}.$$

Hence

$$\lambda_c(1+A) \le \frac{17}{32} \cdot \frac{7}{4} = \frac{119}{128} < 1,$$

whereas

$$\lambda_c(1-A) \ge \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} \ge \chi^{-1},$$

because $\chi^{-1} \le a/8 \le 1/128$. Thus $\lambda_t \in [\chi^{-1}, 1]$ for either phase. Moreover, since the eigenbasis is fixed and only λ_t varies,

$$\|L_{t+1} - L_t\| = |\lambda_{t+1} - \lambda_t| = 2A\lambda_c = 3(1+a)\sqrt{a} \le 4\sqrt{a}.$$

Transfer matrix. Set

$$r_{\mathrm{HB}} := (1+\gamma)A.$$

Fix the phase $s = +1$ and set

$$M_- := \begin{pmatrix} -r_{\mathrm{HB}} & -\gamma \\ 1 & 0 \end{pmatrix}, \qquad M_+ := \begin{pmatrix} r_{\mathrm{HB}} & -\gamma \\ 1 & 0 \end{pmatrix}.$$

These are the even- and odd-step transfer matrices of (42), respectively. Consequently,

$$\begin{pmatrix} e_{2k+2}^{\mathrm{osc}} \\ e_{2k+1}^{\mathrm{osc}} \end{pmatrix} = T_2 \begin{pmatrix} e_{2k}^{\mathrm{osc}} \\ e_{2k-1}^{\mathrm{osc}} \end{pmatrix}, \qquad T_2 := M_+ M_-.$$

Explicitly,

$$T_2 = \begin{pmatrix} -r_{\mathrm{HB}}^2 - \gamma & -r_{\mathrm{HB}}\gamma \\ -r_{\mathrm{HB}} & -\gamma \end{pmatrix},$$
$$\det T_2 = \gamma^2, \qquad \operatorname{tr} T_2 = -(r_{\mathrm{HB}}^2 + 2\gamma).$$

The eigenvalues of T_2 are

$$\lambda_{\mathrm{st}} := -\frac{1}{2}\left(r_{\mathrm{HB}}^2 + 2\gamma - r_{\mathrm{HB}}\sqrt{r_{\mathrm{HB}}^2 + 4\gamma}\right), \qquad \lambda_{\mathrm{un}} := -\frac{1}{2}\left(r_{\mathrm{HB}}^2 + 2\gamma + r_{\mathrm{HB}}\sqrt{r_{\mathrm{HB}}^2 + 4\gamma}\right).$$

They are real and negative because $r_{\mathrm{HB}}, \gamma > 0$ and $(r_{\mathrm{HB}}^2 + 2\gamma)^2 - r_{\mathrm{HB}}^2(r_{\mathrm{HB}}^2 + 4\gamma) = 4\gamma^2 > 0$. The eigenvalue of larger modulus is λ_{un}, with

$$|\lambda_{\mathrm{un}}| = \frac{1}{2}\left(r_{\mathrm{HB}}^2 + 2\gamma + r_{\mathrm{HB}}\sqrt{r_{\mathrm{HB}}^2 + 4\gamma}\right). \tag{43}$$

Exact instability threshold. By (43),

$$|\lambda_{\mathrm{un}}| > 1$$

if and only if

$$r_{\mathrm{HB}}\sqrt{r_{\mathrm{HB}}^2 + 4\gamma} > 2(1-\gamma) - r_{\mathrm{HB}}^2.$$

If the right-hand side is positive, squaring both sides gives

$$r_{\mathrm{HB}}^2(r_{\mathrm{HB}}^2+4\gamma) > \big(2(1-\gamma)-r_{\mathrm{HB}}^2\big)^2,$$

which simplifies to

$$r_{\mathrm{HB}}^2 > (1-\gamma)^2.$$

Since $r_{\mathrm{HB}} > 0$, this is equivalent to $r_{\mathrm{HB}} > 1-\gamma$. If the right-hand side is nonpositive, the displayed inequality is automatic and again implies $r_{\mathrm{HB}} > 1-\gamma$. Therefore

$$|\lambda_{\mathrm{un}}| > 1 \quad \Longleftrightarrow \quad r_{\mathrm{HB}} > 1-\gamma.$$

At $r_{\mathrm{HB}} = 1-\gamma$, one has $|\lambda_{\mathrm{un}}| = 1$ exactly.

Since $r_{\mathrm{HB}} = (1+\gamma)A$,

$$|\lambda_{\mathrm{un}}| > 1 \quad \Longleftrightarrow \quad A > \frac{1-\gamma}{1+\gamma}.$$

Using

$$1-\gamma = \frac{4\sqrt{a}}{(1+\sqrt{a})^2}, \qquad \frac{1-\gamma}{1+\gamma} = \frac{2\sqrt{a}}{1+a},$$

we obtain the exact threshold

$$A > \frac{2\sqrt{a}}{1+a}.$$

In terms of the drift amplitude $\beta = 2A\lambda_c$, this is equivalent to

$$\beta > 2\sqrt{a}.$$

Our choice $A = 3\sqrt{a}$ lies strictly above this threshold.

Quantitative rate. We claim that

$$|\lambda_{\mathrm{un}}| \geq 1+\sqrt{a} \qquad \text{for all } a \in \left(0, \frac{1}{16}\right].$$

Since

$$\sqrt{r_{\mathrm{HB}}^2+4\gamma} \geq 2\sqrt{\gamma} = 2q_a,$$

equation (43) gives

$$|\lambda_{\mathrm{un}}| \geq \gamma + r_{\mathrm{HB}}q_a + \frac{r_{\mathrm{HB}}^2}{2}.$$

Using

$$r_{\mathrm{HB}} = 3\sqrt{a}(1+\gamma) = 3\sqrt{a}(1+q_a^2)$$

and

$$1-\gamma = \frac{4\sqrt{a}}{(1+\sqrt{a})^2},$$

the desired inequality reduces, after dividing by $\sqrt{a}$, to

$$3q_a(1+q_a^2) + \frac{9}{2}\sqrt{a}\,(1+q_a^2)^2 \geq 1 + \frac{4}{(1+\sqrt{a})^2}.$$

Substitute

$$t := \sqrt{a} \in \left(0, \frac{1}{4}\right], \qquad q_a = \frac{1-t}{1+t}, \qquad 1+q_a^2 = \frac{2(1+t^2)}{(1+t)^2}.$$

Multiplying by $(1+t)^4 > 0$, the required inequality becomes

$$\begin{aligned} P(t) &:= 6(1-t^4) + 18t(1+t^2)^2 - (1+t)^4 - 4(1+t)^2 \\ &= 1 + 6t - 10t^2 + 32t^3 - 7t^4 + 18t^5 \geq 0. \end{aligned}$$

For $t \in [0, 1/4]$,

$$10t^2 \leq \frac{10}{4}t, \qquad 7t^4 \leq \frac{7}{4}t^3 \leq 32t^3.$$

Consequently,

$$P(t) \geq 1 + 3.5t > 0,$$

which proves $|\lambda_{\mathrm{un}}| \geq 1 + \sqrt{a}$.

Divergence of the iterate. Since

$$\lambda_{\mathrm{st}}\lambda_{\mathrm{un}} = \det T_2 = \gamma^2,$$

we have

$$|\lambda_{\mathrm{st}}| = \frac{\gamma^2}{|\lambda_{\mathrm{un}}|} < 1 < |\lambda_{\mathrm{un}}|.$$

Thus T_2 has one-dimensional stable and unstable eigenlines.

Under the initialization $e_{-1} = 0$,

$$\begin{pmatrix} e_{2t}^{\mathrm{osc}} \\ e_{2t-1}^{\mathrm{osc}} \end{pmatrix} = e_0^{\mathrm{osc}} T_2^t \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Moreover,

$$T_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -r_{\mathrm{HB}}^2 - \gamma \\ -r_{\mathrm{HB}} \end{pmatrix},$$

which is not collinear with $(1, 0)^\top$ because $r_{\mathrm{HB}} > 0$. Hence $(1, 0)^\top$ does not belong to the stable eigenline of T_2.

Let v_{un} and v_{st} be eigenvectors corresponding to λ_{un} and λ_{st}, respectively. We may therefore write

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = a_{\mathrm{un}} v_{\mathrm{un}} + a_{\mathrm{st}} v_{\mathrm{st}}, \qquad a_{\mathrm{un}} \neq 0.$$

The first coordinate $(v_{\mathrm{un}})_1$ is also nonzero. Indeed, if $(v_{\mathrm{un}})_1 = 0$, then the first row of the eigenvalue equation would give

$$-r_{\mathrm{HB}}\gamma(v_{\mathrm{un}})_2 = 0,$$

forcing $v_{\mathrm{un}} = 0$, a contradiction.

It follows that

$$e_{2t}^{\mathrm{osc}} = e_0^{\mathrm{osc}} \left(a_{\mathrm{un}}(v_{\mathrm{un}})_1 \lambda_{\mathrm{un}}^t + a_{\mathrm{st}}(v_{\mathrm{st}})_1 \lambda_{\mathrm{st}}^t \right).$$

Since

$$\frac{|\lambda_{\mathrm{st}}|}{|\lambda_{\mathrm{un}}|} < 1,$$

the unstable term dominates for all sufficiently large t. Therefore there exists a constant $C_a > 0$, depending only on a, such that

$$|e_{2t}^{\mathrm{osc}}| \geq C_a |e_0^{\mathrm{osc}}| |\lambda_{\mathrm{un}}|^t \geq C_a |e_0^{\mathrm{osc}}| (1 + \sqrt{a})^t$$

for all sufficiently large t.

Finally, since e_{2t}^{osc} is one orthonormal coordinate of e_{2t},

$$\|e_{2t}\| \geq |e_{2t}^{\mathrm{osc}}|.$$

Thus

$$\|e_{2t}\| \geq C_a |e_0^{\mathrm{osc}}| (1 + \sqrt{a})^t.$$

□

H SUPPLEMENTARY NUMERICAL EXPERIMENTS

The supplementary material contains the code, fixed seeds, and raw data needed to reproduce every reported result.

H.1 Protocol and full statistics for Figure 1

Methods and performance measure. We compare WAVE with the classical Chebyshev semi-iteration (Golub & Varga, 1961), the time-varying recurrence of Montijano et al. (2013), minimax Richardson iteration, and gossip. Classical Chebyshev is run continuously without restarts and is included to isolate the effect of the finite windows and restarts in WAVE. All methods process the same normalized operators and spectral bounds. We evaluate the exact operator norm r_k, rather than sampling an initial state. The threshold 10^{-6} is used only to report first-passage rounds and is not supplied to any method.

Metric-drift stress test. Panel (a) uses the six-dimensional invariant block of the weighted-graph family in Appendix H.4. We set $s = 400$, so $\chi = s^2 = 160000$, and $\beta = 120/[s(4s^2 + 9)]$. Hence $\beta\chi^{3/2} \approx 30$, and the window rule $m = \lfloor \min\{\sqrt{\chi}, (3\beta\chi)^{-1}\} \rfloor$ gives $m = 4$. The methods run for $1.25 \cdot 10^6$ rounds, and r_k is recorded every 2500 rounds. WAVE reaches 10^{-6} after 277500 rounds, whereas Montijano and Richardson require 1107500 rounds and gossip does not reach the target. Classical Chebyshev leaves the plotted range within 50000 rounds.

Piecewise-constant graph pairs. For each of 32 fixed test seeds, we independently sample two connected random geometric graphs on $n = 100$ vertices in the unit square, with connection radius $\sqrt{1.25\ln(n)/(\pi n)}$. Both Laplacians are divided by the larger of their two largest eigenvalues, and χ is computed from the smallest positive eigenvalue of the normalized pair. Across the test set, $\chi \in [339, 2873]$, with median 905. The operator alternates between the two graphs every $\mathrm{round}(\sqrt{\chi})$ rounds. Each change is reported to WAVE immediately, but the interval between changes is not supplied to the algorithm. Curves show the pointwise median of r_k over the 32 pairs; the reported round counts are medians of the per-instance first-passage times. The Montijano width parameter $q_{\mathrm{M}} = 0.99$ is selected on validation instances disjoint from this test set and then held fixed.

The intersection graph of every sampled pair is disconnected: it has only 5 to 24 edges on 100 vertices and cannot contain a connected spanning subgraph. The assumptions of the non-recoverable-links method (NRL) of Metelev et al. (2023) are therefore not satisfied on this benchmark.

Lemma H.1 (Exact evaluation of the consensus gap)**.** *Let Φ_k be the linear map generated after k rounds by any method above. Since these methods preserve averages and consensus states, $\mathbf{\Pi}\Phi_k = \Phi_k\mathbf{\Pi} = \mathbf{\Pi}$. Therefore, a run initialized with $X_0 = \mathbf{\Pi}_\perp$ satisfies*

$$X_k = \mathbf{\Pi}_\perp \Phi_k \mathbf{\Pi}_\perp, \qquad \|X_k\| = r_k.$$

Proof. Average preservation and invariance of consensus give $\mathbf{1}_n^\top \Phi_k = \mathbf{1}_n^\top$ and $\Phi_k \mathbf{1}_n = \mathbf{1}_n$, which are equivalent to $\mathbf{\Pi}\Phi_k = \Phi_k\mathbf{\Pi} = \mathbf{\Pi}$. Hence $X_k = \Phi_k\mathbf{\Pi}_\perp = \mathbf{\Pi}_\perp\Phi_k\mathbf{\Pi}_\perp$. For every $\boldsymbol{e}_0 \in \mathbf{1}_n^\perp$, we have $\boldsymbol{e}_k = X_k\boldsymbol{e}_0$, and X_k vanishes on $\mathrm{span}\{\mathbf{1}_n\}$. Its spectral norm is therefore the worst-case ratio defining r_k. □

Table 2: Paired results for the piecewise-constant experiment in Figure 1(b). The Chebyshev median is conditional on its 12 successful runs. Ratios are baseline rounds divided by WAVE rounds on successful pairs; brackets give the interquartile range. A missing first-passage time counts as later than WAVE in the last column.

Method	Successes	Median rounds	Median ratio	WAVE earlier
WAVE	32/32	180	–	–
Chebyshev, no restarts	12/32	516	3.42 [2.52, 3.56]	32/32
Montijano	32/32	344	1.98 [1.75, 2.18]	32/32
Richardson	32/32	344	1.98 [1.77, 2.18]	32/32
Gossip	32/32	422	2.48 [2.35, 2.66]	32/32

H.2 Sensitivity to the interval between changes

We vary the graph density, the number of operators in the sequence, and the interval p between changes. A random geometric graph on $n = 100$ nodes uses radius $\sqrt{c_{\mathrm{r}}\ln(n)/(\pi n)}$, with $c_{\mathrm{r}} =$

2 for RGG and $c_r = 1.25$ for sparse RGG. Each sequence contains two or four independently sampled connected graphs, and the operator changes every $p = \max\{1, \mathrm{round}(c\sqrt{\chi})\}$ rounds for $c \in \{1/8, 1/4, 1/2, 3/4, 1, 3/2, 2, 3\}$.

Each configuration uses 32 fixed test seeds. Montijano's parameter is selected from $\{0.5, 0.7, 0.9, 0.99\}$ on six disjoint validation sequences. Classical Chebyshev is never restarted. The round budget is $\min\{20\chi, 60000\}$, and the target is $r_k \le 10^{-8}$. WAVE has the smallest median in 30 of the 32 settings, ties once, and is within 1% of the smallest median in the remaining setting. All four switching-stable methods reach the target in every run, whereas Chebyshev without restarts frequently fails when changes are rapid; the success counts are given in Table 3.

Because independently sampled graphs can have complementary edges, more frequent changes can improve the observed mixing rate. This sweep is therefore a robustness comparison across change frequencies, rather than an empirical equality for the worst-case term χ/τ in Theorem 3.3.

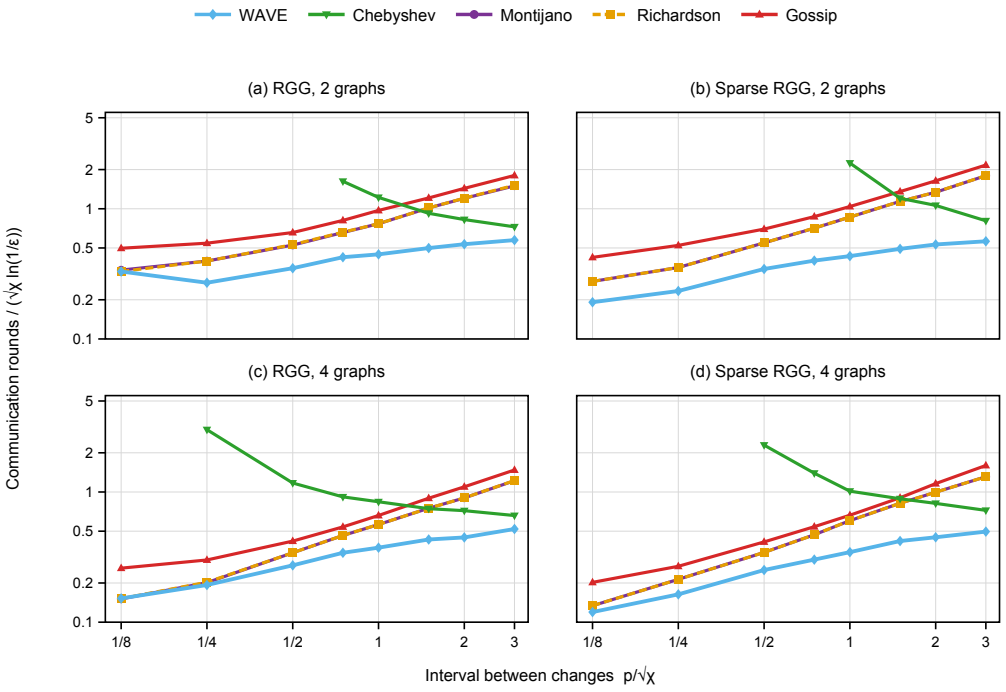

Figure 3: Median rounds to reach $r_k \le 10^{-8}$, normalized by $\sqrt{\chi}\ln(1/\epsilon)$, over 32 independently generated sequences. Classical Chebyshev is run without restarts; its curve is omitted when fewer than 17 runs reach the target.

Table 3: Number of the 32 runs in which classical Chebyshev without restarts reaches $r_k \le 10^{-8}$ within the round budget. WAVE, Montijano, Richardson, and gossip succeed in all runs.

	$p/\sqrt{\chi}$							
Graphs	1/8	1/4	1/2	3/4	1	3/2	2	3
RGG, 2 graphs	0	0	16	29	30	32	32	32
RGG, 4 graphs	0	25	31	32	32	32	32	32
Sparse RGG, 2 graphs	0	0	6	13	22	31	32	32
Sparse RGG, 4 graphs	0	0	25	32	32	32	32	32

H.3 SQUARE-ROOT SCALING IN χ

The comparison above covers a limited range of condition numbers. Corollary 3.2 makes a sharper prediction: when $\beta \le (3\chi^{3/2})^{-1}$, the number of rounds grows as $\sqrt{\chi}$, as for a fixed network. We test this prediction over a 64-fold range of χ.

Design. We draw one support graph from $G(40, 0.2)$ and accept it only when the graph and both halves of a fixed balanced cut are connected; the accepted support has 166 edges. Edges inside the halves have base weight 1, and all cut edges share a base weight w. For each $\chi \in \{25, 50, 100, 200, 400, 800, 1600\}$ we form 16 endpoint pairs by multiplying every base weight by $1 + 0.18s_e$ in L_A and by $1 - 0.18s_e$ in L_B, with independent random signs $s_e \in \{\pm 1\}$, and tune w by bisection so that the pair has condition bound exactly χ under the normalization of Appendix H.1. The two endpoints do not commute. The operator moves between them as in the metric-drift regime, with $\beta = (400\chi^{3/2})^{-1}$, and WAVE uses windows of length $\lfloor\sqrt{\chi}\rfloor$. As a control, the same method is run on the fixed operator at the starting point of the path.

Result. The median number of rounds to reach $r_k \le 10^{-6}$ grows from 51 at $\chi = 25$ to 421 at $\chi = 1600$ (Figure 4). A least-squares fit of $K = C\chi^p$ to the seven medians gives $p = 0.505$ with $R^2 = 0.9998$. The medians under metric drift and on the fixed network coincide at every χ. Thus, at the drift scale of Corollary 3.2, the observed communication counts retain the fixed-network $\sqrt{\chi}$ dependence.

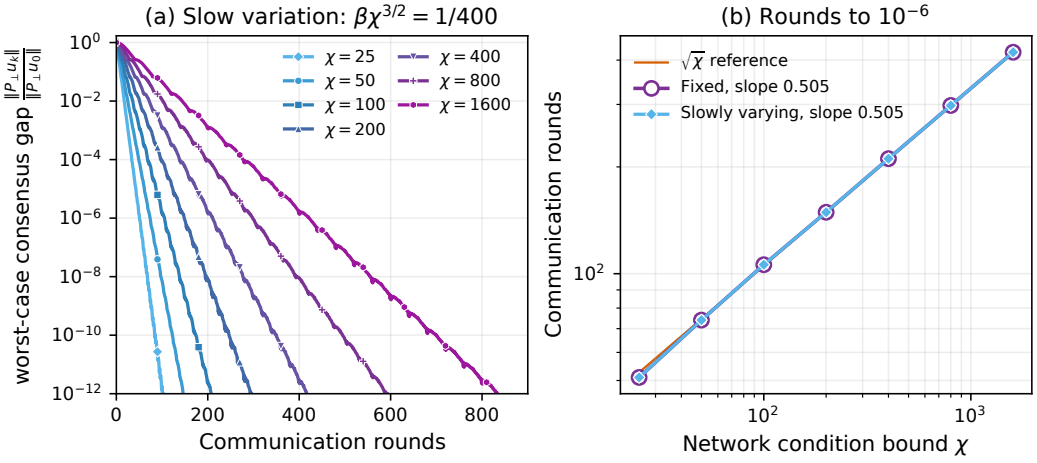

Figure 4: Condition-number scaling of WAVE at $\beta = (400\chi^{3/2})^{-1}$. (a) Median worst-case consensus gap r_k over 16 noncommuting endpoint pairs for each χ. (b) Median number of rounds to reach $r_k \le 10^{-6}$ for the metric-drift runs and the fixed-network control; the two coincide, and both fitted slopes equal 0.505.

H.4 CONTROLLED HIGH-DRIFT FAMILY

We describe the weighted-graph family used in Figure 1(a). This is a controlled stress test of the metric-drift bound rather than a random graph model. Fix $s \ge 400$, let $a = s^{-2}$ and $\chi = s^2$, and take $b = 3$. Define $J_b := \mathbf{1}_b\mathbf{1}_b^\top/b$, $P_b := I_b - J_b$, and, for $j = 0, \dots, b-1$, set $u_j := \sqrt{2/b}\cos(2\pi j/b)$ and $v_j := \sqrt{2/b}\sin(2\pi j/b)$. Let

$$\theta := 2\arctan\frac{3}{2s}, \qquad z_k := \sin(k\theta)u + \cos(k\theta)v.$$

For $P_4 := I_4 - \mathbf{1}_4\mathbf{1}_4^\top/4$, set $h_1 = (1, 1, -1, -1)^\top/2$, $h_2 = (1, -1, 1, -1)^\top/2$, and $w_k := \cos(\pi k/(2s))h_1 + \sin(\pi k/(2s))h_2$. The normalized Laplacian on four groups of b nodes is

$$L_k = I_4 \otimes (P_b - 10az_kz_k^\top) + a(P_4 + w_kw_k^\top) \otimes J_b. \tag{44}$$

An entrywise calculation shows that (44) is the Laplacian of a connected weighted graph. Its spectrum consists of 0, a twice, $2a$, $1 - 10a$ four times, and 1 four times. Hence the condition bound is

$\chi = s^2$. Moreover,

$$\|L_{k+1} - L_k\| = \frac{120}{s(4s^2+9)} =: \beta, \qquad \beta\chi^{3/2} = \frac{120s^2}{4s^2+9} \approx 30.$$

The disagreement dynamics splits into repeated invariant blocks. One representative block, together with the group-mean block, is the six-dimensional matrix used by the code:

$$B_k = \mathrm{diag}(I_2, aI_2, a, 1) - 10a\,(z_k z_k^\top \oplus 0_4) + a\,(0_2 \oplus \widetilde{w}_k \widetilde{w}_k^\top \oplus 0_2),$$

where $\widetilde{w}_k = (\cos(\pi k/(2s)), \sin(\pi k/(2s)))^\top$. The repeated blocks do not change the spectral norm, so propagating B_k gives the exact worst-case consensus gap of the full graph family. The current window rule selects

$$m_s = \left\lfloor (3\beta\chi)^{-1} \right\rfloor = \left\lfloor \frac{4s^2+9}{360s} \right\rfloor.$$

For Figure 1(a), $s = 400$ and $m_s = 4$. All five methods are applied to the same block sequence. The reported values are exact up to the sampling interval of 2500 rounds. Classical Chebyshev is run without restarts and exceeds a gap of 10^6 within 50000 rounds, whereas WAVE reaches $r_k \le 10^{-6}$ after 277500 rounds.