High-Efficiency Distributed Nonconvex Optimization Design with Certified Stopping Rule

Ling Yao†, Xiangyun Rao†, Tao Xu†, Pangkit Fong† and Jianping He†

***Abstract*—Distributed nonconvex optimization with certification constraints is a crucial demand in multi-agent networks, especially sparse ones. Most existing methods require high communication or computational complexity, and there is no rigorously certified rule for terminating redundant iterations. To address these issues, we develop a Chebyshev-Accelerated Surrogate Optimization (CASO) algorithm for distributed optimization of univariate objectives. Inspired by Chebyshev Polynomial-based Consensus Algorithm (CPCA), CASO employs Chebyshev approximation to build local surrogates. To synchronize these surrogates, it applies accelerated coefficient consensus. With a distributedly verifiable stopping rule, CASO provides an ε-global optimality guarantee, and reduces communication costs across sparse networks. Finally, it utilizes algebraic colleague-matrix minimization to extract the global minimizer. Simulations verify the effectiveness of CASO.**

***Index Terms*—Distributed Optimization, Nonconvex Optimization, Chebyshev Surrogate, Sparse Network System.**

I. Introduction

Distributed optimization over multi-agent networks has attracted significant attention because of its robustness, privacy preservation, and communication scalability [1]. It has been widely used in robotic coordination, vehicle platoons, and smart grids [2]–[4]. In this paper, we study a class of constrained distributed nonconvex optimization problems with univariate objective functions. Typical examples include a shared coupling gain, a safety headway, or a calibration coefficient [5]. In these problems, local objectives and constraints may differ across agents, and the resulting global problem is often nonconvex. For such tasks, convergence to a local stationary point is not sufficient. A globally certified solution is required, especially in safety-critical applications.

Although there has been substantial progress in distributed optimization, most existing methods for nonconvex problems still follow a local-search paradigm. Representative first-order, gradient-tracking, successive-convex-approximation, and zeroth-order methods can usually guarantee only stationarity, asymptotic consensus, or convergence to local critical points [6]–[13]. These guarantees are not strong enough for the class of problems considered here. In our setting, the main goal is not to find a stationary point, but to compute a solution with a global optimality certificate.

This work was supported in part by the National Natural Science Foundation of China under Grant 62373247.

†: The Department of Automation, School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai Key Laboratory of Perception and Control in Industrial Network Systems, Shanghai 200240, China. E-mails: {zero-in-ur-area, raoxiangyun, Zerken, fpjgaoge, jphe}@sjtu.edu.cn

This limitation suggests a different route. Instead of optimizing the original nonconvex objective by local descent, one can first build a global surrogate and then optimize that surrogate. For one-dimensional Lipschitz functions, Chebyshezv approximation is a natural choice. It provides high approximation accuracy, compact coefficient representations, and a convenient basis for certified global optimization. A notable example is the Chebyshev Proxy and Consensus-based Algorithm [5], [14]. CPCA replaces each local objective with a Chebyshev polynomial surrogate and exchanges surrogate coefficients across the network, rather than gradients or function values. This surrogate-based pipeline makes distributed global optimization possible for general univariable nonconvex Lipschitz objectives. The core idea of using Chebyshev surrogates directly motivates our work, and it serves as the starting point of our approach.

Different from CPCA, we focus on its communication limitations, especially over sparse networks [15]. In CPCA, surrogate coefficients are synchronized by standard average consensus. This step can be slow when the network is poorly connected. In addition, CPCA does not provide a stopping rule that can be verified in a distributed manner. As a result, one often has to use conservative iteration bounds, which may cause unnecessary communication.

To address these issues, we propose the Chebyshev-Accelerated Surrogate Optimization (CASO) algorithm. CASO leverages the function-approximation strengths of polynomial surrogates but extends the paradigm with two key technical components. (i) Chebyshev filtering [16] for near-optimal consensus acceleration on sparse graphs; and (ii) a distributedly verifiable stopping criterion for adaptive convergence. In this way, CASO preserves the benefit of surrogate-based global optimization while achieving higher communication efficiency on sparse graphs. The core contributions of this paper are summarized as follows.

1) **CASO Architecture.** We develop a distributed surrogate-optimization algorithm (CASO) for distributed optimization of univariate objectives. This algorithm employs Chebyshev approximation to build local surrogates. To synchronize these surrogates, it applies accelerated coefficient consensus. Finally, it utilizes algebraic colleague-matrix minimization to extract the global minimizer.
2) **Certified Adaptive Stopping.** We propose a certified adaptive stopping mechanism for the accelerated consensus stage. The stopping rule is distributedly verifiable and helps avoid unnecessary communication rounds, which is especially important in sparse networked systems.
3) **End-to-End Guarantees and Optimal Communication.**

We provide a rigorous analysis of the output error for the algorithm. This analysis establishes an ε-global optimality guarantee. Using CASO, Stage 2 achieves a square-root spectral-gap dependence in the coefficient-mixing complexity. Simulations verify the effectiveness of CASO.

The remainder of this work is organized as follows. Section II presents the problem formulation and preliminaries. Section III elaborates on Chebyshev-Accelerated Surrogate Optimization algorithm. Section IV presents simulation results and relevant analysis. Finally, Section V concludes the paper.

II. PROBLEM FORMULATION AND PRELIMINARIES

Consider a connected undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with N nodes. Each node $i \in \mathcal{V}$ has access only to its local objective $f_i : X \to \mathbb{R}$, where $X = [a, b] \subset \mathbb{R}$ is a compact interval. The network aims to solve

$$\min_{x \in X} \ f(x) = \frac{1}{N} \sum_{i=1}^{N} f_i(x), \tag{1}$$

where x is a globally shared decision variable. Specifically, we define the global minimum of $f(x)$ as f^* and let $x^* \in \arg\min_{x \in X} f(x)$ denote an arbitrary global minimizer. Let $\tilde{x}_i$ denote the distributed output of the algorithm. The objective of the algorithm is to guarantee the output error

$$|f(\tilde{x}_i) - f^*| \leq \varepsilon, \qquad \forall i \in \mathcal{V}. \tag{2}$$

Now we have the following assumptions.

Assumption 1 (Function regularity)**.** *Each local objective f_i is Lipschitz continuous on X, i.e., there exists $L_i > 0$ such that*

$$|f_i(x) - f_i(y)| \leq L_i \, |x - y| \,, \qquad \forall x, y \in X. \tag{3}$$

Assumption 2 (Mixing matrix)**.** *The network uses a nonnegative symmetric doubly stochastic mixing matrix $W \in \mathbb{R}^{N \times N}$ compatible with $\mathcal{G}$, where $w_{ii} > 0$ and $w_{ij} = 0$ whenever $i \neq j$ and $\{i, j\} \notin \mathcal{E}$. Its eigenvalues satisfy*

$$1 = \lambda_1(W) > \lambda_2(W) \geq \cdots \geq \lambda_N(W) > -1. \tag{4}$$

Remark 1 (Feasibility of Assumption 2)**.** *Assumption 2 is satisfied on any connected undirected graph by, for example, the Metropolis weights, which is defined by*

$$w_{ij} = \frac{1}{1 + \max\{d_i, d_j\}}, \quad j \in \mathcal{N}_i, \qquad w_{ii} = 1 - \sum_{j \in \mathcal{N}_i} w_{ij},$$

where $\mathcal{N}_i$ denotes the neighbor set of i, and all other entries are set to zero. The resulting matrix is nonnegative, symmetric, doubly stochastic, and has a positive diagonal. Connectivity and positive diagonal entries imply $\lambda_2(W) < 1$ and $\lambda_N(W) > -1$, respectively. Each node can construct its weights using its own degree and its neighbors' degrees. However, the exact spectral values are not verifiable from one-hop information alone and are estimated in Stage 0 or computed offline.

Considering that the proposed algorithm uses fully distributed information without any convexity assumption, achieving the strict global certification in (2) is challenging.

III. CASO ALGORITHM AND THEORETICAL ANALYSIS

In this section, we propose a distributed surrogate-optimization algorithm (CASO). We design CASO as a four-stage algorithm with certified adaptive stopping mechanism. We provide a rigorous analysis of the output error, which establishes an ε-global optimality guarantee. Using CASO, the accelerated consensus is achieved with a square-root spectral-gap communication complexity.

A. CASO algorithm

CASO is a distributed surrogate-optimization algorithm. Specifically, we design CASO as a four-stage algorithm. As shown in Fig. 1, CASO begins with Stage 0, using distributed power iteration [17] to estimate the spectral quantities needed to tune acceleration. Each node then constructs a local Chebyshev surrogate in Stage 1. In Stage 2, CASO iteratively synchronizes the surrogate coefficients through Chebyshev-accelerated consensus with certified stopping. Finally, CASO extracts the global minimizer algebraically in Stage 3. Algorithm 1 summarizes the complete algorithm. In the following subsections, we detail the methodology of CASO.

B. Estimation of Network Spectral Gap

In Stage 0, CASO obtains the spectral quantities of the mixing matrix needed to configure the Chebyshev acceleration parameters. Specifically, let $z^t \in \mathbb{R}^N$ denote the network state at time step t. The distributed power iteration is given by

$$z^{t+1} = W z^t. \tag{5}$$

We define the corresponding Rayleigh quotient as

$$\lambda^t = \frac{(z^t)^\top W z^t}{(z^t)^\top z^t}. \tag{6}$$

Applied on the non-consensus subspace, distributed power iteration can be used to obtain spectral estimates for tuning the accelerated consensus. In practice, estimating $\lambda_2(W)$ and $\lambda_N(W)$ requires an additional distributed evaluation of Rayleigh quotients, and if needed, shifted or sign-flipped operators. In this paper, we assume that the spectral quantities $\lambda_2(W)$ and $\lambda_N(W)$ needed for acceleration tuning are available after Stage 0.

Remark 2 (Cost of Stage 0)**.** *Stage 0 is a preprocessing step for acceleration tuning, so its cost is separated from the Stage 2 coefficient-mixing complexity. For a fixed network topology, it can be performed once and amortized across repeated optimization tasks on the same graph. When centralized preprocessing is acceptable, the spectral parameters may also be computed offline. In Table I, $T_{\mathrm{spec}}(\eta)$ denotes the communication/computation cost of the distributed spectral-estimation routine used to reach precision η.*

C. Chebyshev Polynomial Surrogate

In Stage 1, CASO constructs Chebyshev polynomial surrogate from local zero-order queries, replacing the objective function.

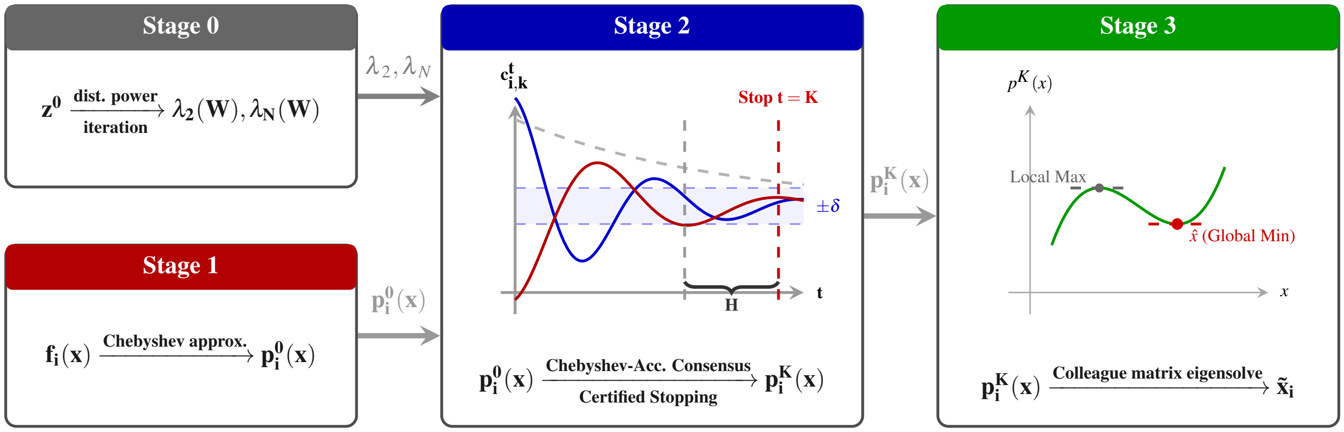

Fig. 1. Visual overview of the four-stage CASO algorithm, from distributed spectral estimation to algebraic extraction of the final minimizer.

1) Construction of Chebyshev Polynomial Surrogate: Define the normalized variable as

$$\xi(x) = \frac{2x - a - b}{b - a} \in [-1, 1]. \tag{7}$$

Let $T_j(\cdot)$ denote the Chebyshev polynomial of the first kind, and let $p_i^{(r)}(x)$ denote the local surrogate held by node i at the r-th adaptive interpolation step. Let

$$S_q = \left\{ \cos\left(\frac{(2\ell - 1)\pi}{2q} \right) \right\}_{\ell=1}^{q}$$

denote the q Chebyshev nodes on $[-1, 1]$, mapped back to $X = [a, b]$ through the inverse of $\xi(x)$ when needed. Each node initializes its local polynomial degree by $m_i^{(0)} = 2$. At adaptive step r, node i constructs a degree-$m_i^{(r)}$ surrogate

$$p_i^{(r)}(x) = \sum_{j=0}^{m_i^{(r)}} c_{i,j}^{(r)} T_j(\xi(x)). \tag{8}$$

In the adaptive Chebyshev interpolation procedure, node i then evaluates the approximation error on the newly added Chebyshev nodes in $S_{2m_i^{(r)}} \setminus S_{m_i^{(r)}}$.

If

$$\max_{x \in S_{2m_i^{(r)}} \setminus S_{m_i^{(r)}}} |f_i(x) - p_i^{(r)}(x)| \le \bar{\varepsilon}_{\text{app}}, \tag{9}$$

where $\bar{\varepsilon}_{\text{app}}$ denotes the tolerance of surrogate approximation error ε_{app} (which will be specified in Sec. III-F), then the adaptive construction stops and the resulting surrogate is taken as the Stage 1 output. Otherwise, CASO doubles the degree and repeats the procedure.

Let $m = \max_{i \in \mathcal{V}} m_i$ be the maximum final degree across all nodes after Stage 1. Each node zero-pads its final coefficient sequence to dimension $m + 1$ and defines $\mathbf{c}_i^0 := [c_{i,0}^0, \ldots, c_{i,m}^0]^\top \in \mathbb{R}^{m+1}$. The resulting initial surrogate at node i is written as

$$p_i^0(x) = \sum_{j=0}^{m} c_{i,j}^0 T_j(\xi(x)). \tag{10}$$

Accordingly, the network-average surrogate is

$$\bar{p}(x) = \frac{1}{N} \sum_{i=1}^{N} p_i^0(x) = \sum_{j=0}^{m} \bar{c}_j T_j(\xi(x)), \tag{11}$$

where $\bar{c}_j := \frac{1}{N} \sum_{i=1}^{N} c_{i,j}^0$ is the j-th entry of $\bar{\mathbf{c}}$, which denotes the network-average coefficient vector. Specifically, $\bar{\mathbf{c}} := \frac{1}{N} \sum_{i=1}^{N} \mathbf{c}_i^0 \in \mathbb{R}^{m+1}$ so that $\bar{\mathbf{c}} = [\bar{c}_0, \ldots, \bar{c}_m]^\top$.

2) Approximation Performance Analysis: To demonstrate the performance of approximation by Chebyshev polynomial surrogate, we give the following theorem.

Theorem 1 (Approximation Performance)**.** *Using approximation by Chebyshev polynomial surrogate gives*

$$\lim_{m \to \infty} \varepsilon_{\text{app}} = 0. \tag{12}$$

Proof. By Assumption 1 and Jackson's Theorem [18], we have

$$E_m(f_i) = \inf_{q \in \Pi_m} \|f_i - q\|_\infty \le \frac{C_J L_i (b - a)}{m}. \tag{13}$$

Chebyshev interpolation [19] is near-optimal for it, thus

$$\|f_i - p_i^0\|_\infty \le (1 + \Lambda_m) E_m(f_i), \qquad \Lambda_m = O(\log m). \tag{14}$$

Averaging gives

$$\varepsilon_{\text{app}} = \max_{x \in X} |\bar{p}(x) - f(x)| \le \frac{1}{N} \sum_{i=1}^{N} \|p_i^0 - f_i\|_\infty, \tag{15}$$

Hence, when m goes to ∞, ε_{app} reduces to 0. $\square$

Theorem 1 demonstrates the effectiveness of approximation by Chebyshev polynomial surrogate. That is, we get an efficient approximation in Stage 2 ideally.

D. Chebyshev-Accelerated Surrogate Consensus

In Stage 2, CASO synchronizes the surrogates constructed in Stage 1 using Chebyshev-accelerated surrogate consensus.

1) Chebyshev-Accelerated Surrogate Consensus: Let $\mathbf{c}_i^t := [c_{i,0}^t, \ldots, c_{i,m}^t]^\top \in \mathbb{R}^{m+1}$ be the coefficient vector held by node i at consensus iteration t. Stack the node states into the matrix gives

$$P^t = \left[(\mathbf{c}_1^t)^\top \quad \cdots \quad (\mathbf{c}_N^t)^\top\right]^\top \in \mathbb{R}^{N\times(m+1)}. \tag{16}$$

CASO uses the three-term recurrence, which is given by

$$P^{t+1} = (1+\gamma_t)WP^t - \gamma_t P^{t-1}, \tag{17}$$

with

$$\begin{aligned} \sigma &= \frac{\lambda_2(W) - \lambda_N(W)}{2 - \lambda_2(W) - \lambda_N(W)}, \\ \gamma_0 &= 0, \qquad \gamma_{t+1} = \frac{1}{4\sigma^{-2} - \gamma_t}. \end{aligned} \tag{18}$$

CASO employs a matrix-valued analogue of Chebyshev polynomial filtering in Stage 2. It accelerates the synchronization of the entire coefficient matrix simultaneously.

2) Consensus Error Analysis: Now we give the following lemma.

Lemma 1. *The recurrence* (17) *preserves the network average and satisfies*

$$\left\|P^t - \mathbf{1}\bar{\mathbf{c}}^\top\right\|_\infty \le 2\rho_{\mathrm{AC}}^t \left\|P^0 - \mathbf{1}\bar{\mathbf{c}}^\top\right\|_\infty, \tag{19}$$

where

$$\rho_{\mathrm{AC}} = \frac{\sigma}{1+\sqrt{1-\sigma^2}}. \tag{20}$$

Consequently, the number of communication rounds needed to reach coefficient tolerance δ scales as

$$K = O\left(\frac{\log(C_0/\delta)}{\sqrt{1-\lambda_2(W)}}\right), \tag{21}$$

where $C_0 = \left\|P^0 - \mathbf{1}\bar{\mathbf{c}}^\top\right\|_\infty$. Since the certified stopping rule requires $(m+1)\delta \le \bar{\varepsilon}_{\mathrm{con}}$ and $\bar{\varepsilon}_{\mathrm{con}} = \Theta(\varepsilon)$ under the error-budget allocation, the Stage 2 communication complexity can be rewritten as $O(\log(m/\varepsilon)/\sqrt{1-\lambda_2(W)})$.

Proof sketch. Because W is doubly stochastic and $P^{-1} = P^0$, multiplying (17) by $N^{-1}\mathbf{1}^\top$ and using induction gives

$$\frac{1}{N}\mathbf{1}^\top P^t = \bar{\mathbf{c}}^\top, \qquad t \ge 0. \tag{22}$$

Hence, $\mathbf{1}\bar{\mathbf{c}}^\top$ in (19) is exactly the invariant consensus target. The proof follows the minimax property of Chebyshev polynomials on the spectral interval $[\lambda_N(W), \lambda_2(W)]$ [16]. The recurrence (17) implements the optimal polynomial filter for suppressing the nontrivial spectrum of W while preserving the eigencomponent associated with $\lambda_1(W) = 1$. Since the update acts columnwise, the same contraction applies simultaneously to all $(m+1)$ coefficient channels. □

Lemma 1 gives a coefficient-level contraction bound, which can be translated into a function-level error bound. With Lemma 1, now we analyze the consensus error versus communication rounds. For any $x \in X$, we have $|T_k(\xi(x))| \le 1$ for all $k = 0, \ldots, m$. Hence,

$$\varepsilon_{\mathrm{con}}(t) = \max_{x\in X} |p_i^t(x) - \bar{p}(x)| \le (m+1)\|P^t - \mathbf{1}\bar{c}^\top\|_\infty. \tag{23}$$

where $\varepsilon_{\mathrm{con}}(t)$ denotes the consensus disagreement error at iteration t (which will be specified in Sec. III-F). Combining it with Lemma 1 yields

$$\varepsilon_{\mathrm{con}}(t) \le 2(m+1)C_0\rho_{\mathrm{AC}}^t. \tag{24}$$

Therefore, the practical consensus error decays geometrically with the number of communication rounds. To guarantee $\varepsilon_{\mathrm{con}}(t) \le \bar{\varepsilon}_{\mathrm{con}}$, the tolerance of the consensus disagreement error, we choose the number of communication rounds by

$$t \ge \frac{\log(2(m+1)C_0/\bar{\varepsilon}_{\mathrm{con}})}{\log(1/\rho_{\mathrm{AC}})}. \tag{25}$$

3) Certified Stopping Under Momentum Oscillations: In Stage 2, acceleration destroys the monotonicity available in standard average consensus. To prevent premature termination, we enforce a sustained consensus condition over a consecutive time window, i.e., the following stopping rule.

Definition 1 (Spectrally-coupled stopping rule)**.** For each iteration t, define the network spread by

$$\mathcal{S}(t) = \max_{0\le k\le m}\left(\max_{i\in\mathcal{V}} c_{i,k}^t - \min_{i\in\mathcal{V}} c_{i,k}^t\right). \tag{26}$$

Given a tolerance $\delta > 0$, CASO stops at the first iteration K such that

$$\mathcal{S}(\tau) \le \delta, \qquad \forall \tau \in \{K-H+1, \ldots, K\}, \tag{27}$$

for a prescribed window length H.

The window length H should be long enough to suppress momentum-induced pseudo-convergence. Here we give a threshold $H^*(\hat{C}_0)$ for window length. Let

$$H^*(\hat{C}_0) = \left\lceil \frac{\log(4\hat{C}_0/\delta)}{\log(1/\rho_{\mathrm{AC}})} \right\rceil, \tag{28}$$

where $\hat{C}_0$ is any conservative upper bound satisfying $\left\|P^0 - \mathbf{1}\bar{c}^\top\right\|_\infty \le \hat{C}_0$ and ρ_{AC} is given by (20). Now we give the following theorem.

Theorem 2 (Stopping correctness)**.** *If $H \ge H^*(\hat{C}_0)$ holds, then at the stopping time K,*

$$\max_{x\in X} \left|p_i^K(x) - \bar{p}(x)\right| \le (m+1)\delta, \qquad \forall i \in \mathcal{V}. \tag{29}$$

Proof sketch. By Lemma 1, the disagreement envelope decays geometrically at rate ρ_{AC}. Replacing the unknown initial disagreement by any upper bound $\hat{C}_0$ only enlarges the resulting window lower bound conservatively, so the consecutive-H test still ensures that the observed spread remains below δ over a window. The window is long enough to dominate any momentum-induced oscillation consistent with that envelope. Since $|T_k(\xi(x))| \le 1$ for all $x \in X$, coefficient-wise disagreement bounded by δ across $(m+1)$ channels implies the function-value bound (29). □

Remark 3 (Flooding Communication)**.** *To evaluate the spread $\mathcal{S}(t)$ in a fully distributed manner, the nodes invoke an auxiliary* *max/min flooding routine over the coefficient channels. Specifically, each stopping check performs U rounds of neighbor-wise max/min propagation, where U is any known upper bound on the graph diameter.*

Remark 4 (Parameter Analysis)**.** *The threshold δ determines the target disagreement level at termination. By Theorem 2, it directly induces the bound. Hence, a smaller δ leads to a tighter consensus error bound. The window length H is used to prevent premature stopping caused by momentum-induced oscillations. Therefore, H should be sufficiently large ($> H^*(\hat{C}_0)$) such that pseudo-convergence can be filtered out.*

E. Algebraic Global Minimization

In Stage 3, once the final surrogate is certified, CASO minimizes it algebraically. Specifically, let

$$p_i^K(x) = \sum_{k=0}^{m} \hat{c}_{i,k} T_k(\xi(x)) \tag{30}$$

denote the final global surrogate at node i. Its stationary points are the roots of the derivative $\frac{d}{dx} p_i^K(x) = 0$. The derivative remains a polynomial. Its coefficients can be directly computed via a standard linear recurrence mapping $\mathscr{D}(\cdot)$ where $\hat{c}_i = [\hat{c}_{i,0}, \ldots, \hat{c}_{i,m}]^\top$. Consequently, finding these roots is mathematically equivalent to solving a matrix eigenvalue problem.

Specifically, the roots of the derivative polynomial within $[-1, 1]$ are the real eigenvalues of the colleague matrix $\mathbf{C}_i$ associated with $d_i = \mathscr{D}(\hat{c}_i)$ [20], [21], constructed as

$$\mathbf{C}_i = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{d_{i,0}}{2d_{i,m-1}} & -\frac{d_{i,1}}{2d_{i,m-1}} & \cdots & \frac{1}{2} - \frac{d_{i,m-3}}{2d_{i,m-1}} & -\frac{d_{i,m-2}}{2d_{i,m-1}} \end{pmatrix}. \tag{31}$$

With the definitions above, Stage 3 reduces the nonconvex minimization problem to deterministic eigenvalue analysis with the following formal steps:

1) Compute the derivative coefficient vector $d_i = \mathscr{D}(\hat{c}_i)$, and build the colleague matrix $\mathbf{C}_i$ by (31).
2) Find all real eigenvalues λ of $\mathbf{C}_i$ within $[-1, 1]$, and map them back to the original domain $X = [a, b]$ to form the candidate set $\mathcal{X}_{\text{cand}}^{(i)}$, which is given by

$$\mathcal{X}_{\text{cand}}^{(i)} = \left\{ \frac{b-a}{2}\lambda + \frac{a+b}{2} \;\middle|\; \lambda \in \operatorname{eig}(\mathbf{C}_i) \cap [-1, 1] \right\}. \tag{32}$$

3) Evaluate $p_i^K(x)$ at all points in $\mathcal{X}_{\text{cand}}^{(i)} \cup \{a, b\}$ and return

$$\tilde{x}_i \in \arg \min_{x \in \mathcal{X}_{\text{cand}}^{(i)} \cup \{a,b\}} p_i^K(x). \tag{33}$$

All these steps yield an $O(m^3)$ complexity of Stage 3 with a standard dense eigensolver, while keeping ε_{opt} (which will be specified in Sec. III-F) analytically separate from surrogate and consensus errors.

Algorithm 1 CASO for Distributed Nonconvex Optimization

Require: Local black-box objectives $\{f_i\}_{i=1}^N$, interval $X = [a, b]$, target accuracy ε.

Stage 0: Spectral Estimation
Initialize $z^0 \in \mathbf{1}^\perp$ and run the distributed spectral-estimation routine based on (5) and (6);
Obtain the spectral quantities $\lambda_2(W)$ and $\lambda_N(W)$;
Set tolerance budgets for ε_{app}, ε_{con} and ε_{opt} by (41);
Stage 1: Chebyshev Polynomial Surrogate
for each node i in parallel **do**
  Normalize the interval by (7) and adaptively construct $p_i^0(x)$ using (8)–(9), with the final form given by (10);
end for
Compute $m = \max_{i \in \mathcal{V}} m_i$ and zero-pad the local coefficient vectors to dimension $m + 1$;
Set the coefficient threshold δ and choose the stopping window $H \geq H^*(\hat{C}_0)$ according to (28);
Stage 2: Chebyshev-Accelerated Surrogate Consensus
Form P^0 by (16) and set $P^{-1} = P^0$;
for $t = 0, 1, 2, \ldots$ **do**
  Calculate P^{t+1} by (17);
  Run U rounds of distributed max/min flooding on the current coefficient vectors and evaluate $\mathscr{s}(t + 1)$;
  if $\mathscr{s}(\tau) \leq \delta$ for all $\tau \in \{t - H + 2, \ldots, t + 1\}$ **then**
    Set $K = t + 1$;
    Break;
  end if
end for
Stage 3: Algebraic Global Minimization
for each node i in parallel **do**
  Form the final surrogate $p_i^K(x)$;
  Build colleague matrix $\mathbf{C}_i$ by (31);
  Construct the candidate set $\mathcal{X}_{\text{cand}}^{(i)}$ by (32);
  Evaluate $p_i^K(x)$ on $\mathcal{X}_{\text{cand}}^{(i)} \cup \{a, b\}$;
  Output the local minimizer $\tilde{x}_i$ according to (33);
end for

F. Optimality Analysis

As described in (2), CASO is expected to calculate with a tolerable output error. Specifically, the output error of CASO consists of surrogate approximation error (ε_{app}), consensus disagreement error (ε_{con}), and algebraic solution error (ε_{opt}).

Definition 2 (Error Sources)**.** Let $\tilde{x}_i$ denote the final result output by any node $i \in \mathcal{V}$ using CASO. Define

$$\varepsilon_{\text{app}} = \max_{x \in X} |\bar{p}(x) - f(x)|, \tag{34}$$

$$\varepsilon_{\text{con}} = \max_{i \in \mathcal{V}} \max_{x \in X} \left| p_i^K(x) - \bar{p}(x) \right|, \tag{35}$$

$$\varepsilon_{\text{opt}} = \max_{i \in \mathcal{V}} \left| p_i^K(\tilde{x}_i) - \min_{x \in X} p_i^K(x) \right|, \tag{36}$$

where p_i^K is the final surrogate of node i after the certified stopping time K.

Correspondingly, $\varepsilon_{\rm app}$ comes from surrogate bias, $\varepsilon_{\rm con}$ comes from distributed disagreement, and $\varepsilon_{\rm opt}$ comes from algebraic minimization. Now we give the following theorem.

Theorem 3 (Error Decomposition)**.** *Let $\tilde{x}_i$ denote the output of node i after the stopping time K. Then, for every $i \in \mathcal{V}$,*

$$|f(\tilde{x}_i) - f^*| \leq 2\varepsilon_{\rm app} + 2\varepsilon_{\rm con} + \varepsilon_{\rm opt}. \tag{37}$$

Proof. Let $x_p^* \in \arg\min_{x\in X} p_i^K(x)$ denote one of the global minimizers of the final surrogate of node i, then we have

$$\begin{aligned} f(\tilde{x}_i) - f^* \leq & \left|f(\tilde{x}_i) - p_i^K(\tilde{x}_i)\right| + \left|p_i^K(\tilde{x}_i) - p_i^K(x_p^*)\right| \\ & + (p_i^K(x_p^*) - p_i^K(x^*)) \\ & + \left|p_i^K(x^*) - \bar{p}(x^*)\right| + \left|\bar{p}(x^*) - f^*\right|. \end{aligned} \tag{38}$$

By Definition 2, the second term in (38) is at most $\varepsilon_{\rm opt}$. Since x_p^* is defined as a global minimizer of the surrogate $p_i^K(x)$, the third term is strictly non-positive. For the remaining terms, we give the following inequalities

$$\begin{aligned} \left|f(\tilde{x}_i) - p_i^K(\tilde{x}_i)\right| \leq & \left|f(\tilde{x}_i) - \bar{p}(\tilde{x}_i)\right| \\ & + \left|\bar{p}(\tilde{x}_i) - p_i^K(\tilde{x}_i)\right| \leq \varepsilon_{\rm app} + \varepsilon_{\rm con}, \\ \left|p_i^K(x^*) - \bar{p}(x^*)\right| & \leq \varepsilon_{\rm con}, \\ \left|\bar{p}(x^*) - f(x^*)\right| & \leq \varepsilon_{\rm app}. \end{aligned} \tag{39}$$

Above all yield (37). □

Theorem 3 shows how all three terms combine into the final optimality gap. These three terms are introduced in different stages, which can be controlled separately. Once the stopping criterion is satisfied and the stage-wise tolerance budgets are chosen such that

$$2\varepsilon_{\rm app} + 2\varepsilon_{\rm con} + \varepsilon_{\rm opt} \leq \varepsilon. \tag{40}$$

The returned value is an ε-global approximation for the original objective in (1). In practice, we select the stage-wise tolerances of error ($\bar{\varepsilon}_{\rm app}$, $\bar{\varepsilon}_{\rm con}$ and $\bar{\varepsilon}_{\rm opt}$) as

$$2\bar{\varepsilon}_{\rm app} \leq \varepsilon/3, 2\bar{\varepsilon}_{\rm con} \leq \varepsilon/3, \bar{\varepsilon}_{\rm opt} \leq \varepsilon/3. \tag{41}$$

G. Algorithm Complexity

Table I summarizes the stage-wise complexity of CASO. To make the notation explicit, let U denote a known upper bound on the graph diameter, let η denote the target absolute precision of the Stage 0 distributed spectral estimation procedure [17], and let F_0 denote the cost of one local function evaluation used in Stage 1 when constructing the Chebyshev surrogate [5], [18]. The key message is that Stage 2 enjoys square-root spectral-gap acceleration [16], while Stage 3 is realized by an explicit polynomial-time eigensolve [21]. Since Stage 0 is a preprocessing step, the total line reports the online cost of Stages 1–3 only. The high-efficiency design of CASO establishes a solid foundation for the superior communication and computational efficiency.

IV. Simulations

A. Environmental Setup

1) Settings: The experiments are designed to validate the full CASO causal chain rather than isolated modules. Accordingly, each experiment is tied to one theoretical claim. Notably, the optimization domain is $X = [-5, 5]$, the target global accuracy is $\varepsilon = 10^{-3}$, and the network size is fixed at $N = 30$. All reported curves and statistics are averaged over 50 random seeds.

The main five-baseline comparison uses a cycle graph, the spectral-scaling study scans complete, Erdős–Rényi $G(N, p)$ random graphs [23], truncated grid, and cycle graphs, and the stopping ablation uses Erdős–Rényi and random geometric graphs. These choices span a broad range of spectral gaps, from dense well-connected networks to sparse topologies where acceleration is expected to matter most.

2) Test Objectives: We test two main classes of objectives. The first is a strongly convex baseline described by

$$f_i(x) = a_i(x - b_i)^2 + c_i,$$

where the parameters are heterogeneous across nodes. Since first-order methods are expected to behave well here, this setting is used to compare communication cost rather than global optimality.

The second is randomly generated from one of three one-dimensional multi-well families: Rastrigin-like, Ackley-like [24], and a custom sinusoidal-composite family. Let

$$z_i = c_i(x - s_i),$$

where $s_i, c_i, b_i, r_i, \phi_i$ are randomly sampled instance parameters. The local objectives are assigned cyclically across agents as

$$\begin{aligned} f_i^{(1)}(x) &= 0.12z_i^2 - \cos(2.5\pi z_i + \phi_i) + r_i \sin(5z_i) + b_i, \\ f_i^{(2)}(x) &= -1.8e^{-0.22|z_i|} - \exp(0.5\cos(2\pi z_i + \phi_i)) \\ &\quad + 2.8 + r_i \cos(4z_i) + b_i, \\ f_i^{(3)}(x) &= \sin(2.8z_i + \phi_i) + 0.8\cos(5.2z_i - 0.4\phi_i) + 0.06z_i^2 \\ &\quad + r_i \sin(8z_i + 0.7) + b_i. \end{aligned}$$

This setting is particularly suitable for evaluating CASO, because these functions contain multiple local minima and directly expose the distinction between stationary-point convergence and globally certified optimization. For nonconvex experiments, each method is initialized from the same seed-dependent start to ensure that the observed differences are attributable to the optimization mechanism rather than to favorable initialization.

3) Comparison Baselines: We compare CASO with four baselines: CPCA as the closest surrogate-based baseline [5], [14]; SONATA as a representative distributed successive-convex-approximation scheme [10]; DIGing as a classical distributed gradient-tracking algorithm [8]; and ZO-GT as the most relevant distributed zeroth-order competitor in black-box settings [13]. This collection covers surrogate, first-order, and zeroth-order paradigms.

TABLE I
COMPLEXITY SUMMARY OF CASO

Stage	Zero-order queries	Local computation per node	Communication rounds	Per-round communication
Stage 0:	0	$O(T_{\text{spec}}(\eta))$	$O(T_{\text{spec}}(\eta))$	$O(1)$
Stage 1:	$O(m)$	$O(m \max\{m, F_0\})$	0	–
Stage 2:	0	$O\Big(m \log(m/\varepsilon)/\sqrt{1-\lambda_2(W)}\Big)$	$O\Big(\log(m/\varepsilon)/\sqrt{1-\lambda_2(W)}\Big)^*$	$O(m)$
Stage 3:	0	$O(m^3)$ with a generic eigensolver; $O(m^2)$ with a Hessenberg solver	0	–
Total (Stages 1–3)	$O(m)$	$O\Big(\max\Big\{m^3, \frac{m\log(m/\varepsilon)}{\sqrt{1-\lambda_2(W)}}, mF_0\Big\}\Big)$	$O\Big(\log(m/\varepsilon)/\sqrt{1-\lambda_2(W)}\Big)^*$	$O(m)$

*Exact stopping checks incur additional max/min flooding rounds proportional to the graph diameter bound U. [22]

B. Performance on Convex Objectives

We first evaluate all methods on heterogeneous quadratic objectives, where the global objective has a unique minimizer and gradient-based schemes are in their most favorable regime. In this setting, the experiment serves mainly as a diagnostic check. It verifies that CASO remains competitive on convex objectives. The communication-acceleration claim is evaluated more directly in the spectral-scaling study.

Consistent with that role, all methods can eventually reach high accuracy on the quadratic benchmark, and the dominant difference is communication cost rather than final solution quality. The main pattern is that the benefit of CASO grows as the topology becomes sparser, which is exactly the regime isolated more directly in the next subsection.

C. Performance on Nonconvex Objectives

We next evaluate the methods on heterogeneous multimodal nonconvex objectives. Figure 2 reports the main five-baseline comparison on a sparse cycle graph. CASO and CPCA both succeed on all seeds and finish with nearly identical final optimality gaps. However, CASO reaches the target band much earlier, with a mean ε-hitting time of 42.8 Stage 2 iterations. This compares with 176.8 Stage 2 iterations for CPCA, i.e., about a $4.1\times$ improvement in the flagship sparse nonconvex setting.

The three local-search baselines behave qualitatively differently. DIGing, SONATA, and ZO-GT often reduce the gap early but then stall at much larger residual errors. Their success rates are low and require far more rounds than CASO to successfully reach the target ε-band. The shaded variability bands in Fig. 2 are also visibly wider for these baselines, demonstrating their seed sensitivity and instability relative to the tightly concentrated certified trajectories. This verifies the failure mode predicted by the theory: local-search dynamics approach stationary points while not providing a global certificate for the original objective.

This experiment validates the algorithm in the target nonconvex setting. The surrogate makes exhaustive search possible, the accelerated Stage 2 sharply reduces the communication burden relative to CPCA, and the final algebraic extraction avoids the local traps that dominate first-order and zeroth-order stationarity-driven baselines.

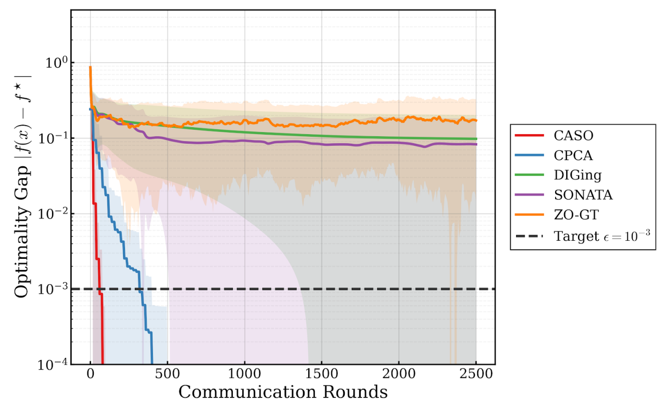

Fig. 2. Flagship five-baseline comparison on sparse nonconvex network system over a cycle graph with $N = 30$. CASO reaches the certified target band rapidly, while DIGing, SONATA, and ZO-GT frequently stall in local-search traps with much larger residual gaps. Curves show means over 50 seeds, and shaded regions denote variability bands.

D. Communication Scaling Across Spectral Gaps

To directly validate the Stage 2 complexity claim, we scan graph families with different values of $\lambda_2(W)$ and record the number of communication rounds required to satisfy the certified stopping rule at the same target accuracy. Figure 3 plots the resulting communication cost against graph sparsity.

The observed trend closely follows the theory. On dense graphs, acceleration is not expected to help much and even incur a mild constant-factor overhead. For the complete graph, CASO requires 24.4 rounds on average versus 20.4 for CPCA.

As the graph becomes sparser, however, the advantage appears and grows rapidly. The mean improvement is about $1.31\times$ on ER-0.30, $1.87\times$ on ER-0.18, $2.89\times$ on the truncated grid, and $5.53\times$ on the cycle graph. This verifies the qualitative transition predicted by Lemma 1: the acceleration pays off when the network spectrum becomes the dominant bottleneck.

E. Certified Adaptive Stopping versus Fixed-Round Execution

To validate the contribution of the adaptive stopping rule, we compare the certified stopping mechanism with a fixed-round strategy based on conservative worst-case communication budgets. Figure 3 summarizes the result. The final optimality gaps are numerically indistinguishable between the two strategies.

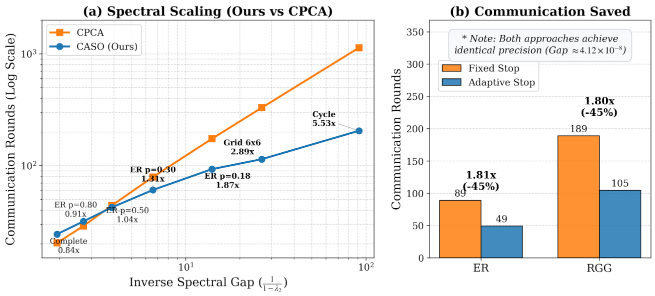

Fig. 3. Stage 2 communication behavior. Left: CASO remains comparable to CPCA on dense graphs, but its communication advantage grows as the network becomes sparser. Right: certified adaptive stopping preserves the same final accuracy as conservative fixed-round execution while substantially reducing redundant communication.

Both adaptive and fixed execution achieve mean gaps of about 4.12×10^{-8} on both ER and random geometric graphs.

The communication overhead, however, is substantially different. On ER graphs, adaptive stopping reduces the average communication count from 88.96 to 49.22 rounds, a mean saving of 44.68%. On random geometric graphs, the rounds drop from 188.72 to 104.62, a mean saving of 44.57%. This verifies that the stopping theorem is not only correct in an asymptotic sense but also delivers a useful reduction in redundant communication while preserving the same accuracy.

V. Conclusion

In this work, we propose a Chebyshev-Accelerated Surrogate Optimization (CASO) algorithm. This algorithm resolves a critical issue in distributed nonconvex optimization: while standard decentralized methods typically guarantee only local stationarity, these tasks demand a rigorous global accuracy certificate. CASO delivers a mathematically certified optimization algorithm rather than another local-search heuristic. It integrates Chebyshev surrogate construction, Chebyshev-accelerated coefficient consensus, distributedly verifiable stopping, and algebraic global extraction. This architecture guarantees ε-global optimality for the original objective and reduces communication bottlenecks on sparse graphs. Simulations verify the effectiveness of CASO. The current algorithm focuses on univariate problems. We expect to extend it to high-dimensional scenarios and dynamic networks in the future.

References

[1] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu, "Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017, pp. 5330–5340.

[2] K.-K. Oh, M.-C. Park, and H.-S. Ahn, "A survey of multi-agent formation control," *Automatica*, vol. 53, pp. 424–440, 2015.

[3] Y. Zheng, S. E. Li, K. Li, F. Borrelli, and J. K. Hedrick, "Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 899–910, 2017.

[4] A. Al-Tawaha, E. Cibaku, S. Park, J. Lavaei, and M. Jin, "Distributed optimization and distributed learning: A paradigm shift for power systems," *IEEE Systems Journal*, vol. 19, no. 4, pp. 1038–1051, 2025.

[5] Z. He, J. He, C. Chen, and X. Guan, "Distributed nonconvex optimization: Gradient-free iterations and ϵ-globally optimal solution," *IEEE Transactions on Control of Network Systems*, vol. 11, no. 4, pp. 2239–2251, 2024.

[6] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[7] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.

[8] A. Nedić, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2597–2633, 2017.

[9] P. Di Lorenzo and G. Scutari, "NEXT: In-network nonconvex optimization," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 120–136, 2016.

[10] G. Scutari and Y. Sun, "Distributed nonconvex constrained optimization over time-varying digraphs," *Mathematical Programming*, vol. 176, no. 1–2, pp. 497–544, 2019.

[11] P. Bianchi and J. Jakubowicz, "Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 391–405, 2013.

[12] D. Hajinezhad, M. Hong, and A. Garcia, "ZONE: Zeroth-order nonconvex multiagent optimization over networks," *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 3995–4010, 2019.

[13] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multiagent optimization," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 269–281, 2021.

[14] Z. He, J. He, C. Chen, and X. Guan, "CPCA: A Chebyshev proxy and consensus based algorithm for general distributed optimization," in *2020 American Control Conference (ACC)*, 2020, pp. 94–99.

[15] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network topology and communication-computation tradeoffs in decentralized optimization," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 953–976, 2018.

[16] E. Kokiopoulou and P. Frossard, "Polynomial filtering for fast convergence in distributed consensus," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 342–354, 2009.

[17] M. Franceschelli, A. Gasparri, A. Giua, and C. Seatzu, "Decentralized estimation of Laplacian eigenvalues in multi-agent systems," *Automatica*, vol. 49, no. 4, pp. 1031–1036, 2013.

[18] E. W. Cheney, *Introduction to Approximation Theory*, ser. International Series in Pure and Applied Mathematics. New York, NY, USA: McGraw-Hill Book Company, 1966.

[19] J. C. Mason and D. C. Handscomb, *Chebyshev Polynomials*. Boca Raton, FL, USA: Chapman & Hall/CRC, 2002.

[20] L. N. Trefethen, *Approximation Theory and Approximation Practice*. SIAM, 2013.

[21] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Johns Hopkins University Press, 2013.

[22] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003, pp. 482–491.

[23] P. Erdős and A. Rényi, "On random graphs. I," *Publicationes Mathematicae Debrecen*, vol. 6, no. 3–4, pp. 290–297, 1959.

[24] M. Jamil and X.-S. Yang, "A literature survey of benchmark functions for global optimisation problems," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 4, no. 2, pp. 150–194, 2013.