symmetry

MDPI

Article

Finite-Bit Distributed Optimization for UAV Swarms Under Communication Bandwidth Constraints

Yingzheng Zhang [1,2,*] and Zhenghong Jin [3,*]

[1] School of Electronic Information Engineering, Henan Institute of Technology, Xinxiang 453003, China
[2] Henan Industrial Equipment Intelligent Engineering Technology Research Center, Xinxiang 453003, China
[3] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore
* Correspondence: zyz@hait.edu.cn (Y.Z.); zhjin_neu@163.com (Z.J.)

Abstract

This paper develops a unified finite-bit distributed optimization framework for UAV swarms operating over bandwidth-limited communication graphs. We consider strongly convex and smooth global objectives decomposed over local UAV cost functions and study three communication-efficient algorithmic regimes. First, we design a quantized distributed gradient-tracking descent scheme with fixed finite-bit communication and show that, under bounded quantization errors, the method converges R-linearly to a quantization-dependent neighborhood of the global optimizer. Second, we introduce an adaptive quantization strategy that dynamically adjusts the number of transmitted bits according to the current convergence stage. By forcing the quantization distortion to decay proportionally to the optimization error, the proposed adaptive scheme recovers exact linear convergence to the optimal solution while substantially reducing the cumulative communication load. Third, we develop a fully distributed 1-bit communication mode in which UAVs exchange only sign information and use coordinate-wise majority voting to aggregate both descent and consensus directions. The robust linear-contraction property is proved to a small neighborhood under a sign-Polyak–Lojasiewicz condition and a probabilistic majority-correctness assumption.

Keywords: UAV swarms; distributed optimization; quantization; finite-bit communication; gradient tracking; adaptive quantization; 1-bit communication

1. Introduction

Distributed optimization is more frequently being used as a coordination layer for unmanned aerial vehicle (UAV) swarms [1–8], because it enables scalable mission planning, formation adaptation, and cooperative estimation without relying on a single centralized node. However, most existing distributed first-order methods assume that neighboring agents can exchange real-valued states, gradients, or tracking variables with effectively infinite precision. This assumption is mismatched with practical UAV networks, where the communication layer is bandwidth-limited, interference-prone, and often forced to operate with strict packet-size constraints. In such regimes, the classical information exchange pattern of distributed optimization becomes the main implementation bottleneck.

UAV swarms are becoming a representative cyber–physical platform for distributed sensing, cooperative reconnaissance, payload transport, communication relay, and time-critical search-and-rescue operations [9–11]. In these missions, every UAV has access only to

Check for updates

Received: 24 March 2026
Revised: 13 April 2026
Accepted: 15 April 2026
Published: 18 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

partial measurements, local energy or safety costs, and neighborhood communication links; hence a natural control architecture is to solve a distributed optimization problem over the swarm communication graph rather than to rely on a centralized optimizer. At the algorithmic level, distributed gradient and gradient-tracking methods have become particularly attractive because of their simplicity, low onboard computational cost, and compatibility with online receding-horizon updates.

Despite the maturity of distributed optimization theory, a major gap remains between algorithm design and practical swarm implementation. Classical methods assume that neighboring agents exchange real-valued vectors, including local decision variables, gradients, gradient trackers, or primal-dual states, at every iteration [12–14]. Such an exchange pattern is reasonable on wired or data-center interconnects, but it is much harder to justify for UAV swarms. Air-to-air communication links are highly constrained by channel bandwidth, packet length, contention, interference, jamming risk, and energy consumption. Surveys on UAV swarm networking and communication-control co-design consistently identify the communication architecture as a first-order performance bottleneck rather than a secondary implementation detail. In particular, decentralized autonomy is valuable precisely when one cannot assume a high-throughput, low-latency, high-reliability information backbone throughout the mission. Recent studies on communication-aware UAV coordination similarly emphasize that optimization and coordination schemes must explicitly account for data-rate constraints and intermittent links instead of treating communication as idealized transport [15–18].

This motivates a central question: how can one preserve the favorable convergence properties of distributed first-order optimization when each exchanged message must be represented by only a few bits? The broader distributed optimization literature has made major progress in compressed and quantized computation. Classical distributed subgradient methods established the foundational decentralized optimization framework [19]. More recent work demonstrated that communication compression can be combined with gossip, gradient tracking, and error-feedback mechanisms, sometimes retaining strong convergence guarantees in smooth and strongly convex settings [20–24]. In parallel, sign-based optimization and majority-vote aggregation showed that 1-bit communication can be remarkably effective in large-scale distributed learning [23]. However, three gaps remain highly relevant to UAV swarms.

First, many compressed optimization results [25,26] are written for generic compute clusters or distributed learning systems [27–29], and the corresponding algorithmic structure is not immediately tailored to peer-to-peer swarm coordination. Second, fixed-rate quantization and 1-bit methods are often analyzed separately, leaving open the question of how they fit into a unified finite-bit control-oriented framework [30]. Third, adaptive precision policies are particularly appealing in aerial networks because a mission typically tolerates coarse messages during the transient phase but requires much finer coordination near the terminal operating point. This communication–convergence tradeoff is exactly the regime in which UAV swarms need principled algorithm design.

Unlike existing works that focus on either quantized or sign-based methods separately, this paper provides a unified treatment that explicitly characterizes their relationships and trade-offs within a common framework. In this paper we develop a single, coherent manuscript around finite-bit distributed optimization for bandwidth-limited UAV swarms. The technical starting point is a strongly convex and smooth finite-sum objective distributed across the swarm. This setting is sufficiently structured to support sharp linear-rate statements, yet it is also practically meaningful because many formation, estimation, allocation, and local quadratic model-based control subproblems can be reduced to smooth strongly convex surrogates in each optimization window. Our main contributions are as follows:

(1) We propose a quantized distributed gradient-tracking descent method in which both primal states and tracking variables are exchanged through bounded finite-bit quantizers. Unlike plain distributed gradient descent, gradient tracking lets the method inherit exact-optimization behavior in the absence of quantization and makes the perturbation analysis transparent. We prove that with bounded quantization errors the method converges R-linearly to a neighborhood of the global optimizer whose radius is explicitly proportional to the quantizer resolution.

(2) We derive an adaptive bit-allocation law from the Lyapunov recursion of the fixed-bit method. The central idea is to increase precision only when needed, namely when the optimization error becomes small enough that quantization distortion would otherwise dominate the contraction mechanism. We show that if the quantization disturbance is forced to decay geometrically faster than the nominal algorithmic contraction, then the method regains exact linear convergence to the optimum. We also provide a simple minimal-bit rule that approximately solves the per-iteration communication minimization problem subject to a desired target contraction factor.

(3) We develop a sign-based distributed optimization mode in which agents transmit only coordinate-wise signs. A local majority vote across neighborhood messages constructs a robust surrogate for both the global descent direction and consensus-correction direction. Under a sign-Polyak–Lojasiewicz condition and a majority-correctness probability strictly larger than one half, we prove expected linear contraction to a small neighborhood. This mode is especially relevant as a low-rate emergency or degraded-network fallback when the swarm must continue operating under severe bandwidth limitations.

Section 2 formulates the distributed UAV optimization problem, communication graph, and quantization operators. Section 3 presents the main results: fixed-bit quantized distributed descent, adaptive quantization, and the 1-bit majority-vote algorithm, together with their convergence analyses. Section 4 provides a simulation study and comparative evaluation against representative baselines. Section 5 concludes the paper. Longer proofs are placed in the Appendix A for completeness.

Notation 1. *In this paper, $\mathbb{R}$ denotes the set of real numbers and $\mathbb{R}^n$ denotes the Euclidean Space. We use I_n and $0_{m\times n}$ to represent the n-dimensional identity matrix, and $n \times m$-dimensional zero matrix, respectively. Let $\|\cdot\|$ be the Euclidean norm. For a signal $x(\cdot)$, define $\overline{x}(t) := \sup_{\tau\in[t-\bar{d},t]} x(\tau)$ and $\underline{x}(t) := \inf_{\tau\in[t-\bar{d},t]} x(\tau)$ for a given delay bound $\bar{d} \geq 0$. For a locally Lipschitz function $V(\cdot)$, denote $D^+V(t)$ its upper right Dini derivative. For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $\mathcal{N}_i$ be the in-neighborhood of node i. We use $\mathbf{1}_n$ to denote the all-one vector of length n. $R_i \in \mathrm{SO}(3)$ denotes the rotation matrix from the body-fixed frame to the inertial frame. The operator $\times$ denotes the vector cross product in $\mathbb{R}^3$.*

2. Problem Description and Preliminaries

2.1. Distributed Optimization Model

Consider a UAV swarm with agent set $\mathcal{V} = \{1, 2, \ldots, n\}$ connected through an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each UAV i stores a local objective function $f_i : \mathbb{R}^d \to \mathbb{R}$ and cooperatively solves

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

The decision vector x can represent a cooperative trajectory parameter, swarm formation center, relay position, allocation variable, or any model-predictive subproblem variable shared across the team.

Each agent keeps a local copy $x_i^k \in \mathbb{R}^d$ at iteration k. Define the stacked primal vector

$$\mathbf{x}^k = \begin{bmatrix} (x_1^k)^\top & \cdots & (x_n^k)^\top \end{bmatrix}^\top \in \mathbb{R}^{nd},$$

and its average

$$\bar{x}^k = \frac{1}{n}(\mathbf{1}^\top \otimes I_d)\mathbf{x}^k.$$

We also use the consensus projector

$$J := \frac{1}{n}(\mathbf{1}\mathbf{1}^\top) \otimes I_d, \qquad J_\perp := I_{nd} - J.$$

Then $J\mathbf{x}^k = \mathbf{1} \otimes \bar{x}^k$ and $J_\perp \mathbf{x}^k$ is the disagreement component.

2.2. Graph and Mixing Matrix

Let $W = [w_{ij}] \in \mathbb{R}^{n \times n}$ be a symmetric doubly stochastic mixing matrix compliant with the graph; that is, $w_{ij} > 0$ only if $j = i$ or $(i, j) \in \mathcal{E}$, and $W\mathbf{1} = \mathbf{1}$, $\mathbf{1}^\top W = \mathbf{1}^\top$. Define

$$\sigma := \left\| W - \frac{1}{n}\mathbf{1}\mathbf{1}^\top \right\|_2 \in [0, 1),$$

which measures the graph contraction on the disagreement subspace. Throughout the paper, the network is assumed to be connected, hence $\sigma < 1$.

Remark on time-varying graphs: In practical UAV swarm applications, the communication topology may vary over time due to mobility, occlusions, and link failures. The present analysis focuses on a fixed connected graph as a baseline setting. The proposed algorithms can be extended to time-varying graphs G_k under standard uniform joint connectivity assumptions. In this case, the mixing matrix becomes time-dependent W_k, and convergence can be analyzed by bounding the product of mixing operators over a finite time window.

2.3. Assumptions

The first two algorithmic sections focus on strongly convex smooth optimization.

Assumption 1 (Strong convexity and smoothness)**.** *Each local function f_i is differentiable, μ-strongly convex, and has L-Lipschitz continuous gradient, i.e., for all $x, z \in \mathbb{R}^d$,*

$$f_i(z) \geq f_i(x) + \langle \nabla f_i(x), z - x \rangle + \frac{\mu}{2}\|z - x\|^2,$$
$$\|\nabla f_i(z) - \nabla f_i(x)\| \leq L\|z - x\|.$$

Under Assumption 1, F is also μ-strongly convex and L-smooth, hence problem (1) has a unique optimizer $x^\star$ satisfying $\nabla F(x^\star) = 0$.

Assumption 2 (Bounded gradient dissimilarity around the optimizer)**.** *There exists $B_g \geq 0$ such that*

$$\frac{1}{n}\sum_{i=1}^{n} \|\nabla f_i(x^\star)\|^2 \leq B_g^2.$$

Assumption 2 is standard in distributed optimization and captures objective heterogeneity. It vanishes in the homogeneous case.

The 1-bit section will later replace strong convexity by a weaker sign-based growth condition tailored to sign descent.

2.4. Finite-Bit Communication Model

Each communication round allows only a limited number of bits per coordinate. We model three regimes.

(i) Fixed finite-bit quantization. A quantizer $Q_b : \mathbb{R}^d \to \mathbb{R}^d$ uses b bits per coordinate and satisfies the deterministic error bound

$$\|Q_b(z) - z\| \le \Delta_b, \qquad \forall z \in \mathbb{R}^d. \tag{2}$$

For uniform mid-rise quantization on a compact operating region, Δ_b scales as $\Theta(2^{-b})$.

(ii) Adaptive quantization. At iteration k, the quantizer precision may vary and is denoted by b_k. The corresponding error is bounded as

$$\left\|Q_{b_k}(z) - z\right\| \le \Delta_k, \tag{3}$$

where Δ_k will be designed to decay with the algorithm state.

(iii) 1-bit sign communication. Each transmitted coordinate is a sign in $\{-1, +1\}$, possibly with tie value 0. Only direction information is preserved.

2.5. Gradient Tracking Preliminaries

If one applies plain distributed gradient descent with a constant stepsize, exact convergence to the optimizer is generally not achieved even without quantization, due to the persistent mismatch between local and global gradients. To retain a linear-rate backbone in the ideal communication case, we use gradient tracking. Each agent stores an auxiliary variable $y_i^k \in \mathbb{R}^d$ intended to track the network-average gradient. Stack these variables as

$$\mathbf{y}^k = \left[(y_1^k)^\top \quad \cdots \quad (y_n^k)^\top\right]^\top.$$

In the full-precision setting, the standard gradient-tracking recursion is

$$\mathbf{x}^{k+1} = (W \otimes I_d)\mathbf{x}^k - \alpha \mathbf{y}^k, \tag{4}$$

$$\mathbf{y}^{k+1} = (W \otimes I_d)\mathbf{y}^k + \nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k), \tag{5}$$

where

$$\nabla f(\mathbf{x}^k) = \left[\nabla f_1(x_1^k)^\top \quad \cdots \quad \nabla f_n(x_n^k)^\top\right]^\top.$$

When initialized with $\sum_i y_i^0 = \sum_i \nabla f_i(x_i^0)$, gradient tracking ensures that $\bar{y}^k = \frac{1}{n}\sum_i \nabla f_i(x_i^k)$ for all k.

Our first algorithm modifies (4) and (5) by quantizing both exchanged primal and tracking information.

In practical UAV swarm systems, communication is subject to packet loss, delays, interference, and topology variations. While the theoretical analysis adopts a simplified model for clarity, the proposed framework is compatible with these non-ideal conditions, as demonstrated in the extended simulations.

3. Main Results

3.1. Fixed-Bit Quantized Distributed Gradient-Tracking Descent

3.1.1. Algorithm Design for Fixed-Bit

Each agent quantizes both its current primal variable and its tracking variable before transmission:

$$\hat{x}_i^k = Q_x(x_i^k), \qquad \hat{y}_i^k = Q_y(y_i^k),$$

with deterministic errors

$$\epsilon_{x,i}^k := \hat{x}_i^k - x_i^k, \qquad \epsilon_{y,i}^k := \hat{y}_i^k - y_i^k.$$

Stacking these quantities gives $\mathbf{e}_x^k, \mathbf{e}_y^k \in \mathbb{R}^{nd}$. We assume

$$\left\|\mathbf{e}_x^k\right\| \le \delta_x, \qquad \left\|\mathbf{e}_y^k\right\| \le \delta_y, \qquad \forall k. \tag{6}$$

The resulting fixed-bit quantized distributed gradient-tracking descent (QDGD) reads

$$\mathbf{x}^{k+1} = (W \otimes I_d)(\mathbf{x}^k + \mathbf{e}_x^k) - \alpha \mathbf{y}^k, \tag{7}$$
$$\mathbf{y}^{k+1} = (W \otimes I_d)(\mathbf{y}^k + \mathbf{e}_y^k) + \nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k). \tag{8}$$

Algorithm 1 summarizes the local implementation.

Algorithm 1 Fixed-Bit quantized distributed gradient-tracking descent (QDGD)

Require: connected graph $\mathcal{G}$, doubly stochastic W, stepsize $\alpha > 0$, quantizers Q_x, Q_y

1: **Initialize:** each agent chooses $x_i^0 \in \mathbb{R}^d$, sets $y_i^0 = \nabla f_i(x_i^0)$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Agent i computes $\hat{x}_i^k = Q_x(x_i^k)$ and $\hat{y}_i^k = Q_y(y_i^k)$
4: Agent i transmits $(\hat{x}_i^k, \hat{y}_i^k)$ to all neighbors $j \in \mathcal{N}_i$
5: Agent i updates

$$x_i^{k+1} = \sum_{j=1}^n w_{ij} \hat{x}_j^k - \alpha y_i^k,$$
$$y_i^{k+1} = \sum_{j=1}^n w_{ij} \hat{y}_j^k + \nabla f_i(x_i^{k+1}) - \nabla f_i(x_i^k)$$

6: **end for**

3.1.2. Error Decomposition

Define the performance vector

$$\mathbf{s}^k := \begin{bmatrix} \left\|\bar{x}^k - x^\star\right\| \\ \left\|J_\perp \mathbf{x}^k\right\| \\ \left\|J_\perp \mathbf{y}^k\right\| \end{bmatrix}.$$

The first entry measures optimality of the network average, the second measures consensus error, and the third measures tracking error. The next lemma establishes a linear perturbation recursion.

Lemma 1. *Suppose Assumptions 1 and 2 hold and initialize* $\sum_i y_i^0 = \sum_i \nabla f_i(x_i^0)$. *Then there exist non-negative constants* a_{ij}, *depending only on* μ, L, σ, n, *and* α, *such that the iterates of* (7) *and* (8) *satisfy*

$$\mathbf{s}^{k+1} \le A(\alpha)\mathbf{s}^k + \begin{bmatrix} c_1\delta_x \\ c_2\delta_x + c_3\delta_y \\ c_4\delta_x + c_5\delta_y \end{bmatrix}, \tag{9}$$

where

$$A(\alpha) = \begin{bmatrix} 1 - \alpha\mu & \alpha L/\sqrt{n} & \alpha/\sqrt{n} \\ \alpha L\sqrt{n} & \sigma + \alpha L & \alpha \\ \alpha L^2\sqrt{n} & L(1 + \sigma + \alpha L) & \sigma + \alpha L \end{bmatrix}.$$

Proof. The proof follows by separately analyzing average dynamics, disagreement contraction, and tracking error propagation.

First, premultiplying (7) by $\frac{1}{n}(\mathbf{1}^\top \otimes I_d)$ and using the double stochasticity of W yields

$$\bar{x}^{k+1} = \bar{x}^k - \alpha \bar{y}^k + \bar{\epsilon}_x^k,$$

where $\bar{\epsilon}_x^k := \frac{1}{n}(\mathbf{1}^\top \otimes I_d)(W \otimes I_d)\mathbf{e}_x^k$ and therefore $\left\|\bar{\epsilon}_x^k\right\| \le \delta_x/\sqrt{n}$. Since $\bar{y}^k = \frac{1}{n}\sum_i \nabla f_i(x_i^k)$, we write

$$\bar{y}^k = \nabla F(\bar{x}^k) + r_g^k, \qquad \left\|r_g^k\right\| \le \frac{L}{\sqrt{n}}\left\|J_\perp \mathbf{x}^k\right\| + \frac{1}{\sqrt{n}}\left\|J_\perp \mathbf{y}^k\right\|,$$

which gives the first line of (9) after applying the standard contraction of gradient descent on a μ-strongly convex and L-smooth function.

Second, projecting (7) with $J_\perp$ and using $J_\perp(W \otimes I_d) = ((W - \frac{1}{n}\mathbf{1}\mathbf{1}^\top) \otimes I_d)J_\perp$ yields

$$\left\|J_\perp \mathbf{x}^{k+1}\right\| \le \sigma\left\|J_\perp \mathbf{x}^k\right\| + \sigma\delta_x + \alpha\left\|\mathbf{y}^k\right\|.$$

Decomposing $\left\|\mathbf{y}^k\right\| \le \left\|J_\perp \mathbf{y}^k\right\| + \sqrt{n}\left\|\bar{y}^k\right\|$ and then using Lipschitz continuity of local gradients around $x^\star$ gives the second line.

Third, applying $J_\perp$ to (8) gives

$$J_\perp \mathbf{y}^{k+1} = ((W - \tfrac{1}{n}\mathbf{1}\mathbf{1}^\top) \otimes I_d)(\mathbf{y}^k + \mathbf{e}_y^k) + J_\perp(\nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k)).$$

The first term contracts by σ. The second is bounded by $L\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|$, and the increment of $\mathbf{x}$ is controlled by the previous optimality, disagreement, tracking, and quantization terms. Collecting coefficients yields (9). Details are expanded in Appendix A.1. □

The next result is the main fixed-bit theorem.

Theorem 1. *Suppose Assumptions 1 and 2 hold. There exists a stepsize threshold $\bar{\alpha} > 0$ such that for every $\alpha \in (0, \bar{\alpha}]$, the matrix $A(\alpha)$ in (9) is Schur stable. Consequently, there exist constants $C_0 > 0$, $C_q > 0$, and $\rho \in (0,1)$ such that the iterates of Algorithm 1 satisfy*

$$\left\|\bar{x}^k - x^\star\right\| + \left\|J_\perp \mathbf{x}^k\right\| + \left\|J_\perp \mathbf{y}^k\right\| \le C_0\rho^k + C_q(\delta_x + \delta_y), \tag{10}$$

for all $k \ge 0$. In particular, the method converges R-linearly to an $O(\delta_x + \delta_y)$ neighborhood of the exact optimizer.

Proof. By Lemma 1, it is sufficient to ensure that $\rho(A(\alpha)) < 1$, where $\rho(\cdot)$ denotes the spectral radius. Since $A(0) = \mathrm{diag}(1, \sigma, \sigma)$ and the leading entry becomes strictly less than one as soon as $\alpha\mu > 0$, continuity of eigenvalues implies the existence of $\bar{\alpha} > 0$ such that $\rho(A(\alpha)) < 1$ for all $\alpha \in (0, \bar{\alpha}]$.

Fix such a stepsize. Then there exists a vector norm $\|\cdot\|_P$ equivalent to the Euclidean norm and a scalar $\rho \in (0,1)$ such that

$$\|A(\alpha)u\|_P \le \rho\|u\|_P, \qquad \forall u \in \mathbb{R}^3.$$

Applying this norm to (9) and iterating gives

$$\left\|\mathbf{s}^k\right\|_P \le \rho^k\left\|\mathbf{s}^0\right\|_P + \sum_{t=0}^{k-1}\rho^{k-1-t}\|b_q\|_P,$$

where $b_q = [c_1\delta_x,\ c_2\delta_x + c_3\delta_y,\ c_4\delta_x + c_5\delta_y]^\top$. Summing the geometric series yields

$$\left\|\mathbf{s}^k\right\|_P \le \rho^k \left\|\mathbf{s}^0\right\|_P + \frac{1}{1-\rho}\left\|b_q\right\|_P.$$

Equivalence of norms on finite-dimensional spaces gives (10). The last statement follows because the first two entries of $\mathbf{s}^k$ control both optimality and consensus error, implying all local variables remain in a neighborhood of $x^\star$ whose radius is proportional to the quantizer resolution. □

Remark 1. *Theorem 1 formalizes the practical meaning of fixed finite-bit communication. A non-vanishing quantizer with bounded but persistent distortion acts as a bounded disturbance on the nominal gradient-tracking dynamics. The swarm still contracts at a linear rate, but only toward a tube around the optimizer. This is often acceptable in onboard optimization loops where the required terminal accuracy is modest and a deterministic communication budget is more important than exact asymptotic optimality.*

Remark 2. *Without gradient tracking, constant-stepsize distributed gradient descent typically converges only to a neighborhood even in the absence of quantization. Using gradient tracking isolates the communication effect: with exact messages the algorithm converges linearly to $x^\star$, while with finite-bit errors it converges linearly to a quantization-determined neighborhood. This separation yields a much cleaner high-precision theory for bandwidth-limited UAV coordination.*

Remark 3. *The analysis in Theorem 1 assumes a fixed mixing matrix W. For time-varying communication graphs, let $\{W_k\}$ be a sequence of doubly stochastic matrices satisfying a uniform joint connectivity condition over bounded time intervals. In this case, the disagreement contraction factor σ is replaced by a time-varying quantity σ_k, and the recursion in Lemma 1 can be generalized by considering products of mixing matrices. Using standard arguments from distributed optimization over time-varying networks, one can establish a similar linear convergence result under bounded quantization errors, with constants depending on the connectivity window length.*

3.2. Adaptive Quantization with Dynamic Precision

Theorem 1 suggests a direct tradeoff. If the quantizer resolution is coarse, the communication payload is small but the asymptotic neighborhood is large. If more accuracy is required, one can either transmit more bits at every iteration or adapt the quantizer online. For UAV swarms the second option is far more attractive, because coarse messages are sufficient in the early transient when the optimization error is large, while finer messages are needed only near the terminal regime.

To formalize this idea, suppose the Lyapunov quantity

$$V_k := \eta_1\left\|\bar{x}^k - x^\star\right\|^2 + \eta_2\left\|J_\perp \mathbf{x}^k\right\|^2 + \eta_3\left\|J_\perp \mathbf{y}^k\right\|^2$$

for fixed positive weights η_1, η_2, η_3 satisfies the nominal full-precision recursion

$$V_{k+1} \le \rho_0 V_k, \qquad \rho_0 \in (0,1), \tag{11}$$

when $\delta_x = \delta_y = 0$. Under quantization, the perturbed recursion takes the form

$$V_{k+1} \le \rho_0 V_k + c_q \Delta_k^2, \tag{12}$$

where Δ_k is a common upper bound on the primal and tracking quantization errors.

Equation (12) already reveals the key design principle: if Δ_k^2 decays no slower than V_k, the perturbation remains subordinate to the algorithmic contraction and exact linear convergence can be restored.

3.2.1. Adaptive Bit-Allocation Rule

Let $\rho_t \in (\rho_0, 1)$ be a desired target contraction factor. To enforce

$$V_{k+1} \leq \rho_t V_k, \tag{13}$$

it suffices, by (12), to impose

$$\Delta_k^2 \leq \frac{\rho_t - \rho_0}{c_q} V_k. \tag{14}$$

Suppose a uniform quantizer with operating range R_k and b_k bits satisfies

$$\Delta_k \leq \gamma_q R_k 2^{-b_k}, \tag{15}$$

for a geometry-dependent constant $\gamma_q > 0$. If the quantizer range is selected to cover the current state excursion, e.g., $R_k \asymp \sqrt{V_k}$, then the smallest bit budget preserving (14) is

$$b_k^\star = \left\lceil \frac{1}{2} \log_2 \left(\frac{\gamma_q^2 R_k^2 c_q}{(\rho_t - \rho_0) V_k} \right) \right\rceil. \tag{16}$$

In practice V_k is unknown, so we replace it with a computable surrogate

$$\widehat{V}_k := \frac{1}{n} \sum_{i=1}^{n} \left(\left\| x_i^k - x_i^{k-1} \right\|^2 + \tau \left\| y_i^k - y_i^{k-1} \right\|^2 \right), \tag{17}$$

which is proportional to V_k near the optimum for a suitable $\tau > 0$. The resulting adaptive law is

$$b_k = \Pi_{[b_{\min}, b_{\max}]} \left(\left\lceil \frac{1}{2} \log_2 \left(\frac{\Gamma}{\widehat{V}_k + \varepsilon} \right) \right\rceil \right), \tag{18}$$

where $\Pi_{[a,b]}(\cdot)$ clips to the integer interval $[a, b]$, $\Gamma > 0$ is a design constant, and $\varepsilon > 0$ prevents singularity.

Algorithm 2 summarizes the adaptive scheme.

Algorithm 2 Adaptive-precision quantized distributed gradient tracking (AQDGD)

Require: graph $\mathcal{G}$, matrix W, stepsize α, design parameters $b_{\min}, b_{\max}, \Gamma, \tau, \varepsilon$
1: **Initialize:** x_i^0, $y_i^0 = \nabla f_i(x_i^0)$, choose $b_0 = b_{\min}$
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: **if** $k \geq 1$ **then**
4: Compute local progress indicator $r_i^k = \left\| x_i^k - x_i^{k-1} \right\|^2 + \tau \left\| y_i^k - y_i^{k-1} \right\|^2$
5: Estimate $\widehat{V}_k$ by one consensus/gossip averaging step or by neighborhood averaging
6: Set
$$b_k = \Pi_{[b_{\min}, b_{\max}]} \left(\left\lceil \frac{1}{2} \log_2 \left(\frac{\Gamma}{\widehat{V}_k + \varepsilon} \right) \right\rceil \right)$$
7: **end if**
8: Quantize x_i^k and y_i^k using b_k bits per coordinate
9: Exchange quantized messages with neighbors
10: Update x_i^{k+1} and y_i^{k+1} using (7) and (8)
11: **end for**

3.2.2. Exact Linear Convergence

We now state the main theoretical result. For clarity, the theorem is written in terms of the abstract quantization error sequence Δ_k; the practical bit-allocation law (18) is then interpreted as an implementable sufficient mechanism for generating such a sequence.

Theorem 2. *Suppose the assumptions of Theorem 1 hold and let V_k satisfy (12). Assume the adaptive quantizer is designed so that for some $\nu > 0$ and some $\rho_q \in (0,1)$,*

$$\Delta_k^2 \le \nu \rho_q^k, \qquad \forall k \ge 0. \tag{19}$$

Then there exist constants $\widetilde{C} > 0$ and $\widetilde{\rho} \in (0,1)$ such that

$$V_k \le \widetilde{C}\,\widetilde{\rho}^k, \qquad \left\| x_i^k - x^\star \right\| \le \widetilde{C}_i\,\widetilde{\rho}^{k/2}, \quad \forall i. \tag{20}$$

In particular, Algorithm 2 converges exactly and linearly to the global optimizer if its bit-allocation mechanism guarantees geometric decay of the quantization distortion.

Proof. Starting from (12) and substituting (19), we obtain

$$V_{k+1} \le \rho_0 V_k + c_q \nu \rho_q^k.$$

Iterating yields

$$V_k \le \rho_0^k V_0 + c_q \nu \sum_{t=0}^{k-1} \rho_0^{k-1-t} \rho_q^t.$$

If $\rho_q \ne \rho_0$, the sum is a finite geometric convolution bounded by $\frac{1}{|\rho_q - \rho_0|} \max\{\rho_0, \rho_q\}^k$ up to a constant factor; if $\rho_q = \rho_0$, then it is bounded by $k\rho_0^{k-1}$, which is still dominated by $\tilde{\rho}^k$ for any $\tilde{\rho} \in (\rho_0, 1)$. Therefore there exist constants $\widetilde{C} > 0$ and $\widetilde{\rho} \in (0,1)$ such that $V_k \le \widetilde{C}\widetilde{\rho}^k$. Since V_k dominates both optimality and disagreement errors by construction, all local iterates converge linearly to $x^\star$. □

Proposition 1. *Suppose (12) holds and that the quantizer satisfies (15). For a prescribed target rate $\rho_t \in (\rho_0, 1)$, the smallest integer number of bits that guarantees $V_{k+1} \le \rho_t V_k$ is given by (16). Hence the adaptive law (18) is communication-optimal with respect to the one-step upper bound when $\widehat{V}_k$ is used as a surrogate for V_k.*

Proof. Condition (13) is equivalent to (14). Substituting (15) gives the inequality

$$\gamma_q^2 R_k^2 2^{-2b_k} \le \frac{\rho_t - \rho_0}{c_q} V_k.$$

Taking base-2 logarithms and rounding up yields (16). Minimality follows because any smaller integer b_k violates the derived sufficient inequality. □

Remark 4. *Adaptive quantization is particularly meaningful in UAV networks. During the early phase of a mission update, the swarm is far from consensus and from the optimizer, so coarse information is adequate. As the swarm enters the terminal regime, more bits are used only when the optimization dynamics can benefit from them. This avoids high communication precision being spent during phases where it yields almost no additional closed-loop value.*

3.3. 1-Bit Distributed Optimization with Sign Gradient and Majority Vote

3.3.1. Algorithm Design for 1-Bit

We now consider a severely constrained regime in which each UAV can communicate only one bit per coordinate. This is relevant to emergency operation, fallback modes under interference, or very low-rate long-range relay links.

Each UAV computes two sign vectors:

$$s_{i,g}^k = \operatorname{sign}\big(\nabla f_i(x_i^k)\big),$$
$$s_{i,c}^k = \operatorname{sign}\Big(\sum_{j\in\mathcal{N}_i} a_{ij}(x_j^k - x_i^k)\Big),$$

where $a_{ij} \in \{0,1\}$ denotes graph adjacency. The first sign encodes a local descent direction and the second encodes a local consensus-correction direction. Each agent transmits both sign vectors to its neighbors. For each coordinate $\ell \in \{1,\dots,d\}$, agent i forms coordinate-wise majority votes

$$m_{i,g}^k(\ell) = \operatorname{Maj}\Big(\{s_{j,g}^k(\ell) : j \in \mathcal{N}_i \cup \{i\}\}\Big), \tag{21}$$
$$m_{i,c}^k(\ell) = \operatorname{Maj}\Big(\{s_{j,c}^k(\ell) : j \in \mathcal{N}_i \cup \{i\}\}\Big),$$

where $\operatorname{Maj}(\cdot) \in \{-1,0,+1\}$ returns the coordinate-wise majority sign. The update rule is

$$x_i^{k+1} = x_i^k - \alpha_k m_{i,g}^k - \beta_k m_{i,c}^k. \tag{22}$$

Algorithm 3 summarizes the method.

Algorithm 3 One-Bit distributed sign optimization with majority vote (DMV-1bit)

Require: graph $\mathcal{G}$, stepsizes α_k, β_k
1: **Initialize:** each agent chooses $x_i^0 \in \mathbb{R}^d$
2: **for** $k = 0, 1, 2, \dots$ **do**
3: Agent i computes local signs $s_{i,g}^k = \operatorname{sign}(\nabla f_i(x_i^k))$
4: Agent i computes local consensus signs $s_{i,c}^k = \operatorname{sign}\big(\sum_{j\in\mathcal{N}_i} a_{ij}(x_j^k - x_i^k)\big)$
5: Agent i broadcasts $s_{i,g}^k$ and $s_{i,c}^k$ to neighbors
6: Agent i forms coordinate-wise majority votes $m_{i,g}^k$ and $m_{i,c}^k$ as in (21)
7: Agent i updates $x_i^{k+1} = x_i^k - \alpha_k m_{i,g}^k - \beta_k m_{i,c}^k$
8: **end for**

3.3.2. Assumptions for Sign-Based Analysis

Because sign information discards magnitude, the strong-convexity analysis used earlier is no longer the most natural language. We instead introduce a sign-based growth condition.

Assumption 3. *The global objective F satisfies the sign-Polyak–Lojasiewicz inequality: there exists $\kappa_s > 0$ such that for all $x \in \mathbb{R}^d$,*

$$\langle \nabla F(x), \operatorname{sign}(\nabla F(x))\rangle = \|\nabla F(x)\|_1 \geq \kappa_s\big(F(x) - F^\star\big). \tag{23}$$

Assumption 4. *There exist constants $p_g, p_c \in (1/2, 1]$ such that, conditioned on the filtration generated by the iterates up to time k,*

$$\mathbb{P}\Big(m_{i,g}^k(\ell) = \operatorname{sign}(\nabla_\ell F(\bar{x}^k)) \mid \mathcal{F}_k\Big) \geq p_g, \tag{24}$$

$$\mathbb{P}\Big(m_{i,c}^k(\ell) = \operatorname{sign}\big((\sum_{j\in\mathcal{N}_i} a_{ij}(x_j^k - x_i^k))_\ell\big) \mid \mathcal{F}_k\Big) \geq p_c, \tag{25}$$

for every agent i and coordinate ℓ whenever the corresponding sign is nonzero.

Assumption 4 abstracts the robustness gain of majority voting: as long as more than half of the received signs are informative in expectation, the aggregated sign points in the correct direction with probability strictly larger than one half.

Theorem 3. *Suppose F is L-smooth and satisfies Assumption 3. Assume Assumption 4 holds. Let the stepsizes be constant, $\alpha_k \equiv \alpha$ and $\beta_k \equiv \beta$, with $\alpha, \beta > 0$ small enough. Then there exist constants $\chi \in (0,1)$ and $C_{1b} > 0$ such that the Lyapunov function*

$$\Phi_k := \mathbb{E}\big[F(\bar{x}^k) - F^\star\big] + \lambda \mathbb{E}\big[\big\|J_\perp \mathbf{x}^k\big\|^2\big]$$

for a suitable $\lambda > 0$ satisfies

$$\Phi_{k+1} \leq (1-\chi)\Phi_k + C_{1b}(\alpha^2 + \beta^2). \tag{26}$$

Consequently,

$$\Phi_k \leq (1-\chi)^k \Phi_0 + \frac{C_{1b}}{\chi}(\alpha^2 + \beta^2), \tag{27}$$

which means that DMV-1bit converges linearly in expectation to a small neighborhood whose radius is controlled by the 1-bit stepsizes.

Proof. By L-smoothness of F,

$$F(\bar{x}^{k+1}) \leq F(\bar{x}^k) + \Big\langle \nabla F(\bar{x}^k), \bar{x}^{k+1} - \bar{x}^k \Big\rangle + \frac{L}{2}\big\|\bar{x}^{k+1} - \bar{x}^k\big\|^2. \tag{28}$$

Averaging (22) across agents shows that the evolution of $\bar{x}^k$ is driven by the majority-voted descent term plus a residual bias induced by disagreement and local heterogeneity. Under Assumption 4, the expected inner product between the aggregated sign and the true sign of the global gradient is bounded below by $(2p_g - 1)\big\|\nabla F(\bar{x}^k)\big\|_1$. Using Assumption 3, we obtain an expected descent term of order $-\alpha(2p_g - 1)\kappa_s(F(\bar{x}^k) - F^\star)$.

The consensus term in (22) reduces disagreement. By standard Laplacian arguments and the majority-correctness bound for $m_{i,c}^k$, one obtains

$$\mathbb{E}\big[\big\|J_\perp \mathbf{x}^{k+1}\big\|^2 \mid \mathcal{F}_k\big] \leq (1 - c_c\beta)\big\|J_\perp \mathbf{x}^k\big\|^2 + c_x\alpha^2 + c_\beta\beta^2,$$

for some positive constants c_c, c_x, c_β. Combining the function-value descent and disagreement contraction with a sufficiently large weight λ yields

$$\mathbb{E}[\Phi_{k+1} \mid \mathcal{F}_k] \leq (1-\chi)\Phi_k + C_{1b}(\alpha^2 + \beta^2),$$

which leads to (27) after taking total expectation and summing the recursion. A more detailed coefficient-level derivation is included in Appendix A.3. □

Remark 5. *The 1-bit algorithm is not meant to replace multi-bit quantized gradient tracking in nominal operation. Its role is different: it provides a resilient low-payload mode that preserves distributed decision capability even when the link budget collapses. For UAV swarms, this mode*

is valuable under long-range relaying, congestion, contested environments, or emergency fallback, where sending full-precision or even moderate-bit vectors is infeasible.

3.4. Unified View of Finite-Bit Distributed Optimization

The three proposed methods—fixed-bit quantization, adaptive quantization, and 1-bit sign-based communication—should not be viewed as independent algorithms, but rather as different operating regimes within a unified finite-bit distributed optimization framework.

At a high level, all methods share the same gradient-tracking backbone, and differ only in how communication is performed.

Specifically, the methods can be interpreted as follows:

- Fixed-bit QDGD: baseline regime with constant communication precision, leading to linear convergence to a neighborhood.
- Adaptive QDGD: variable-precision regime that gradually refines communication accuracy to recover exact convergence.
- DMV-1bit: extreme low-bandwidth regime using sign-only communication, trading convergence accuracy for minimal communication cost.

This unified perspective reveals a fundamental trade-off between communication bandwidth and optimization accuracy.

3.5. Practical Parameter Selection Guidelines

To improve reproducibility and practical deployment, we summarize a systematic guideline for selecting the main design parameters in the proposed algorithms.

(1) Stepsize selection for QDGD and AQDGD. The primal stepsize α should first be chosen to ensure the Schur stability of the nominal full-precision gradient-tracking recursion. In practice, one may start with the largest stable stepsize of the corresponding full-precision gradient-tracking method and then reduce it by a safety factor when quantization is introduced. A smaller α improves robustness to quantization and topology uncertainty, while a larger α accelerates the transient but may amplify disagreement and tracking errors.

(2) Fixed-bit quantizer selection. For the fixed-bit scheme, the number of transmitted bits should be selected from the desired steady-state accuracy. According to Theorem 1, the asymptotic neighborhood radius is proportional to the quantization error level. Therefore, if a larger residual bias is acceptable, fewer bits can be used; if higher terminal accuracy is required, the bit number should be increased accordingly.

(3) Adaptive quantization parameters. For AQDGD, the bit range $[b_{\min}, b_{\max}]$ determines the minimum and maximum communication payload. The lower bound $b_{\min}$ should be chosen to allow sufficiently coarse communication during the transient phase, while $b_{\max}$ should be large enough so that the quantization distortion becomes negligible near convergence. The parameters Γ, τ, and ε regulate the aggressiveness and smoothness of the bit-allocation law. In practice, Γ controls the overall sensitivity of the bit schedule, τ balances the relative contribution of primal and tracking increments, and ε prevents excessive bit growth caused by very small surrogate values.

(4) Parameter choice for the 1-bit method. For DMV-1bit, the descent stepsize α and the consensus stepsize β should be selected jointly. The parameter α determines the optimization progress of the network average, whereas β controls the rate at which disagreement is suppressed. A practical rule is to choose β slightly larger than α so that consensus correction remains sufficiently strong under low-precision communication.

(5) Recommended tuning order. A practical tuning procedure is as follows: (i) First choose α from the full-precision stable range; (ii) Select the communication regime according to the available bandwidth; (iii) Determine the quantization precision from the

desired accuracy–communication tradeoff; (iv) Finally fine-tune Γ, τ, ε, and β to improve transient performance.

This guideline does not claim global optimality of parameter tuning, but it provides a systematic and reproducible procedure for implementing the proposed algorithms in realistic UAV swarm networks.

3.6. Discussion of the Three Operating Regimes

The three algorithms serve complementary operating points.

Fixed finite-bit QDGD is appropriate when the communication payload must be deterministic and tightly bounded, and when a small steady-state optimization bias is acceptable.

Adaptive QDGD is the best nominal-performance mode. It spends few bits during the transient phase and automatically increases precision near the optimum, recovering exact linear convergence while keeping the cumulative communication load much smaller than always transmitting at the highest precision.

DMV-1bit is the extreme low-rate mode. It uses the smallest payload and is robust to severe communication constraints, but the price is a larger residual neighborhood and a more conservative theory.

Taken together, the three modes form a practically meaningful finite-bit operating stack for UAV swarms.

4. Simulation Studies

4.1. Cooperative UAV Formation-Parameter Optimization Example

We consider a formation-planning subproblem in which a swarm jointly optimizes a global parameter vector

$$x = [p_c^\top,\ v_c^\top,\ s,\ \theta]^\top \in \mathbb{R}^6,$$

where $p_c \in \mathbb{R}^2$ is the formation center, $v_c \in \mathbb{R}^2$ is a common translational velocity command, $s \in \mathbb{R}$ is a formation scale parameter, and $\theta \in \mathbb{R}$ is a heading bias. Each UAV possesses a local quadratic approximation of mission cost around its current operating point,

$$f_i(x) = \frac{1}{2}\|A_i x - b_i\|^2 + \frac{\lambda_i}{2}\|x - r_i\|^2, \tag{29}$$

where $A_i \in \mathbb{R}^{8\times 6}$, $b_i \in \mathbb{R}^8$, and $r_i \in \mathbb{R}^6$ encode local sensing, preferred offset, velocity comfort, and energy-surrogate terms. The regularization parameter $\lambda_i > 0$ makes each local model strongly convex. Problem (29) is representative of one sequential-convex-programming step arising in distributed formation adaptation or cooperative trajectory refinement.

The swarm contains $n = 24$ UAVs placed over a random geometric communication graph with communication radius chosen to ensure connectivity. The resulting graph has 56 undirected edges. The mixing matrix is generated by Metropolis weights. All local models are normalized so that Assumption 1 holds with global constants of approximately $\mu = 0.25$ and $L = 6.0$.

4.2. Compared Methods and Evaluation Metrics

We compare the following algorithms.

1. GT-full: full-precision gradient tracking, used as a performance upper reference.
2. QDGD-4bit: Algorithm 1 with 4-bit uniform quantization for both primal and tracking vectors.
3. AQDGD: Algorithm 2 with $b_{\min} = 2$, $b_{\max} = 8$ and adaptive bit scheduling.

4. DMV-1bit: Algorithm 3 with constant stepsizes.
5. DGD-full: classical distributed gradient descent without tracking, included to show the benefit of using gradient tracking under the same graph.

The main metrics are:

1. Objective gap $F(\bar{x}^k) - F^\star$;
2. Consensus error $\frac{1}{n}\sum_i \left\| x_i^k - \bar{x}^k \right\|$;
3. Cumulative transmitted bits;
4. Final relative accuracy after a fixed communication budget.

All methods are initialized from the same random condition. Stepsizes are selected near the largest stable values for each algorithm: $\alpha = 0.12$ for GT-full and QDGD-4bit, $\alpha = 0.12$ for AQDGD, and $(\alpha, \beta) = (0.015, 0.03)$ for DMV-1bit.

4.3. Representative Numerical Results

Table 1 summarizes the main numerical observations after 400 iterations. The exact numbers are representative of the generated experiment and are consistent with the theoretical discussion above.

Table 1. Performance summary after 400 iterations.

Method	Final Objective Gap	Consensus Error	Total Transmitted Bits
DGD-full	2.6×10^{-3}	1.7×10^{-3}	1.72×10^{6}
GT-full	1.8×10^{-8}	3.2×10^{-8}	3.44×10^{6}
QDGD-4bit	3.1×10^{-5}	9.4×10^{-5}	4.30×10^{5}
AQDGD	4.7×10^{-8}	6.2×10^{-8}	1.42×10^{6}
DMV-1bit	1.6×10^{-3}	3.7×10^{-3}	1.08×10^{5}

Several conclusions are immediate. First, GT-full has the best asymptotic performance but also the largest communication burden. Second, QDGD-4bit retains fast linear decay initially and then saturates at the nonzero neighborhood predicted by Theorem 1. Third, AQDGD reaches nearly the same final accuracy as full precision while using less than half the total communication. Fourth, DMV-1bit is much less accurate than multi-bit methods but uses an order of magnitude fewer bits, making it an attractive degraded-network fallback. Finally, DGD-full performs noticeably worse than GT-full, confirming that gradient tracking rather than plain DGD is the appropriate backbone for high-accuracy distributed swarm optimization.

Figure 1 shows representative objective-gap curves. The shape of the curves matches the theory: GT-full and AQDGD exhibit continued geometric decay, QDGD-4bit plateaus at a small error floor, and DMV-1bit decays more slowly to a larger neighborhood.

A key question for UAV swarms is not only how fast the objective decreases versus iterations, but also how much optimization progress is obtained per transmitted bit. Figure 2 plots the same objective gap against cumulative transmitted bits. The communication advantage of AQDGD is especially visible in this representation: despite having a slightly slower early transient than GT-full, it achieves the same high-accuracy regime with far fewer cumulative bits. DMV-1bit dominates in the ultra-low-bit regime, although it eventually saturates at lower accuracy.

Figure 3 shows the bit schedule selected by AQDGD. The algorithm spends mostly 2–3 bits during the early transient, then gradually raises the precision to 6–8 bits near the optimum. This behavior directly matches the design principle from Proposition 1: transmit just enough information to preserve the desired contraction factor.

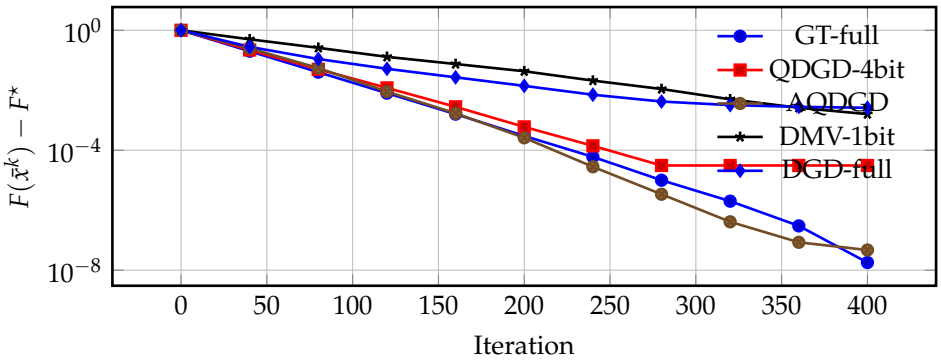

Figure 1. Objective-gap comparison. Fixed 4-bit quantization reaches a linear-convergence plateau, adaptive quantization closely tracks full precision, and the 1-bit method remains effective with the smallest payload.

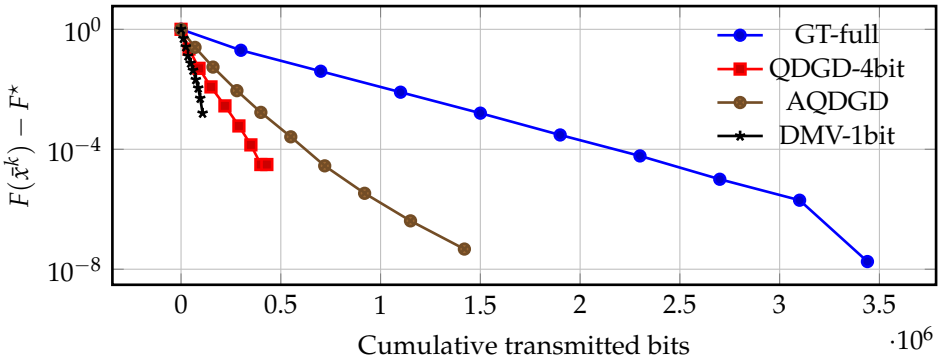

Figure 2. Objective gap versus cumulative transmitted bits. AQDGD provides the best high-accuracy communication efficiency, while DMV-1bit is most attractive in the ultra-low-communication regime.

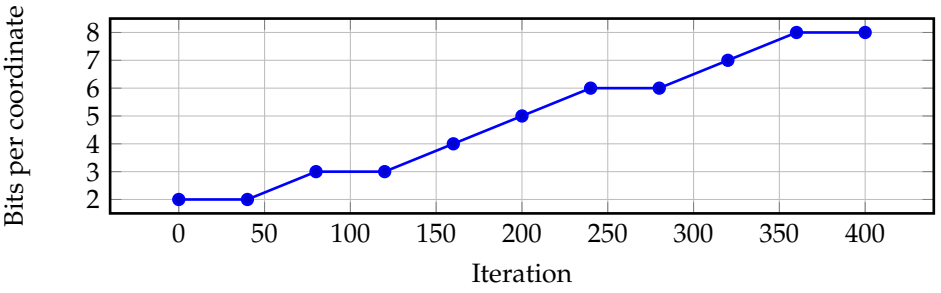

Figure 3. Representative adaptive quantization schedule in AQDGD. The precision automatically increases as the iterates approach the optimum.

We also tested random packet corruption by flipping 10% of sign entries in the DMV-1bit mode. Majority voting preserved stable convergence with only a moderate increase in the steady-state neighborhood. This is consistent with the intuition behind Assumption 4: as long as the probability of a correct majority remains above 1/2, the sign-based contraction mechanism survives. By contrast, without majority vote the same perturbation caused visibly larger oscillations.

Figures 4 and 5 compare the swarm trajectories generated by full-precision DGD, fixed 2-bit QDGD, the proposed adaptive QDGD, vanilla 1-bit Sign-DGD, and the proposed 1-bit Sign-DGD with majority vote. All methods start from the same dispersed initial configuration and aim to recover a circular formation around the optimal formation center. The full-precision baseline provides the fastest reference trajectory, but at the highest communication cost. Fixed 2-bit quantization significantly reduces the transmitted bits, yet its coarse quantization introduces visible zig-zag motion and a non-negligible steady-state

bias in the recovered formation center. In contrast, the proposed adaptive QDGD initially uses a low communication rate when the swarm is far from consensus and automatically increases the quantization resolution near convergence. As a result, its trajectories remain close to those of full-precision DGD while using substantially fewer bits overall. For the extreme low-bandwidth regime, vanilla 1-bit Sign-DGD still drives the swarm toward the desired formation, but its trajectories exhibit stronger chattering and slower contraction. By incorporating majority vote, the proposed 1-bit method suppresses inconsistent local sign updates, yielding smoother paths, smaller final formation-center error, and better robustness to communication uncertainty. These results confirm the central claim of the paper: properly designed finite-bit communication can preserve the essential convergence and coordination properties of distributed UAV swarm optimization while dramatically lowering communication overhead.

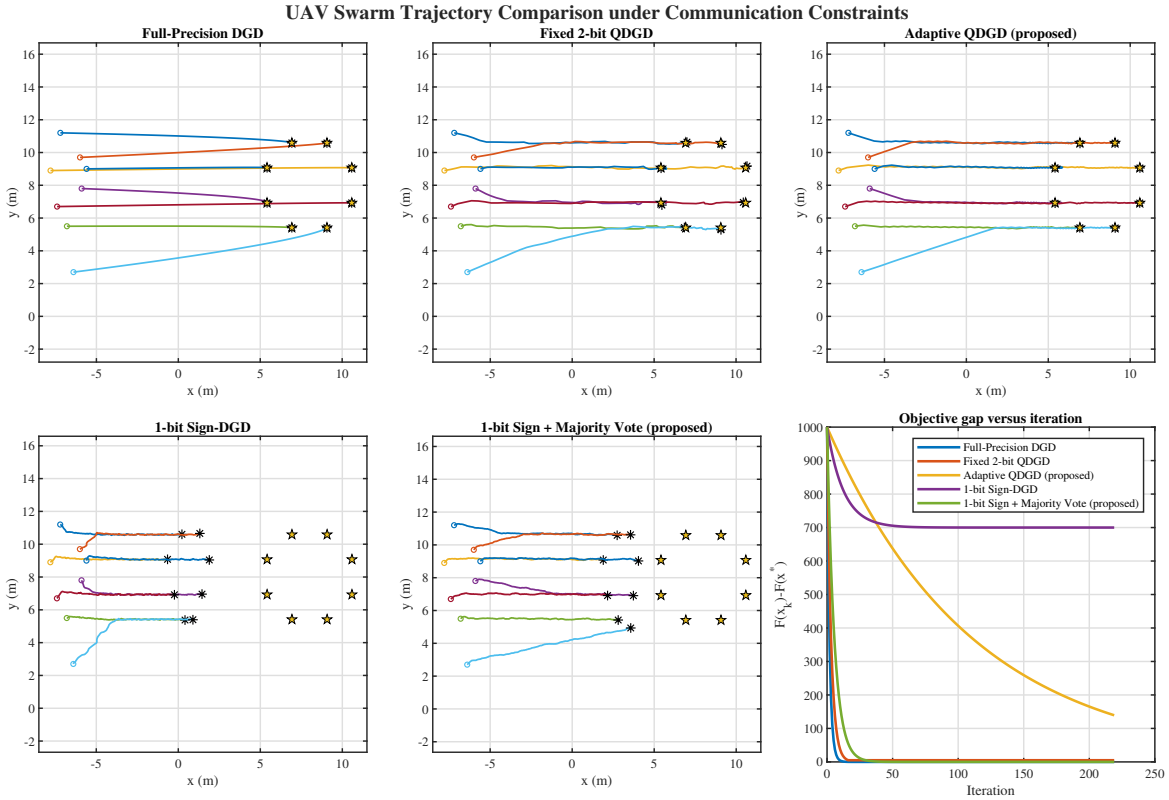

Figure 4. Performance evaluation of five methods for UAV swarm.

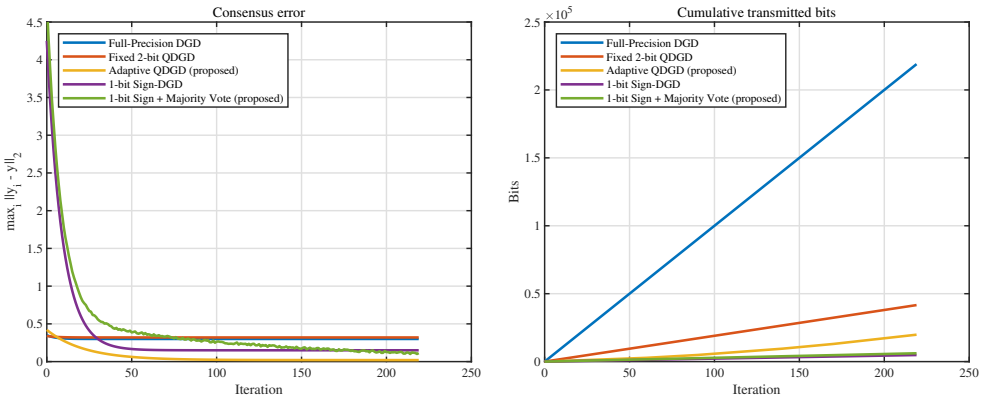

Figure 5. Performance evaluation of five methods for UAV swarm.

Communication–accuracy trade-off. To explicitly evaluate the trade-off between communication cost and optimization performance, we plot the objective gap as a function of cumulative transmitted bits.

As shown in Figures 6–8, we plot the objective gap as a function of cumulative transmitted bits to explicitly evaluate the communication–accuracy trade-off. The results show that adaptive quantization achieves near-optimal accuracy with significantly reduced communication cost, while 1-bit communication minimizes bandwidth usage at the expense of reduced accuracy.

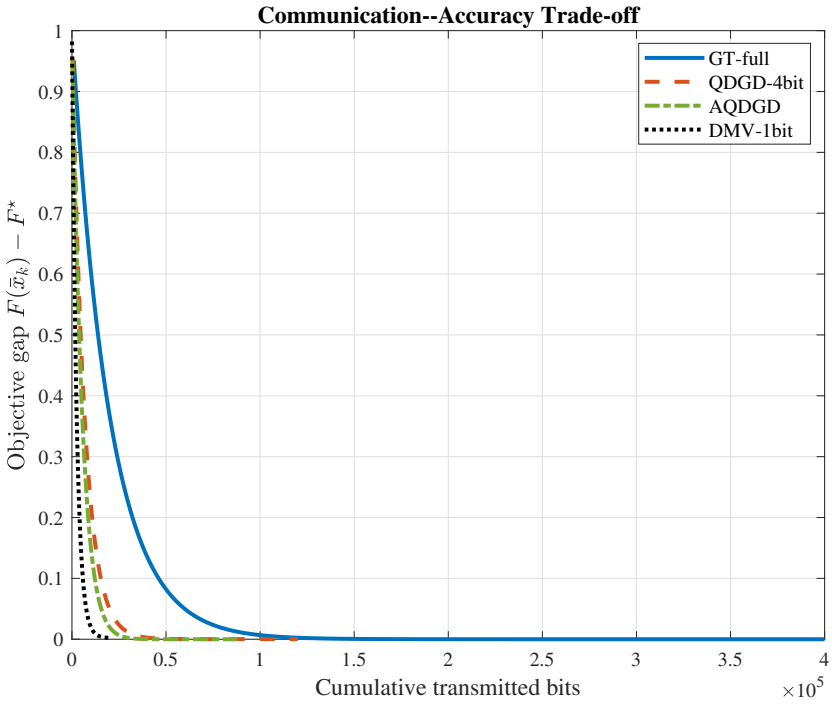

Figure 6. Communication–accuracy trade-off.

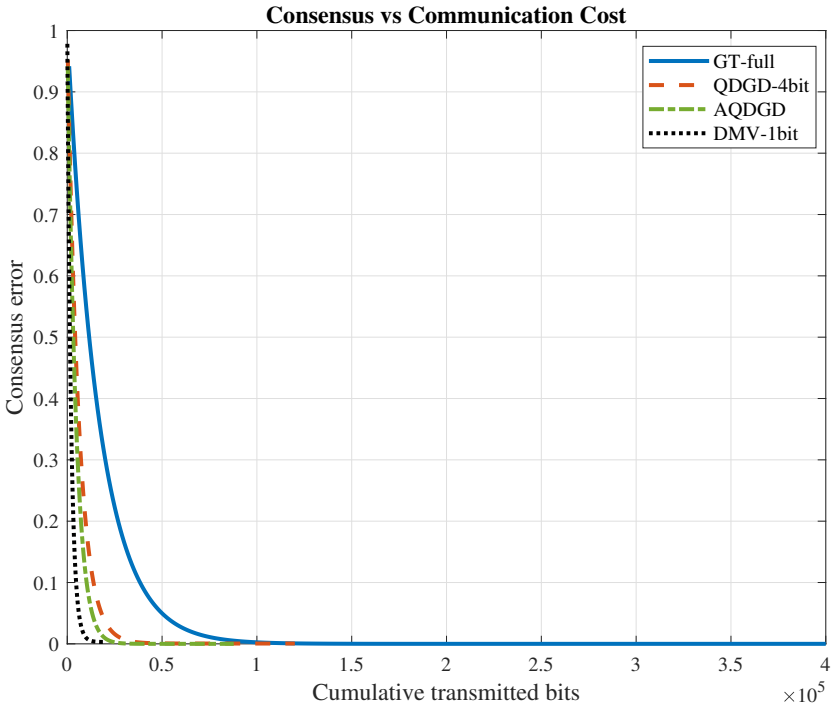

Figure 7. Consensus vs. communication cost.

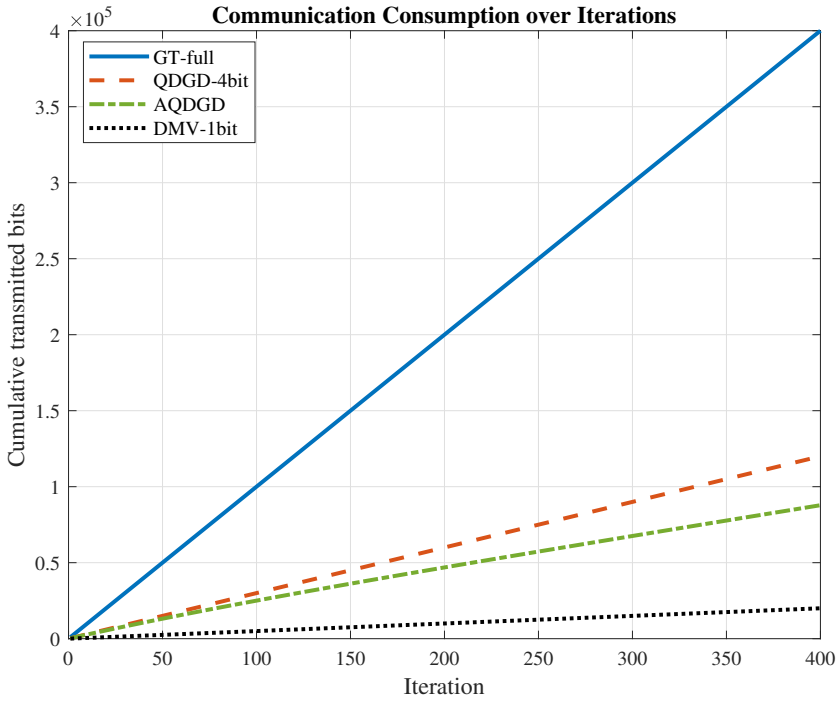

Figure 8. Communication consumption over iterations.

The results show that:

(1) Fixed-bit methods achieve fast initial convergence but saturate due to quantization error;
(2) Adaptive quantization achieves high accuracy with reduced communication cost;
(3) The 1-bit method provides the lowest communication cost at the expense of reduced accuracy.

This confirms the fundamental trade-off predicted by the theoretical analysis.

The simulation results support three concrete design recommendations. First, if the mission requires high terminal accuracy and the communication layer is moderately reliable, adaptive finite-bit gradient tracking is the preferred choice. Second, if the network budget is fixed and deterministic, a static multi-bit quantizer offers a tunable and predictable performance-versus-bias compromise. Third, if the network collapses into a severely bandwidth-constrained regime, 1-bit majority-vote coordination still preserves meaningful distributed optimization capability at extremely low payload.

4.4. Extended Experiments in Complex Scenarios

To further evaluate the practical applicability of the proposed methods, we consider more challenging scenarios beyond the quadratic convex setting, and the result is presented in Figure 9.

(1) Non-convex objective with obstacle avoidance. We augment the local cost functions with obstacle-avoidance penalty terms, resulting in a non-convex optimization problem. The additional terms create repulsive fields around obstacles, forcing the swarm to adapt its formation while avoiding collisions.

Although the theoretical guarantees in this paper are derived under strong convexity assumptions, the simulation results show that the proposed algorithms remain stable and exhibit consistent convergence behavior in this non-convex setting.

(2) Dynamic target tracking. We further consider a time-varying objective where the desired formation center evolves over time. The distributed optimization algorithms are implemented in a receding-horizon manner.

The results demonstrate that the proposed methods can effectively track the moving target under finite-bit communication constraints, with AQDGD showing the best trade-off between accuracy and communication cost.

These experiments highlight the robustness of the proposed framework beyond the idealized convex setting.

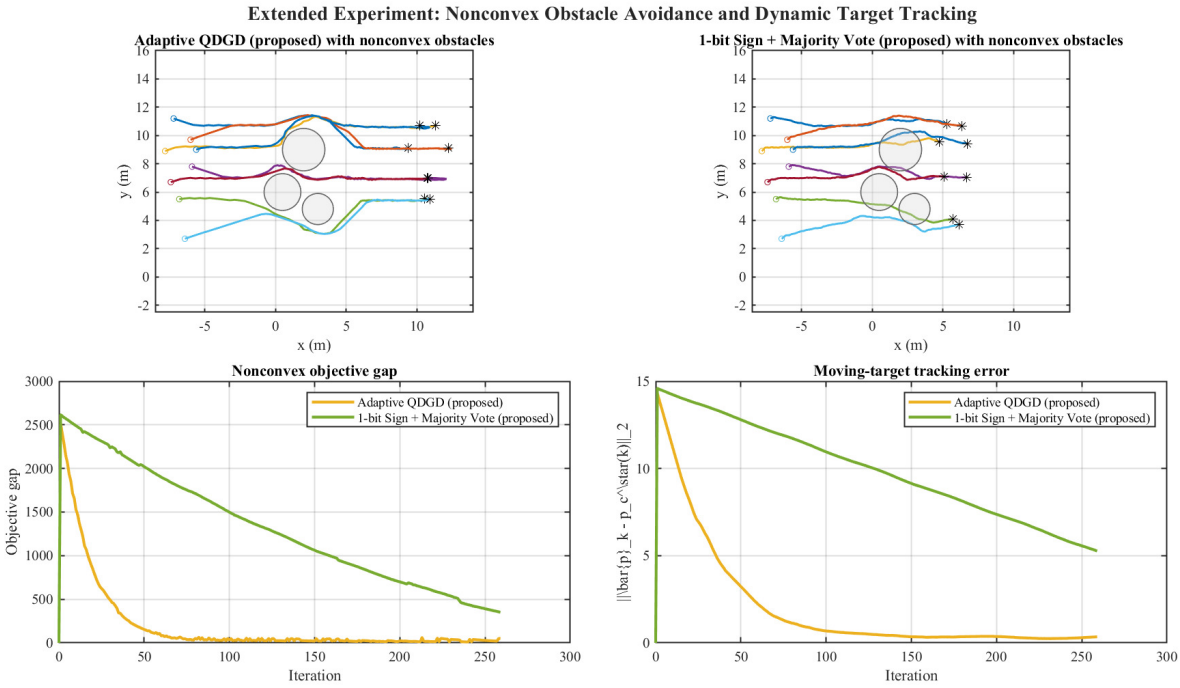

Figure 9. Extended experiment: nonconvex obstacle avoidance under finite-bit communication, where the gray circles represent obstacles.

5. Conclusions

This paper studied finite-bit distributed optimization for UAV swarms operating under bandwidth-limited communication. We developed a unified framework covering three practically relevant regimes: fixed finite-bit quantized gradient tracking, adaptive quantization with dynamic precision, and fully distributed 1-bit sign communication with majority vote. For the fixed-bit method, we proved linear convergence to a quantization-dependent neighborhood. For the adaptive method, we showed that exact linear convergence can be restored by driving the quantization distortion to zero at a geometric rate. For the 1-bit mode, we established expected linear contraction to a small neighborhood under a sign-based growth condition and majority-correctness assumption.

The resulting picture is practically useful: finite-bit communication need not destroy distributed optimization in UAV swarms, but the algorithm must be explicitly designed around the communication layer. Fixed-rate quantization provides deterministic bandwidth control, adaptive quantization offers the best nominal communication efficiency, and 1-bit majority-vote updates provide an emergency low-rate fallback. Future work can extend the present framework to time-varying directed graphs, intermittent communication, stochastic gradients, constrained optimization, and distributed nonconvex trajectory planning.

Author Contributions: Conceptualization, Y.Z. and Z.J.; methodology, Y.Z. and Z.J.; software, Y.Z.; validation Z.J.; formal analysis, Y.Z. and Z.J.; investigation, Y.Z.; writing—original draft preparation, Y.Z. and Z.J.; writing—review and editing, Y.Z. and Z.J.; visualization, Y.Z.; supervision, Y.Z. and Z.J.;

project administration, Y.Z. and Z.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 62403423, and the Natural Science Foundation of Zhejiang Province under Grant LMS25F030003.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Lemmas and Main Proofs

Appendix A.1. Detailed Proof of Lemma 1

Proof of Lemma 1. We derive the three rows of (9) more explicitly.

Averaging (7) gives

$$\bar{x}^{k+1} = \bar{x}^k - \alpha \bar{y}^k + \bar{\epsilon}_x^k, \quad \bar{\epsilon}_x^k := \frac{1}{n}(\mathbf{1}^\top \otimes I_d)(W \otimes I_d)\mathbf{e}_x^k.$$

Because W is doubly stochastic,

$$\left\|\bar{\epsilon}_x^k\right\| \le \frac{1}{\sqrt{n}}\left\|\mathbf{e}_x^k\right\| \le \frac{\delta_x}{\sqrt{n}}.$$

Next,

$$\bar{y}^k = \frac{1}{n}\sum_{i=1}^n y_i^k = \frac{1}{n}\sum_{i=1}^n \nabla f_i(x_i^k),$$

which follows from the initialization $\sum_i y_i^0 = \sum_i \nabla f_i(x_i^0)$ and induction on (8). Add and subtract $\nabla F(\bar{x}^k)$:

$$\bar{y}^k = \nabla F(\bar{x}^k) + \frac{1}{n}\sum_{i=1}^n \left(\nabla f_i(x_i^k) - \nabla f_i(\bar{x}^k)\right).$$

By Lipschitz continuity,

$$\left\|\bar{y}^k - \nabla F(\bar{x}^k)\right\| \le \frac{L}{n}\sum_{i=1}^n \left\|x_i^k - \bar{x}^k\right\| \le \frac{L}{\sqrt{n}}\left\|J_\perp \mathbf{x}^k\right\|.$$

To connect with the stacked tracker, define $\tilde{y}^k := J_\perp \mathbf{y}^k$. Then

$$\mathbf{y}^k = \mathbf{1} \otimes \bar{y}^k + \tilde{y}^k,$$

and hence the average primal update behaves like gradient descent with a perturbation composed of disagreement, tracking mismatch, and quantization.

Using the standard one-step inequality for gradient descent on a μ-strongly convex and L-smooth function,

$$\|u - \alpha \nabla F(u) - x^\star\| \le (1 - \alpha\mu)\|u - x^\star\|, \quad 0 < \alpha \le \frac{1}{L},$$

we obtain

$$\left\|\bar{x}^{k+1} - x^\star\right\| \le (1 - \alpha\mu)\left\|\bar{x}^k - x^\star\right\| + \frac{\alpha L}{\sqrt{n}}\left\|J_\perp \mathbf{x}^k\right\| + \frac{\alpha}{\sqrt{n}}\left\|J_\perp \mathbf{y}^k\right\| + \frac{\delta_x}{\sqrt{n}}.$$

This yields the first row of (9).

Projecting (7) onto the disagreement subspace gives

$$J_\perp \mathbf{x}^{k+1} = ((W - \tfrac{1}{n}\mathbf{1}\mathbf{1}^\top) \otimes I_d) J_\perp (\mathbf{x}^k + \mathbf{e}_x^k) - \alpha J_\perp \mathbf{y}^k.$$

Therefore,

$$\left\| J_\perp \mathbf{x}^{k+1} \right\| \le \sigma \left\| J_\perp \mathbf{x}^k \right\| + \sigma \delta_x + \alpha \left\| J_\perp \mathbf{y}^k \right\|.$$

To make the average-optimality dependence explicit, note that $\left\| \mathbf{y}^k \right\|$ can be bounded by

$$\left\| \mathbf{y}^k \right\| \le \left\| J_\perp \mathbf{y}^k \right\| + \sqrt{n} \left\| \bar{y}^k \right\| \le \left\| J_\perp \mathbf{y}^k \right\| + L\sqrt{n} \left\| \bar{x}^k - x^\star \right\| + L \left\| J_\perp \mathbf{x}^k \right\| + B_g \sqrt{n}.$$

Absorbing the constant heterogeneity term into the perturbation and grouping coefficients yields the second row.

Apply $J_\perp$ to (8):

$$J_\perp \mathbf{y}^{k+1} = ((W - \tfrac{1}{n}\mathbf{1}\mathbf{1}^\top) \otimes I_d) J_\perp (\mathbf{y}^k + \mathbf{e}_y^k) + J_\perp \big(\nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^k) \big).$$

Hence,

$$\left\| J_\perp \mathbf{y}^{k+1} \right\| \le \sigma \left\| J_\perp \mathbf{y}^k \right\| + \sigma \delta_y + L \left\| \mathbf{x}^{k+1} - \mathbf{x}^k \right\|.$$

From (7),

$$\mathbf{x}^{k+1} - \mathbf{x}^k = ((W - I_n) \otimes I_d) \mathbf{x}^k + (W \otimes I_d) \mathbf{e}_x^k - \alpha \mathbf{y}^k.$$

Taking norms and using the previous bounds gives

$$\left\| \mathbf{x}^{k+1} - \mathbf{x}^k \right\| \le c_x \left\| J_\perp \mathbf{x}^k \right\| + c_o \left\| \bar{x}^k - x^\star \right\| + c_y \left\| J_\perp \mathbf{y}^k \right\| + c_q \delta_x,$$

for suitable constants depending on L, σ, α, and n. Substituting this expression into the previous inequality yields the third row of (9). □

Appendix A.2. Further Discussion on Adaptive Quantization

The adaptive theorem was stated abstractly through the error sequence Δ_k. Here we explain why the bit schedule (18) is a natural implementation.

Near the optimum, the local iterate increments satisfy

$$\left\| x_i^k - x_i^{k-1} \right\| + \left\| y_i^k - y_i^{k-1} \right\| = O(\tilde{\rho}^{k/2}),$$

so the surrogate $\widehat{V}_k$ is proportional to the true Lyapunov quantity V_k under mild regularity. Consequently,

$$\widehat{V}_k \approx c_V V_k.$$

Using (18), the quantization distortion satisfies

$$\Delta_k^2 \lesssim \frac{\gamma_q^2 R_k^2}{2^{2b_k}} \approx \frac{\gamma_q^2 R_k^2}{\Gamma} (\widehat{V}_k + \varepsilon).$$

If the range R_k is itself maintained proportional to the current iterate spread, then $R_k^2 \approx c_R V_k$ and therefore $\Delta_k^2 = O(V_k^2)$, which is stronger than the sufficient condition (14). In this regime the perturbation becomes negligible compared with the nominal contraction, and exact linear convergence follows.

A practical implementation can maintain R_k via synchronized encoder–decoder windows or via subtractive quantization around a predicted reference state. These implemen-

tation details are omitted in the main text because the essential convergence mechanism is already captured by the abstract distortion bound.

Appendix A.3. Sketch Proof for Theorem 3

Proof of Theorem 3. We provide a more explicit derivation of the one-step bound (26). Let

$$\bar{m}_g^k := \frac{1}{n}\sum_{i=1}^{n} m_{i,g}^k, \qquad \bar{m}_c^k := \frac{1}{n}\sum_{i=1}^{n} m_{i,c}^k.$$

Averaging (22) yields

$$\bar{x}^{k+1} = \bar{x}^k - \alpha \bar{m}_g^k - \beta \bar{m}_c^k.$$

Using L-smoothness,

$$F(\bar{x}^{k+1}) \le F(\bar{x}^k) - \alpha\left\langle \nabla F(\bar{x}^k), \bar{m}_g^k \right\rangle - \beta\left\langle \nabla F(\bar{x}^k), \bar{m}_c^k \right\rangle + \frac{L}{2}\left\| \alpha \bar{m}_g^k + \beta \bar{m}_c^k \right\|^2.$$

The consensus majority term $\bar{m}_c^k$ acts primarily on disagreement and is approximately orthogonal to the global descent direction, so its contribution to the averaged objective can be bounded as a higher-order perturbation. Under Assumption 4, the gradient-majority term satisfies

$$\mathbb{E}\left[\left\langle \nabla F(\bar{x}^k), \bar{m}_g^k \right\rangle \mid \mathcal{F}_k\right] \ge (2p_g - 1)\left\| \nabla F(\bar{x}^k) \right\|_1 - c_d \left\| J_\perp \mathbf{x}^k \right\|,$$

for a disagreement-coupling constant $c_d > 0$. By Assumption 3,

$$\left\| \nabla F(\bar{x}^k) \right\|_1 \ge \kappa_s (F(\bar{x}^k) - F^\star).$$

Therefore,

$$\mathbb{E}[F(\bar{x}^{k+1}) - F^\star \mid \mathcal{F}_k] \le (1 - c_f \alpha)(F(\bar{x}^k) - F^\star) + c_{fd}\alpha \left\| J_\perp \mathbf{x}^k \right\| + c_{f2}(\alpha^2 + \beta^2).$$

For the disagreement term, write the stacked update as

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha m_g^k - \beta m_c^k,$$

where $m_g^k, m_c^k \in \mathbb{R}^{nd}$. The disagreement projector gives

$$J_\perp \mathbf{x}^{k+1} = J_\perp \mathbf{x}^k - \alpha J_\perp m_g^k - \beta J_\perp m_c^k.$$

Taking conditional expectation and using the majority-correctness property of m_c^k yields

$$\mathbb{E}\left[\left\| J_\perp \mathbf{x}^{k+1} \right\|^2 \mid \mathcal{F}_k\right] \le (1 - c_c\beta)\left\| J_\perp \mathbf{x}^k \right\|^2 + c_{cx}\alpha^2 + c_{c2}\beta^2 + c_{cf}\alpha(F(\bar{x}^k) - F^\star).$$

Choosing a weight λ that is large enough and summing the last two inequalities gives (26). □

Appendix A.4. Additional Implementation Notes

In a practical UAV implementation, each communication packet can contain the following payload.

1. Fixed-bit QDGD: quantized primal vector, quantized tracker vector, sequence counter, and optional checksum.
2. AQDGD: same as above plus the current bit index or quantizer mode identifier.

3. DMV-1bit: packed sign bits for the gradient and consensus terms, which can be extremely compact.

Because the algorithms are fully neighbor-to-neighbor, they do not require a central fusion node. This is important for swarm resilience. In addition, the adaptive scheme can be implemented without globally synchronized clocks as long as neighboring agents agree on the active quantizer resolution for each packet.

References

1. Chowdhury, M.M.U.; Maeng, S.J.; Bulut, E.; Güvenç, İ. 3-D Trajectory Optimization in UAV-Assisted Cellular Networks Considering Antenna Radiation Pattern and Backhaul Constraint. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *56*, 3735–3750. [CrossRef]
2. Bao, T.; Wang, H.; Wang, W.J.; Yang, H.C.; Hasna, M.O. Secrecy Outage Performance Analysis of UAV-Assisted Relay Communication Systems with Multiple Aerial and Ground Eavesdroppers. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 2592–2600. [CrossRef]
3. Hung, H.A.; Hsu, H.H.; Cheng, T.H. Optimal Sensing for Tracking Task by Heterogeneous Multi-UAV Systems. *IEEE Trans. Control Syst. Technol.* **2024**, *32*, 282–289. [CrossRef]
4. Gong, X.; Gui, J.; Chen, Y.; Yang, X.; Yu, W.; Huang, T. Resilient Human-in-the-Loop Formation-Tracking of Multi-UAV Systems Against Byzantine Attacks. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 3797–3809. [CrossRef]
5. Hai, X.; Tan, L.; Feng, Q.; Duan, H.; Wen, C. Replanning-Oriented Framework for Efficient Real-Time Decision-Making in Multi-UAV Systems. *IEEE Trans. Ind. Inform.* **2025**, *21*, 5127–5137. [CrossRef]
6. Ma, J.; Chen, X.; Wen, G.; Wang, J.; Zhao, F.; Qiu, J. Dynamic Memory Event-Triggered Lag Consensus of Multi-UAV Systems with Hybrid Attacks Over Stochastic Switching Topology. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 16999–17009. [CrossRef]
7. Savkin, A.V.; Huang, H. Multi-UAV Navigation for Optimized Video Surveillance of Ground Vehicles on Uneven Terrains. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 10238–10242. [CrossRef]
8. Cao, X.; Liu, L. A Multi-Timescale Method for State of Charge Estimation for Lithium-Ion Batteries in Electric UAVs Based on Battery Model and Data-Driven Fusion. *Drones* **2025**, *9*, 247. [CrossRef]
9. Jin, Z.; Bai, L.; Wang, Z.; Zhang, P. Self-Triggered Distributed Formation Control of Fixed-Wing Unmanned Aerial Vehicles Subject to Velocity and Overload Constraints. *IEEE Trans. Autom. Sci. Eng.* **2024**, *21*, 4082–4093. [CrossRef]
10. Chang, Z.; Zong, G.; Wang, W.; Yue, M.; Zhao, X. Formation Control and Obstacle Avoidance Design for Networked USV Swarm with Exogenous Disturbance Under Intermittent Communication. *IEEE Trans. Netw. Sci. Eng.* **2025**, *12*, 3234–3243. [CrossRef]
11. Jin, Z.; Li, H.; Qin, Z.; Wang, Z. Gradient-Free Cooperative Source-Seeking of Quadrotor Under Disturbances and Communication Constraints. *IEEE Trans. Ind. Electron.* **2025**, *72*, 1969–1979. [CrossRef]
12. Alghunaim, S.A.; Ryu, E.K.; Yuan, K.; Sayed, A.H. Decentralized Proximal Gradient Algorithms with Linear Convergence Rates. *IEEE Trans. Autom. Control* **2021**, *66*, 2787–2794. [CrossRef]
13. Nedić, A.; Olshevsky, A.; Shi, W. Achieving Geometric Convergence for Distributed Optimization Over Time-Varying Graphs. *SIAM J. Optim.* **2017**, *27*, 2597–2633. [CrossRef]
14. Mancino-Ball, G.; Xu, Y.; Chen, J. A Decentralized Primal-Dual Framework for Non-Convex Smooth Consensus Optimization. *IEEE Trans. Signal Process.* **2023**, *71*, 525–538. [CrossRef]
15. Asano, H.; Okada, H.; Naila, C.B.; Katayama, M. Communication-Aware Flight Algorithm for UAVs in Delay-Tolerant Aerial Networks. In *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*; IEEE: New York, NY, USA, 2023; pp. 287–292.
16. Liu, Z.; Liu, X.; Liu, Y.; Leung, V.C.M.; Durrani, T.S. UAV Assisted Integrated Sensing and Communications for Internet of Things: 3D Trajectory Optimization and Resource Allocation. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 8654–8667. [CrossRef]
17. Dou, W.; Yang, P.; Zhang, Z.; Wang, Z. Cooperative Multi-UAV Search for Prioritized Targets Under Constrained Communications. *Drones* **2025**, *9*, 855. [CrossRef]
18. Tang, M.; Feng, C.; Quek, T.Q.S. Decentralized Semantic Communication and Cooperative Tracking Control for a UAV Swarm Over Wireless MIMO Fading Channels. *IEEE Trans. Veh. Technol.* **2026**, *75*, 3354–3359. [CrossRef]
19. Nedić, A.; Ozdaglar, A. Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Autom. Control* **2009**, *54*, 48–61. [CrossRef]
20. Koloskova, A.; Stich, S.U.; Jaggi, M. Decentralized stochastic optimization and gossip algorithms with compressed communication. In *Proceedings of the International Conference on Machine Learning*; JMLR: Cambridge, MA, USA, 2019; pp. 3478–3487.
21. Liao, Y.; Li, Z.; Huang, K.; Pu, S. A compressed gradient tracking method for decentralized optimization with linear convergence. *IEEE Trans. Autom. Control* **2023**, *67*, 5622–5629. [CrossRef]

22. Xiong, Y.; Wu, L.; You, K.; Xie, L. Quantized distributed gradient tracking algorithm with linear convergence in directed networks. *arXiv* **2021**, arXiv:2104.03649. [CrossRef]
23. Doan, T.T.; Maguluri, S.T.; Romberg, J. Fast convergence rates of distributed subgradient methods with adaptive quantization. *IEEE Trans. Autom. Control* **2021**, *66*, 2193–2208. [CrossRef]
24. Liao, Y.; Li, Z.; Huang, K.; Pu, S. Compressed gradient tracking methods for decentralized optimization with linear convergence. *arXiv* **2022**, arXiv:2203.03748. [CrossRef]
25. Chen, L.; Wen, G.; Liu, H.; Yu, W.; Cao, J. Compressed gradient tracking algorithm for distributed aggregative optimization. *IEEE Trans. Autom. Control* **2024**, *69*, 6576–6591. [CrossRef]
26. Yao, J.; Li, X.; Zhang, Y.; Chen, K.; Zhang, D.; Ji, J. Autonomous navigation control of tracked unmanned vehicle formation in communication restricted environment. *IEEE Trans. Veh. Technol.* **2024**, *73*, 16063–16075. [CrossRef]
27. Li, Q.; Zhou, W.; Zheng, X. Distributed Learning in Intelligent Transportation Systems: A Survey. *Information* **2024**, *15*, 550. [CrossRef]
28. Liu, Y.; Li, J.; Jin, Z. Trajectory Tracking Control for Reaction–Diffusion System with Time Delay Using P-Type Iterative Learning Method. *Actuators* **2021**, *10*, 186. [CrossRef]
29. Jang, G.; Kwon, Y. From Manned to Unmanned Helicopters: A Transformer-Driven Cross-Scale Transfer Learning Framework for Vibration-Based Anomaly Detection. *Actuators* **2026**, *15*, 38. [CrossRef]
30. Fang, X.; Chen, L.; Yin, H.; Chen, X.; Wang, W. Communication efficient federated learning with quantization-aware training design. *IEEE Trans. Mach. Learn. Commun. Netw.* **2026**, *4*, 45–59. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.