Local Acceptance without Global Agreement: Convergence of Decentralized Trust-Region Methods

Hong Wang*

October 30, 2026

Abstract

We study decentralized smooth nonconvex optimization with a lower-bounded average objective and independent local trust-region acceptance decisions. Successful local trials can increase the average objective, while rejection produces a zero accepted inverse block at a nonzero gradient tracker. Our main contribution is a convergence analysis that permits such genuine local rejections. The method combines gradient tracking, positive definite quadratic models, and adaptive radii proportional to tracker norms. Under uniform communication contraction and explicit accepted-operator stability conditions, we prove vanishing stationarity, consensus, and tracking residuals and stationary consensus accumulation points. Finite acceptance windows and a Schur decomposition give a best-iterate iteration bound quadratic in the inverse residual tolerance, including all-rejected rounds. We give explicit admissible ranges for the step size and residual weight, and an exact heterogeneous four-node quadratic family in which every node rejects infinitely many nonzero trials. Numerical experiments examine the rejection mechanism and compare fixed-parameter behavior with independently tuned baselines. The tests show that the trust-region safeguard prevents divergence of model-based steps at a common parameter choice, but radius adaptation gives no general performance advantage.

Keywords: Decentralized optimization · Trust-region method · Gradient tracking · Local rejection · Convergence analysis
Mathematics Subject Classification: 90C30 · 90C26 · 65K05

1 Introduction

We consider $n \geq 2$ agents cooperating through neighbor communication to solve

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

where each agent evaluates only its local objective and derivatives. The functions may be nonconvex and need not share a stationary point. Trust-region (TR) methods use actual-to-predicted reduction ratios to accept steps and update radii [4, 14]. Gradient-scaled radii also appear in centralized complexity analyses [5]. Here the tracker defines the linear model term, the matching correction in actual reduction, and the radius scale. Agents mix accepted trial points and update trackers by vector exchanges. Each agent accepts or rejects its own trial from local function, gradient, tracker and model values, so no average-objective evaluation or network-wide agreement on acceptance is required. This locality, however, complicates the analysis.

*Corresponding author. School of Artificial Intelligence, Shenzhen Technology University, 3002 Lantian Road, Pingshan District, Shenzhen, Guangdong 518118, China. Email: hitwanghong@163.com

Successful local trials can increase the average objective even with exact models and perfect mixing (Example 3.1). Rejection also produces a zero step at a nonzero tracker. Both effects involve the accepted inverse blocks $A_i = a_i(B_i + \lambda_i I)^{-1}$, where B_i, λ_i, and a_i denote the model, TR multiplier, and acceptance indicator. Model similarity alone cannot control their heterogeneity.

1.1 Related Work and the Gap Addressed

First-order methods include EXTRA [17] and gradient tracking (GT) [25, 13, 15]. Nonconvex GT analyses give network-dependent stationarity bounds for stochastic gradients [24] and refine spectral dependence, including the noiseless case [9]. With fixed symmetric nonnegative doubly stochastic mixing, our iteration has the algebraic form of Wu and Wang's variable-metric framework [22]. Their Theorem 2.4 covers all-accepted trajectories with our default initialization under additional restrictions on the model and step-size constants. At a rejected trial with $y_i^k \neq 0$, however, $(y_i^k)^\mathsf{T} A_i^k y_i^k = 0$, violating their strictly positive nodewise directional lower bound. Neither step-size bound is claimed to imply the other.

Curvature-based methods include Network Newton [11], decentralized approximate Newton [21], and quasi-Newton methods [8, 18]. NEXT [7] and its successors [16, 19] combine surrogates with tracking. Newton Tracking [27] tracks corrected Newton directions, and Network-GIANT [10, 6] uses harmonic Hessian consensus. Our accepted inverses additionally depend on multipliers and local decisions.

For TR methods over networks, Armacki et al. [1] use first-order models and evaluate the ratio at a mixed trial. DOFFO_TR [3] uses objective-function-free steps and gradient-based radius updates within penalty continuation, with complexity guarantees for fixed penalty subproblems. Here local corrected ratios select steps, and convergence concerns the original average objective, consensus, and tracking errors.

ASY-DSCA [20], an asynchronous decentralized successive convex approximation (SCA) method, permits bounded activation windows and delays, which play a role similar to our acceptance windows. The difference lies in the inactive state: rejected nodes in our method still mix and track, whereas inactive nodes in its event-indexed algorithm keep these states unchanged. In the smooth unconstrained setting, an active surrogate step at a nonzero tracker also cannot equal our zero rejected increment. Indeed, gradient consistency makes the corrected surrogate gradient at the current point proportional to the nonzero tracker, so that point cannot be its exact unconstrained minimizer. These comparisons concern direct application of the stated algorithms and theorems.

1.2 Contributions and Scope

Our analysis addresses local rejection at a nonzero tracker, for which an accepted inverse block vanishes even though the model remains positive definite. We make three contributions.

1. Theorem 4.3 proves vanishing stationarity, consensus, and tracking residuals under explicit uniform geometry and stability conditions. Its best-iterate $\mathcal{O}(\epsilon^{-2})$ bound counts every communication round, including all-rejected rounds.

2. The accepted-block variance in Proposition 3.2 and the Schur decomposition identify how heterogeneous decisions couple tracking error to the mean direction. Finite acceptance windows recover stationarity when the current mean-direction energy alone cannot do so. The associated results provide explicit tests for selecting the step size and residual weight.

3. Proposition 4.6 proves compatibility of the stability condition with positive gradient heterogeneity and infinitely many genuine rejections at every node. The uniform bounds it requires are verified along the actual ratio-tested trajectory, rather than under an externally prescribed acceptance schedule.

A round is called *productive* if at least one node accepts its local trial. The positive definite mean inverse at such a round defines the mean-direction energy used in the main proof. This energy can vanish on an all-rejected round without stationarity. The proof therefore joins the Schur decomposition, which controls the interaction of the mean direction and disagreement, with acceptance windows that recover stationarity across rejected rounds. The numerical study checks the analytic mechanism and evaluates fixed-parameter behavior and independently tuned comparisons on unselected problems.

The analysis assumes exact local subproblem solutions, uniformly positive definite models, and a uniform contraction bound for each communication round. Its accumulation-point statement does not assert the existence of a bounded subsequence. If such a subsequence exists, each of its limits is a stationary consensus.

Sections 2–4 present the algorithm, accepted-operator geometry, and convergence analysis, respectively. Section 5 reports the numerical verification and paired comparisons. Section 6 concludes the paper.

1.3 Notation

Matrices $X, Y, D \in \mathbb{R}^{n\times d}$ have rows x_i^T, $y_i^\mathsf{T}, d_i^\mathsf{T}$. Let $\mathbf{1} \in \mathbb{R}^n$ be the all-ones vector, $J = \mathbf{1}\mathbf{1}^\mathsf{T}/n$, and $J_\perp = I_n - J$. For row-stacked Z, including $Z = X, Y, D$, write $\bar{z} = n^{-1}Z^\mathsf{T}\mathbf{1}$ and $Z_\perp = J_\perp Z$. The symbols I and $\mathcal{I}$ are compatible matrix and Hilbert-space identities, with subscripts specifying dimensions or restrictions. The rows of $G(X)$ are $\nabla f_i(x_i)^\mathsf{T}$, and $\iota_{\{E\}}$ denotes the indicator of an event E. The normalized Hilbert inner product and norm are $\langle U, V\rangle_n = \frac{1}{n}\operatorname{tr}(U^\mathsf{T}V)$ and $\|U\|_n = \frac{1}{\sqrt{n}}\|U\|_F$, so $\left\|\mathbf{1}x^\mathsf{T}\right\|_n = \|x\|$. The norm $\|\cdot\|_{\mathrm{op}}$ is induced by these Hilbert norms and the Euclidean norm on $\mathbb{R}^d$. For blocks H_i, the row-wise operator $\mathcal{H}[D]$ has rows $(H_i d_i)^\mathsf{T}$. Symmetric blocks make $\mathcal{H}$ self-adjoint, with $\|\mathcal{H}\|_{\mathrm{op}} = \max_i \|H_i\|_2$. Products of operators denote composition. Vector norms are Euclidean. Matrix norms $\|\cdot\|$ and $\|\cdot\|_2$ are spectral, while $\|\cdot\|_F$ is Frobenius. We use $M \succeq N$ for the semidefinite order, $\lambda_{\min}, \lambda_{\max}$ for extreme eigenvalues, and $(t)_+ = \max\{t, 0\}$. The per-round radius extrema $\delta_k^{\min} = \min_i \delta_i^k$ and $\delta_k^{\max} = \max_i \delta_i^k$ are distinct from the fixed bounds $\delta_{\min}, \delta_{\max}$.

2 The Gradient-Tracked Trust-Region Iteration

We first specify the local test and radius update. Their accepted steps produce the inverse operator studied in Section 3.

Assumption 2.1 (Smooth objective and uniformly bounded models)**.** *Each f_i has a globally L-Lipschitz gradient, with $L > 0$, and $f(x) \geq f_{\mathrm{low}} > -\infty$. For fixed $0 < b_{\min} \leq b_{\max}$, $b_{\min} I \preceq B_i^k \preceq b_{\max} I$ for all i, k.*

The assumption concerns the models, while the objective Hessians may be indefinite. Global smoothness avoids an implicit invariance requirement on the trial and mean-step segments.

Let the rows of X^k be the base points and those of Y^k the gradient trackers. The local models are symmetric positive definite (SPD) matrices B_i^k, for example regularized Hessian or Gauss–Newton models. Communication is a linear operator $\mathcal{T}_k$ on $\mathbb{R}^{n\times d}$ that preserves column sums, $\mathbf{1}^\mathsf{T}\mathcal{T}_k[Z] = \mathbf{1}^\mathsf{T}Z$. Its standard realization is $\mathcal{T}_k[Z] = W_k Z$ with a

doubly stochastic mixing matrix W_k, and the contraction assumed in the analysis is stated in Section 3.1.

Choose $0 < \alpha \le 1$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_{\mathrm{dec}} < 1 < \gamma_{\mathrm{inc}}$, and $0 < \delta_i^0 \le \delta_{\max} < \infty$. Each agent solves

$$p_i^k = \operatorname*{argmin}_{\|p\| \le \delta_i^k \|y_i^k\|} \left\{ (y_i^k)^{\mathsf{T}} p + \tfrac{1}{2} p^{\mathsf{T}} B_i^k p \right\}, \qquad s_i^k = \alpha p_i^k. \tag{2}$$

The predicted and corrected actual reductions of the relaxed trial are

$$\mathrm{Pred}_i^k = -(y_i^k)^{\mathsf{T}} s_i^k - \tfrac{1}{2} (s_i^k)^{\mathsf{T}} B_i^k s_i^k, \tag{3}$$

$$\mathrm{Ared}_i^k = f_i(x_i^k) - f_i(x_i^k + s_i^k) - (y_i^k - \nabla f_i(x_i^k))^{\mathsf{T}} s_i^k. \tag{4}$$

The correction matches the model to the local surrogate $f_i(x_i^k + s) + (y_i^k - \nabla f_i(x_i^k))^{\mathsf{T}} s$, whose gradient at $s = 0$ is the tracker. For $y_i^k \neq 0$, form $\rho_i^k = \mathrm{Ared}_i^k \,/\, \mathrm{Pred}_i^k$ and set

$$a_i^k = \iota_{\{\rho_i^k \ge \eta_1\}}, \qquad d_i^k = a_i^k s_i^k, \qquad \widetilde{x}_i^{k+1} = x_i^k + d_i^k. \tag{5}$$

If $y_i^k = 0$, take $p_i^k = s_i^k = d_i^k = 0$, set $a_i^k = 1$, and treat the trial as very successful without forming a ratio. A round k is *productive* if $\sum_{i=1}^n a_i^k \ge 1$ and *all-rejected* otherwise. Accepted zero trials are included, so a productive round need not have a nonzero applied step.

For the operator representation below, fix the virtual multiplier $\lambda_i^k = [(\delta_i^k)^{-1} - \lambda_{\min}(B_i^k)]_+$ when $y_i^k = 0$. This convention changes no step. The relative radius factor is updated by

$$\delta_i^{k+1} = \begin{cases} \min\{\gamma_{\mathrm{inc}} \delta_i^k, \delta_{\max}\}, & y_i^k = 0 \text{ or } \rho_i^k \ge \eta_2, \\ \delta_i^k, & \eta_1 \le \rho_i^k < \eta_2, \\ \gamma_{\mathrm{dec}} \delta_i^k, & \rho_i^k < \eta_1. \end{cases} \tag{6}$$

Collecting the accepted increments as the rows of D^k, communication and tracking follow the local decisions:

$$X^{k+1} = \mathcal{T}_k[X^k + D^k], \qquad Y^{k+1} = \mathcal{T}_k[Y^k] + G(X^{k+1}) - G(X^k). \tag{7}$$

The tracker update is the dynamic-average recursion of GT [25, 13, 15], driven here by the accepted base points. For $\mathcal{T}_k[Z] = W_k Z$, each agent exchanges the pair $(\widetilde{x}_i^{k+1}, y_i^k)$ with its neighbors, and no matrix is transmitted.

Algorithm 1 Gradient-tracked decentralized trust-region method

Require: X^0, factors δ_i^0 and parameters as above, communication $\mathcal{T}_k$.

Initialize $Y^0 = G(X^0)$.
for $k = 0, 1, \ldots$ **do**
  for all agents i in parallel **do**
    Form B_i^k under Assumption 2.1 and solve (2).
    If $y_i^k \neq 0$, form ρ_i^k using (3)–(4) and set a_i^k, d_i^k by (5).
    If $y_i^k = 0$, set $a_i^k = 1$, $d_i^k = 0$ without forming ρ_i^k.
    Set $\widetilde{x}_i^{k+1} = x_i^k + d_i^k$ and update δ_i^{k+1} by (6).
  end for
  All nodes exchange $(\widetilde{x}_i^{k+1}, y_i^k)$ and apply (7).
end for

For zero trackers, (2) gives $p_i^k = s_i^k = 0$ and (6) uses the capped increase without evaluating ρ_i^k.

What the corrected test measures. For a fixed node and round, suppress the indices and write the local Taylor remainder as

$$r(s) = f_i(x+s) - f_i(x) - \nabla f_i(x)^\mathsf{T} s, \qquad |r(s)| \le \tfrac{L}{2}\|s\|^2 .$$

The two reductions then satisfy

$$\mathrm{Ared} = -y^\mathsf{T} s - r(s), \qquad \mathrm{Ared} - \mathrm{Pred} = \tfrac{1}{2} s^\mathsf{T} B s - r(s).$$

Thus the correction removes the first-order tracker discrepancy. Without it, the reduction error contains a term linear in s with coefficient $y - \nabla f_i(x)$, which can remain nonzero at a stationary point of the average objective because local gradients need not vanish. The corrected test compares the curvature remainder with the model on the actual relaxed trial and requires only the local function, gradient, tracker and model, with no average-objective evaluation or agreement on acceptance.

Radius factors and relaxation. The factor δ_i^k controls $\left\|p_i^k\right\| \le \delta_i^k \left\|y_i^k\right\|$, while α scales the applied trial $s_i^k = \alpha p_i^k$. At a fixed state, changing α changes the ratio evaluation point but not the subproblem solution or multiplier. Since $\alpha \le 1$, the relaxed trial remains in the same trust region. The physical radius $\delta_i^k \left\|y_i^k\right\|$ can decrease even when its factor increases. At a zero tracker, the feasible set is $\{0\}$. The prescribed successful zero trial avoids an undefined ratio, and the virtual multiplier serves only the operator representation and bounds.

Communication after rejection. Each round mixes the accepted trial points and the old trackers, which can be exchanged as one pair of d-vectors. The new gradient is evaluated at the mixed point before its difference is added to the mixed tracker. A rejecting node supplies $\widetilde{x}_i^{k+1} = x_i^k$, but its next base point and tracker can change through its neighbors. If all nodes reject, the mean stays fixed while disagreement can contract. The complexity bound therefore counts every communication round.

Lemma 2.1 (Mean tracking and trust-region KKT relations)**.** *If* $\mathbf{1}^\mathsf{T} Y^0 = \mathbf{1}^\mathsf{T} G(X^0)$, *then* $\bar{y}^k = n^{-1}\sum_i \nabla f_i(x_i^k)$ *for every* k, *irrespective of the acceptance decisions. Each subproblem* (2) *has a unique solution and a Karush–Kuhn–Tucker (KKT) multiplier* $\lambda_i^k \ge 0$ *satisfying*

$$(B_i^k + \lambda_i^k I)p_i^k = -y_i^k, \qquad \lambda_i^k\left(\left\|p_i^k\right\| - \delta_i^k \left\|y_i^k\right\|\right) = 0. \tag{8}$$

For $y_i^k \neq 0$,

$$\mathrm{Pred}_i^k = \alpha(1-\alpha/2)(p_i^k)^\mathsf{T} B_i^k p_i^k + \alpha\lambda_i^k \left\|p_i^k\right\|^2 > 0. \tag{9}$$

At $y_i^k = 0$, *the virtual multiplier specified above also satisfies* (8).

Proof. Sum the tracker update over nodes and telescope to obtain mean tracking. For $y_i^k \neq 0$ the radius is positive, the origin is strictly feasible, and the strictly convex subproblem has a unique KKT solution [12, 4]. A sufficiently short step along $-y_i^k$ has negative model value, so its minimizer is nonzero. Substitution of $(B_i^k + \lambda_i^k I)p_i^k = -y_i^k$ into the prediction gives (9), which is strictly positive. If $y_i^k = 0$, the feasible set is $\{0\}$ and the prescribed virtual multiplier satisfies (8). □

The accepted inverse blocks are

$$\widetilde{B}_i^k = B_i^k + \lambda_i^k I, \qquad A_i^k = a_i^k(\widetilde{B}_i^k)^{-1}, \tag{10}$$

and $\mathcal{A}_k$ denotes their row-wise operator, so that $\mathcal{A}_k[Y]$ has rows $(A_i^k y_i)^\mathsf{T}$. Then $D^k = -\alpha\mathcal{A}_k[Y^k]$ by (8) and (5). Rejection sets the inverse block to zero, although the local model remains positive definite.

Remark 2.2 (A capped isotropic case). If every trial is accepted, $B_i^k = b_i^k I$ with $b_i^k \le b_{\max}$, and $\delta_i^0 = \delta_{\max} = \delta < 1/b_{\max}$, the cap keeps every factor equal to δ. The constraint is active at a nonzero tracker, so $\lambda_i^k = \delta^{-1} - b_i^k$ and $p_i^k = -\delta y_i^k$. The zero-tracker convention gives the same applied step. Thus $D^k = -\alpha\delta Y^k$ and the full recursion is GT with step size $\alpha\delta$. An active constraint alone does not distinguish trajectories.

Model, multiplier and acceptance determine the increment's shape, length and retention. The next section quantifies the inverse blocks' interaction with consensus and tracking errors.

3 Geometry of Accepted Steps

The convergence analysis uses the inverse blocks after the local ratio tests. Rejection can make these blocks heterogeneous even when the local models coincide. The coupling quantities and residuals introduced in this section and in Section 4 serve only the analysis and are not computed by the agents.

3.1 Consensus Geometry and Communication

In the normalized inner product $\langle\cdot,\cdot\rangle_n$, define

$$\mathcal{C} = \{\mathbf{1}x^\mathsf{T} : x \in \mathbb{R}^d\}, \qquad \mathcal{P}[Z] = JZ, \qquad \mathcal{Q}[Z] = J_\perp Z. \tag{11}$$

These are orthogonal projectors [2]. The map $Sx = \mathbf{1}x^\mathsf{T}$ is an isometry with $S^*Z = n^{-1}Z^\mathsf{T}\mathbf{1}$, $S^*S = I_d$ and $SS^* = \mathcal{P}$, so that $\mathcal{P}[X] = S\bar{x}$ and $\mathcal{Q}[X] = X_\perp$. Thus

$$\|X\|_n^2 = \|\bar{x}\|^2 + \|X_\perp\|_n^2 . \tag{12}$$

We allow linear communication operators satisfying

$$\mathcal{T}_k\mathcal{P} = \mathcal{P}\mathcal{T}_k = \mathcal{P}, \qquad \|\mathcal{Q}\mathcal{T}_k\mathcal{Q}\|_{\mathrm{op}} \le \beta < 1. \tag{13}$$

A fixed symmetric, nonnegative, doubly stochastic matrix W with positive diagonal and $W_{ij} > 0$ for every edge $\{i,j\}$ of a connected undirected graph gives $\mathcal{T}_k[Z] = WZ$ and $\beta = \|W - J\|_2 < 1$ [23]. The state update consequently preserves the mean increment and contracts disagreement:

$$\bar{x}^{k+1} = \bar{x}^k + \bar{d}^k, \qquad X_\perp^{k+1} = \mathcal{Q}\mathcal{T}_k\mathcal{Q}[X_\perp^k + D_\perp^k]. \tag{14}$$

3.2 Accepted Inverses and Normalized Coupling

Example 3.1 (Local acceptance does not ensure average descent)**.** *For two scalar agents, take*

$$f_1(x) = 2x^2 + 2x, \qquad f_2(x) = x^2/4 - x, \qquad f(x) = 9x^2/8 + x/2. \tag{15}$$

At $X = 0$, initialize $Y = (2,-1)^\mathsf{T}$ and use exact models $B_1 = 4$, $B_2 = 1/2$, radius factors at least 2, and $\alpha = 1$. The feasible unconstrained trials are $d_1 = -1/2$ and $d_2 = 2$. At each node the tracker correction vanishes, and actual and predicted reductions coincide, with values $1/2$ and 1. Both ratios equal one, yet perfect mixing gives $\bar{x}^+ = 3/4$ and

$$f(\bar{x}^+) - f(0) = 129/128 > 0. \tag{16}$$

This one-round increase does not preclude subsequent convergence.

Example 3.1 shows that the mean step can be an ascent direction of f. To locate the source, recall from Section 2 the actual increment and introduce the absolute coupling

$$D^k = -\alpha \mathcal{A}_k[Y^k], \qquad \varepsilon_{A,k} = \|\mathcal{P}\mathcal{A}_k\mathcal{Q}\|_{\mathrm{op}}. \tag{17}$$

The map $(x, T) \mapsto Sx + T$ is an isometry from $\mathbb{R}^d \times \mathcal{C}^\perp$ onto $\mathbb{R}^{n\times d}$. In these coordinates the self-adjoint operator $\mathcal{A}_k$ is represented by

$$\begin{pmatrix} \bar{A}_k & K_{A,k}^* \\ K_{A,k} & \mathcal{Q}\mathcal{A}_k\mathcal{Q} \end{pmatrix},$$

whose mean and off-diagonal blocks, together with the normalized coupling, are

$$\bar{A}_k = S^*\mathcal{A}_k S, \qquad K_{A,k} = \mathcal{Q}\mathcal{A}_k S, \qquad \chi_{A,k} = \left\| K_{A,k}\bar{A}_k^{-1/2} \right\|_{\mathrm{op}}. \tag{18}$$

Explicitly, $\bar{A}_k = n^{-1}\sum_i A_i^k$ is the average accepted inverse, and $K_{A,k}x$ has rows $((A_i^k - \bar{A}_k)x)^\mathsf{T}$. The last quantity is defined on productive rounds, when $\bar{A}_k \succ 0$. On an all-rejected round set $\chi_{A,k} = 0$ without taking an inverse. Self-adjointness gives $\|K_{A,k}\|_{\mathrm{op}} = \varepsilon_{A,k}$, and

$$-\alpha^{-1}\bar{d}^k = \bar{A}_k\bar{y}^k + K_{A,k}^* Y_\perp^k. \tag{19}$$

Thus tracking disagreement enters the mean step through $K_{A,k}$. In Example 3.1, $\bar{A}\bar{y} = 9/16$ and $K_A^* Y_\perp = -21/16$, so the cross term reverses the mean step. The factor $\bar{A}_k^{-1/2}$ measures the disagreement generated by the mean direction relative to the mean-direction energy used in Section 4. If $\bar{A}_k \succeq \theta I$, then $\chi_{A,k} \le \varepsilon_{A,k}/\sqrt{\theta}$.

Proposition 3.2 (Variance of accepted inverse blocks)**.** *Suppress k, write $R_i = (B_i + \lambda_i I)^{-1} \succ 0$ and $A_i = a_i R_i$, and let $\mathcal{I}_+ = \{i : a_i = 1\}$, $m = |\mathcal{I}_+|$, and $\bar{a} = m/n$. For $m \ge 1$, set*

$$M_+ = \frac{1}{m}\sum_{i\in\mathcal{I}_+} R_i, \qquad V_+ = \frac{1}{m}\sum_{i\in\mathcal{I}_+}(R_i - M_+)^2. \tag{20}$$

Then $\bar{A} = \bar{a}M_+$ and

$$\varepsilon_A^2 = \lambda_{\max}\left(\frac{1}{n}\sum_i A_i^2 - \bar{A}^2\right) = \lambda_{\max}\left(\bar{a}V_+ + \bar{a}(1-\bar{a})M_+^2\right), \tag{21}$$

$$\chi_A^2 = \lambda_{\max}\left(M_+^{-1/2}V_+M_+^{-1/2} + (1-\bar{a})M_+\right). \tag{22}$$

These identities require no commutativity. If all accepted inverse blocks equal R, then

$$\left\|\bar{A}\right\|_2 = \bar{a}\,\|R\|_2, \qquad \chi_A^2 = (1-\bar{a})\,\|R\|_2. \tag{23}$$

For $m = 0$, $\mathcal{A} = 0$ and $\varepsilon_A = \chi_A = 0$.

Proof. Using $\mathcal{Q} = \mathcal{I} - SS^*$ gives

$$K_A^*K_A = S^*\mathcal{A}^2S - \bar{A}^2 = \frac{1}{n}\sum_i A_i^2 - \bar{A}^2 = \bar{a}V_+ + \bar{a}(1-\bar{a})M_+^2.$$

The last equality follows by expanding the squares in V_+ in their given order. Taking the largest eigenvalue proves the first identity. Congruence by $\bar{A}^{-1/2}$ and substitution of $\bar{A} = \bar{a}M_+$ give the second. For common blocks $V_+ = 0$. □

The conditional variance measures differences among the accepted inverses, while $1 - \bar{a}$ measures the accepted/rejected split. In particular, with common models $B_i = bI$, zero multipliers and two scalar nodes whose decisions are $(1, 0)$, the inverse coupling is $\varepsilon_A = 1/(2b)$ despite identical models. A smaller coupling alone need not improve progress, since the mean inverse can also shrink.

Interpretation of accepted-set variance. The normalization in (22) compares tracking contamination with the mean inverse that produces progress. For a common accepted block R, decreasing the accepted fraction $\bar{a}$ shrinks $\bar{A} = \bar{a}R$ while $\chi_A^2 = (1-\bar{a})\,\|R\|_2$. A small absolute coupling therefore need not be small relative to mean-direction progress. For different accepted blocks, the two positive semidefinite terms in (22) describe within-set variation and the accepted/rejected split. Their dominant directions need not coincide, which is why the exact formula retains a matrix sum.

The following bound separates the model, multiplier, and decision contributions before this normalization. It explains why model similarity alone cannot control the accepted inverse blocks.

Proposition 3.3 (Three sources of accepted inverse coupling)**.** *Suppress k and let $B_i \succeq b_{\min} I$, $\lambda_i \geq 0$, and $a_i \in \{0,1\}$. Let $\mathcal{B}$ and $\widetilde{\mathcal{B}}$ be generated by B_i and $B_i + \lambda_i I$, respectively, and set*

$$\varepsilon_B = \|\mathcal{P}\mathcal{B}\mathcal{Q}\|_{\mathrm{op}}, \qquad \bar{a} = \frac{1}{n}\sum_i a_i,$$

$$\bar{\lambda} = \frac{1}{n}\sum_i \lambda_i, \qquad \sigma_\lambda = \left(\frac{1}{n}\sum_i (\lambda_i - \bar{\lambda})^2\right)^{1/2}.$$

Then, for arbitrary acceptance decisions,

$$\varepsilon_A \leq \bar{a}\, b_{\min}^{-2}(\varepsilon_B + \sigma_\lambda) + b_{\min}^{-1}\sqrt{\bar{a}(1-\bar{a})}. \tag{24}$$

If all shifted models equal $\widetilde{B}_0$, then

$$\varepsilon_A = \left\|\widetilde{B}_0^{-1}\right\|_2 \sqrt{\bar{a}(1-\bar{a})}. \tag{25}$$

Proof. For a self-adjoint $\mathcal{H} \succeq b\mathcal{I}$ with $b > 0$, the orthogonal splitting $\mathcal{C} \oplus \mathcal{C}^\perp$ gives $\|[\mathcal{P}, \mathcal{H}]\|_{\mathrm{op}} = \|\mathcal{P}\mathcal{H}\mathcal{Q}\|_{\mathrm{op}}$. It follows from $[\mathcal{P}, \mathcal{H}^{-1}] = -\mathcal{H}^{-1}[\mathcal{P}, \mathcal{H}]\mathcal{H}^{-1}$ that

$$\left\|\mathcal{P}\mathcal{H}^{-1}\mathcal{Q}\right\|_{\mathrm{op}} \leq b^{-2}\,\|\mathcal{P}\mathcal{H}\mathcal{Q}\|_{\mathrm{op}}.$$

For the scalar blocks $\lambda_i I$, direct multiplication gives $\left\|\mathcal{P}(\widetilde{\mathcal{B}} - \mathcal{B})\mathcal{Q}\right\|_{\mathrm{op}} = \sigma_\lambda$. Hence $\left\|\mathcal{P}\widetilde{\mathcal{B}}^{-1}\mathcal{Q}\right\|_{\mathrm{op}} \leq b_{\min}^{-2}(\varepsilon_B + \sigma_\lambda)$. Let $\Pi = \mathrm{diag}(a_1, \ldots, a_n)$. Since $\mathcal{A}[U] = \Pi\widetilde{\mathcal{B}}^{-1}[U]$ and $\|J(\Pi - \bar{a}I_n)\|_2 = \sqrt{\bar{a}(1-\bar{a})}$, we have

$$\mathcal{P}\mathcal{A}\mathcal{Q}[U] = \bar{a}J\widetilde{\mathcal{B}}^{-1}[J_\perp U] + J(\Pi - \bar{a}I_n)\widetilde{\mathcal{B}}^{-1}[J_\perp U].$$

Taking norms and using $\left\|\widetilde{\mathcal{B}}^{-1}\right\|_{\mathrm{op}} \leq b_{\min}^{-1}$ proves (24). For common shifted blocks, $\mathcal{P}\mathcal{A}\mathcal{Q}[U] = (J\Pi J_\perp)U\widetilde{B}_0^{-1}$, whose norm is (25). □

The bound concerns the absolute coupling ε_A. The main theorem instead uses uniform bounds on the normalized coupling χ_A and the Schur residual. Computing these quantities on a finite trajectory provides a posteriori checks for those rounds, not a guarantee valid for all rounds.

3.3 Schur Residual Energy

At a productive round define the Schur complement of $\bar{A}_k$ in the block form of $\mathcal{A}_k$, as an operator on $\mathcal{C}^\perp$, by

$$Z_k = \mathcal{Q}\mathcal{A}_k\mathcal{Q} - K_{A,k}\bar{A}_k^{-1}K_{A,k}^*. \tag{26}$$

It separates mean-direction progress from disagreement energy [26].

Lemma 3.4 (Residual energy of the Schur complement)**.** *At a productive round, suppress k and put $E = \mathcal{A}^{1/2}S\bar{A}^{-1/2}$. Then $E^*E = I_d$, and the extension of Z by zero on $\mathcal{C}$ is*

$$\widehat{Z} = \mathcal{A}^{1/2}(\mathcal{I} - EE^*)\mathcal{A}^{1/2} \succeq 0. \tag{27}$$

For $T \in \mathcal{C}^\perp$,

$$\langle T, ZT\rangle_n = \min_x \langle Sx + T, \mathcal{A}[Sx + T]\rangle_n, \qquad x_* = -\bar{A}^{-1}K_A^*T. \tag{28}$$

If the $m \geq 1$ accepted inverse blocks all equal R, write $a = (a_1, \ldots, a_n)^\mathsf{T}$ and $\Pi = \operatorname{diag}(a_1, \ldots, a_n)$. Then

$$Z[T] = \left(\Pi - \frac{aa^\mathsf{T}}{m}\right)TR, \qquad \|Z\|_{\mathrm{op}} = \begin{cases} 0, & m = 1, \\ \|R\|_2, & m \geq 2. \end{cases} \tag{29}$$

Proof. The identity $E^*E = \bar{A}^{-1/2}(S^*\mathcal{A}S)\bar{A}^{-1/2} = I_d$ makes EE^* an orthogonal projector. Expanding (27) gives $\mathcal{A} - \mathcal{A}S\bar{A}^{-1}S^*\mathcal{A}$, which vanishes on $\mathcal{C}$ and restricts to Z on $\mathcal{C}^\perp$. This proves positivity despite zero rejected blocks. Completing the square gives

$$\langle Sx + T, \mathcal{A}[Sx + T]\rangle_n = (x + \bar{A}^{-1}K_A^*T)^\mathsf{T}\bar{A}(x + \bar{A}^{-1}K_A^*T) + \langle T, ZT\rangle_n,$$

so $\bar{A} \succ 0$ proves the minimization formula. For common accepted blocks, substitute $\mathcal{A}[U] = \Pi UR$ and $\bar{A} = (m/n)R$ in the expanded operator. The matrix $\Pi - aa^\mathsf{T}/m$ is the orthogonal projector onto zero-sum vectors supported on the accepted set. It is zero for $m = 1$ and has norm one otherwise, giving the stated norm. □

4 Convergence with Local Rejections

Radius contraction supplies bounded acceptance windows, and Schur geometry controls rounds with nonzero accepted inverse blocks.

4.1 Radius Bounds and Acceptance Windows

The relative radius enters the analysis only through the KKT multiplier. For $\delta > 0$ define

$$\theta(\delta) = \left[b_{\max} + (\delta^{-1} - b_{\min})_+\right]^{-1}, \qquad \Theta(\delta) = \left[b_{\min} + (\delta^{-1} - b_{\max})_+\right]^{-1}, \tag{30}$$

both nondecreasing, with $\theta(\delta) \leq \Theta(\delta) \leq b_{\min}^{-1}$ and $\Theta(\delta) < b_{\min}^{-1}$ exactly when $\delta < 1/b_{\max}$.

Lemma 4.1 (Bounds on the multiplier and inverse model)**.** *Let $b_{\min}I \preceq B \preceq b_{\max}I$, $y \neq 0$, $\delta > 0$, and let λ be the KKT multiplier of (2) for these data. Then*

$$(\delta^{-1} - b_{\max})_+ \leq \lambda \leq (\delta^{-1} - b_{\min})_+, \qquad \theta(\delta)I \preceq (B + \lambda I)^{-1} \preceq \Theta(\delta)I, \tag{31}$$

and the constraint is active whenever $\delta < 1/b_{\max}$. The same bounds hold at zero trackers under the virtual-multiplier convention. In particular, $0 \preceq A_i^k \preceq \Theta(\delta_{\max})I$.

Proof. If $\lambda = 0$, feasibility of $p = -B^{-1}y$ gives $\delta\|y\| \geq \|y\|/b_{\max}$ and hence the lower multiplier bound. If $\lambda > 0$, complementarity gives $\|p\| = \delta\|y\|$, and

$$(b_{\min} + \lambda)\|p\| \leq \|y\| \leq (b_{\max} + \lambda)\|p\|.$$

These inequalities give both multiplier bounds and, by inversion, the inverse bounds. At $y = 0$, the chosen $(\delta^{-1} - \lambda_{\min}(B))_+$ lies between the same endpoints. Multiplication by the acceptance indicator gives the final bound. □

Under Assumption 2.1, every nonzero trial satisfies

$$\rho_i^k \geq 1 - \frac{\alpha(L - b_{\min})_+}{(2-\alpha)b_{\min} + 2\lambda_i^k} \geq 1 - \frac{\alpha(L - b_{\min})_+}{(2-\alpha)b_{\min} + 2\ell}, \qquad \ell := (\delta_{\max}^{-1} - b_{\max})_+. \tag{32}$$

Indeed, smoothness gives $\mathrm{Ared}_i^k - \mathrm{Pred}_i^k \geq -\alpha^2(L - b_{\min})_+\|p_i^k\|^2/2$. Divide by (9) and use $\lambda_i^k \geq \ell$ to obtain (32).

Lemma 4.2 (Radius lower bound and bound on consecutive rejections)**.** *Under Assumption 2.1, fix $\alpha \in (0, 1]$ and set*

$$\Lambda = \left[\frac{\alpha(L - b_{\min})_+}{2(1-\eta_1)} - (1 - \alpha/2)b_{\min}\right]_+. \tag{33}$$

If $\Lambda = 0$, every trial is accepted, and we set $\delta_{\min} = \delta_0^{\min}$ and $K_{\mathrm{rej}} = 0$. Otherwise let

$$\begin{aligned} \delta_{\mathrm{acc}} = (b_{\max} + \Lambda)^{-1}, \qquad \delta_{\min} = \min\{\delta_0^{\min}, \gamma_{\mathrm{dec}}\delta_{\mathrm{acc}}\}, \\ K_{\mathrm{rej}} = \left\lceil \frac{[\log(\delta_{\max}/\delta_{\mathrm{acc}})]_+}{\log(1/\gamma_{\mathrm{dec}})} \right\rceil. \end{aligned} \tag{34}$$

Then $\delta_i^k \geq \delta_{\min}$ for all i, k, and every shifted inverse block satisfies

$$\theta_0 I \preceq (\widetilde{B}_i^k)^{-1} \preceq \Theta(\delta_{\max})I, \qquad \theta_0 = \theta(\delta_{\min}), \tag{35}$$

and each node has at most K_{rej} consecutive rejected trials. Thus each node accepts at least once in every $K_{\mathrm{rej}} + 1$ consecutive rounds.

Proof. If $\Lambda = 0$, (32) guarantees acceptance and no factor decreases. Otherwise, $\delta_i^k \leq \delta_{\mathrm{acc}}$ implies $\lambda_i^k \geq \Lambda$ at a nonzero tracker, hence $\rho_i^k \geq \eta_1$. Zero trackers are accepted by convention. A rejected factor therefore shrinks from above δ_{acc}, which proves $\delta_i^k \geq \delta_{\min}$ by induction. Lemma 4.1 gives (35). After K_{rej} consecutive contractions from at most $\delta_{\max}$ the factor is at most δ_{acc}, so the next trial is accepted. This proves the window bound, including $K_{\mathrm{rej}} = 0$. □

At a productive round at least one inverse block is bounded below by $\theta_0 I$, so $\bar{A}_k \succeq (\theta_0/n)I$.

4.2 Residuals and the Convergence Theorem

We use the scalar residuals

$$g_k = \left\|\nabla f(\bar{x}^k)\right\|, \qquad c_k = \left\|X_\perp^k\right\|_n, \qquad t_k = \left\|Y_\perp^k\right\|_n. \tag{36}$$

Mean tracking and Assumption 2.1 give $\left\|Y^k\right\|_n \leq g_k + Lc_k + t_k$ and $\left\|D^k\right\|_n \leq \alpha\Theta(\delta_{\max})\left\|Y^k\right\|_n$. The variation of the gradient heterogeneity $\mathcal{Q}G(Sx)$ under a common displacement is measured by its Lipschitz modulus

$$\kappa_f := \mathrm{Lip}(\mathcal{Q}GS) = \sup_{x \neq z} \frac{\|\mathcal{Q}[G(Sx) - G(Sz)]\|_n}{\|x - z\|}. \tag{37}$$

The projector is nonexpansive and S is an isometry, so $\kappa_f \le L$. The rows of $\mathcal{Q}G(Sx)$ are $\nabla(f_i - f)(x)^\mathsf{T}$, so $\kappa_f = 0$ whenever the local objectives differ by affine functions, including identical objectives. These facts use only smoothness.

Theorem 4.3 (Convergence with local rejections)**.** *Assume (13), Assumption 2.1, mean tracking at initialization, $0 < \delta_i^0 \le \delta_{\max}$, and the radius update (6). Fix $\alpha \in (0, 1]$, let θ_0 and the window length $T_{\mathrm{acc}} = K_{\mathrm{rej}} + 1$ be given by Lemma 4.2, and let κ_f be defined by (37), with $\kappa_f \le L$. Suppose every productive round satisfies, for fixed $\Gamma_A > 0$, $\chi_A, \sigma_A \ge 0$,*

$$\left\|\bar{A}_k\right\| \le \Gamma_A, \qquad \left\|K_{A,k}\bar{A}_k^{-1/2}\right\|_{\mathrm{op}} \le \chi_A, \qquad 0 \preceq Z_k \preceq \sigma_A \mathcal{I}_{\mathcal{C}^\perp}. \tag{38}$$

For some $\omega > 0$, define

$$\begin{aligned} q_M &= \max\{\beta + \omega L(1+\beta),\ \beta + \alpha\sigma_A\beta(1+\omega L)/\omega\}, \\ a_M &= \alpha(1 - \alpha L\Gamma_A/2), \qquad j_M = \alpha \max\{L\sqrt{\Gamma_A}, \chi_A/\omega\}, \\ d_M &= \alpha\{\beta(1+\omega L)\chi_A + \omega\kappa_f\sqrt{\Gamma_A}\}, \\ & q_M < 1, \qquad a_M(1 - q_M) > j_M d_M. \end{aligned} \tag{39}$$

Then $g_k, c_k, t_k, \left\|Y^k\right\|_n, \left\|D^k\right\|_n \to 0$. Every accumulation point of (X^k) has the form $\mathbf{1}(x^*)^\mathsf{T}$ with $\nabla f(x^*) = 0$. For the quantitative bound, choose the energy as follows. Set $r_k^S = c_k + \omega t_k$. If $d_M > 0$, take $w = j_M/(2d_M)$. If $d_M = 0$, take any $w > j_M^2/[4a_M(1 - q_M^2)]$. Define

$$N_M = \begin{pmatrix} a_M - w d_M^2 & -j_M/2 - w q_M d_M \\ -j_M/2 - w q_M d_M & w(1 - q_M^2) \end{pmatrix}, \qquad \nu = \lambda_{\min}(N_M). \tag{40}$$

Then $\nu > 0$ and we set $\mathcal{V}_k^S = f(\bar{x}^k) + w(r_k^S)^2$. For explicit constants

$$\begin{aligned} C_1 &= 3/\theta_0 + 3L^2(T_{\mathrm{acc}} - 1)\alpha^2\Gamma_A, \\ C_2 &= \max\{6L^2, (6 + 3\sigma_A/\theta_0)/\omega^2\} + \max\{1, \omega^{-2}\}, \qquad C = \max\{C_1, C_2\}, \end{aligned} \tag{41}$$

every integer $K \ge T_{\mathrm{acc}}$ satisfies

$$\min_{0 \le k \le K - T_{\mathrm{acc}}} (g_k^2 + c_k^2 + t_k^2) \le \frac{C T_{\mathrm{acc}}(\mathcal{V}_0^S - f_{\mathrm{low}})}{\nu(K - T_{\mathrm{acc}} + 1)}. \tag{42}$$

Proof. At a productive round, suppress k, put $T = Y_\perp$, $v = S^*\mathcal{A}[Y]$, and $u = \left\|\bar{A}^{-1/2}v\right\|$. Block elimination gives

$$\mathcal{Q}\mathcal{A}[Y] = K_A\bar{A}^{-1}v + ZT, \qquad \langle Y, \mathcal{A}[Y]\rangle_n = u^2 + \langle T, ZT\rangle_n. \tag{43}$$

Indeed, $v = \bar{A}\bar{y} + K_A^* T$, and substitution proves both identities. Lemma 3.4 gives $Z \succeq 0$ even though some blocks of $\mathcal{A}$ are zero. Thus (38) implies $\|v\| \le \sqrt{\Gamma_A}u$ and $\|\mathcal{Q}\mathcal{A}[Y]\|_n \le \chi_A u + \sigma_A t$. Mean tracking gives $\|\bar{y} - \nabla f(\bar{x})\| \le Lc$, so

$$\nabla f(\bar{x})^\mathsf{T} v \ge u^2 - (L\sqrt{\Gamma_A}c + \chi_A t)u.$$

Since $\bar{x}^+ = \bar{x} - \alpha v$, the descent lemma yields the first inequality below. For the second, consensus contracts as $c^+ \le \beta c + \alpha\beta(\chi_A u + \sigma_A t)$, while $t^+ \le \beta t + L(c^+ + c) + \alpha\kappa_f \|v\|$. To obtain this last bound, insert $G(S\bar{x}^+)$ and $G(S\bar{x})$ in the gradient difference. The two disagreement terms are bounded by Lc^+ and Lc. Definition (37) bounds the projected consensus term by $\kappa_f \left\|\bar{x}^+ - \bar{x}\right\| = \alpha\kappa_f \|v\|$. Consequently,

$$f(\bar{x}^{k+1}) - f(\bar{x}^k) \le -a_M u_k^2 + j_M u_k r_k^S, \qquad r_{k+1}^S \le q_M r_k^S + d_M u_k. \tag{44}$$

At an all-rejected round set $u_k = 0$ and $Z_k = 0$. Then $\mathcal{A}_k = 0$, the mean is unchanged, $c_{k+1} \le \beta c_k$, and $t_{k+1} \le \beta t_k + L(1+\beta)c_k$, so (44) remains valid without inverting $\bar{A}_k$.

If $d_M > 0$, the lower-right entry of N_M is positive and

$$\det N_M = \frac{j_M(1+q_M)}{2d_M}[a_M(1-q_M) - j_M d_M] > 0.$$

If $d_M = 0$, positivity follows from the stated lower bound on w. Thus $N_M \succ 0$ in both cases, and (44) gives

$$\mathcal{V}_{k+1}^S \le \mathcal{V}_k^S - \nu\{u_k^2 + (r_k^S)^2\}, \qquad \sum_{k=0}^{\infty}\{u_k^2 + (r_k^S)^2\} \le \frac{\mathcal{V}_0^S - f_{\text{low}}}{\nu}. \tag{45}$$

Next we bound the stationarity residual at every round. For each k and node i, choose an accepted round $j_i \in [k, k+T_{\text{acc}}-1]$, as guaranteed by Lemma 4.2. On that round $\theta_0 \left\|y_i^{j_i}\right\|^2 \le (y_i^{j_i})^{\mathsf{T}} A_i^{j_i} y_i^{j_i}$. Apply $\|a+b+c\|^2 \le 3(\|a\|^2 + \|b\|^2 + \|c\|^2)$ to

$$\nabla f(\bar{x}^k) = y_i^{j_i} + [\nabla f(\bar{x}^{j_i}) - y_i^{j_i}] + [\nabla f(\bar{x}^k) - \nabla f(\bar{x}^{j_i})]$$

and average over i. The middle term is bounded using $\left\|y_i^j - \nabla f(\bar{x}^j)\right\|^2 \le 2\left\|y_i^j - \bar{y}^j\right\|^2 + 2L^2 c_j^2$. For the last term, Cauchy–Schwarz and $\left\|\bar{x}^{j+1} - \bar{x}^j\right\|^2 \le \alpha^2 \Gamma_A u_j^2$ give a factor $T_{\text{acc}} - 1$. Grouping equal j_i and using (43) therefore gives

$$g_k^2 \le \sum_{j=k}^{k+T_{\text{acc}}-1}\{C_1 u_j^2 + 6L^2 c_j^2 + (6 + 3\sigma_A/\theta_0)t_j^2\}.$$

Since $c_k^2 + t_k^2 \le \max\{1, \omega^{-2}\}(r_k^S)^2$, $g_k^2 + c_k^2 + t_k^2 \le C\sum_{j=k}^{k+T_{\text{acc}}-1}\{u_j^2 + (r_j^S)^2\}$. Summing for $0 \le k \le K - T_{\text{acc}}$ counts each summand at most T_{acc} times and proves (42). The same window bound and (45) give $g_k, c_k, t_k \to 0$. Finally, $\left\|Y^k\right\|_n \le g_k + Lc_k + t_k$ and $\left\|D^k\right\|_n \le \alpha\Theta(\delta_{\max})\left\|Y^k\right\|_n$ give the remaining limits, and continuity identifies stationary consensus accumulation points. □

Iteration complexity. For any tolerance $\epsilon > 0$, set

$$K_\epsilon = T_{\text{acc}} - 1 + \max\left\{1, \left\lceil \frac{CT_{\text{acc}}(\mathcal{V}_0^S - f_{\text{low}})}{\nu\epsilon^2} \right\rceil\right\}.$$

It follows from (42) with $K = K_\epsilon \ge T_{\text{acc}}$ that $(g_k^2 + c_k^2 + t_k^2)^{1/2} \le \epsilon$ for some $0 \le k \le K_\epsilon - T_{\text{acc}}$. Hence at most $K_\epsilon = \mathcal{O}(\epsilon^{-2})$ total rounds suffice as $\epsilon \downarrow 0$ with the problem and algorithm parameters fixed. This is a best-iterate guarantee for the combined stationarity, consensus, and tracking residual, counting both productive and all-rejected rounds.

Complexity constants and computational work. The inverse-square tolerance exponent agrees with nonconvex first-order bounds after consistent conversion of squared-residual tolerances [9]. Their assumptions and constants can differ. Here

$$C = C(\theta_0, T_{\text{acc}}, L, \alpha, \Gamma_A, \sigma_A, \omega),$$

and the full coefficient is $CT_{\text{acc}}(\mathcal{V}_0^S - f_{\text{low}})/\nu$. Through w and ν it also depends on β, χ_A and κ_f. The radius and model parameters enter through θ_0, T_{acc} and the geometry bounds. No improvement of this coefficient from curvature information is established.

Every round exchanges two d-vectors per node, including all-rejected rounds. Local work includes model construction, a quadratic solve, objective and gradient evaluations, reusing the base gradient from tracking. Hence a round bound does not bound floating-point work. The experiments report communication and local computation separately. Condition (39) requires uniform trajectory bounds. Any valid upper bound on κ_f may be used in d_M.

4.3 Stability Conditions and Parameter Selection

The two branches in q_M control the passage from consensus error to tracker error and the Schur residual contribution, respectively. For fixed uniform geometry bounds, $q_M < 1$ is equivalent to

$$\omega L(1+\beta) < 1-\beta, \qquad \alpha\sigma_A\beta(1+\omega L) < \omega(1-\beta).$$

A small ω helps the first inequality but increases χ_A/ω and $\alpha\sigma_A\beta/\omega$. The feedback test $a_M(1-q_M) > j_M d_M$ is also required. It combines the descent loss caused by disagreement with the disagreement generated by the mean direction. The resulting $\nu = \lambda_{\min}(N_M)$ enters the iteration coefficient, so a small positive scalar margin can still produce a weak quantitative bound.

Proposition 4.4 (Explicit parameter ranges)**.** *Let $U = \Theta(\delta_{\max})$. The choices $\Gamma_A = \sigma_A = U$ and $\chi_A = \sqrt{U}$ satisfy (38) uniformly over productive rounds. For any fixed geometry bounds and $\alpha > 0$, a weight $\omega > 0$ with $q_M < 1$ exists if and only if*

$$2\alpha L\sigma_A\beta < (1-\beta)^2. \tag{46}$$

When this condition holds, the admissible weights are exactly

$$\frac{\alpha\sigma_A\beta}{1-\beta-\alpha L\sigma_A\beta} < \omega < \frac{1-\beta}{L(1+\beta)}. \tag{47}$$

For the automatic bounds, fix $0 < \omega < (1-\beta)/[L(1+\beta)]$ and any fixed upper bound $\bar\kappa \in [\kappa_f, L]$. Define

$$\begin{aligned}
Q_\omega &= \beta + \omega L(1+\beta), & B_\omega &= U\beta(1+\omega L)/\omega,\\
c_\Gamma &= LU/2, & J_\omega &= \max\{L\sqrt{U}, \sqrt{U}/\omega\},\\
D_\omega &= \{\beta(1+\omega L) + \omega\bar\kappa\}\sqrt{U},\\
E_\omega &= c_\Gamma(1-Q_\omega) + J_\omega D_\omega, && (48)\\
\alpha_\omega &= \min\left\{1, \frac{Q_\omega-\beta}{B_\omega}, \frac{1-Q_\omega}{2E_\omega}\right\} > 0. && (49)
\end{aligned}$$

The middle entry is $+\infty$ when $B_\omega = 0$. Then every $0 < \alpha \le \alpha_\omega$ satisfies (39) with these automatic bounds.

Proof. It follows from Lemma 4.1 that $0 \preceq \mathcal{A} \preceq U\mathcal{I}$, so $\bar A \preceq UI$ and $\mathcal{A}^2 \preceq U\mathcal{A}$. Hence

$$K_A^*K_A = S^*\mathcal{A}^2S - \bar A^2 \preceq U\bar A - \bar A^2, \qquad 0 \preceq Z \preceq \mathcal{Q}\mathcal{A}\mathcal{Q} \preceq U\mathcal{I}_{\mathcal{C}^\perp}.$$

Congruence by $\bar A^{-1/2}$ gives $\left\|K_A\bar A^{-1/2}\right\|_{\mathrm{op}}^2 \le U - \lambda_{\min}(\bar A) \le U$, which proves the automatic bounds.

Put $r = 1-\beta > 0$. The first branch of $q_M < 1$ is equivalent to $0 < \omega < r/[L(1+\beta)]$. If $\beta\sigma_A = 0$, the second branch is $\beta < 1$. Condition (46) then holds automatically, and (47) reduces to the same interval for ω. Otherwise, the second branch requires

$r - \alpha L\sigma_A\beta > 0$ and $\omega > \alpha\sigma_A\beta/(r - \alpha L\sigma_A\beta)$. The resulting interval is nonempty exactly when $2\alpha L\sigma_A\beta < r^2$. Conversely, this last inequality gives $\alpha L\sigma_A\beta < r^2/2 < r$, proving positivity of the denominator and the stated equivalence. Equality gives no strictly feasible weight.

For the automatic bounds, $\beta < Q_\omega < 1$, $c_\Gamma, J_\omega > 0$, and $B_\omega, D_\omega \geq 0$. If $0 < \alpha \leq \alpha_\omega$, then $\beta + \alpha B_\omega \leq Q_\omega$, so $q_M = Q_\omega$. Moreover, $a_M = \alpha(1 - \alpha c_\Gamma)$, $j_M = \alpha J_\omega$, and $d_M \leq \alpha D_\omega$. Therefore

$$\begin{aligned} a_M(1 - q_M) - j_M d_M &\geq \alpha\left[(1 - Q_\omega) - \alpha\{c_\Gamma(1 - Q_\omega) + J_\omega D_\omega\}\right] \\ &\geq \alpha(1 - Q_\omega)/2 > 0. \end{aligned}$$

The denominator in (49) remains positive when $D_\omega = 0$, because $c_\Gamma(1 - Q_\omega) > 0$. □

To apply the theorem, first fix the problem, model, radius, and communication constants, then obtain θ_0 and T_{acc} from Lemma 4.2. For an a priori choice, $\bar{\kappa} = L$ and the automatic bounds make (49) depend only on these fixed data. Sharper geometry bounds provide a second route, requiring a separate argument valid at every productive round of the actual trajectory at the chosen step size. Both routes must satisfy the full stability test, including its feedback inequality. Finite recorded rounds can check their own geometry but cannot verify the uniform bounds over an infinite trajectory. A change in α can change the decisions, so bounds measured at one step size cannot be reused without justification.

The acceptance windows recover stationarity from the mean-direction energy. An all-rejected round has $u_k = 0$ even when $g_k > 0$, but each node accepts within the next T_{acc} rounds. The proof compares the gradient at the beginning of the window with these accepted trackers at a common index, accounting for disagreement and mean movement. Overlapping windows explain the factor T_{acc} in the complexity bound. All constants remain fixed as ϵ decreases, and the finite-horizon rate concerns the best combined residual rather than the last iterate.

Remark 4.5 (Small steps and absence of rejection). With ℓ from (32), every trial is accepted if

$$0 < \alpha \leq \alpha_{\mathrm{nr}} := \min\left\{1, \frac{2(1 - \eta_1)(b_{\min} + \ell)}{(L - b_{\min})_+ + (1 - \eta_1)b_{\min}}\right\}. \tag{50}$$

Indeed, rearranging (32) under this condition gives $\rho_i^k \geq \eta_1$ at every nonzero tracker, while zero trackers are accepted by convention. Hence automatic small-step feasibility alone need not demonstrate a regime with rejection. The following family verifies the full condition at $\alpha = 1$ with infinitely many genuine nonzero rejections.

Proposition 4.6 (An explicit family with recurring genuine rejections)**.** *Let $n = 4$, $d = 1$, $v = (1, 1, -1, -1)^\mathsf{T}$, and $0 \leq \varepsilon \leq 1/100$. Set $f_i(x) = (1 + \varepsilon v_i)x^2/2$, $B_i^k = 1/4$, and let*

$$W = \frac{1}{3}\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

be the equal-weight mixing matrix on the cycle with edges $\{1,2\}$, $\{2,3\}$, $\{3,4\}$ and $\{4,1\}$. Choose

$$\alpha = 1, \quad (\eta_1, \eta_2) = (1/10, 3/4), \quad (\gamma_{\mathrm{inc}}, \gamma_{\mathrm{dec}}, \delta_{\max}) = (8, 1/8, 2),$$

and initialize $X^0 = \mathbf{1}$, $Y_i^0 = 1 + \varepsilon v_i$, and $\delta^0 = (1/4, 1/4, 2, 2)^\mathsf{T}$. Every tracker remains positive, each node alternates between acceptance and rejection of a nonzero trial, and $X^k, Y^k \to 0$. Condition (39) holds uniformly with

$$L = 1 + \varepsilon, \quad \kappa_f = \varepsilon, \quad \beta = 1/3, \qquad (\Gamma_A, \chi_A, \sigma_A) = (1/8, 1/(2\sqrt{2}), 1/4).$$

Proof. The average objective is $f(x) = x^2/2$ and $Y^0 = G(X^0)$. Since $\mathcal{Q}G(Sx) = \varepsilon vx$ and $\|v\|_n = 1$, the Lipschitz modulus of the gradient heterogeneity is exactly $\kappa_f = \varepsilon$. At either radius and any nonzero scalar tracker, the subproblem solution is $p_i = -\delta_i y_i$, with multiplier $\delta_i^{-1} - 1/4 > 0$. Direct substitution into the corrected ratio gives

$$\rho_i(1/4) = \frac{28 - 4\varepsilon v_i}{31} > 3/4, \qquad \rho_i(2) = -\frac{4}{3}\varepsilon v_i < 1/10.$$

We verify the nonzero trackers and these decisions jointly. Put $s_k = (-1)^k$ and represent the states as

$$X^k = m_k \mathbf{1} + s_k z_k v, \qquad Y^k = (m_k + \varepsilon s_k z_k)\mathbf{1} + s_k h_k v.$$

The initial triple is $(m_0, z_0, h_0) = (1, 0, \varepsilon)$. Suppose (m, z, h) lies in $\mathcal{K} = \{m > 0,\ 0 \le z \le m/12,\ 0 \le h \le m/8\}$ and nodes with $v_i = s_k$ have radius factor $1/4$, the others 2. Then every tracker is at least $(7/8 - \varepsilon/12)m > 0$. The actual ratios above interchange the radius factors and give $D^k = -(m + h + \tau z)(\mathbf{1} + s_k v)/8$, where $\tau = \varepsilon s_k$. Using $W\mathbf{1} = \mathbf{1}$, $Wv = v/3$, and (7) yields the next represented triple, with the sign reversed:

$$\begin{pmatrix} m' \\ z' \\ h' \end{pmatrix} = \begin{pmatrix} \frac{7}{8} & -\frac{\tau}{8} & -\frac{1}{8} \\ \frac{1}{24} & -\frac{1}{3} + \frac{\tau}{24} & \frac{1}{24} \\ \frac{1}{24} + \frac{\tau}{8} & \frac{2}{3} + \frac{\tau}{24} + \frac{\tau^2}{8} & -\frac{7}{24} + \frac{\tau}{8} \end{pmatrix} \begin{pmatrix} m \\ z \\ h \end{pmatrix}. \tag{51}$$

The cone bounds and $|\tau| \le 1/100$ imply

$$\frac{4}{5}m \le m' \le \frac{9}{10}m, \qquad 0 \le z' \le m/16 \le m'/12, \qquad 0 \le h' \le m/10 \le m'/8.$$

For example, substitution uses $|\tau z| \le m/1200$, $|\tau h| \le m/800$, and $0 \le \tau^2 z \le m/120000$ in (51). Thus $\mathcal{K}$ is invariant. It contains the initial triple, so induction establishes the states, positive trackers, radius pattern and actual decisions at every round. Moreover $m_k \le (9/10)^k$, which gives the asserted limits.

Exactly two accepted inverse blocks equal $1/4$ and the others are zero. Equations (23) and (29) give the stated geometry. With $\omega = 1/4$, substitution into (39) gives

$$q_M = (9 + \varepsilon)/12 \le 901/1200 < 1,$$
$$a_M(1 - q_M) - j_M d_M = \frac{5 - 50\varepsilon + \varepsilon^2}{192} \ge \frac{45001}{1920000} > 0. \qquad \square$$

The case $\varepsilon = 0$ is homogeneous. For $\varepsilon > 0$, the heterogeneous gradients participate in the actual trajectory. The family establishes compatibility of the sufficient assumptions with infinitely many genuine rejections within this structured class. Here $U = 2$. At the same $\alpha = 1$ and $\omega = 1/4$, the automatic bounds in Proposition 4.4 give $q_M = (11 + 2\varepsilon)/3 > 1$, whereas the exact accepted-block bounds give $q_M = (9 + \varepsilon)/12 < 1$. Thus the sufficient test can fail with coarse bounds even when finer geometry verifies it for the actual run.

5 Numerical Verification and Comparisons

We check implementation, fixed-parameter behavior and independently tuned comparisons in serial MATLAB R2025a runs on a MacBook Pro with an M1 processor and 16 GB of memory. The implementation stores states as d-by-n columns, the transpose of the paper convention. We use the combined residual $R_k = (g_k^2 + c_k^2 + t_k^2)^{1/2}$ throughout.

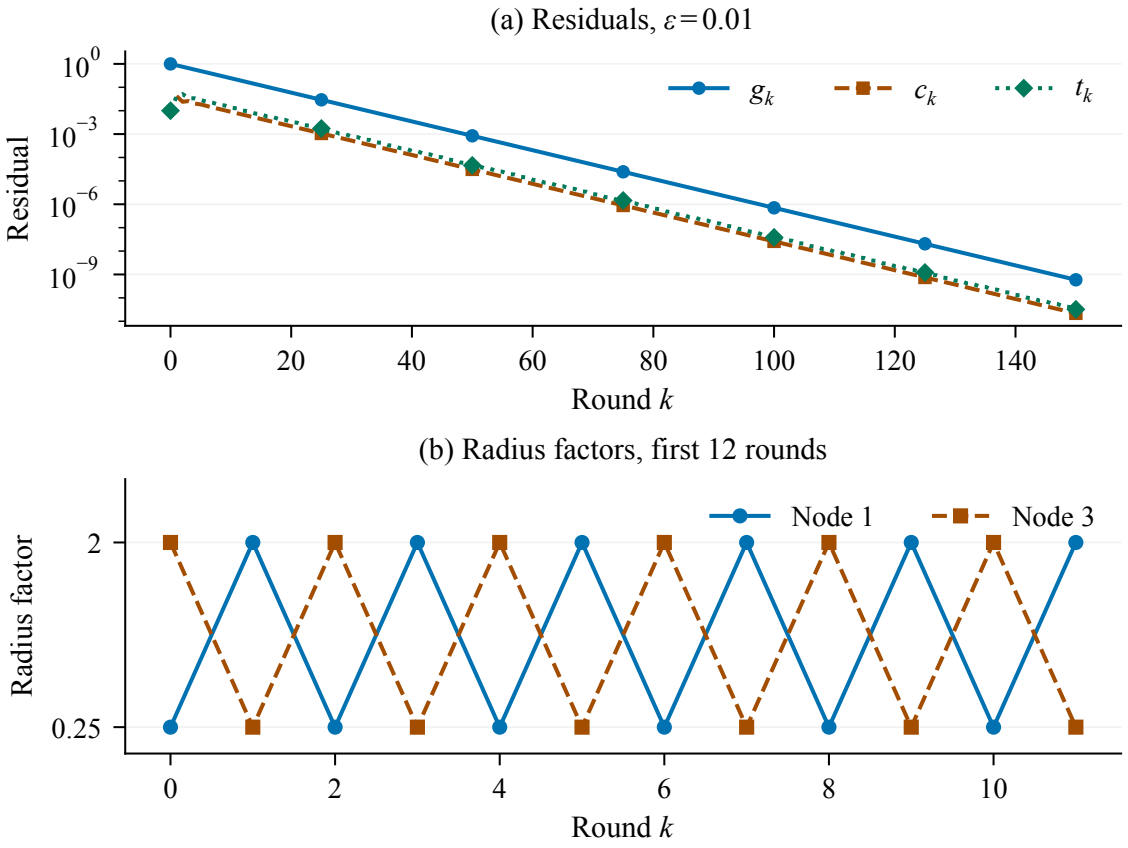


Figure 1: The four-node ring of Proposition 4.6 with $\varepsilon = 0.01$. (a) Stationarity, consensus and tracking residuals over 150 rounds. The initial zero consensus residual is omitted only from the logarithmic axis. (b) Radius factors at nodes 1 and 3 over the first 12 rounds.

5.1 A Numerical Check of the Rejection Mechanism

For $\varepsilon \in \{0, 0.005, 0.01\}$ in Proposition 4.6, each 150-round execution produces 300 genuine nonzero rejections. At $\varepsilon = 0.01$, the final residual is 5.92×10^{-10} and the constant stability margin is 0.0234380208. Figure 1 plots residuals and alternating radius factors. These runs check the subproblem, corrected ratio and radius updates. The invariant-cone proof supplies the infinite-time guarantee.

5.2 Test Families and Evaluation Protocol

The local objective families are

$$f_i^{\mathrm{quad}}(x) = \tfrac{1}{2} x^{\mathsf{T}} H_i x + c_i^{\mathsf{T}} x,$$

$$f_i^{\cos}(x) = \tfrac{q}{2} \|x\|^2 + a[1 - \cos(x_1)] + \frac{1}{2d} \sum_{\ell=1}^{2d} [1 - \cos(v_{i\ell}^{\mathsf{T}} x)] + c_i^{\mathsf{T}} x,$$

We set $q = 0.05$ and $a = 2.5$. The quadratic Hessian eigenvalues are uniform on $[0.1, 1]$. Independent Gaussian–QR orthogonal matrices and eigenvalues uniform on $[0.25, 1]$ define the base SPD models $\widetilde{B}_i$. The actual models are $B_i = s\widetilde{B}_i$ for $s \in \{1, 1/4\}$. Each $v_{i\ell}$ is a normalized Gaussian vector of norm $\sqrt{2}$. The vectors c_i are 0.15 times standard Gaussian vectors, centered across nodes. Initial states are independent Gaussian columns with covariance $4I_d$, and $Y^0 = G(X^0)$. We use $W = (I + \mathrm{Adj}(C_n))/3$, where $\mathrm{Adj}(C_n)$ is

the adjacency matrix of the cycle graph C_n with edges $\{i, i+1\}$, indices taken modulo n. Each node thus mixes its own state and those of its two neighbors with equal weights $1/3$.

The common cosine term guarantees nonconvexity for every sampled average objective. Indeed, at $x = \pi e_1$,

$$e_1^\top \nabla^2 f(x) e_1 = q - a + \frac{1}{2dn} \sum_{i,\ell} \cos(\pi v_{i\ell,1}) v_{i\ell,1}^2 \le q - a + 2 = -0.45.$$

The cosine gradients are globally 4.55-Lipschitz, since the quadratic, shared cosine and random cosine Hessian norms are bounded by q, a and 2, respectively. The average objective is bounded below by $q\|x\|^2/2$, because the centered linear terms cancel and all cosine terms are nonnegative. The shared term was fixed during development before validation and testing.

Eight cells combine the two objective families, sizes $(n, d) = (12, 8), (32, 20)$ and model scales $s = 1, 1/4$. Development, validation and test instances use disjoint sets of one, five and twenty seeds. All test seeds are retained, giving 160 problem settings per protocol. The model scales share base fixtures, so repeated settings and methods are paired observations rather than independent replications.

We compare GT, preconditioned GT (PGT), frozen-radius TR and adaptive TR. GT applies $D = -\alpha Y$, and PGT applies $d_i = -\alpha B_i^{-1} y_i$. Both TR variants use the corrected ratio and $(\eta_1, \eta_2) = (0.1, 0.75)$, and adaptive TR uses $(\gamma_{\mathrm{dec}}, \gamma_{\mathrm{inc}}) = (1/2, 2)$. Both solve (2) in the eigenbasis of B_i, with the multiplier computed by bisection on the secular equation and taken at the feasible end of the final bracket. The fixed-parameter protocol sets $\alpha = 1$ and $\delta_i^0 = \delta_{\max} = 5$. It examines behavior at a common parameter choice without assigning a tuned performance ranking.

The tuned protocol allocates 12 candidates per method and cell. GT and PGT use the step-size grid

$$\mathcal{A}_{\mathrm{GT}} = \{0.01, 0.02, 0.025, 0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1\}.$$

Both TR variants use $\alpha \in \{0.05, 0.2, 0.5, 1\}$ and $\delta_i^0 = \delta_{\max} \in \{0.25, 1, 5\}$. Hence frozen TR also selects its constant radius on validation data. For each candidate, five validation runs execute 350 rounds without stopping at the residual tolerance. The residual is first clipped to $[10^{-15}, 10^{12}]$. The score is the mean of its base-ten logarithm over the five validation seeds. A diverged or nonfinite run receives score 12. The first candidate within 10^{-12} of the minimum mean score is selected, in the recorded grid order. This gives 1920 validation runs. The protocol and candidate rules were frozen before these runs.

Both test protocols stop at $R_k \le 10^{-5}$ or after 1500 rounds. The divergence guard is $R_k > 10^{12}$ or a nonfinite state or residual. Each protocol has 640 method runs. Success denominators retain failures. Round medians use successful runs only. Early-stopped terminal residuals are not fixed-budget values.

5.3 Complete Outcomes and Radius Activity

The fixed protocol gives 40, 3, 21 and 17 successes for GT, PGT, frozen TR and adaptive TR, respectively. PGT diverges in 69 of its 160 fixed runs, whereas neither TR variant, which uses the same models, diverges. At this common parameter choice the trust-region safeguard therefore prevents the divergence of model-based steps, although GT, which does not diverge either, succeeds more often. This matches the classical role of trust regions as a globalization safeguard for model-based steps [4]. The tuned counts are 80, 80, 79 and 80, and no tuned run diverges. All tuned 32-node runs exhaust the budget. Table 1 reports their terminal residuals. In cells 3, 4, 7 and 8, adaptive TR has median ratios R_0/R_{1500} of 1.14×10^5, 1.14×10^5, 1.62×10^3 and 1.64×10^3, respectively. Both TR

variants select $\alpha = 0.05$, their grid minimum, in every 32-node cell. Progress within the budget and boundary selections do not identify the cause of missed tolerance or imply future attainment. The comparisons remain conditional on the method-specific grids and do not establish scalability or a general advantage for adaptive TR.

Table 1: Terminal residual medians in the tuned $(n, d) = (32, 20)$ tests. Each entry includes all 20 seeds. All 320 method runs stop at the 1500-round budget without reaching 10^{-5}. Frozen and adaptive TR select the same parameters and no radius factor changes, so their runs coincide seed by seed.

Cell	Family	s	GT	PGT	Frozen/Adaptive TR
3	Quadratic	1	2.75×10^{-4}	2.33×10^{-4}	8.99×10^{-5}
4	Quadratic	1/4	2.75×10^{-4}	9.17×10^{-3}	8.99×10^{-5}
7	Cosine	1	3.22×10^{-3}	4.92×10^{-3}	5.71×10^{-3}
8	Cosine	1/4	3.22×10^{-3}	7.47×10^{-3}	5.58×10^{-3}

Table 2 reports all outcomes and selected parameters.

In the fixed protocol, frozen TR rejects 55.22% of its nonzero trials and adaptive TR 12.51%, with longest all-rejected episodes of 1500 and 1 rounds, respectively, while adaptive TR changes a nodewise radius factor 1312626 times. Radius adaptation therefore changes the rejection behavior, but the shorter all-rejected episodes do not yield a higher success rate on these settings.

In the tuned protocol, adaptive TR accepts every trial and never changes a radius factor, although positive multipliers occur in 120 of its 160 runs. Among the 80 settings where both adaptive TR and GT succeed, adaptive TR needs fewer rounds in 18 and more in 62, with median round ratio 1.066. Against PGT the counts are 28 fewer, 12 more and 40 equal, with median ratio 1. Hence the tuned results do not isolate a performance benefit from radius adaptation. We do not verify the uniform sufficient condition (39) for these random trajectories, and their finite records support no infinite-time conclusion.

Accuracy and work. Across all 640 TR tests, the normalized KKT errors of the computed subproblem solutions are at most 5.54×10^{-13}. Each round exchanges $4|E|d$ scalars. Over seven serial repetitions on the first test seed per cell, the median wall times of the TR variants are 2.8 to 14.4 times those of GT. These times measure implementation cost rather than distributed elapsed time.

6 Conclusions

We in this paper develop a decentralized trust-region method with local corrected acceptance tests and adaptive relative radii. Finite acceptance windows and Schur energy estimates establish vanishing stationarity, consensus, and tracking residuals under explicit sufficient conditions, including genuine local rejections at nonzero trackers. The best-iterate complexity counts productive and all-rejected communication rounds.

The explicit parameter ranges show that the sufficient regime is nonempty. The heterogeneous four-node family proves compatibility of the full stability condition with infinitely many genuine rejections, while the accepted-block formulas explain the effects of model, multiplier, and acceptance differences. In the reported fixed-parameter tests, the trust-region safeguard prevents the divergence that PGT exhibits with the same models, and adaptive radii shorten all-rejected episodes without increasing the success rate. Independently tuned runs provide no general performance advantage from radius adaptation. These findings concern the reported problem families and protocols. Efficiently

Table 2: Complete fixed and tuned outcomes, with 20 settings per cell and method. Cells 1–4 are quadratic and 5–8 cosine. Within each family, (n, d, s) follows $(12, 8, 1)$, $(12, 8, 1/4)$, $(32, 20, 1)$, $(32, 20, 1/4)$. In the tuned protocol every method succeeds in all settings of cells 1, 2, 5 and 6, except frozen TR in cell 5 (19/20), and in none of cells 3, 4, 7 and 8. Tuned TR trials are all accepted except for frozen TR in cell 5, which rejects 60.4% of its nonzero trials. Tuned rounds are medians among successful runs. Dashes indicate no success.

	Quadratic				Cosine			
Cell	1	2	3	4	5	6	7	8
Fixed successes								
GT	20	20	0	0	0	0	0	0
PGT	3	0	0	0	0	0	0	0
Frozen TR	15	5	0	0	1	0	0	0
Adaptive TR	11	3	0	0	3	0	0	0
Tuned median rounds								
GT	265.5	265.5	–	–	491.5	491.5	–	–
PGT	263.5	302.5	–	–	481.5	572	–	–
Frozen TR	263.5	277.5	–	–	439	558	–	–
Adaptive TR	263.5	277.5	–	–	481.5	558	–	–
Tuned step size α								
GT	0.1	0.1	0.02	0.02	0.1	0.1	0.025	0.025
PGT	0.05	0.01	0.01	0.01	0.05	0.01	0.02	0.01
Frozen TR	0.05	0.5	0.05	0.05	0.5	0.5	0.05	0.05
Adaptive TR	0.05	0.5	0.05	0.05	0.05	0.5	0.05	0.05
Tuned radius $\delta_{\max}$								
Frozen TR	5	0.25	0.25	0.25	5	0.25	1	1
Adaptive TR	5	0.25	0.25	0.25	5	0.25	1	1

verifiable uniform bounds for broader heterogeneous problems, full-sequence convergence, last-iterate rates, inexact subproblems, and weaker communication assumptions remain open directions. The identities of Proposition 3.2 and Lemma 3.4 are algebraic and hold for any accepted blocks $a_i R_i$ with $R_i \succ 0$, whereas the acceptance windows of Lemma 4.2 use the trust-region radius update. Pairing these identities with acceptance windows for other local acceptance rules, such as local line searches, is a natural extension.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (grant no. 12201428) and in part by the Natural Science Foundation of Top Talent of SZTU (grant no. GDRC202136).

Declarations

Competing Interests The author declares no competing interests.

Data and Code Availability The code and data supporting the numerical results are available from the author upon reasonable request.

References

[1] Armacki, A., Jakovetić, D., Krejić, N., Krklec Jerinkić, N.: Distributed trust-region method with first order models. In: IEEE EUROCON 2019—18th International Conference on Smart Technologies, pp. 1–6 (2019). https://doi.org/10.1109/EUROCON.2019.8861739

[2] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-48311-5

[3] Bellavia, S., Malaspina, G., Morini, B.: DOFFO_TR: a decentralized objective function-free optimization method with trust-region. arXiv:2609.00878v1 (2026). URL https://arxiv.org/abs/2609.00878v1

[4] Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust Region Methods. SIAM (2000). https://doi.org/10.1137/1.9780898719857

[5] Curtis, F.E., Lubberts, Z., Robinson, D.P.: Concise complexity analyses for trust region methods. Optim. Lett. **12**(8), 1713–1724 (2018). https://doi.org/10.1007/s11590-018-1286-2

[6] Das, S., Schenato, L., Dey, S.: On convergence analysis of Network-GIANT: An approximate Hessian-based fully distributed optimization algorithm. arXiv:2602.14830v2 (2026). URL https://arxiv.org/abs/2602.14830v2

[7] Di Lorenzo, P., Scutari, G.: NEXT: In-network nonconvex optimization. IEEE Trans. Signal Inf. Process. Netw. **2**(2), 120–136 (2016). https://doi.org/10.1109/TSIPN.2016.2524588

[8] Eisen, M., Mokhtari, A., Ribeiro, A.: Decentralized quasi-Newton methods. IEEE Trans. Signal Process. **65**(10), 2613–2628 (2017). https://doi.org/10.1109/TSP.2017.2666776

[9] Koloskova, A., Lin, T., Stich, S.U.: An improved analysis of gradient tracking for decentralized machine learning. In: Advances in Neural Information Processing Systems, vol. 34, pp. 11,422–11,435 (2021). URL https://proceedings.neurips.cc/paper/2021/hash/5f25fbe144e4a81a1b0080b6c1032778-Abstract.html

[10] Maritan, A., Sharma, S., Schenato, L., Dey, S.: Network-GIANT: Fully distributed Newton-type optimization via harmonic Hessian consensus. arXiv:2305.07898 (2023). URL https://arxiv.org/abs/2305.07898

[11] Mokhtari, A., Ling, Q., Ribeiro, A.: Network Newton distributed optimization methods. IEEE Trans. Signal Process. **65**(1), 146–161 (2017). https://doi.org/10.1109/TSP.2016.2617829

[12] Moré, J.J., Sorensen, D.C.: Computing a trust region step. SIAM J. Sci. Stat. Comput. **4**(3), 553–572 (1983). https://doi.org/10.1137/0904038

[13] Nedić, A., Olshevsky, A., Shi, W.: Achieving geometric convergence for distributed optimization over time-varying graphs. SIAM J. Optim. **27**(4), 2597–2633 (2017). https://doi.org/10.1137/16M1084316

[14] Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer (2006). https://doi.org/10.1007/978-0-387-40065-5

[15] Qu, G., Li, N.: Harnessing smoothness to accelerate distributed optimization. IEEE Trans. Control Netw. Syst. **5**(3), 1245–1260 (2018). https://doi.org/10.1109/TCNS.2017.2698261

[16] Scutari, G., Sun, Y.: Distributed nonconvex constrained optimization over time-varying digraphs. Math. Program. **176**, 497–544 (2019). https://doi.org/10.1007/s10107-018-01357-w

[17] Shi, W., Ling, Q., Wu, G., Yin, W.: EXTRA: An exact first-order algorithm for decentralized consensus optimization. SIAM J. Optim. **25**(2), 944–966 (2015). https://doi.org/10.1137/14096668X

[18] Shorinwa, O., Schwager, M.: Distributed quasi-Newton method for multi-agent optimization. IEEE Trans. Signal Process. **72**, 3535–3546 (2024). https://doi.org/10.1109/TSP.2024.3424436

[19] Sun, Y., Scutari, G., Daneshmand, A.: Distributed optimization based on gradient tracking revisited: Enhancing convergence rate via surrogation. SIAM J. Optim. **32**(2), 354–385 (2022). https://doi.org/10.1137/19M1259973

[20] Tian, Y., Sun, Y., Scutari, G.: Asynchronous decentralized successive convex approximation. arXiv:1909.10144v2 (2020). URL https://arxiv.org/abs/1909.10144v2

[21] Wei, H., Qu, Z., Wu, X., Wang, H., Lu, J.: Decentralized approximate Newton methods for convex optimization on networked systems. IEEE Trans. Control Netw. Syst. **8**(3), 1489–1500 (2021). https://doi.org/10.1109/TCNS.2021.3070663

[22] Wu, H., Wang, L.: A unified algorithm for nonconvex decentralized nonlinear optimization. arXiv:2511.19182v3 (2026). URL https://arxiv.org/abs/2511.19182v3

[23] Xiao, L., Boyd, S.: Fast linear iterations for distributed averaging. Syst. Control Lett. **53**(1), 65–78 (2004). https://doi.org/10.1016/j.sysconle.2004.02.022

[24] Xin, R., Khan, U.A., Kar, S.: An improved convergence analysis for decentralized online stochastic non-convex optimization. IEEE Trans. Signal Process. **69**, 1842–1858 (2021). https://doi.org/10.1109/TSP.2021.3062553

[25] Xu, J., Zhu, S., Soh, Y.C., Xie, L.: Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes. In: Proceedings of the 54th IEEE Conference on Decision and Control (CDC), pp. 2055–2060 (2015). https://doi.org/10.1109/CDC.2015.7402509

[26] Zhang, F. (ed.): The Schur Complement and Its Applications, *Numerical Methods and Algorithms*, vol. 4. Springer, New York (2005). https://doi.org/10.1007/b105056

[27] Zhang, J., Ling, Q., So, A.M.-C.: A Newton tracking algorithm with exact linear convergence for decentralized consensus optimization. IEEE Trans. Signal Inf. Process. Netw. **7**, 346–358 (2021). https://doi.org/10.1109/TSIPN.2021.3083975