DECENTRALIZED OPTIMIZATION WITH CROSS-COUPLED MIXED AFFINE CONSTRAINTS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study decentralized optimization with cross-coupled mixed affine constraints, where local and shared variables interact through two affine channels. We show that the intrinsic difficulty of combining separately well-conditioned channels is governed by their Friedrichs angle. This geometry induces a cross-coupling factor that cannot be removed by channelwise preconditioning and governs the additional affine-oracle and communication complexity. We develop an accelerated decentralized method with matching minimax guarantees in the smooth strongly convex regime and extend the framework to smooth and nonsmooth convex objectives. Experiments confirm the predicted dependence on cross-channel geometry and network conditioning.

1 INTRODUCTION

Decentralized first-order optimization is a popular approach for solving problems where the data is distributed across compute nodes, and the communication between nodes is slow, unreliable or restricted by privacy concerns (Lian et al., 2017; Gorbunov et al., 2022). It is best studied in the *consensus optimization* case where the global variable x is shared by locally stored terms $f_i(x)$ of the objective $\min_x \sum_i f_i(x)$, which corresponds to data-parallel learning. This setup has received thorough practical examination (Assran et al., 2019; Ying et al., 2021; Yuan et al., 2022) and has a well-developed complexity theory with optimal algorithms and matching lower complexity bounds (Scaman et al., 2017; 2018; Kovalev et al., 2020; Tyurin, 2026).

While in consensus optimization all compute nodes work with the same shared variable x, many practical applications require more general coupling between variables. Recently, optimal algorithms were established for the case of *coupled constraints*, where local variables x_i are subject to a kind of resource-allocation constraint $\min_x \sum_i f_i(x_i)$ s.t. $\sum_i (A_i x_i - b_i) = 0$ (Yarmoshik et al., 2025; 2026). This is a generalization of consensus optimization, which also makes it possible to express model-parallel learning formulations and different sharing constraints in control and operations research (Yarmoshik et al., 2026; Boyd et al., 2011).

In these settings, optimal complexities are described by the objective, network, and affine condition numbers (Scaman et al., 2017; 2018; Salim et al., 2022; Yarmoshik et al., 2025; 2026). Mathematically, the core question is how the conditioning of one problem component affects the others. We study it for local variables x_i and a shared variable z. Existing optimal methods cover coupled constraints $\sum_i (A_i x_i - b_i) = 0$ (Yarmoshik et al., 2025) and the mixed-affine model in which z enters only node-wise constraints (Yarmoshik et al., 2026). For n agents, we define the cross-coupled mixed-affine problem as

$$\begin{aligned} \min_{x_1,\dots,x_n,z} \quad & \sum_{i=1}^{n} f_i(x_i, z), \\ \text{s.t.} \quad & \sum_{i=1}^{n} (A_i x_i + B_i z - b_i) = 0, \\ & C_i x_i = c_i, \qquad D_i z = e_i, \qquad i = 1, \dots, n. \end{aligned} \tag{CC-MAC}$$

Thus z enters both the aggregate equality and the node-wise shared-variable constraints. The canonical reduction separates these equations into two operators K_1 and K_2. Let $U = \mathrm{Range}(K_1^\top)$

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

and $V = \mathrm{Range}(K_2^\top)$; for any subspace S, let P_S denote its Euclidean orthogonal projector. After exact separate left whitening and stacking the channels as $\widehat{K} = [\widehat{K}_1; \widehat{K}_2]$, the joint normal operator is

$$\widehat{K}^\top \widehat{K} = P_U + P_V. \tag{1}$$

When U and V contain nearly aligned but nonidentical directions, the smallest positive eigenvalue of (1) can be arbitrarily small. Thus two channels can be separately ideal and jointly ill conditioned. Exact overlap is benign—a common direction contributes eigenvalue 2—whereas *near-overlap* is the source of difficulty.

Where the CC-MAC structure arises. The formulation is useful when a system has genuinely local decisions and a shared decision, and both must enter the same aggregate equality. Three representative examples are:

- *Vertical federated learning.* Agent i owns a feature block and its weight x_i, while z is the shared prediction vector, as in vertically partitioned federated learning (Hardy et al., 2017). The identity $Xx - z = 0$ can be written as $\sum_i (X_i x_i - \alpha_i z) = 0$ with $\sum_i \alpha_i = 1$, so $A_i = X_i$ and $B_i = -\alpha_i I$. Consensus or admissibility constraints on copies of z form the second channel. This is also the full-data model used in Section 6.
- *Networked resource allocation and control.* Local dispatch or control variables x_i must satisfy equipment equations $C_i x_i = c_i$, while a shared reserve, reference trajectory, or market schedule z satisfies $D_i z = e_i$. Aggregate balance has the form $\sum_i (A_i x_i + B_i z - b_i) = 0$: the shared schedule changes every node's contribution to the same system-wide balance, as in distributed optimization and ADMM models (Boyd et al., 2011).
- *Personalized models with global calibration.* Local corrections x_i are combined with a common model z (Hanzely and Richtárik, 2020; Liang et al., 2020). A population-level calibration, fairness, or moment-matching equality aggregates local statistics and the common predictor as $\sum_i (A_i x_i + B_i z - b_i) = 0$ (Agarwal et al., 2018); local identifiability and shared-policy equalities remain in the second channel.

In all three cases $B_i \neq 0$ is structural rather than a cosmetic reparameterization: deleting it changes the feasible model, while merging the two equation classes changes which operations are locally available and therefore changes the oracle being charged.

Setting $B_i = 0$ recovers the previous mixed-affine model; in our canonical reformulation the two channel row spaces are then orthogonal and the cross-coupling factor equals one. This means only that no *additional cross-channel* conditioning remains after the channels are normalized separately. It does not make the primitive affine realization unit-cost: products with A_i, C_i, D_i and the communication needed to implement a normalized channel remain charged, while $N_B = 0$.

The two canonical CC-MAC channels are block operators and may each contain arbitrarily many affine equations. After separate normalization, exact common row directions are spectrally benign while near-overlap is hard. If c_F is the Friedrichs cosine after removing exact common directions and $\theta_F := \arccos c_F \in (0, \pi/2]$ (with $\theta_F = \pi/2$ when either reduced subspace is zero), the residual two-channel factor is

$$\chi_\times = \sqrt{\frac{1 + c_F}{1 - c_F}} = \cot\left(\frac{\theta_F}{2}\right). \tag{2}$$

Thus $\chi_\times$ is exactly the conditioning left after within-channel conditioning has been removed for the prescribed channel decomposition. Principal- and Friedrichs-angle tools are classical (Björck and Golub, 1973; Deutsch, 1995); our use identifies the residual factor created by the canonical CC-MAC split and its resource consequences. Separate preprocessing cannot rotate the two row spaces, whereas mixing equations across channel classes changes the charged oracle. The fixed-J extension is in Appendix C.

The resulting picture is summarized by the following contributions.

Contributions.

- **A genuinely cross-coupled decentralized model.** We introduce CC-MAC, in which the shared block enters both the global coupled equality and node-wise affine constraints. The canonical reduction isolates two affine channels whose interaction is absent from the predecessor mixed-affine model: setting $B_i = 0$ makes the channel row spaces orthogonal and recovers $\chi_\times = 1$.

- **A CC-MAC specialization of classical angle geometry.** After separate whitening, the nontrivial spectrum of the prescribed joint normal operator is determined by the Friedrichs angle and the residual condition number is exactly $\chi_\times^2 = (1 + c_F)/(1 - c_F)$. Classical angle and perturbation inequalities (Knyazev and Argentati, 2002) explain the barrier; our contribution is to identify the canonical CC-MAC pair, attain the barrier by channelwise preprocessing, and propagate it through the resource ledger. Exact common directions are deduplicated before the angle is measured.
- **Resource-separated smooth strongly convex guarantees.** For a fixed execution representation with a supplied initial primal–dual energy bound, Cross-APAPC uses $O(\sqrt{\kappa_f}\log_+(R_E/\varepsilon))$ gradients, $O(\chi_\times\sqrt{\kappa_f}\log_+(R_E/\varepsilon))$ normalized-affine calls, and the corresponding supplied-realization communication budget. The matching smooth-SC statement concerns the gradient and normalized-affine resource orders; the physical-communication lower construction is radius-explicit. Appendix I states the oracle and radius scope explicitly.
- **A log-free smooth-convex extension.** A retained-dual continuation with polynomial feasibility repair closes the multiplicative target-accuracy logarithmic gap of the regularization-based mixed-affine analysis. Under a fixed scale-invariant multiplier budget we obtain the matching class-level orders $\Theta(T_{\mathrm{cvx}})$, $\Theta(\chi_\times T_{\mathrm{cvx}})$, and $\Theta(\chi_\times\sqrt{\kappa_W}T_{\mathrm{cvx}})$. We also give nonsmooth resource-separated upper bounds and a fixed-J normalized-oracle extension.
- **Algorithms and full-data validation.** The main text gives Cross-APAPC and the appendix supplies its normalization/filtering primitives; the nonsmooth extension states the additional projection-oracle access required by its recursive-sliding subroutine. On H100 full-data LIBSVM experiments, all 182 controlled geometry/network runs and all 48 robustness runs reach the target accuracy; the measured scaling exponents are 0.988 for normalized affine work in $\chi_\times$ and 0.980 for communication in $\chi_\times\sqrt{\kappa_W}$. Five-seed full-data comparisons report both normalized-oracle and raw primitive ledgers.

Relation to prior and concurrent work. The coupled-constraint method of Yarmoshik et al. (2025) supplies the direct algorithmic and lower-bound lineage for a single global affine mechanism, while Yarmoshik et al. (2026) separates local and shared affine products when the shared block does not enter the coupled equality. CC-MAC keeps those native costs visible but introduces the nonorthogonal canonical pair created by $B_i \neq 0$. The comparison is therefore between different oracle decompositions, not equal-cost matrix units. Concurrent work of Jha (2026) studies one common variable constrained to the intersection of agent-private sets, with projection/constrained-solve access. For affine sets its gap is $1 - c_F^2$ between product feasibility and consensus. Here c_F is instead the angle between two normalized CC-MAC channel row spaces and the oracle separately charges matrix–vector affine products, gradients, and neighbor rounds.

Closest asymptotic comparison. To separate the new cross-channel effect from previously known objective, affine, and network conditioning, define $L_\varepsilon := \log(1/\varepsilon)$, $T_{\mathrm{sc}} := \sqrt{\kappa_f}L_\varepsilon$, $T_{\mathrm{cvx}} := \sqrt{L_f R^2/\varepsilon}$, $\chi_K := \sqrt{\kappa_+(K^\top K)} = \sqrt{\kappa(K)}$, $\Lambda_{\mathrm{cpl}} := \sqrt{\widehat{\kappa}_A}$, $\Lambda_{\mathrm{mix}} := \sqrt{\widetilde{\kappa}_{AC}}(1 + \sqrt{\kappa_C}) + \sqrt{\widehat{\kappa}_{\widetilde{C}^\top}}$, and $\Gamma_{\mathrm{mix}} := (\sqrt{\widetilde{\kappa}_{AC}} + \sqrt{\widehat{\kappa}_{\widetilde{C}^\top}})\sqrt{\kappa_W}$. Table 1 reports the published leading upper orders in the resource units used by each paper; for Yarmoshik et al. (2026), the affine column sums their separately reported $A, C, \widetilde{C}$ products. Our affine entry counts normalized joint-stack calls, while our physical-round realization depends additionally on the supplied channel-normalization implementation. The raw $A/B/C/D$ products and physical-round conversion are expanded in Appendix L. Consequently the affine and communication columns are model/realization dependent and should not be read as equal-unit wall-clock comparisons.

The new dependence is geometric rather than another within-channel condition number: after separate normalization, the residual affine factor is exactly $\chi_\times$. **In the smooth-convex regime all three leading normalized-resource orders are log-free;** the regularization-based mixed-affine bound of Yarmoshik et al. (2026) retains a multiplicative L_ε factor. The table is therefore a leading-order oracle comparison rather than an equal-unit raw-cost ranking; Appendix L gives the explicit conversion for our solver.

Table 1: Closest first-order comparison in the cited works' reported resource models. Affine and communication units are model/realization dependent: predecessor rows use their native accounting, whereas our affine entries count normalized stack products and Appendix L expands our solver into raw matrix products and physical rounds. A dash means that the cited work does not give the corresponding decentralized result in that regime.

Work	regime/model	N_∇	affine work (native model)	N_{comm}
Salim et al. (2022)	smooth SC, $Kx = b$	$O(T_{\mathrm{sc}})$	$O(\chi_K T_{\mathrm{sc}})$	–
Yarmoshik et al. (2025)	smooth SC, coupled	$O(T_{\mathrm{sc}})$	$O(\Lambda_{\mathrm{cpl}} T_{\mathrm{sc}})$	$O(\Lambda_{\mathrm{cpl}}\sqrt{\kappa_W} T_{\mathrm{sc}})$
Yarmoshik et al. (2026)	smooth SC, mixed	$O(T_{\mathrm{sc}})$	$O(\Lambda_{\mathrm{mix}} T_{\mathrm{sc}})$	$O(\Gamma_{\mathrm{mix}} T_{\mathrm{sc}})$
This work	**smooth SC, CC-MAC**	$\mathbf{O}(\mathbf{T}_{\mathrm{sc}})$	$\mathbf{O}(\chi_\times \mathbf{T}_{\mathrm{sc}})$	$\mathbf{O}(\chi_\times \sqrt{\kappa_{\mathbf{W}}} \mathbf{T}_{\mathrm{sc}})$
Yarmoshik et al. (2026)	smooth convex, mixed	$O(T_{\mathrm{cvx}} L_\varepsilon)$	$O(\Lambda_{\mathrm{mix}} T_{\mathrm{cvx}} L_\varepsilon)$	$O(\Gamma_{\mathrm{mix}} T_{\mathrm{cvx}} L_\varepsilon)$
This work	**smooth convex, CC-MAC**	$\mathbf{\Theta}(\mathbf{T}_{\mathrm{cvx}})$	$\mathbf{\Theta}(\chi_\times \mathbf{T}_{\mathrm{cvx}})$	$\mathbf{\Theta}(\chi_\times \sqrt{\kappa_{\mathbf{W}}} \mathbf{T}_{\mathrm{cvx}})$

2 Problem Class and Canonical Reformulation

We follow the notation and oracle model of Yarmoshik et al. (2026). Agent i has $x_i \in \mathbb{R}^{d_i}$, while $z \in \mathbb{R}^q$ is shared, and

$$A_i \in \mathbb{R}^{m\times d_i}, \qquad B_i \in \mathbb{R}^{m\times q}, \qquad b_i \in \mathbb{R}^m,$$
$$C_i \in \mathbb{R}^{p_i\times d_i}, \qquad c_i \in \mathbb{R}^{p_i}, \qquad D_i \in \mathbb{R}^{r_i\times q}, \qquad e_i \in \mathbb{R}^{r_i},$$

where $f_i : \mathbb{R}^{d_i} \times \mathbb{R}^q \to \mathbb{R}$ is local. Agent i stores its displayed data, evaluates ∇f_i, and applies its matrices and adjoints.

Oracle and accuracy conventions. We use the resource-separated linear-span model of Salim et al. (2022); Yarmoshik et al. (2025). A normalized affine call applies the separately normalized joint stack or its adjoint; raw $A/B/C/D$ products, graph polynomials, setup, and physical neighbor rounds are charged in Appendix L. One multiplication by symmetric $W \succeq 0$, $\operatorname{Ker} W = \operatorname{span}\{\mathbf{1}\}$, is one neighbor round, with $\kappa_W := \lambda_{\max}(W)/\lambda^+_{\min}(W)$ (and $\kappa_W = 1$, zero rounds, for $n = 1$).

For nonzero M and positive-semidefinite G, write

$$\kappa(M) := \frac{\sigma_{\max}(M)^2}{\sigma^+_{\min}(M)^2}, \qquad \kappa_+(G) := \frac{\lambda_{\max}(G)}{\lambda^+_{\min}(G)},$$

so $\kappa(M) = \kappa_+(M^\top M)$. Empty channels are omitted. Unless stated otherwise, each f_i is L_f-smooth and μ_f-strongly convex, with $\kappa_f = L_f/\mu_f$.

In the smooth strongly convex regime, ε-accuracy means $\|w_{\mathrm{out}} - w^\star\| \le \varepsilon$ in the execution space, and N_r counts resource-r operations to the first such output. Initialization, certificates, output map, and an upper bound R_E on the APAPC energy of Appendix H belong to the instance. The empirical error in Section 6 is a separate statistic.

2.1 Canonical channels and a decentralized implementation lift

Each agent keeps a local copy $z_i \in \mathbb{R}^q$ of the shared variable. With $d := \sum_i d_i$, write

$$\boldsymbol{x} = \mathrm{col}(x_1, \dots, x_n) \in \mathbb{R}^d, \qquad \boldsymbol{z} = \mathrm{col}(z_1, \dots, z_n) \in \mathbb{R}^{nq}, \qquad u := \mathrm{col}(\boldsymbol{x}, \boldsymbol{z}).$$

Define

$$\begin{aligned} A &:= \mathrm{diag}(A_1, \dots, A_n), & B &:= \mathrm{diag}(B_1, \dots, B_n), \\ C &:= \mathrm{diag}(C_1, \dots, C_n), & D &:= \mathrm{diag}(D_1, \dots, D_n), \\ \boldsymbol{b} &:= \mathrm{col}(b_1, \dots, b_n), & \boldsymbol{c} &:= \mathrm{col}(c_1, \dots, c_n), \qquad \boldsymbol{e} := \mathrm{col}(e_1, \dots, e_n). \end{aligned} \tag{3}$$

The lifted objective is $F(u) := \sum_i f_i(x_i, z_i)$. We use

$$\mathcal{W}_m := W \otimes I_m, \qquad \mathcal{W}_q := W \otimes I_q, \qquad E_m := n^{-1/2}(\mathbf{1}^\top \otimes I_m).$$

Thus $E_m r = 0$ means that the blocks of r sum to zero, while $\mathcal{W}_q \boldsymbol{z} = 0$ is exactly $z_1 = \cdots = z_n$.

Crucially, we define the geometry *before* introducing any auxiliary localization variable. The canonical reduced channels are

$$K_1 := \begin{bmatrix} E_m A & E_m B \\ C & 0 \end{bmatrix}, \qquad K_2 := \begin{bmatrix} 0 & D \\ 0 & \mathcal{W}_q \end{bmatrix}, \qquad K := \begin{bmatrix} K_1 \\ K_2 \end{bmatrix}, \tag{4}$$

with right-hand sides $v_1 := \mathrm{col}(E_m \boldsymbol{b}, \boldsymbol{c})$ and $v_2 := \mathrm{col}(\boldsymbol{e}, 0)$. This system acts only on the original variables and their canonical local copies. The first channel contains the coupled equality and local constraints on x_i; the second contains node-specific constraints on the shared variable and consensus.

Each K_j is a block family. Appendix C gives the fixed-J projector-sum reduction; Appendix B gives an exact graph-local lift whose row-space distortion is controlled by Proposition B.3.

Theorem 2.1 (Canonical reduction and decentralized realization)**.** *A tuple $(x_1, \ldots, x_n, z)$ is feasible for* (CC-MAC) *iff its local-copy vector u satisfies $K_1 u = v_1$ and $K_2 u = v_2$. Appendix B gives an equivalent graph-local lift and proves that the canonical c_F is independent of its auxiliary scalings.*

Each graph-filtered local action costs $O(\sqrt{\kappa_W})$ neighbor rounds before the separate-normalization costs of Appendices G and L. If every $B_i = 0$, the canonical row spaces are orthogonal and $\chi_\times = 1$ (Corollary B.4); only the cross factor and N_B vanish, not the $A/C/D$ or communication costs.

3 The Cross-Angle Geometry

Separately normalized channels yield a sum of row-space projectors (Proposition C.2). For two canonical CC-MAC classes its spectrum is described by principal angles; the Friedrichs angle removes benign exact overlap. Appendix D.1 keeps the feasible problem fixed while making this angle arbitrarily small.

We first isolate the linear-algebraic core. Let $M_1 : \mathbb{R}^d \to \mathbb{R}^{p_1}$ and $M_2 : \mathbb{R}^d \to \mathbb{R}^{p_2}$ be two nonzero linear maps. Set

$$U = \mathrm{Range}(M_1^\top), \qquad V = \mathrm{Range}(M_2^\top).$$

Put $I := U \cap V$. For a consistent affine pair, Lemma C.4 retains exact common equations in one channel without changing feasibility; near-common directions remain. The Friedrichs cosine is

$$c_F(U, V) := \sup\big\{|\langle u, v\rangle| : \ u \in U \cap I^\perp, \ v \in V \cap I^\perp, \ \|u\| = \|v\| = 1\big\}. \tag{5}$$

Set $c_F(U, V) = 0$ if either reduced subspace is zero. When $U \cap V = \{0\}$, this is the largest cosine of the principal angles.

For the canonical pair we henceforth suppress the arguments and write $c_F := c_F(\mathrm{Range}(K_1^\top), \mathrm{Range}(K_2^\top))$. The single cross-coupling factor used in the paper is

$$\chi_\times := \sqrt{\frac{1 + c_F}{1 - c_F}}. \tag{6}$$

In the core model $C = D = 0$, Proposition D.3 reduces c_F to an $m \times m$ generalized eigenproblem in A_i, B_i. Theorem D.1 gives the weighted bounds; after consistent deduplication and exact separate whitening,

$$\kappa\left(\begin{bmatrix} \widehat{M}_1 \\ \widehat{M}_2 \end{bmatrix}\right) = \frac{1 + c_F}{1 - c_F} = \chi_\times^2. \tag{7}$$

Theorem E.2 shows that this value is the infimum over all invertible within-channel preprocessing, so the penalty is not a whitening artifact.

4 Accelerated Method from the Cross-Angle Geometry

We now turn the geometry into an algorithmic statement. The construction reuses the affine-constrained acceleration of Salim et al. (2022) and the decentralized polynomial reductions developed in the line of Kovalev et al. (2020); Yarmoshik et al. (2025; 2026). Algorithmically, the number of channel classes enters only through the spectrum of the normalized joint stack: Proposition C.2 gives condition $O(\chi_{\times,J}^2)$ for J classes. The two-channel Friedrichs factor is the closed-form specialization relevant to CC-MAC.

Exact whitening is unnecessary. Corollary F.3 proves that fixed spectral-equivalence accuracy preserves the two-channel order, while Proposition F.4 and Lemma G.1 give a constructive decentralized realization from the certificates in Assumption 1. The canonical channels K_1, K_2 therefore define the CC-MAC geometry, whereas the graph-filtered local channels $\mathcal{L}_1, \mathcal{L}_2$ implement it; Proposition B.3 controls the corresponding angle distortion. Raw matrix-product and neighbor-round costs are accounted for separately in Appendix L.

Algorithmic organization. Let $M_j = K_j$ in the normalized canonical oracle and $M_j = \mathcal{L}_j$ in the decentralized implementation. Channelwise maps construct $\widehat{M}_j = S_j M_j$ with fixed spectral-equivalence accuracy; thus the hat denotes the normalized execution operator, not a claim that canonical K_j is directly local. Algorithm 1 gives the complete outer iteration here; Appendix H specifies Cheb_N, normalization, certificates, and the proof. The convex continuation and nonsmooth interface are also deferred to their appendices.

Algorithm 1 CROSS-APAPC: accelerated affine optimization after separate channel normalization

Require: w^0, $\widehat{K}$, $\widehat{v}$, smooth strongly convex Φ, L_Φ, μ_Φ, bounds $\lambda_\pm$, horizon T

$N \leftarrow \lceil \sqrt{\lambda_+/\lambda_-} \rceil$, $\zeta \leftarrow \mu_\Phi$

$\tau \leftarrow \min\{1, \frac{1}{2}\sqrt{19/(15\kappa_\Phi)}\}$, $\eta \leftarrow (4\tau L_\Phi)^{-1}$, $\theta \leftarrow 15/(19\eta)$

$w_f^0 \leftarrow w^0$, $p^0 \leftarrow 0$

for $k = 0, 1, \ldots, T-1$ **do**

  $w_g^k \leftarrow \tau w^k + (1-\tau) w_f^k$

  $w^{k+1/2} \leftarrow (1+\eta\zeta)^{-1}[w^k - \eta(\nabla\Phi(w_g^k) - \zeta w_g^k + p^k)]$

  $r^k \leftarrow \theta[w^{k+1/2} - \mathrm{Cheb}_N(w^{k+1/2}; \widehat{K}, \widehat{v})]$

  $p^{k+1} \leftarrow p^k + r^k$

  $w^{k+1} \leftarrow w^{k+1/2} - \eta(1+\eta\zeta)^{-1} r^k$

  $w_f^{k+1} \leftarrow w_g^k + \frac{2\tau}{2-\tau}(w^{k+1} - w^k)$

end for

return w^T and its original-variable coordinate projection

Theorem 4.1 (Accelerated two-channel upper bound)**.** *Fix an execution representation and suppose the inputs include a valid upper bound R_E on the initial APAPC primal–dual energy defined in Appendix H. Under the smooth strongly convex assumptions and fixed-accuracy separate normalization, Cross-APAPC returns an ε-accurate* CC-MAC *primal solution with*

$$N_{\nabla f} = O(T_{\mathrm{sc,E}}), \qquad N_{\mathrm{aff}} = O(\widehat{\chi} T_{\mathrm{sc,E}}),$$
$$N_{\mathrm{comm}} = O\left(\widehat{\chi} T_{\mathrm{sc,E}} \sum_j s_j^{\mathrm{comm}}\right), \quad T_{\mathrm{sc,E}} := \sqrt{\kappa_f}\log_+(R_E/\varepsilon). \tag{8}$$

where $\log_+(t) := \max\{1, \log t\}$ and $\widehat{\chi} := \sqrt{\lambda_+/\lambda_-} = O(\chi_\times)$ for constant-factor-tight certificates. A neighbor-round specialization requires a supplied normalized-channel realization and accounts for all additional communication primitives. This is a fixed-instance energy-controlled guarantee, not a uniform claim over a class that controls only the primal initial radius. The fixed-J normalized-oracle extension and the raw primitive ledger are in Appendices C and L.

Primitive realization. The normalized count in Theorem 4.1 is converted to raw work only after fixing an implementation. For the certified construction of Appendix L, put $T_{\Xi,E} := \sqrt{\kappa_f}\,\Xi \log_+(R_E/\varepsilon)$. Then

$$N_A, N_B = O(T_{\Xi,E}\sqrt{\bar{k}_1}), \quad N_C = O(T_{\Xi,E}\sqrt{\bar{k}_1\bar{\kappa}_C}), \quad N_D = O(T_{\Xi,E}\sqrt{\bar{k}_2}),$$

and

$$N_{\mathrm{aff}}^{\mathrm{prim}} = O(T_{\Xi,E}\Lambda_{\mathrm{in}}), \qquad N_{\mathrm{comm}} = O\Big(T_{\Xi,E}(\sqrt{\bar{k}_1} + \sqrt{\bar{k}_2})\sqrt{\bar{\kappa}_W}\Big).$$

For constant-factor certificates $\Xi = O(\chi_\times)$. If $B = 0$, then $N_B = 0$, but the A, C, D and communication costs remain. This translation is the basis for comparing our normalized theorem-level unit with the native matrix products counted by Yarmoshik et al. (2025; 2026).

4.1 MINIMAX OPTIMALITY

The formal parameter-bounded normalized-channel class and the infimum–supremum resource definition are given in Appendix I. The hard families separately isolate gradient, affine, and physical-communication resources while keeping the other legal resources free.

Theorem 4.2 (Scope of the smooth strongly convex lower bounds)**.** *For the normalized-channel linear-span model with budgets $\kappa_f \le \bar{\kappa}$, $\chi_\times \le \Xi$, $\kappa_W \le \Omega$, and fixed-accuracy normalization, the constructions in Appendix I separately give the standard objective lower order and cross-angle affine lower order,*

$$\mathfrak{C}_\nabla = \Omega\left(\sqrt{\bar{\kappa}}\log\frac{1}{\varepsilon}\right), \qquad \mathfrak{C}_{\mathrm{aff}} = \Omega\left(\sqrt{\bar{\kappa}}\,\Xi\log\frac{1}{\varepsilon}\right), \tag{9}$$

in the stated nontrivial parameter ranges. The current physical-path construction gives the radius-explicit communication lower bound stated there; it is not promoted here to a uniform $\Omega(\Xi\sqrt{\Omega}\log(1/\varepsilon))$ equality at fixed absolute accuracy. Accordingly, Theorem 4.1 and the lower statements use their own explicit radius and oracle hypotheses.

5 CONVEX AND NONSMOOTH EXTENSIONS

The same normalized-channel geometry governs the other convex regimes. Appendix J defines the scale-invariant smooth-convex objective/feasibility criterion and multiplier budget $\Lambda_{\mathrm{cvx}} \le \bar{\Lambda}$; Appendix K states the corresponding full-domain nonsmooth assumptions and criterion. Write

$$T_{\mathrm{cvx}} := \sqrt{L_f R^2/\varepsilon}, \qquad T_{\mathrm{ns}} := M_{3R}R/\varepsilon, \qquad T_{\mathrm{nsc}} := M_{3R}/\sqrt{\mu_f\varepsilon}.$$

Theorem 5.1 (Log-free smooth-convex minimax complexity)**.** *For every fixed multiplier budget $\bar{\Lambda} \ge 1$ and $T_{\mathrm{cvx}} \ge 50$,*

$$\mathfrak{C}^{\mathrm{cvx}}_\nabla = \Theta(T_{\mathrm{cvx}}), \qquad \mathfrak{C}^{\mathrm{cvx}}_{\mathrm{aff}} = \Theta(\Xi T_{\mathrm{cvx}}), \qquad \mathfrak{C}^{\mathrm{cvx}}_{\mathrm{comm}} = \Theta(\Xi\sqrt{\Omega}\,T_{\mathrm{cvx}}). \tag{10}$$

One explicit continuation method attains all three upper orders simultaneously. Its final feasibility repair costs only an additive $O_{\bar{\Lambda}}(\widehat{\chi}\log T_{\mathrm{cvx}})$ affine term, dominated by $O_{\bar{\Lambda}}(\widehat{\chi}T_{\mathrm{cvx}})$; hence there is no multiplicative target-accuracy logarithm.

Theorem 5.2 (Nonsmooth cross-coupled upper bounds)**.** *For the full-domain convex criterion of Appendix K, and in a normalized model that additionally provides exact Euclidean projections onto the phase primal and dual balls, Cross-Sliding satisfies*

$$N_{\partial f} = O(T_{\mathrm{ns}}^2), \quad N_{\mathrm{aff}} = O(\chi_\times T_{\mathrm{ns}}), \tag{11}$$

and after restarting in the μ_f-strongly convex case,

$$N_{\partial f} = O(T_{\mathrm{nsc}}^2), \quad N_{\mathrm{aff}} = O(\chi_\times T_{\mathrm{nsc}}). \tag{12}$$

The first-order counts have the standard centralized lower bounds; no matching $\chi_\times$-dependent nonsmooth affine/communication lower bound is claimed. A physical decentralized round bound additionally requires an implementation and accounting of the global ball projections and scalar reductions, and is not inferred solely from a per-affine-call graph factor. For any fixed $J \ge 2$, the normalized-oracle upper bounds replace $\chi_\times$ by the projector-sum factor $\chi_{\times,J}$; raw primitive and physical-round extensions require a regime-specific realization and are not obtained by analogy with the smooth-SC ledger.

6 EXPERIMENTS

The experiments are designed to test two different claims separately: (i) whether the theorem-level resource dependence on the cross geometry and the network appears with the predicted exponent, and (ii) whether the method remains stable on full, non-toy data and benefits from separate channel normalization under a common implementation. The complete protocol, the controlled scaling plots, primitive-versus-normalized ledgers, and all additional tables are in Appendix A.

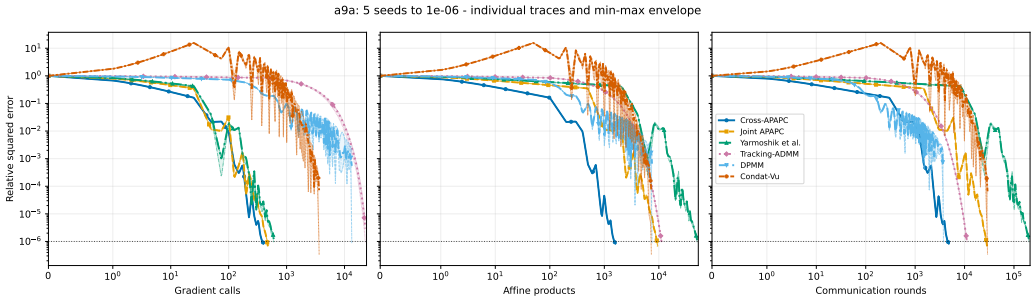

Figure 1: Full-data `a9a` on H100 over five feature-partition seeds. Thin traces are individual seeds, shaded regions and dotted boundaries are the empirical min–max interval, and the thick line is the median. Panels use gradient evaluations, native affine products, and physical neighbor rounds; the horizontal dotted line is the 10^{-6} target.

Full-data CC-MAC benchmark. We use vertically partitioned LIBSVM binary-regression instances (Chang and Lin, 2011). For $X = [X_1, \dots, X_n]$ we solve

$$\min_w \frac{1}{2N}\|Xw - y\|^2 + \frac{\lambda}{2}\|w\|^2, \qquad \lambda = 10^{-2}, \tag{13}$$

through local weights w_i and local copies $z_i \in \mathbb{R}^N$ of the shared prediction vector. The two affine channels are

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n X_i w_i - \sum_{i=1}^n \alpha_i z_i\right) = 0, \qquad z_i - \frac{1}{n}\sum_{j=1}^n z_j = 0, \tag{14}$$

with $\sum_i \alpha_i = 1$. On the consensus subspace, (14) is exactly $Xw - z = 0$ for every admissible α, so changing α changes the two-channel geometry without changing the underlying feasible prediction relation. This gives a representation-preserving controlled test of $\chi_\times$, rather than a comparison of different optimization problems.

All reported runs use the complete training split, deterministic PyTorch FP64 on an NVIDIA H100, column-partitioned features, the same zero original-primal initialization, and a high-accuracy centralized reference. For the scaling and robustness suites, a run is successful when

$$\mathrm{RelErr}(w^k) := \frac{\|w^k - w^\star\|^2}{\|w^0 - w^\star\|^2} \le 10^{-4}. \tag{15}$$

The method-comparison suites use the stricter threshold 10^{-6}.

Full-data robustness. All $48/48$ Cross-APAPC configurations on full `a9a`, `mushrooms`, `w8a`, and `rcv1` reach (15); this includes the 47 236-feature sparse `rcv1` task with both 8 and 16 agents. Appendix A gives the dimensions, topologies, seeds, and primitive-resource ledger.

Five-seed method comparison. We compare six matrix-free implementations on full `a9a` and `mushrooms` with $n = 8$, a ring network, feature-partition seeds $0, \dots, 4$, and the tighter target 10^{-6}. Cross-APAPC and joint APAPC share the same state representation, centralized reference, stopping rule, and normalized-channel interface. The Yarmoshik-type generic coupled reduction, Tracking-ADMM, DPMM, and Condat–Vũ (Condat, 2013; Vũ, 2013) curves are our CUDA ports/adaptations of the corresponding recurrences under this common benchmark interface; they are not presented as the authors' native software or native tuning. Figures 1–2 show every seed trajectory, their full empirical min–max envelope, and a median guide. Gradient evaluations and neighbor rounds are the common cross-method resources; native affine products are included as a diagnostic ledger but are not treated as equal-cost oracle units across different decompositions. All $60/60$ method–seed runs reach the target.

The cleanest controlled algorithmic comparison is Cross versus Joint. At the 10^{-6} threshold, the counts are identical across the five partitions for each method. On `a9a`, Cross uses 400 versus 475

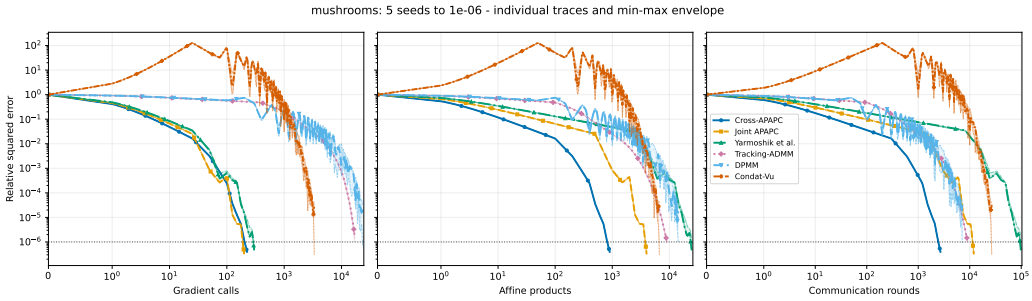

Figure 2: Full-data `mushrooms` under the same five-seed protocol and interval convention as Figure 1. Cross-APAPC and joint APAPC differ only in separate versus joint channel normalization; the remaining curves are common-interface CUDA ports/adaptations detailed in Appendix A.

gradients, 800 versus 4 750 normalized affine calls, and 4 800 versus 28 500 neighbor rounds; on `mushrooms`, the respective counts are 225 versus 200, 450 versus 2 000, and 2 700 versus 12 000. Thus separate normalization reduces the normalized affine and communication ledgers by factors 5.94 and 4.44, without implying uniform gradient or wall-clock dominance. Appendix A reports the raw primitive interpretation.

7 Discussion and Conclusions

The results separate three sources of difficulty that are conflated by a single condition number: objective curvature, network propagation, and cross-channel geometry. After within-channel conditioning is removed, the two affine classes contribute the intrinsic factor $\chi_\times$, while physical communication contributes $\sqrt{\kappa_W}$. Exact overlap is benign because common equations can be deduplicated; nearly coincident but nonidentical directions are hard because they leave a small positive singular value. The log-free smooth-convex result preserves this separation under its stated multiplier and spectral budgets.

This viewpoint is useful when the channel decomposition has operational meaning and cannot be replaced by arbitrary row mixing. Examples include vertical feature partitioning with a shared prediction variable, distributed estimation with local measurements coupled to a global state, and networked control or resource-allocation models combining nodewise physics with shared coordination constraints. In such problems separate normalization is implementable within the access policy, whereas a globally optimal joint preconditioner may require forbidden data aggregation or a different oracle. The Friedrichs factor then identifies when separate processing is sufficient and when the interface between otherwise well-conditioned channels is the actual bottleneck.

The experiments support this interpretation rather than a universal ranking of implementations. The controlled sweeps recover near-unit exponents in $\chi_\times$ and $\chi_\times\sqrt{\kappa_W}$; the full-data comparisons show a reproducible normalized-oracle advantage over joint normalization, including the additional 16-agent `a8a` stress test in Appendix A. Native affine products and wall time remain implementation dependent, and the two-seed appendix interval is descriptive rather than inferential.

The scope is therefore deliberate. The nonsmooth result assumes exact projection access and claims no matching cross-dependent lower bound. A normalized affine call is not a universal raw-matvec or wall-clock unit; Appendix L expands it into primitives. Important open questions are matching nonsmooth bounds, adaptive or cheaper normalization without prior spectral certificates, finite-sample tuning rules, and a sharp multi-channel geometry for $J \geq 3$. These directions would turn the present two-channel structural characterization into a broader design principle for constrained decentralized optimization.

AI USE STATEMENT

In this work, we used generative AI tools to assist with developing and stress-testing theoretical arguments, formulating and refining mathematical claims and hypotheses, checking and editing proofs, reviewing experimental methodology and parameter choices, implementing and debugging research code, constructing synthetic test instances, interpreting diagnostic and experimental results, language translation where needed, surveying and summarizing related literature, and drafting, restructuring, and editing parts of the manuscript. We also used generative AI tools to assist with figure/table preparation, software-code editing, and reference formatting. We did not use generative AI to replace, clean, or reformat the original benchmark datasets or to perform qualitative or thematic data analysis; transcription of research material was not applicable. All AI-assisted mathematical claims and proofs were independently checked by the authors against the stated assumptions, cited literature was verified against the original sources, and AI-assisted code was executed and tested by the authors. We have reviewed all AI-assisted work and take responsibility for the final content of this submission, including text, claims, code, and artifacts produced with the aid of generative AI.

REFERENCES

Alekh Agarwal, Alina Beygelzimer, Miroslav Dudik, John Langford, and Hanna Wallach. A reductions approach to fair classification. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of PMLR, pages 60–69, 2018.

Mahmoud Assran, Nicolas Loizou, Nicolas Ballas, and Mike Rabbat. Stochastic gradient push for distributed deep learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of PMLR, pages 344–353, 2019.

Heinz H. Bauschke, J. Y. Bello Cruz, Tran T. A. Nghia, Hung M. Phan, and Xianfu Wang. The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle. *Journal of Approximation Theory*, 185:63–79, 2014.

Åke Björck and Gene H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973. doi:10.1090/S0025-5718-1973-0348991-3.

Ekaterina Borodich, Alexander Gasnikov, and Dmitry Kovalev. On linear convergence in smooth convex–concave bilinearly-coupled saddle-point optimization: lower bounds and optimal algorithms. In *ICML*, PMLR 267:5045–5100, 2025.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), Article 27:1–27, 2011. doi:10.1145/1961189.1961199.

Laurent Condat. A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158:460–479, 2013.

Laurent Condat, Abdurakhmon Sadiev, and Peter Richtárik. A Nesterov-accelerated primal–dual splitting algorithm for convex nonsmooth optimization. arXiv:2604.09245v1, 2026.

Frank Deutsch. The angle between subspaces of a Hilbert space. In S. P. Singh, editor, *Approximation Theory, Wavelets and Applications*, NATO ASI Series C, volume 454, pages 107–130. Kluwer, 1995. doi:10.1007/978-94-015-8577-4_7.

Alessandro Falsone, Ivano Notarnicola, Giuseppe Notarstefano, and Maria Prandini. Tracking-ADMM for distributed constraint-coupled optimization. *Automatica*, 117:108962, 2020.

Kai Gong and Liwei Zhang. Decentralized proximal method of multipliers for convex optimization with coupled constraints. arXiv:2310.15596, 2023.

Eduard Gorbunov, Alexander Rogozin, Aleksandr Beznosikov, Darina Dvinskikh, and Alexander Gasnikov. Recent theoretical advances in decentralized distributed convex optimization. In *High-Dimensional Optimization and Probability: With a View Towards Data Science*, volume 191 of Springer Optim. Appl., pages 253–325. Springer, Cham, 2022.

Filip Hanzely and Peter Richtárik. Federated learning of a mixture of global and local models. arXiv:2002.05516, 2020.

Stephen Hardy, Wilko Henecka, Hamish Ivey-Law, Richard Nock, Giorgio Patrini, Guillaume Smith, and Brian Thorne. Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption. arXiv:1711.10677, 2017.

Sarvagya Jha. Two spectral gaps: Decentralized optimization over intersections of local convex sets. Optimization Online preprint, 27 August 2026.

Andrew V. Knyazev and Merico E. Argentati. Principal angles between subspaces in an A-based scalar product: Algorithms and perturbation estimates. *SIAM Journal on Scientific Computing*, 23(6):2008–2040, 2002. doi:10.1137/S1064827500377332.

Dmitry Kovalev, Adil Salim, and Peter Richtárik. Optimal and practical algorithms for smooth and strongly convex decentralized optimization. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Guanghui Lan and Yuyuan Ouyang. Accelerated gradient sliding for structured convex optimization. *Computational Optimization and Applications*, 82(2):361–394, 2022.

Guanghui Lan, Soomin Lee, and Yi Zhou. Communication-efficient algorithms for decentralized and stochastic optimization. *Mathematical Programming*, 180:237–284, 2020. doi:10.1007/s10107-018-1355-4.

Xiangru Lian, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang, and Ji Liu. Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B. Allen, Randy P. Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency. Think locally, act globally: Federated learning with local and global representations. arXiv:2001.01523, 2020.

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

Nhat Trung Nguyen and Alexander Gasnikov. Sliding methods for Hölder-smooth convex–concave minimax optimization with bilinear coupling. arXiv:2608.03846, 2026.

Ivano Notarnicola and Giuseppe Notarstefano. Constraint-coupled distributed optimization: a relaxation and duality approach. *IEEE Transactions on Control of Network Systems*, 7(1):483–492, 2020.

Yuyuan Ouyang and Yangyang Xu. Lower complexity bounds of first-order methods for convex–concave bilinear saddle-point problems. *Mathematical Programming*, 185(1–2):1–35, 2021.

Alexander Rogozin, Demyan Yarmoshik, Ksenia Kopylova, and Alexander Gasnikov. Decentralized strongly-convex optimization with affine constraints: primal and dual approaches. In *Advances in Optimization and Applications*, pages 93–105. Springer, 2022.

Adil Salim, Laurent Condat, Dmitry Kovalev, and Peter Richtárik. An optimal algorithm for strongly convex minimization under affine constraints. In *AISTATS*, PMLR 151:4482–4498, 2022.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal algorithms for smooth and strongly convex distributed optimization in networks. In *ICML*, PMLR 70:3027–3036, 2017.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal algorithms for non-smooth distributed optimization in networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Alexander Tyurin. Optimality in decentralized optimization under bandwidth constraints. arXiv:2603.20735, 2026.

Bàng Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013. doi:10.1007/s10444-011-9254-8.

Demyan Yarmoshik, Alexander Rogozin, and Alexander Gasnikov. Decentralized optimization with affine constraints over time-varying networks. *Computational Management Science*, 21:10, 2024.

Demyan Yarmoshik, Alexander Rogozin, Nikita Kiselev, Daniil Dorin, Alexander Gasnikov, and Dmitry Kovalev. Decentralized optimization with coupled constraints. In *ICLR*, 2025.

Demyan Yarmoshik, Nhat Trung Nguyen, Alexander Rogozin, and Alexander Gasnikov. Decentralized optimization with mixed affine constraints. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2026.

Bicheng Ying, Kun Yuan, Hanbin Hu, Yiming Chen, and Wotao Yin. BlueFog: Make decentralized algorithms practical for optimization and deep learning. arXiv:2111.04287, 2021.

Binhang Yuan, Yongjun He, Jared Davis, Tianyi Zhang, Tri Dao, Beidi Chen, Percy S. Liang, Christopher Ré, and Ce Zhang. Decentralized training of foundation models in heterogeneous environments. In *Advances in Neural Information Processing Systems*, volume 35, pages 25464–25477, 2022.

Zhe Zhang and Guanghui Lan. Solving convex smooth function constrained optimization is almost as easy as unconstrained optimization. arXiv:2210.05807, version 5, 2026.

NOTATION REFERENCE

This table retains all notation reused across sections; proof-, algorithm-, experiment-, and construction-local symbols are defined where they appear.

Table 2: Global and recurring notation.

Symbol	Meaning
Problem data and canonical reformulation	
n, i	Number of agents and agent index, $i \in \{1, \dots, n\}$.
x_i, z, z_i	Local variable, shared variable, and local copy of the shared variable; $x_i \in \mathbb{R}^{d_i}$ and $z, z_i \in \mathbb{R}^q$.
d_i, q, m, p_i, r_i	Local/shared dimensions, coupled-equality dimension, and row counts of C_i, D_i.
$A_i, B_i, b_i; C_i, c_i; D_i, e_i$	Coupled-equality, local-affine, and shared-variable affine data in (CC-MAC).
$\boldsymbol{x}, \boldsymbol{z}; A, B, C, D; \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{e}$	Stacked variables, block-diagonal matrix stacks, and stacked right-hand sides.
$u; F(u)$	Canonical stacked primal variable $u = \mathrm{col}(\boldsymbol{x}, \boldsymbol{z})$ and lifted objective $F(u) = \sum_i f_i(x_i, z_i)$.
$w; \Phi(w)$	Execution-space variable and objective after the stated canonical or decentralized lift.
$E_m, \mathcal{W}_m, \mathcal{W}_q$	Normalized global-sum map and graph channels: $E_m = n^{-1/2}(\mathbf{1}^\top \otimes I_m)$, $\mathcal{W}_m = W \otimes I_m$, $\mathcal{W}_q = W \otimes I_q$.
$K_1, v_1; K_2, v_2$	Two canonical affine channels and right-hand sides defined in (4).
K, v	Joint canonical stack $K = \mathrm{col}(K_1, K_2)$ and $v = \mathrm{col}(v_1, v_2)$.
$\mathcal{L}_j$	Executable local channel j in the decentralized lift.
Network and linear-algebra quantities	
$\mathcal{G} = (\mathcal{V}, \mathcal{E}); W; \kappa_W$	Communication graph, symmetric PSD gossip matrix with $\mathrm{Ker}\, W = \mathrm{span}\{\mathbf{1}\}$, and $\kappa_W = \lambda_{\max}(W)/\lambda^+_{\min}(W)$; for $n = 1$, $\kappa_W = 1$ and the round count is zero.
$\mathrm{Range}(M), \mathrm{Ker}(M); P_{\mathcal{S}}$	Range and kernel of a linear map; Euclidean orthogonal projector onto a subspace $\mathcal{S}$.
$\sigma_{\max}(M), \sigma^+_{\min}(M)$	Largest and smallest positive singular values.
$\kappa(M), \kappa_+(G)$	Squared positive singular-value ratio and positive-eigenvalue ratio; $\kappa(M) = \kappa_+(M^\top M)$.
Cross-channel geometry and normalization	
$J; M_j, U_j$	Number of channel classes; channel maps and row spaces $U_j = \mathrm{Range}(M_j^\top)$.
$G_J, \chi_{\times,J}$	Projector sum $G_J = \sum_j P_{U_j}$ and multi-channel spectral factor $\chi_{\times,J} = \sqrt{\kappa_+(G_J)}$.
$U, V; I$	Two-channel row spaces and their exact intersection $I = U \cap V$.
$c_F(U, V), \theta_F$	Friedrichs cosine after removing I and Friedrichs angle $\theta_F = \arccos c_F$.
$\chi_\times$	Canonical factor $\sqrt{(1 + c_F)/(1 - c_F)} = \cot(\theta_F/2)$ for $U = \mathrm{Range}(K_1^\top)$ and $V = \mathrm{Range}(K_2^\top)$.
$\rho_{\mathrm{loc}}, \chi_{\mathrm{loc}}$	Friedrichs cosine and cross factor of the executable local pair.
$S_j; \widehat{K}_j, \widehat{v}_j$	Separate left normalizer and normalized channel/right-hand-side pair.
$\widehat{K}, \widehat{v}; \bar{K}, \bar{v}$	Generic normalized joint stack; $\bar{K}, \bar{v}$ denote the fixed normalized pair in the convex and nonsmooth appendices.
δ_{wh}	Fixed relative whitening accuracy, chosen independently of the target accuracy ε.
Ξ	Supplied upper certificate on the relevant cross factor, in particular $\Xi \geq \chi_\times$ when stated.
Optimization, accuracy, and resources	
L_f, μ_f, κ_f	Objective smoothness, strong convexity, and condition number $\kappa_f = L_f/\mu_f$.
$w^0, w^\star; R_0, R_E, R$	Initialization/solution, primal radius, supplied smooth-SC initial primal–dual energy bound, and convex/nonsmooth radius budget.
$\varepsilon; \log_+(t)$	Target accuracy and the truncated logarithm used in the smooth-SC statements.
$N_{\nabla f}, N_{\partial f}$	Numbers of gradient and subgradient evaluations.
$N_{\mathrm{aff}}, N_{\mathrm{comm}}$	Numbers of normalized affine-oracle calls and physical neighbor communication rounds.
N_A, N_B, N_C, N_D	Raw products with the primitive affine blocks in a fixed executable realization.
s_j^{comm}	Communication cost of one normalized application of channel j in the stated realization.
$T_{\mathrm{sc}}, T_{\mathrm{sc,E}}, T_{\mathrm{cvx}}, T_{\mathrm{ns}}, T_{\mathrm{nsc}}$	Standard first-order scales used in the smooth-SC, smooth-convex, nonsmooth-convex, and nonsmooth-SC statements.
Λ_{cvx}	Scale-invariant smooth-convex multiplier parameter $\sigma^+_{\min}(\bar{K})\|\lambda^\star\|/(L_f R)$.
M_{3R}	Local nonsmooth regularity bound on the Euclidean ball $B(w^\star, 3R)$.
$\bar{\kappa}, \Xi, \Omega$	Class budgets on κ_f, the cross-factor certificate, and κ_W, respectively.

Appendix roadmap. For reviewer accessibility, Appendix A first gives the complete experimental protocol and the theorem-aligned scaling validation. The remaining appendices follow the proof dependency graph: canonical reduction and the decentralized lift; projector/Friedrichs geometry; optimal separate preconditioning; generic normalization and the certified CC-MAC realization; smooth strongly convex upper and lower bounds; smooth-convex and nonsmooth extensions; and primitive resource accounting.

A Experimental Protocol, Scaling Validation, and Full H100 Results

Hardware, arithmetic, and stopping rule. All reported experiments were executed with PyTorch in FP64 on an NVIDIA H100 80GB HBM3 (CUDA 12.8, PyTorch 2.11.0+cu128), with deterministic execution enabled and TF32 disabled. Sparse feature blocks are stored as CUDA CSR tensors unless the density threshold selects a dense representation. All methods start from the same zero original-primal point and are evaluated against a high-accuracy centralized reference solution. We report the relative squared error in (15); the scaling and robustness suites stop at 10^{-4}, while both method-comparison suites use 10^{-6}. The artifact records every run as a JSON summary plus a CSV trace.

Full-data VFL formulation. The benchmark is (13) with the two CC-MAC channels (14). For coefficients $\alpha \in \mathbb{R}^n$ satisfying $\sum_i \alpha_i = 1$, consensus implies the same relation $Xw - z = 0$ independently of α. Hence the controlled geometry experiment varies the representation without changing the original feasible prediction relation or objective. If $\chi_{\mathrm{targ}} \geq 1$ and $c = (\chi_{\mathrm{targ}}^2 - 1)/(\chi_{\mathrm{targ}}^2 + 1)$, we choose a unit zero-sum vector v and

$$\alpha = \frac{1}{n}\mathbf{1} + \frac{c}{\sqrt{n(1-c^2)}}v,$$

which realizes the requested Friedrichs factor up to floating-point roundoff.

Resource accounting. A *normalized affine call* is one application of the separately normalized joint stack or its adjoint, exactly as in the theorem-level oracle. The experiment counter also expands this call into *raw affine products*: the polynomial K_1 normalizer contributes its internal $K_1/K_1^\top$ products, while K_2 actions charge physical neighbor rounds according to the measured graph condition. These are different ledgers. The main-text Cross-versus-Joint statement is made in the common normalized-channel oracle; raw primitive counts are reported separately below. No wall-clock superiority claim is made.

A.1 Controlled geometry and network scaling

The geometry sweep uses the complete `a9a` and `w8a` training sets, $n = 8$ agents, data seed 307, feature-partition seed 0, normalizer degree 28, and

$$\chi_{\mathrm{targ}} \in \{1, 1.25, 1.5, 2, 3, 5, 8\}.$$

We combine these values with path, ring, grid, Erdős–Rényi, and random-regular networks. Deterministic topologies use one graph realization; the two random graph families use five graph seeds, giving 182 runs in total. Every run reaches the target error. The log–log fits in Figure 3 use the nonplateau regime $\chi_\times \geq 1.5$. Resampling graph realizations as blocks gives

$$\alpha_{\mathrm{aff}} = 0.988\ [0.966, 1.009], \qquad \alpha_{\mathrm{comm}} = 0.980\ [0.959, 1.003], \tag{16}$$

for $N_{\mathrm{aff}} \propto \chi_\times^{\alpha_{\mathrm{aff}}}$ and $N_{\mathrm{comm}} \propto (\chi_\times \sqrt{\kappa_W})^{\alpha_{\mathrm{comm}}}$, respectively. A two-factor fit gives exponent 0.988 in $\chi_\times$ and 0.957 in $\sqrt{\kappa_W}$; the graph-block bootstrap interval for the latter is $[0.886, 1.039]$.

A.2 Full-data robustness

We use feature-norm weights $\alpha_i \propto \|X_i\|_F$, polynomial separate normalization of degree 24, and full training sets for `a9a`, `mushrooms`, `w8a`, and `rcv1`. We run $n \in \{8, 16\}$, two feature-partition seeds, and ring/random-regular networks (with two random-regular graph seeds), for 48 total configurations. All 48 runs converge. Table 3 reports medians and interquartile ranges over the available topology/seed realizations for each dataset/agent count; "raw aff." expands the degree-24 channel normalizer.

The high-dimensional `rcv1` instance has 20 242 samples and 47 236 features; it reaches the target for both 8 and 16 agents.

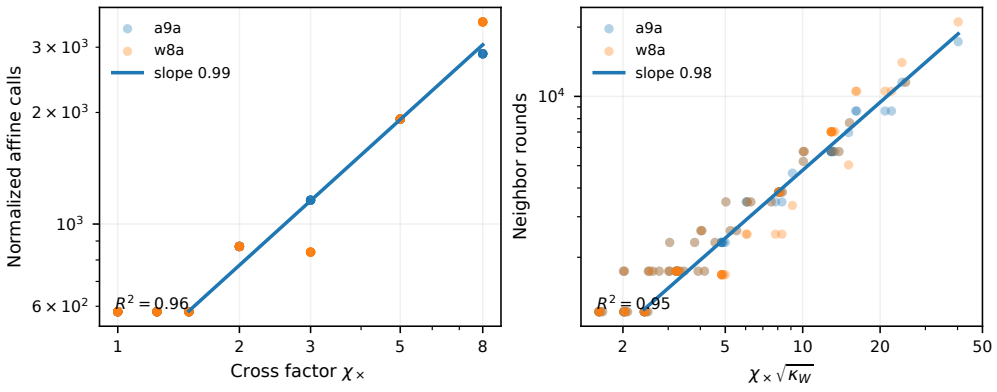

Figure 3: Controlled full-data scaling on `a9a` and `w8a`. Left: normalized affine calls versus the realized cross factor $\chi_\times$. Right: physical neighbor rounds versus $\chi_\times\sqrt{\kappa_W}$. Fits use the nonplateau regime $\chi_\times \geq 1.5$; both slopes are statistically consistent with the unit exponents predicted by the theory.

Table 3: Full-data Cross-APAPC robustness on H100. Brackets denote $[Q_1, Q_3]$; the final column is the worst terminal relative squared error among the corresponding runs.

Dataset	N	d	agents	grad.	norm. aff.	raw aff.	comm.	max err.
`a9a`	32,561	123	8	145	580	29 000	1 160 [1 160, 1 595]	3.9×10^{-5}
`a9a`	32,561	123	16	145	580	29 000	1 740 [1 740, 3 045]	3.9×10^{-5}
`mushrooms`	8,124	112	8	110	440	22 000	880 [880, 1 210]	8.1×10^{-5}
`mushrooms`	8,124	112	16	145	580	29 000	1 740 [1 740, 3 045]	5.6×10^{-5}
`w8a`	49,749	300	8	145	580	29 000	1 160 [1 160, 1 595]	8.3×10^{-5}
`w8a`	49,749	300	16	220	880	44 000	2 640 [2 640, 4 620]	7.9×10^{-5}
`rcv1`	20,242	47,236	8	150	600	30 000	1 200 [1 200, 1 650]	9.7×10^{-5}
`rcv1`	20,242	47,236	16	170	680	34 000	2 040 [2 040, 3 570]	7.4×10^{-5}

A.3 Five-seed full-data method comparison

For the broader comparison we use full `a9a` and `mushrooms`, $n = 8$, a ring network, feature-partition seeds $0, \ldots, 4$, and tolerance 10^{-6}. Cross-APAPC and joint APAPC share the same matrix-free state, reference solution, and normalized-oracle convention. The Yarmoshik-type generic coupled reduction, Tracking-ADMM, DPMM, and Condat–Vũ curves in Figures 1–2 are our CUDA matrix-free ports/adaptations of the corresponding recurrences under the common benchmark interface; they are not claimed to be the authors' native software or native tuning. All 60 method–seed combinations converge; the largest terminal relative squared error is $9.934 \cdot 10^{-7}$.

The figures carry the seed-level information directly: thin curves show all five partitions and the envelope is the full empirical min–max interval. The middle panels report each implementation's native affine-action ledger for reproducibility; unlike gradients and physical neighbor rounds, it is not interpreted as a directly comparable oracle unit across methods.

Normalized versus raw Cross/Joint accounting. Because Cross-APAPC and joint APAPC are two realizations of the same APAPC template, their normalized-call comparison is clean. At tolerance 10^{-6} Cross reduces normalized affine calls and neighbor rounds by factors 5.94 on `a9a` and 4.44 on `mushrooms`. The separate normalizer nevertheless expands each normalized K_1 action into raw primitives, so the native affine-action ledger is reported independently in the middle panels rather than identified with the theorem-level oracle. The figures retain all five traces and their full empirical intervals instead of replacing them by a second median table; no raw-matvec or wall-clock dominance is claimed.

A.4 Additional 16-agent a8a comparison

As a larger-agent stress test, we repeat the six-method comparison on the complete a8a training split (22 696 samples and 123 features) with $n = 16$, a ring network, feature-partition seeds 0 and 1, and tolerance 10^{-6}. All $12/12$ method–seed runs converge. The worst terminal relative squared error is $9.971 \cdot 10^{-7}$. Cross-APAPC reaches the target in 375 gradient evaluations for both partitions, joint APAPC in 400, and the Yarmoshik-type reduction in 650; the remaining methods require between 12 100 and 33 900 outer iterations depending on the seed and recurrence. These iteration numbers are descriptive rather than equal-cost comparisons, so Figure 4 retains the common gradient and neighbor-round ledgers and reports native affine actions separately.

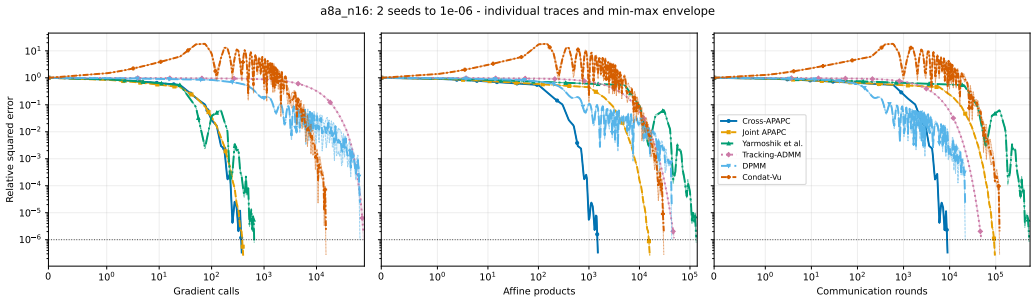

Figure 4: Full-data a8a with 16 agents on H100. Thin curves are the two feature-partition seeds; the shaded band and dotted boundaries show their empirical min–max range, and the thick curve is the median guide. All six methods reach the 10^{-6} target. With only two seeds, the band is a descriptive seed interval, not a confidence interval.

Reproducibility artifact. The accompanying artifact contains the H100 YAML configurations, dataset/partition and graph generators, CUDA sparse/dense operators, explicit counters for gradients/normalized affine calls/raw affine products/neighbor rounds, per-run JSON and CSV traces, and the aggregation scripts used to generate Figures 1–4, Figure 3, and Table 3. The direct smooth-convex ε-sweep and normalization ablation configurations are included in the artifact but are not used for any empirical claim reported here.

B Canonical Reduction and Decentralized Lift

Lemma B.1 (Network range identities)**.** *For every $s \geq 1$, $\mathrm{Ker}(W \otimes I_s) = \mathrm{span}\{\mathbf{1}\} \otimes \mathbb{R}^s$ and $\mathrm{Range}(W \otimes I_s) = \{\mathrm{col}(u_1, \ldots, u_n) : \sum_i u_i = 0\}$.*

For the executable lift introduce $y \in \mathcal{Y}_\perp := \mathrm{Range}(W \otimes I_m)$ and impose

$$A\boldsymbol{x} + B\boldsymbol{z} + \alpha \mathcal{W}_m y = \boldsymbol{b}, \quad \alpha > 0, \tag{17}$$

with optional consensus-row scale $\gamma > 0$, giving

$$K^{\mathrm{loc}}_{\alpha,\gamma} = \begin{bmatrix} A & B & \alpha\mathcal{W}_m \\ C & 0 & 0 \\ 0 & D & 0 \\ 0 & \gamma\mathcal{W}_q & 0 \end{bmatrix}. \tag{18}$$

Proposition B.2 (Lift-scaling invariance)**.** *The canonical pair $\mathrm{Range}(K_1^\top), \mathrm{Range}(K_2^\top)$ is independent of α, γ and of the positive spectrum/edge weights of any connected symmetric gossip matrix with the same consensus kernel.*

Proposition B.3 (Canonical-to-local decentralized bridge)**.** *Assume $n \geq 2$. Let $\widetilde{W} = P_W(W)$, where $P_W(0) = 0$, and suppose*

$$0 < \underline{\sigma}_W \leq \sigma^+_{\min}(\widetilde{W}) \leq \sigma_{\max}(\widetilde{W}) \leq \overline{\sigma}_W, \qquad \overline{\sigma}_W / \underline{\sigma}_W = O(1).$$

Put $H = [A\ B]$. For $H \neq 0$, choose a certified $\bar{h} \geq \|H\|^2$ and set $\alpha = \sqrt{3\bar{h}}/\underline{\sigma}_W$, with $\gamma > 0$. Writing $\widetilde{W}_s := \widetilde{W} \otimes I_s$, define

$$L_1 = \begin{bmatrix} A & B & \alpha\widetilde{W}_m \\ C & 0 & 0 \end{bmatrix}, \qquad L_2 = \begin{bmatrix} 0 & D & 0 \\ 0 & \gamma\widetilde{W}_q & 0 \end{bmatrix}. \tag{19}$$

Let $U_{\mathrm{loc}} := \mathrm{Range}(L_1^\top)$, $V_{\mathrm{loc}} := \mathrm{Range}(L_2^\top)$, $\rho_{\mathrm{loc}} := c_F(U_{\mathrm{loc}}, V_{\mathrm{loc}})$, and $\chi_{\mathrm{loc}} := \sqrt{(1+\rho_{\mathrm{loc}})/(1-\rho_{\mathrm{loc}})}$. On the lifted space with $y \in \mathcal{Y}_\perp$,

$$\rho_{\mathrm{loc}}^2 \leq \tfrac{1}{4} + \tfrac{3}{4}c_F^2, \qquad \chi_{\mathrm{loc}} \leq \sqrt{3}\,\chi_\times. \tag{20}$$

The feasible-set penalty below has $\kappa_{\Phi_{\mathrm{loc}}} = O(\kappa_f)$. Each literal $L_j/L_j^\top$ call uses local products and one $\widetilde{W}$ action, costing $\deg P_W$ neighbor rounds. The certified graph polynomial is specified in Appendix G. If $H = 0$, remove the vacuous coupled equation and auxiliary variable, take $\Phi_{\mathrm{loc}} = F$, and retain only the local C channel and the shared/consensus channel; the two row spaces are orthogonal.

Corollary B.4 (Recovery of the previous mixed model)**.** *If $B_i = 0$ for all i, then* $\mathrm{Range}(K_1^\top) \perp \mathrm{Range}(K_2^\top)$ *and* $\chi_\times = 1$.

B.1 Proof of Lemma B.1

Since $W = W^\top \succeq 0$ is connected, $\operatorname{Ker} W = \mathrm{span}\{\mathbf{1}\}$. The Kronecker identity gives $\mathrm{Ker}(W \otimes I_s) = \mathrm{span}\{\mathbf{1}\} \otimes \mathbb{R}^s$. Symmetry implies $\mathrm{Range}(W \otimes I_s) = \mathrm{Ker}(W \otimes I_s)^\perp$, which is exactly the set of block vectors with zero block sum. □

B.2 Proof of Theorem 2.1

Let $r := A\boldsymbol{x} + B\boldsymbol{z} - \boldsymbol{b}$. The original coupled equality is $E_m r = 0$. Since $\operatorname{Ker} E_m = \{r : \sum_i r_i = 0\} = \mathrm{Range}(\mathcal{W}_m)$ by Lemma B.1,

$$E_m r = 0 \quad \Longleftrightarrow \quad r \in \mathrm{Range}(\mathcal{W}_m) \quad \Longleftrightarrow \quad \exists y:\ r + \alpha\mathcal{W}_m y = 0$$

for every $\alpha > 0$. Likewise $\mathcal{W}_q \boldsymbol{z} = 0$ iff $z_1 = \cdots = z_n$, and multiplying this row by $\gamma > 0$ changes neither its nullspace nor feasibility. The C and D rows are unchanged. Therefore the original CC-MAC system, the canonical reduced system $K_1 u = v_1, K_2 u = v_2$, and every local lift (18) have the same feasible original variables and the same objective value. □

B.3 Proof of Proposition B.2

The canonical matrices in (4) contain neither α nor γ, so their row spaces are manifestly invariant to these auxiliary scalings. If W' is any other symmetric positive-semidefinite gossip matrix on a connected graph with $\operatorname{Ker} W' = \mathrm{span}\{\mathbf{1}\}$, then

$$\mathrm{Range}((W' \otimes I_q)^\top) = \mathrm{Range}(W' \otimes I_q) = (\mathrm{span}\{\mathbf{1}\} \otimes \mathbb{R}^q)^\perp = \mathrm{Range}(\mathcal{W}_q).$$

Thus replacing the consensus realization changes its positive spectrum and hence its implementation cost, but not the row subspace entering the Friedrichs angle. The E_m block is the canonical normalized global-sum map and is independent of the graph altogether. Hence c_F and $\chi_\times$ depend only on the CC-MAC affine data and the fixed Euclidean geometry of $(\boldsymbol{x}, \boldsymbol{z})$. □

B.4 Proof of Proposition B.3

If $H = 0$, then $A = B = 0$ and feasibility makes the original coupled equation vacuous. Removing its auxiliary variable leaves F, local C constraints on $\boldsymbol{x}$, and shared/consensus constraints on $\boldsymbol{z}$. Their row spaces are orthogonal, so $\chi_{\mathrm{loc}} = \chi_\times = 1$. Hence assume $H \neq 0$.

Let $\mathcal{H}_0$ be the original-coordinate subspace and $\mathcal{Y}$ the auxiliary-coordinate subspace. Set $U_\alpha = \mathrm{Range}(L_1^\top)$ and $V = \mathrm{Range}(L_2^\top)$. Then $V \subset \mathcal{H}_0$. A row-space vector of L_1 is

$$(A^\top a + C^\top c,\ B^\top a,\ \alpha\widetilde{W}_m a).$$

Decompose $a = a_0 + a_\perp$ into consensus and disagreement. Its auxiliary component is zero precisely when $a_\perp = 0$, and the vectors generated by (a_0, c) form the embedded canonical row space $U_0 = \mathrm{Range}(K_1^\top)$. Thus

$$U_\alpha \cap \mathcal{H}_0 = U_0. \tag{21}$$

Write $U_\alpha = U_0 \oplus U_e$ orthogonally. Every $e \in U_e$ is the row $(H^\top a_\perp, \alpha\widetilde{W}_m a_\perp)$ minus its projection onto U_0. Therefore

$$\|P_{\mathcal{H}_0} e\| \le \|H\|\|a_\perp\|, \qquad \|P_{\mathcal{Y}} e\| \ge \alpha\underline{\sigma}_W \|a_\perp\|.$$

Consequently

$$\|P_{\mathcal{H}_0} e\| \le \delta\|e\|, \qquad \delta := \frac{\|H\|}{\sqrt{\|H\|^2 + \alpha^2\underline{\sigma}_W^2}} \le \frac{1}{2}. \tag{22}$$

The intersections $U_\alpha \cap V$ and $U_0 \cap V$ agree. For a unit $v \in V$ orthogonal to this intersection, put $t = \|P_{U_0} v\|^2 \le c_F^2$. Since $P_{\mathcal{H}_0} U_e \subset U_0^\perp$,

$$\|P_{U_e} v\| \le \delta\sqrt{1-t}, \qquad \|P_{U_\alpha} v\|^2 \le t + \delta^2(1-t) \le \tfrac{1}{4} + \tfrac{3}{4}c_F^2.$$

This proves the first bound in (20). For $s \in [0, 1)$,

$$(1+2s)^2 - \frac{(2+s)^2}{4}(1+3s^2) = \frac{3}{4}s(1-s^2)(s+4) \ge 0.$$

Taking $s = c_F$ gives $\rho_{\mathrm{loc}}(2 + c_F) \le 1 + 2c_F$, which rearranges to $\chi_{\mathrm{loc}}^2 \le 3\chi_\times^2$.

Define

$$\Phi_{\mathrm{loc}}(\xi) := F(\boldsymbol{x}, \boldsymbol{z}) + \frac{r}{2}\|A\boldsymbol{x} + B\boldsymbol{z} + \alpha\widetilde{W}_m y - \boldsymbol{b}\|^2, \qquad \xi = \mathrm{col}(\boldsymbol{x}, \boldsymbol{z}, y), \quad r := \frac{\mu_f}{2\bar{h}}. \tag{23}$$

The penalty vanishes on the localized equality. For an increment (d_u, d_y) with $d_y \in \mathcal{Y}_\perp$, its Bregman divergence $D_{\Phi_{\mathrm{loc}}}(\xi + (d_u, d_y), \xi)$ satisfies

$$D_{\Phi_{\mathrm{loc}}}(\xi + (d_u, d_y), \xi) \ge \frac{\mu_f}{2}\|d_u\|^2 + \frac{r}{2}\|Hd_u + \alpha\widetilde{W}_m d_y\|^2 \ge \frac{\mu_f}{4}\|d_u\|^2 + \frac{3\mu_f}{8}\|d_y\|^2.$$

Here we used $\|a+b\|^2 \ge \frac{1}{2}\|b\|^2 - \|a\|^2$ and $\|H\|^2 \le \bar{h}$. Thus $\mu_{\Phi_{\mathrm{loc}}} \ge \mu_f/2$, whereas

$$L_{\Phi_{\mathrm{loc}}} \le L_f + r(\|H\|^2 + \alpha^2\overline{\sigma}_W^2) \le L_f + \frac{\mu_f}{2}\left(1 + 3\frac{\overline{\sigma}_W^2}{\underline{\sigma}_W^2}\right).$$

Hence $\kappa_{\Phi_{\mathrm{loc}}} = O(\kappa_f)$. Initialization $y^0 = 0$ and the stated channel and gradient actions preserve $\mathcal{Y}_\perp$. A polynomial of degree $\deg P_W$ is evaluated with that many raw W products. The map E_m defines the canonical geometry and is not evaluated by this lift. □

B.5 Proof of Corollary B.4

When $B = 0$, every vector in $\mathrm{Range}(K_1^\top)$ has support only in the $\boldsymbol{x}$ coordinates, whereas every vector in $\mathrm{Range}(K_2^\top)$ has support only in the $\boldsymbol{z}$ coordinates. The spaces are orthogonal, so $c_F = 0$ and $\chi_\times = 1$. □

C Projector Geometry, Multi-Channel Reduction, and Affine Deduplication

C.1 Two weighted projectors

Lemma C.1 (Spectrum of two weighted projectors)**.** *Let $a, b > 0$. On a principal-angle plane of U, V with angle $\theta \in (0, \pi/2]$, the eigenvalues of $aP_U + bP_V$ are*

$$\lambda_\pm(\theta; a, b) = \frac{a + b \pm \sqrt{(a-b)^2 + 4ab\cos^2\theta}}{2}. \tag{24}$$

The common component $U \cap V$ has eigenvalue $a + b$; exclusive directions have eigenvalue a or b. Among principal planes the smallest eigenvalue occurs at the Friedrichs cosine.

C.2 Proof of Lemma C.1

Use the canonical principal-angle decomposition of $U + V$. Exact intersections, directions belonging to only one subspace, and two-dimensional principal-angle planes are mutually orthogonal. On a nontrivial plane choose an orthonormal basis (u, w) such that $u \in U$ and a unit vector spanning the corresponding direction of V is

$$v = \cos\theta\, u + \sin\theta\, w, \qquad \theta \in (0, \pi/2].$$

Then

$$P_U = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad P_V = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}.$$

Therefore

$$aP_U + bP_V = \begin{bmatrix} a + b\cos^2\theta & b\cos\theta\sin\theta \\ b\cos\theta\sin\theta & b\sin^2\theta \end{bmatrix}.$$

Its trace is $a + b$ and its determinant is $ab\sin^2\theta$. The quadratic formula gives

$$\lambda_\pm = \frac{a + b \pm \sqrt{(a+b)^2 - 4ab\sin^2\theta}}{2} = \frac{a + b \pm \sqrt{(a-b)^2 + 4ab\cos^2\theta}}{2},$$

which is (24). The derivative of λ_- with respect to $|\cos\theta|$ is nonpositive, so among nontrivial principal-angle planes its smallest value occurs at the largest nonredundant cosine, i.e. the Friedrichs cosine. One-dimensional exclusive components have eigenvalue a or b and therefore cannot create a smaller value than the formula with $\cos\theta = 0$. □

C.3 Multi-channel reduction

For J separately normalized channel classes with row spaces $U_1, \ldots, U_J$, define

$$G_J := \sum_{j=1}^{J} P_{U_j}, \qquad \chi_{\times,J} := \sqrt{\kappa_+(G_J)}. \tag{25}$$

If all channels are vacuous we set $\chi_{\times,J} = 1$ and charge no affine calls. Pairwise angles need not determine G_J when $J \geq 3$.

Proposition C.2 (Multi-channel spectral reduction)**.** *Suppose separate normalizations satisfy, for universal $0 < a \leq b$,*

$$aP_{U_j} \preceq \widehat{M}_j^\top \widehat{M}_j \preceq bP_{U_j}, \qquad j = 1, \ldots, J,$$

and let $\widehat{M} = \mathrm{col}(\widehat{M}_1, \ldots, \widehat{M}_J)$. Then

$$aG_J \preceq \widehat{M}^\top \widehat{M} \preceq bG_J, \qquad \kappa(\widehat{M}) \leq \frac{b}{a}\,\chi_{\times,J}^2. \tag{26}$$

Exact separate whitening gives equality $\kappa(\widehat{M}) = \chi_{\times,J}^2$. For $J = 1$, $\chi_{\times,1} = 1$. For $J = 2$, if $U = U_1$, $V = U_2$ and $\chi_\times$ is the Friedrichs factor defined in Section 3, then

$$\chi_\times \leq \chi_{\times,2} \leq \sqrt{2}\,\chi_\times, \tag{27}$$

with $\chi_{\times,2} = \chi_\times$ when $U \cap V = \{0\}$.

Proof. Summing the J spectral-equivalence inequalities gives the displayed PSD sandwich; exact whitening has $\widehat{M}_j^\top \widehat{M}_j = P_{U_j}$. The $J = 2$ comparison follows from Lemma C.1: exact common directions contribute only eigenvalue 2, while every nontrivial principal-angle block contributes $1 \pm \cos\theta$. Hence $\kappa_+(P_U + P_V)$ lies between $\chi_\times^2$ and $2\chi_\times^2$, with equality to $\chi_\times^2$ in the irredundant case. □

Lemma C.3 (Pairwise angles do not determine multi-channel conditioning)**.** *For every $c \in (0, 1/2)$ there are two triples of one-dimensional subspaces with identical pairwise Friedrichs cosines, all equal to c, but different spectra of $G_3 = \sum_{j=1}^{3} P_{U_j}$.*

Proof. Choose unit generators whose Gram matrices are

$$H_+ = \begin{bmatrix} 1 & c & c \\ c & 1 & c \\ c & c & 1 \end{bmatrix}, \qquad H_- = \begin{bmatrix} 1 & c & c \\ c & 1 & -c \\ c & -c & 1 \end{bmatrix}.$$

Both are positive definite for $0 < c < 1/2$, and all off-diagonal absolute inner products equal c. Their eigenvalues are respectively $\{1-c, 1-c, 1+2c\}$ and $\{1-2c, 1+c, 1+c\}$. The nonzero eigenvalues of $\sum_j P_{U_j}$ equal those of the generator Gram matrix, so the two projector sums have different condition numbers despite identical pairwise angles. □

C.4 Exact common affine directions

Lemma C.4 (Affine deduplication of exact common row directions)**.** *Let $M_j \in \mathbb{R}^{p_j \times d}$ and suppose*

$$\mathcal{F} := \{x : M_1 x = v_1,\ M_2 x = v_2\} \neq \varnothing.$$

Set

$$U = \mathrm{Range}(M_1^\top), \qquad V = \mathrm{Range}(M_2^\top), \qquad I = U \cap V,$$

and choose orthogonal decompositions

$$U = I \oplus U_0, \qquad V = I \oplus V_0$$

with orthonormal basis matrices Q_I, Q_1, Q_2 for I, U_0, V_0, respectively. For any $x_\star \in \mathcal{F}$, put

$$c_I = Q_I^\top x_\star, \qquad c_1 = Q_1^\top x_\star, \qquad c_2 = Q_2^\top x_\star.$$

Then

$$\mathcal{F} = \{x : Q_I^\top x = c_I,\ Q_1^\top x = c_1,\ Q_2^\top x = c_2\}.$$

In particular, there is a left map R_2 with $R_2 M_2 = Q_2^\top$ and $R_2 v_2 = c_2$ such that

$$\mathcal{F} = \{x : M_1 x = v_1,\ R_2 M_2 x = R_2 v_2\}.$$

Thus an exact common row direction need be retained in only one affine channel.

Proof. Since $x_\star$ is feasible,

$$M_1 x = v_1 \iff M_1(x - x_\star) = 0 \iff x - x_\star \in \mathrm{Ker}(M_1) = U^\perp.$$

Because $U = I \oplus U_0$, this is equivalent to $Q_I^\top x = c_I$ and $Q_1^\top x = c_1$. Likewise, $M_2 x = v_2$ is equivalent to $Q_I^\top x = c_I$ and $Q_2^\top x = c_2$. The same common right-hand side appears in both descriptions precisely because the joint system is consistent.

Moreover, $\mathrm{Range}(Q_2) \subseteq \mathrm{Range}(M_2^\top)$, so $Q_2^\top = R_2 M_2$ for some R_2. Since $v_2 = M_2 x_\star$, $R_2 v_2 = Q_2^\top x_\star = c_2$. The first channel still imposes the common equations, hence replacing the second by its V_0 component leaves $\mathcal{F}$ unchanged. □

The lemma is an algebraic equivalence statement; it does not assert a decentralized or zero-cost procedure for constructing R_2. A merely near-common direction cannot be removed by this argument: if its principal angle is $\eta > 0$, it belongs to neither exact intersection and its contribution to the cross factor diverges as $\eta \downarrow 0$.

D Friedrichs Geometry and Cross-Angle Conditioning

D.1 Representation-equivalent two-agent example

Take $f_i(x_i, z) = \frac{1}{2}(x_i^2 + z^2)$, no local equations, and $x_1 + x_2 + 2z = 1$. With copies z_1, z_2, every $t \in \mathbb{R}$ gives the equivalent constraints

$$x_1 + x_2 + (1+t)z_1 + (1-t)z_2 = 1, \qquad z_1 - z_2 = 0.$$

The term $t(z_1 - z_2)$ vanishes at consensus, so the original feasible set and objective are unchanged. Using one normalized consensus row, the two channels are

$$M_1(t) = \tfrac{1}{\sqrt{2}}[1\ \ 1\ \ 1+t\ \ 1-t], \qquad M_2 = \tfrac{1}{\sqrt{2}}[0\ \ 0\ \ 1\ \ -1],$$

with right-hand sides $1/\sqrt{2}$ and 0. Their row spaces are the spans of these rows: they describe constraint normals, not feasible points. Normalizing each row to unit length gives

$$c_F(t) = \frac{|t|}{\sqrt{2+t^2}}, \qquad \kappa([\widehat{M}_1(t); \widehat{M}_2]) = \frac{1+c_F(t)}{1-c_F(t)}.$$

At $t = 0$ the channels are orthogonal; increasing $|t|$ makes their normalized rows nearly parallel without changing the optimization problem. Separate normalization cannot rotate these row spaces, whereas an equivalent redistribution that mixes the two channel classes can. Such cross-channel redistribution changes the resource-separated oracle representation and is not counted as channelwise preconditioning. The Friedrichs cosine below captures this distinction.

D.2 Full weighted conditioning statement

Theorem D.1 (Conditioning and optimal scalar balancing of two cross-coupled affine channels)**.** *For the actual nonzero maps M_1, M_2, let*

$$\mu_j = \sigma^+_{\min}(M_j)^2, \qquad L_j = \sigma_{\max}(M_j)^2, \qquad j \in \{1,2\},$$

put $I = U \cap V$ and $c = c_F(U,V) < 1$, and define $M_\eta = [M_1; \eta M_2]$ for $\eta > 0$. Its nonzero singular values satisfy

$$\sigma_{\max}(M_\eta)^2 \le \Lambda_\eta := \begin{cases} \dfrac{L_1 + \eta^2 L_2 + \sqrt{(L_1 - \eta^2 L_2)^2 + 4\eta^2 L_1 L_2 c^2}}{2}, & I = \{0\}, \\ L_1 + \eta^2 L_2, & I \neq \{0\}, \end{cases} \tag{28}$$

$$\sigma^+_{\min}(M_\eta)^2 \ge \delta_\eta := \frac{\mu_1 + \eta^2 \mu_2 - \sqrt{(\mu_1 - \eta^2 \mu_2)^2 + 4\eta^2 \mu_1 \mu_2 c^2}}{2}. \tag{29}$$

Hence $\kappa(M_\eta) \le \Lambda_\eta/\delta_\eta$. If $I = \{0\}$, this envelope, which uses only (μ_j, L_j, c), is minimized over scalar balances by

$$\eta_\star^2 = \sqrt{\frac{\mu_1 L_1}{\mu_2 L_2}}. \tag{30}$$

In that irredundant case, if $c = 0$, then

$$\kappa(M_{\eta_\star}) \le \max\{\kappa(M_1), \kappa(M_2)\}. \tag{31}$$

Still assuming $I = \{0\}$, if also $M_1^\top M_1 = P_U$ and $M_2^\top M_2 = P_V$, then

$$\kappa\left(\begin{bmatrix} M_1 \\ M_2 \end{bmatrix}\right) = \frac{1+c}{1-c}. \tag{32}$$

Corollary D.2 (Removing exact overlap)**.** *Let $I = U \cap V$, $U_0 = U \cap I^\perp$, and $V_0 = V \cap I^\perp$. Exact common directions contribute the separate benign eigenvalue $a + b$ to $aP_U + bP_V$ and may be consistently deduplicated by Lemma C.4. The cross-channel conditioning is therefore governed by the reduced pair (U_0, V_0), for which $U_0 \cap V_0 = \{0\}$. The scalar-balancing and exact-whitening conclusions of Theorem D.1 apply to that reduced stack; the unreduced stack retains the common-direction eigenvalue.*

D.3 Proof of Theorem D.1

Common directions in the unreduced stack. Lemma C.1 shows that an exact common direction contributes weight $a + b$ to $aP_U + bP_V$. Hence the unreduced stack cannot use the irredundant $1 + c_F$ upper edge: common directions remain genuine eigen-directions until the affine system itself is deduplicated by Lemma C.4. We therefore bound the actual stack on $S = U + V$ and invoke irredundancy only for the scalar-balancing and exact-whitening identities below.

Work on $S = U + V$, the range of $M_\eta^\top$, without discarding $I = U \cap V$. Let P_U, P_V be the orthogonal projectors onto U, V. Since $\mathrm{Ker}(M_1) = U^\perp$, $\mathrm{Ker}(M_2) = V^\perp$, and the nonzero eigenvalues of $M_j^\top M_j$ lie in $[\mu_j, L_j]$,

$$\mu_1 P_U \preceq M_1^\top M_1 \preceq L_1 P_U, \qquad \mu_2 P_V \preceq M_2^\top M_2 \preceq L_2 P_V. \tag{33}$$

Consequently

$$\mu_1 P_U + \eta^2 \mu_2 P_V \preceq M_\eta^\top M_\eta \preceq L_1 P_U + \eta^2 L_2 P_V.$$

If $I = \{0\}$, Lemma C.1 bounds the largest eigenvalue of the right-hand side by the first branch of (28). Exclusive one-dimensional components have weights L_1 or $\eta^2 L_2$, no larger than that envelope. If $I \neq \{0\}$, the common component has eigenvalue $L_1 + \eta^2 L_2$ and attains the projector sum's upper bound, proving the second branch. For the lower bound, the principal-plane values are at least δ_η; the exclusive weights $\mu_1, \eta^2\mu_2$ and the common weight $\mu_1 + \eta^2\mu_2$ are also at least δ_η. Both comparison operators are positive definite on S, so the Rayleigh bounds prove (29), including cases without a nontrivial principal plane.

For the scalar-balance and exact-whitening claims, assume from now on that $I = \{0\}$. It remains to optimize the ratio Λ_η/δ_η. For $a, b > 0$ write

$$\lambda_+(a, b; c) := \frac{a + b + \sqrt{(a-b)^2 + 4abc^2}}{2}.$$

Lemma C.1 also gives $\lambda_+(a, b; c)\lambda_-(a, b; c) = ab(1 - c^2)$. Set

$$G_c(t) := \cosh t + \sqrt{\sinh^2 t + c^2}.$$

Then

$$\lambda_+(a, b; c) = \sqrt{ab}\, G_c\left(\frac{1}{2}\log\frac{b}{a}\right), \qquad \lambda_-(a, b; c) = \frac{\sqrt{ab}(1 - c^2)}{G_c(\frac{1}{2}\log(b/a))}.$$

Consequently, with $\vartheta := \eta^2$,

$$\frac{\Lambda_\eta}{\delta_\eta} = \frac{\sqrt{\kappa(M_1)\kappa(M_2)}}{1 - c^2} G_c(t_L) G_c(t_\mu), \quad t_L := \frac{1}{2}\log\frac{\vartheta L_2}{L_1}, \quad t_\mu := \frac{1}{2}\log\frac{\vartheta\mu_2}{\mu_1}. \tag{34}$$

The difference $t_L - t_\mu$ is independent of ϑ. Moreover $\log G_c$ is even and convex: for $c > 0$,

$$\frac{d}{dt}\log G_c(t) = \frac{\sinh t}{\sqrt{\sinh^2 t + c^2}}, \qquad \frac{d^2}{dt^2}\log G_c(t) = \frac{c^2 \cosh t}{(\sinh^2 t + c^2)^{3/2}} \geq 0,$$

and for $c = 0$ the limiting function is $\log G_0(t) = |t|$. Hence the product in (34) is minimized when $t_L + t_\mu = 0$, which is equivalent to $\vartheta^2 = \mu_1 L_1/(\mu_2 L_2)$ and proves (30); write $\vartheta_\star := \eta_\star^2$ for this minimizer.

If $c = 0$, the projector eigenvalues are simply the two weights. Suppose without loss of generality $\kappa(M_1) \geq \kappa(M_2)$. At $\vartheta = \vartheta_\star$, one has $\vartheta L_2 \leq L_1$ and $\vartheta\mu_2 \geq \mu_1$, so $\Lambda_{\eta_\star} = L_1$ and $\delta_{\eta_\star} = \mu_1$, proving (31). Finally, in the exactly whitened case $M^\top M = P_U + P_V$; Lemma C.1 gives extremal eigenvalues $1 - c$ and $1 + c$, which yields (7). □

D.4 Canonical core-model heterogeneity characterization

Proposition D.3 (Exact heterogeneity characterization of the canonical cross angle)**.** *Consider the canonical core model $C_i = D_i = 0$ and assume* $\mathrm{Range}(K_1^\top) \cap \mathrm{Range}(K_2^\top) = \{0\}$. *Let*

$$H_0 := \sum_{i=1}^n (A_i A_i^\top + B_i B_i^\top), \qquad H_{\mathrm{het}} := \sum_{i=1}^n (B_i - \bar{B})(B_i - \bar{B})^\top, \qquad \bar{B} := \frac{1}{n}\sum_{i=1}^n B_i. \tag{35}$$

Then the canonical Friedrichs cosine is exactly

$$c_F^2 = \sup_{h:\, h^\top H_0 h > 0} \frac{h^\top H_{\mathrm{het}} h}{h^\top H_0 h} = \lambda_{\max}\left(H_0^{\dagger/2} H_{\mathrm{het}} H_0^{\dagger/2}\right), \tag{36}$$

where the eigenvalue is taken on $\mathrm{Range}(H_0)$. *Moreover $0 \preceq H_{\mathrm{het}} \preceq H_0$, so the spectrum in* (36) *lies in $[0, 1)$ under irredundancy. Thus $\chi_\times$ is computable from an $m \times m$ generalized eigenvalue problem; in particular, $c_F \to 1$ iff the largest heterogeneity Rayleigh quotient tends to one.*

D.5 Proof of Proposition D.3

For $C = D = 0$, the canonical channels in (4) are $K_1 = [E_m A \; E_m B]$ and $K_2 = [0 \; \mathcal{W}_q]$. Hence

$$V := \mathrm{Range}(K_2^\top) = \{0\} \times \mathrm{Range}(\mathcal{W}_q),$$

and, because K_1 has only the coupled rows, every vector in $U := \mathrm{Range}(K_1^\top)$ is of the form

$$\frac{1}{\sqrt{n}} g(h), \qquad g(h) := \mathrm{col}\big(A_i^\top h, B_i^\top h\big)_{i=1}^n, \qquad h \in \mathbb{R}^m.$$

Thus the parameterization by h is exhaustive, not merely a restricted family of row-space directions. Under $U \cap V = \{0\}$,

$$c_F = \sup_{u \in U \setminus \{0\}} \frac{\|P_V u\|}{\|u\|}.$$

The projector onto $\mathrm{Range}(\mathcal{W}_q)$ subtracts the block mean, so

$$P_V g(h) = \mathrm{col}\big(0, (B_i - \bar{B})^\top h\big)_{i=1}^n.$$

Consequently

$$\frac{\|P_V g(h)\|^2}{\|g(h)\|^2} = \frac{\sum_i \|(B_i - \bar{B})^\top h\|^2}{\sum_i (\|A_i^\top h\|^2 + \|B_i^\top h\|^2)} = \frac{h^\top H_{\mathrm{het}} h}{h^\top H_0 h}.$$

Taking the supremum proves the first equality in (36). Also

$$H_{\mathrm{het}} = \sum_i B_i B_i^\top - n \bar{B} \bar{B}^\top \preceq \sum_i B_i B_i^\top \preceq H_0,$$

which implies $\mathrm{Range}(H_{\mathrm{het}}) \subseteq \mathrm{Range}(H_0)$ and converts the Rayleigh quotient into the pseudoinverse generalized-eigenvalue formula in (36). Irredundancy excludes eigenvalue one. □

E Sharpness and Optimal Separate Preconditioning

Proposition E.1 (Sharp two-dimensional obstruction)**.** *For every $\theta \in (0, \pi/2]$ there are rank-one $M_1, M_2 : \mathbb{R}^2 \to \mathbb{R}$ with $\kappa(M_1) = \kappa(M_2) = 1$ but $\kappa([M_1; M_2]) = \cot^2(\theta/2)$.*

E.1 Proof of Proposition E.1

Take

$$M_1 = [1 \quad 0], \qquad M_2 = [\cos\theta \quad \sin\theta].$$

Both rows have unit norm, so each channel has condition number one and their row-space angle is θ. Lemma C.1 with $a = b = 1$ gives joint Gram eigenvalues $1 \pm \cos\theta$, hence

$$\kappa([M_1; M_2]) = \frac{1 + \cos\theta}{1 - \cos\theta} = \cot^2(\theta/2).$$

In particular, $\kappa([M_1; M_2]) \sim 4/\theta^2$ as $\theta \downarrow 0$. □

Theorem E.2 (Optimality of the cross-angle barrier under arbitrary channelwise preconditioning)**.** *Let $U = \mathrm{Range}(M_1^\top)$ and $V = \mathrm{Range}(M_2^\top)$ be irredundant, $U \cap V = \{0\}$, and let S_j range over all linear maps that are injective on $\mathrm{Range}(M_j)$. Then*

$$\inf_{S_1, S_2} \kappa\left(\begin{bmatrix} S_1 M_1 \\ S_2 M_2 \end{bmatrix}\right) = \frac{1 + c_F(U, V)}{1 - c_F(U, V)} = \chi_\times^2. \tag{37}$$

The infimum is attained by exact separate whitening,

$S_j = (M_j M_j^\top)^{\dagger/2}$ on $\mathrm{Range}(M_j)$, where $\dagger$ denotes the Moore–Penrose pseudoinverse. Thus no invertible linear processing performed separately inside the two affine channels can improve the intrinsic cross-angle condition number.

E.2 Proof of Theorem E.2

Write $\widetilde{M_j} = S_j M_j$ and

$$\mathsf{G}_1 := \widetilde{M_1}^\top \widetilde{M_1}, \qquad \mathsf{G}_2 := \widetilde{M_2}^\top \widetilde{M_2}, \qquad \mathsf{G} := \mathsf{G}_1 + \mathsf{G}_2.$$

Injectivity of S_j on $\mathrm{Range}(M_j)$ preserves the row spaces, so G_1 is positive definite on U and vanishes on $U^\perp$, while G_2 is positive definite on V and vanishes on $V^\perp$. Since $U \cap V = \{0\}$, G is positive definite on $\mathcal{S} := U + V$.

Let Q_U, Q_V be orthonormal basis matrices for U, V and set $Q_{UV} := [Q_U\ Q_V]$. The map $Q_{UV} : \mathbb{R}^{\dim U} \oplus \mathbb{R}^{\dim V} \to \mathcal{S}$ is an isomorphism. For some positive definite matrices H_U, H_V,

$$\mathsf{G} = Q_{UV} \begin{bmatrix} H_U & 0 \\ 0 & H_V \end{bmatrix} Q_{UV}^\top \quad \text{on } \mathcal{S}.$$

Hence

$$\mathsf{G}^{-1} = Q_{UV}^{-\top} \begin{bmatrix} H_U^{-1} & 0 \\ 0 & H_V^{-1} \end{bmatrix} Q_{UV}^{-1},$$

and therefore every $u \in U$ and $v \in V$ are orthogonal in the G^{-1}-metric:

$$u^\top \mathsf{G}^{-1} v = 0. \tag{38}$$

Choose a Friedrichs principal pair of Euclidean unit vectors $u \in U$, $v \in V$, with $u^\top v = c := c_F(U, V)$ (flip the sign of v if needed). Let $0 < \underline{g} \le \overline{g}$ be the smallest and largest eigenvalues of G^{-1} on $\mathcal{S}$. By (38),

$$(u+v)^\top \mathsf{G}^{-1}(u+v) = (u-v)^\top \mathsf{G}^{-1}(u-v).$$

Using the Rayleigh bounds on the two sides gives

$$2\underline{g}(1+c) \le (u+v)^\top \mathsf{G}^{-1}(u+v) = (u-v)^\top \mathsf{G}^{-1}(u-v) \le 2\overline{g}(1-c).$$

Thus

$$\kappa_+(\mathsf{G}) = \kappa_+(\mathsf{G}^{-1}|_{\mathcal{S}}) = \frac{\overline{g}}{\underline{g}} \ge \frac{1+c}{1-c} = \chi_\times^2.$$

Since G is the Gram matrix of the stacked preconditioned operator, this proves the lower bound in (37). Exact separate whitening gives $\mathsf{G}_1 = P_U$ and $\mathsf{G}_2 = P_V$; the exact-whitening part of Theorem D.1 then gives equality. □

F Generic Separate Normalization

Lemma F.1 (Row-space invariance under left spectral preconditioning)**.** *If S maps $\mathrm{Range}(M)$ bijectively onto itself, then $\mathrm{Range}((SM)^\top) = \mathrm{Range}(M^\top)$.*

Lemma F.2 (Residual condition after separate normalization)**.** *Let $U = \mathrm{Range}(\widehat{K}_1^\top)$ and $V = \mathrm{Range}(\widehat{K}_2^\top)$, with $U + V \neq \{0\}$. Put $I = U \cap V$, and define $g_I = 1$ if $I = \{0\}$ and $g_I = 2$ otherwise. Suppose $0 < a \le b$ and*

$$aP_U \preceq \widehat{K}_1^\top \widehat{K}_1 \preceq bP_U, \qquad aP_V \preceq \widehat{K}_2^\top \widehat{K}_2 \preceq bP_V.$$

Put $c = c_F(U, V) < 1$, with the convention in (5)*. Then*

$$\frac{a}{b}\frac{1+c}{1-c} \le \kappa([\widehat{K}_1; \widehat{K}_2]) \le g_I \frac{b}{a}\frac{1+c}{1-c}. \tag{39}$$

In all cases the upper bound $2b/[a(1-c)]$ also holds.

F.1 Proof of Lemma F.1

Because S is injective on $\mathrm{Range}(M)$,

$$SMx = 0 \iff Mx = 0,$$

so $\mathrm{Ker}(SM) = \mathrm{Ker}(M)$. In finite-dimensional Euclidean spaces, $\mathrm{Range}(M^\top) = \mathrm{Ker}(M)^\perp$. Therefore

$$\mathrm{Range}((SM)^\top) = \mathrm{Ker}(SM)^\perp = \mathrm{Ker}(M)^\perp = \mathrm{Range}(M^\top).$$

The same argument applies to any polynomial left preconditioner whose scalar polynomial is nonzero on the positive spectrum of $MM^\top$. Hence the Friedrichs angle is invariant. □

F.2 Proof of Lemma F.2

Set $P = P_U + P_V$ and $\mathsf{G} = \widehat{K}_1^\top \widehat{K}_1 + \widehat{K}_2^\top \widehat{K}_2$. Then $aP \preceq \mathsf{G} \preceq bP$ and $\operatorname{Ker} \mathsf{G} = (U+V)^\perp$. The principal-angle decomposition gives eigenvalue 2 on I, eigenvalue 1 on exclusive directions, and $1 \pm \cos\theta$ on each nontrivial principal plane. Consequently $\lambda_{\max}(P) \le 2$ and $\lambda^+_{\min}(P) \ge 1 - c$. If $I = \{0\}$, the upper bound sharpens to $\lambda_{\max}(P) \le 1 + c$. These facts prove both asserted upper bounds. If $c > 0$, a principal plane attaining c supplies unit vectors with Rayleigh quotients $1 + c$ and $1 - c$; comparison with G gives the lower bound in (39). If $c = 0$, that bound follows from $\kappa \ge 1 \ge a/b$. No common-direction eigenvalue is discarded from the full stack. □

F.3 Constant-accuracy inexact channel solves

Corollary F.3 (Inexact whitening preserves cross-optimality)**.** *Let $\widetilde{K}_1, \widetilde{K}_2$ have row spaces U, V, with $U + V \ne \{0\}$, and let I, g_I be as in Lemma F.2. Define $\chi_{U,V} := \sqrt{(1 + c_F(U,V))/(1 - c_F(U,V))}$. Suppose, for $0 \le \delta_{\mathrm{wh}} < 1$,*

$$(1 - \delta_{\mathrm{wh}})P_U \preceq \widetilde{K}_1^\top \widetilde{K}_1 \preceq (1 + \delta_{\mathrm{wh}})P_U, \qquad (1 - \delta_{\mathrm{wh}})P_V \preceq \widetilde{K}_2^\top \widetilde{K}_2 \preceq (1 + \delta_{\mathrm{wh}})P_V.$$

Then

$$\chi^2_{U,V} \le \kappa([\widetilde{K}_1; \widetilde{K}_2]) \le g_I \frac{1 + \delta_{\mathrm{wh}}}{1 - \delta_{\mathrm{wh}}} \chi^2_{U,V}. \tag{40}$$

Thus fixed whitening accuracy preserves the angle dependence. An outer certified interval of this order gives joint Chebyshev degree $O(\chi_{U,V})$.

Proof. Apply Lemma F.2 with $a = 1 - \delta_{\mathrm{wh}}$ and $b = 1 + \delta_{\mathrm{wh}}$ for the upper bound. For $I = \{0\}$ and both row spaces nonzero, Theorem E.2 gives the lower bound. For $I \ne \{0\}$ and $c_F(U,V) > 0$, put $U_0 = U \cap I^\perp$, $V_0 = V \cap I^\perp$ and $S_0 = U_0 + V_0$. The compressed maps $\widetilde{K}_j P_{S_0}$ have irredundant row spaces U_0, V_0 and the same Friedrichs cosine. Theorem E.2 bounds their stacked condition below by $\chi^2_{U,V}$. The extremal Rayleigh quotients of the full positive Gram on $U + V$ bound those of its compression to S_0 from outside, so its condition number is at least as large. For $c_F(U,V) = 0$, including a zero row space, the lower bound is simply $\kappa \ge 1$. Since $g_I \le 2$, taking square roots of the certified condition bound proves the degree claim. Fixed inner accuracy suffices independently of the target outer accuracy. □

F.4 Constructive polynomial channel normalization

Proposition F.4 (Polynomial fixed-accuracy channel normalization)**.** *Let $M \ne 0$, $Q = MM^\top$, $U = \operatorname{Range}(M^\top)$, and suppose $\operatorname{spec}(Q) \subset \{0\} \cup [a,b]$, where $0 < a \le b$. Put $\kappa_{\mathrm{poly}} = b/a$. For $0 < \delta_{\mathrm{wh}} < 1$ there is a real polynomial p_N with*

$$N + 1 = O\left(\sqrt{\kappa_{\mathrm{poly}}} \log \frac{6}{\delta_{\mathrm{wh}}}\right)$$

such that $\widehat{M} = p_N(Q)M$ satisfies

$$(1 - \delta_{\mathrm{wh}})P_U \preceq \widehat{M}^\top \widehat{M} \preceq (1 + \delta_{\mathrm{wh}})P_U. \tag{41}$$

Moreover $\operatorname{Ker} \widehat{M} = \operatorname{Ker} M$ and $\operatorname{Range}(\widehat{M}^\top) = U$. For a feasible equation $Mw = v$, its equivalent normalized equation is $\widehat{M}w = p_N(Q)v$. One forward or adjoint normalized action uses at most $2N + 1$ actions of $M, M^\top$.

Proof. Set $\tau = \delta_{\mathrm{wh}}/3$. If $a = b$, take $N = 0$ and $p_0 = a^{-1/2}$. Otherwise put $c = (a+b)/2$, $s = (b-a)/2$. Let T_k be the degree-k Chebyshev polynomial of the first kind and set

$$N + 1 = \left\lceil \frac{\sqrt{\kappa_{\mathrm{poly}}}}{2} \log \frac{6}{\delta_{\mathrm{wh}}} \right\rceil, \qquad R_{N+1}(x) = \mathsf{T}_{N+1}((c - x)/s).$$

Let p_N interpolate $x^{-1/2}$ at the $N + 1$ roots of R_{N+1} in (a,b). To bound its relative error, use

$$x^{-1/2} = \frac{1}{\pi} \int_0^\infty \frac{t^{-1/2}}{x + t}\, dt.$$

For $t \geq 0$, the polynomial

$$q_t(x) = \frac{1 - R_{N+1}(x)/R_{N+1}(-t)}{x + t}$$

has degree at most N and interpolates $(x+t)^{-1}$ at the same roots. Integration of the finite interpolation formula therefore gives

$$\left| x^{-1/2} - p_N(x) \right| \leq \frac{x^{-1/2}}{\mathsf{T}_{N+1}(c/s)} \quad (a \leq x \leq b),$$

because $|R_{N+1}(x)| \leq 1$ and $R_{N+1}(-t) = \mathsf{T}_{N+1}((c+t)/s) \geq \mathsf{T}_{N+1}(c/s) > 0$. Writing $\rho = (\sqrt{\kappa_{\mathrm{poly}}} - 1)/(\sqrt{\kappa_{\mathrm{poly}}} + 1)$,

$$\frac{1}{\mathsf{T}_{N+1}(c/s)} = \frac{2\rho^{N+1}}{1 + \rho^{2(N+1)}} \leq 2\exp(-2(N+1)/\sqrt{\kappa_{\mathrm{poly}}}) \leq \tau.$$

Thus $|\sqrt{x}\, p_N(x) - 1| \leq \tau$. For a thin SVD $M = U_M \Sigma V_M^\top$ with positive singular values σ_i,

$$\widehat{M}^\top \widehat{M} = V_M \operatorname{diag}\big(\sigma_i^2 p_N(\sigma_i^2)^2\big)\, V_M^\top.$$

Its positive eigenvalues lie in $[(1-\tau)^2, (1+\tau)^2]$, which is contained in $[1 - \delta_{\mathrm{wh}}, 1 + \delta_{\mathrm{wh}}]$. This proves (41) and both kernel and row-space assertions.

Exact whitening is $(Q^\dagger)^{1/2} M$, with the Moore–Penrose inverse acting as zero on $\operatorname{Ker} Q$. The polynomial need not approximate that operator at zero: M maps into $\operatorname{Range}(Q)$, so $p_N(0)$ is immaterial. Feasibility implies $v \in \operatorname{Range}(M)$, and $p_N(Q)$ is injective on that range, proving equivalence of the affine equations. A forward action uses one M followed by N products $Qz = M(M^\top z)$; the adjoint reverses the order. These use $(N+1, N)$ forward/adjoint actions, or $(N, N+1)$. There is no additional $\log \kappa_{\mathrm{poly}}$ factor. For fixed δ_{wh}, $N + 1 = O(\sqrt{\kappa_{\mathrm{poly}}})$. □

G CERTIFIED CC-MAC LOCAL REALIZATION

This section specializes the generic normalization results above to the executable CC-MAC lift and records the spectral certificates used by the primitive-cost ledger.

Structural parameters. In the original variable ordering put $H = [A\ B]$ and $T = [C\ 0]$. A fixed permutation groups these into the local blocks $H_i = [A_i\ B_i]$ and $\bar{C}_i = [C_i\ 0]$. Define

$$\begin{aligned} h &:= \|H\|^2 = \max_i \lambda_{\max}(A_i A_i^\top + B_i B_i^\top), \\ S_{HC} &:= \frac{1}{n} \sum_i (A_i P_{\operatorname{Ker} C_i} A_i^\top + B_i B_i^\top), \\ d_D &:= \max_i \lambda_{\max}(D_i^\top D_i), \qquad S_D := \frac{1}{n} \sum_i D_i^\top D_i. \end{aligned} \tag{42}$$

The subscript on d_D distinguishes this scale from the primal dimension d. Set $\mu_{HC} = \lambda^+_{\min}(S_{HC})$ if $S_{HC} \neq 0$, and $\mu_{HC} = 0$ otherwise. Set $\mu_D = \lambda^+_{\min}(S_D)$ if $D \neq 0$, and $\mu_D = 0$ otherwise. The ideal condition parameters are

$$k_1 := \widetilde{\kappa}_{HC} = \begin{cases} h/\mu_{HC}, & S_{HC} \neq 0, \\ 1, & S_{HC} = 0, \end{cases} \qquad k_2 := \widehat{\kappa}_{D^\top} = \begin{cases} d_D/\mu_D, & D \neq 0, \\ 1, & D = 0. \end{cases}$$

For $C \neq 0$ define

$$\kappa_C := \frac{\max_i \sigma_{\max}(C_i)^2}{\min_{i: C_i \neq 0} \sigma^+_{\min}(C_i)^2},$$

and use $\kappa_C = 1$ when $C = 0$. The graph convention remains $\kappa_W = \lambda_{\max}(W)/\lambda^+_{\min}(W)$. For the one-agent case use the bookkeeping value $\kappa_W = 1$ and perform no communication.

Assumption 1 (Supplied numerical spectral certificates)**.** *For $H \neq 0$ the implementation receives $\bar{h} \geq h$ and $0 < \underline{\mu}_{HC} \leq \bar{h}$ satisfying*

$$\operatorname{spec}(S_{HC}) \subset \{0\} \cup [\underline{\mu}_{HC}, \bar{h}].$$

For $D \neq 0$ it receives $\bar{d}_D \geq d_D$ and $0 < \underline{\mu}_D \leq \bar{d}_D$ satisfying $\operatorname{spec}(S_D) \subset \{0\} \cup [\underline{\mu}_D, \bar{d}_D]$. For $C \neq 0$ it receives $0 < \ell_C \leq u_C$ with $\operatorname{spec}(CC^\top) \subset \{0\} \cup [\ell_C, u_C]$. For $n \geq 2$ it receives $0 < \ell_W \leq u_W$ with $\operatorname{spec}(W) \subset \{0\} \cup [\ell_W, u_W]$. The outer method receives the objective bounds and either $\Xi \geq \chi_\times$ or a valid positive-spectrum interval for the final normalized stack.

Define the corresponding certified ratios by

$$\bar{k}_1 = \bar{h}/\underline{\mu}_{HC}, \qquad \bar{k}_2 = \bar{d}_D/\underline{\mu}_D, \qquad \bar{\kappa}_C = u_C/\ell_C, \qquad \bar{\kappa}_W = u_W/\ell_W. \tag{43}$$

The respective bookkeeping values are 1 for $H = 0$, $D = 0$, $C = 0$, and $n = 1$, so no absent-block ratio is evaluated. If $S_{HC} = 0$ is known and $H \neq 0$, one may take $\underline{\mu}_{HC} = \bar{h}$. The excluded-gap certificates allow rank deficiency and do not require the rank as an input.

These are supplied problem/setup parameters, as in spectrally tuned Chebyshev preconditioning (Salim et al., 2022; Yarmoshik et al., 2025). Exact eigenvalues or eigendecompositions are not required. Conservative bounds suffice; their ratios determine the costs. Constant-factor-tight certificates recover the ideal orders. Obtaining certificates from arbitrary black-box access is a separate setup operation, outside the solve counts below. Upper bounds from explicitly available local matrix entries may be used, with any acquisition and aggregation costs charged separately; access to such entries is not implicit in a matvec oracle.

Preliminary graph and local normalization. For $n \geq 2$ and $\ell_W < u_W$, put $c_W = (u_W + \ell_W)/2$, $s_W = (u_W - \ell_W)/2$ and

$$\nu = \left\lceil \frac{\sqrt{\bar{\kappa}_W}}{2} \log 4 \right\rceil, \qquad P_W(x) = 1 - \frac{\mathsf{T}_\nu((c_W - x)/s_W)}{\mathsf{T}_\nu(c_W/s_W)}.$$

For $\ell_W = u_W$ take $P_W(x) = x/\ell_W$. Then $\widetilde{W} = P_W(W)$ has the same kernel as W and positive spectrum in $[1/2, 3/2]$. Thus its positive Gram spectrum lies in $[1/4, 9/4]$, while one action costs $O(\sqrt{\bar{\kappa}_W})$ raw neighbor rounds. The filter is applied to W, not W^2.

For $C \neq 0$, apply Proposition F.4 with accuracy $1/2$, and write

$$C_\star = p_C(CC^\top)C, \qquad c_\star = p_C(CC^\top)\boldsymbol{c}.$$

Then $\frac{1}{2}P_{\operatorname{Range}(C^\top)} \preceq C_\star^\top C_\star \preceq \frac{3}{2}P_{\operatorname{Range}(C^\top)}$ and $\operatorname{Ker} C_\star = \operatorname{Ker} C$. This block-diagonal operation costs $O(\sqrt{\bar{\kappa}_C})$ raw $C/C^\top$ products per action and no communication. If $C = 0$, omit this layer and its rows.

Lemma G.1 (Certified spectra of the balanced executable channels)**.** *Suppose $H, D \neq 0$ and $n \geq 2$. With $\widetilde{W}_s = \widetilde{W} \otimes I_s$, choose*

$$\alpha^2 = 12\bar{h}, \qquad \beta^2 = 6\bar{h}, \qquad \gamma^2 = 8\bar{d}_D,$$

and form the final executable channels

$$\mathcal{L}_1 = \begin{bmatrix} A & B & \alpha\widetilde{W}_m \\ \beta C_\star & 0 & 0 \end{bmatrix}, \qquad \mathcal{L}_2 = \begin{bmatrix} 0 & D & 0 \\ 0 & \gamma\widetilde{W}_q & 0 \end{bmatrix}. \tag{44}$$

Their right-hand sides are $\operatorname{col}(\boldsymbol{b}, \beta c_\star)$ and $\operatorname{col}(\boldsymbol{e}, 0)$, respectively. Their positive Gram spectra satisfy

$$\begin{aligned} \operatorname{spec}^+(\mathcal{L}_1\mathcal{L}_1^\top) &\subset [\underline{\mu}_{HC}/4, 37\bar{h}], \\ \operatorname{spec}^+(\mathcal{L}_2\mathcal{L}_2^\top) &\subset [\underline{\mu}_D/2, 19\bar{d}_D]. \end{aligned} \tag{45}$$

If $S_{HC} = 0$, the first lower endpoint improves to $\bar{h}$. The final row spaces obey $\chi_{\mathrm{loc}} \leq \sqrt{3}\,\chi_\times$, and the penalty objective in Proposition B.3, with $r = \mu_f/(2\bar{h})$, has $\kappa_{\Phi_{\mathrm{loc}}} = O(\kappa_f)$.

Proof. This is the coupled/local spectral reduction $A_i \mapsto H_i$, $C_i \mapsto \bar{C}_i$ of Yarmoshik et al. (2026, Lemma 4.4 and Appendix H); the following estimates verify the compatible bridge scaling.

Let $\Pi = (\mathbf{1}\mathbf{1}^\top/n) \otimes I_m$, $T_\star = [C_\star\ 0]$ and $\mathcal{J} = \mathrm{col}(\Pi H, \beta T_\star)$. The nonzero spectrum of $\Pi H P_{\mathrm{Ker}\, T} H^\top \Pi$ equals that of S_{HC}. For $t \in \mathrm{Range}(\mathcal{J}^\top)$, decompose $t = y_0 + g_\perp$ orthogonally with $y_0 \in \mathrm{Range}(T^\top)$ and $g_\perp \in \mathrm{Range}(P_{\mathrm{Ker}\, T} H^\top \Pi)$. Using $\|a+b\|^2 \geq \frac{1}{2}\|b\|^2 - \|a\|^2$ gives

$$\|\mathcal{J}t\|^2 \geq (\beta^2/2 - h)\|y_0\|^2 + (\underline{\mu}_{HC}/2)\|g_\perp\|^2 \geq (\underline{\mu}_{HC}/2)\|t\|^2.$$

If $S_{HC} = 0$, then $g_\perp = 0$ and the last bound improves to $2\bar{h}\|t\|^2$.

For $z = (a, c) \in \mathrm{Range}(\mathcal{L}_1)$ write $a_0 = \Pi a$, $a_\perp = (\mathrm{Id} - \Pi)a$. Representing $z = \mathcal{L}_1(u, y)$ shows $(a_0, c) = \mathcal{J}u \in \mathrm{Range}(\mathcal{J})$. Hence

$$\|\mathcal{L}_1^\top z\|^2 \geq \tfrac{1}{2}\|\mathcal{J}^\top(a_0, c)\|^2 + (\alpha^2/4 - h)\|a_\perp\|^2 \geq (\underline{\mu}_{HC}/4)\|z\|^2.$$

When $S_{HC} = 0$, this gives $\bar{h}\|z\|^2$. The upper bound is $\|\mathcal{L}_1\|^2 \leq h + (9/4)\alpha^2 + (3/2)\beta^2 \leq 37\bar{h}$.

For channel 2, work on its active shared coordinates. Its kernel is $\mathbf{1} \otimes \mathrm{Ker}\, S_D$. For z orthogonal to this kernel, decompose $z = z_0 + z_\perp$ into consensus and disagreement. Then $\|Dz_0\|^2 \geq \underline{\mu}_D\|z_0\|^2$, and

$$\|Dz\|^2 + \gamma^2\|\widetilde{W}_q z\|^2 \geq \tfrac{1}{2}\underline{\mu}_D\|z_0\|^2 + (\gamma^2/4 - d_D)\|z_\perp\|^2 \geq \tfrac{1}{2}\underline{\mu}_D\|z\|^2.$$

The upper bound is $d_D + (9/4)\gamma^2 \leq 19\bar{d}_D$.

At fixed $\alpha, \widetilde{W}$, replacing C by $C_\star$ and positive row scaling preserve channel 1's row space. Channel 2 graph filtering and positive scaling preserve its row space. Replacing W inside the lifted channel 1 generally changes its full lifted row space. Apply Proposition B.3 to this final lift: $\underline{\sigma}_W = 1/2$ and $\alpha^2\underline{\sigma}_W^2 = 3\bar{h}$ give the asserted angle bound. The same proposition gives $\mu_{\Phi_{\mathrm{loc}}} \geq \mu_f/2$ and $L_{\Phi_{\mathrm{loc}}} \leq L_f + 14\mu_f$. □

Zero blocks and reductions. If $H = 0$, feasibility makes the original coupled equation vacuous: remove that equation and the auxiliary variable, take $\Phi_{\mathrm{loc}} = F$, and use only $[C_\star\ 0]$ for channel 1. Its nonzero Gram spectrum is in $[1/2, 3/2]$. If $D = 0$, omit the D rows but retain $\widetilde{W}_q$ with $\gamma = 1$; its nonzero Gram spectrum is in $[1/4, 9/4]$. If $C = 0$, omit its rows and the β scaling. For $n = 1$, omit both graph terms and the auxiliary variable and take $\Phi_{\mathrm{loc}} = F$; the bounds (45) still hold for the remaining balanced nonzero H, D branches. Completely empty channels need no calls. When $B = 0$, S_{HC} and k_1 reduce to the predecessor coupled/local quantities and $\chi_\times = 1$. When $C = 0$, $S_{HC} = n^{-1}\sum_i H_i H_i^\top$ and the local normalization layer disappears. The substitution is a channel spectral reduction, not a reduction of the full cross-coupled problem.

Final normalization and outer certificate. For each nonempty channel, set $Q_j = \mathcal{L}_j\mathcal{L}_j^\top$, $U_j = \mathrm{Range}(\mathcal{L}_j^\top)$ and $\widehat{L}_j = p_j(Q_j)\mathcal{L}_j$. For fixed $0 < \delta_{\mathrm{wh}} < 1$, independent of ε, Proposition F.4 and (45) give degrees $N_j^{\mathrm{wh}} + 1 = O(\sqrt{\bar{k}_j})$ and

$$(1 - \delta_{\mathrm{wh}})P_{U_j} \preceq \widehat{L}_j^\top \widehat{L}_j \preceq (1 + \delta_{\mathrm{wh}})P_{U_j}.$$

Transform each final right-hand side by the same $p_j(Q_j)$. This last transformation preserves each final channel row space. These spectral estimates concern $\mathcal{L}_j$, not automatically the literal unbalanced L_j of (19). For $\widehat{L} = [\widehat{L}_1; \widehat{L}_2]$, Lemma F.2 permits common rows to remain and a supplied $\Xi \geq \chi_\times$ gives

$$\mathrm{spec}^+(\widehat{L}^\top\widehat{L}) \subset \left[\frac{1 - \delta_{\mathrm{wh}}}{3\Xi^2},\ 2(1 + \delta_{\mathrm{wh}})\right]. \tag{46}$$

The positive graph condition is already constant inside $\mathcal{L}_j$. Its physical cost $O(\sqrt{\kappa_W})$ is therefore charged once per graph action, not again inside N_j^{wh}. The fixed normalization accuracy contributes no further $\log(1/\varepsilon)$ factor.

H Cross-APAPC Algorithm and Smooth-SC Upper Bound

H.1 Chebyshev correction used inside Cross-APAPC

Let $Q = \widehat{K}^\top \widehat{K}$, $c = \widehat{K}^\top \widehat{v}$, and suppose $0 < \lambda_- \le \lambda_{\min}^+(Q) \le \lambda_{\max}(Q) \le \lambda_+$. Algorithm 2 is the classical Chebyshev semi-iteration for the consistent normal equation $Qz = c$ (Algorithm 2 of Salim et al. (2022) with K, b replaced by $\widehat{K}, \widehat{v}$). It is an inner correction routine, not a second outer optimization method.

Algorithm 2 Chebyshev normal-equation correction $\mathrm{Cheb}_N(z_0; \widehat{K}, \widehat{v})$

Require: z_0, $\widehat{K}$, $\widehat{v}$, interval $[\lambda_-, \lambda_+]$, degree $N \ge 1$
1: $\varrho \leftarrow (\lambda_+ - \lambda_-)^2/16$, $\nu \leftarrow (\lambda_+ + \lambda_-)/2$, $\gamma_0 \leftarrow -\nu/2$
2: $p_0 \leftarrow -\widehat{K}^\top(\widehat{K} z_0 - \widehat{v})/\nu$, $z_1 \leftarrow z_0 + p_0$
3: **for** $i = 1, \dots, N-1$ **do**
4: $\quad \beta_{i-1} \leftarrow \varrho/\gamma_{i-1}$, $\gamma_i \leftarrow -(\nu + \beta_{i-1})$
5: $\quad p_i \leftarrow [\widehat{K}^\top(\widehat{K} z_i - \widehat{v}) + \beta_{i-1} p_{i-1}]/\gamma_i$
6: $\quad z_{i+1} \leftarrow z_i + p_i$
7: **end for**
8: **return** z_N

Each inner step uses one multiplication by $\widehat{K}$ and one by $\widehat{K}^\top$. The degree $N = \lceil \sqrt{\lambda_+/\lambda_-} \rceil$ used by Cross-APAPC is what converts the joint normalized condition number into the affine-oracle factor.

H.2 Complete Cross-APAPC specification

Choose J execution channels $M_1, \dots, M_J$ with right-hand sides v_j. For canonical or local CC-MAC take $J = 2$ and respectively $M_j = K_j$ or $M_j = \mathcal{L}_j$, including the preliminary transformations in Lemma G.1. Let $U_j := \mathrm{Range}(M_j^\top)$ and choose channelwise maps S_j such that

$$\widehat{K}_j := S_j M_j, \qquad a P_{U_j} \preceq \widehat{K}_j^\top \widehat{K}_j \preceq b P_{U_j}, \qquad 0 < a \le b,$$

with universal a, b. Set $\widehat{v}_j = S_j v_j$ and

$$\widehat{K} := \mathrm{col}(\widehat{K}_1, \dots, \widehat{K}_J), \qquad \widehat{v} := \mathrm{col}(\widehat{v}_1, \dots, \widehat{v}_J).$$

Injectivity of S_j on $\mathrm{Range}(M_j)$ preserves feasibility. Supply bounds

$$0 < \lambda_- \le \lambda_{\min}^+(\widehat{K}^\top \widehat{K}), \qquad \lambda_{\max}(\widehat{K}^\top \widehat{K}) \le \lambda_+.$$

For a generic J-channel instance the certified ratio is $O(\chi_{\times,J}^2)$. For two-channel CC-MAC, any supplied $\Xi \ge \chi_\times$ permits

$$\lambda_- = a/(3\Xi^2), \qquad \lambda_+ = 2b,$$

by Lemma F.2 and the canonical-to-local bridge. Wider certified intervals are also valid; their endpoint ratio determines the Chebyshev degree.

For the standard APAPC estimate, let P be the Chebyshev-preconditioned normal operator, $p^\star = -\nabla\Phi(w^\star) \in \mathrm{Range}(P)$, and $y^\star = P^{\dagger/2} p^\star$. A sufficient supplied initial-energy bound is

$$R_E^2 \ge \|w^0 - w^\star\|^2 + \frac{\eta}{\theta}\|y^0 - y^\star\|^2 + \frac{2\eta(1-\tau)}{\tau} D_\Phi(w_f^0, w^\star), \qquad y^0 = 0,$$

where $D_\Phi(a, b) := \Phi(a) - \Phi(b) - \langle \nabla\Phi(b), a - b\rangle$. This energy, rather than the primal radius alone, controls the convergence prefactor.

The complete outer iteration is Algorithm 1 in Section 4; only its inner Chebyshev correction is repeated here because its affine-call count is used explicitly in the proof.

H.3 Proof of Theorem 4.1

Fix the execution channels $M_1, \dots, M_J$ and normalized stack specified above. Feasibility gives $v_j \in \mathrm{Range}(M_j)$; injectivity of S_j there implies $S_j(M_j w - v_j) = 0$ iff $M_j w - v_j = 0$. Proposition C.2 gives

$$\kappa(\widehat{K}) \le \frac{b}{a}\,\chi_{\times,J}^2. \tag{47}$$

For the canonical CC-MAC oracle, $J = 2$, $M_j = K_j$, $\Phi = F$, and $\chi_{\times,2} \le \sqrt{2}\,\chi_\times$. For the local implementation, $M_j = \mathcal{L}_j$; Proposition B.3 gives $\chi_{\mathrm{loc}} \le \sqrt{3}\,\chi_\times$, while Proposition C.2 then gives $\chi_{\times,2}^{\mathrm{loc}} \le \sqrt{2}\chi_{\mathrm{loc}} = O(\chi_\times)$ and $\kappa_\Phi = \kappa_{\Phi_{\mathrm{loc}}} = O(\kappa_f)$. The zero-block branches use the reduced channels specified there.

Algorithm 1 is the affine-constrained accelerated method of Salim et al. (2022) for Φ and $\widehat{K}w = \widehat{v}$. With the supplied joint interval its Chebyshev degree is $N = \lceil \sqrt{\lambda_+/\lambda_-} \rceil$. With the energy certificate above and $T = O(\sqrt{\kappa_\Phi}\log_+(R_E/\varepsilon))$, it uses

$$O\left(\sqrt{\kappa_\Phi}\log_+\frac{R_E}{\varepsilon}\right) \quad \text{gradients and} \quad O\left(N\sqrt{\kappa_\Phi}\log_+\frac{R_E}{\varepsilon}\right)$$

normalized joint-stack products. The assumed ratio $O(\chi_{\times,J}^2)$ gives the normalized affine bound in (8); in the two-channel CC-MAC specialization, using (46) gives $N = O(\Xi)$.

Each stack action uses each normalized channel once. The supplied graph realization converts the same stack-product count into the communication bound in (8) after summing the per-channel communication costs. In the local CC-MAC case each evaluation of $\nabla\Phi_{\mathrm{loc}}$ additionally uses $O(1)$ actions of $H, H^\top$ and $\widetilde{W}$; these are dominated by the explicit channel costs in Appendix L. The gradient count refers to the original ∇f_i. This proves Theorem 4.1. □

I Smooth-SC Lower Bounds and Their Scope

Oracle and radius conventions. The lower bounds use a linear-span oracle with zero initial vector memories. New primal and dual vectors may be formed only by scalar linear combinations of stored vectors and the specified gradient, channel, adjoint-channel, and neighbor-communication operations. Access to matrix entries, arbitrary new basis vectors, exact factorizations, and unconstrained global projections is not included. Free local operations in a communication lower bound mean closure under these allowed local maps, not unrestricted coordinate generation.

For reference, if $\mathfrak{P}$ is an explicitly specified class and $r \in \{\nabla, \mathrm{aff}, \mathrm{comm}\}$, define

$$\mathfrak{C}_r(\varepsilon; \mathfrak{P}) := \inf_{\mathcal{A}} \sup_{\mathcal{I}\in\mathfrak{P}} N_r(\mathcal{A}, \mathcal{I}, \varepsilon). \tag{48}$$

The energy-controlled upper class of Theorem 4.1 and the primal-radius-normalized hard families below are not silently identified. The statements that follow isolate particular resource mechanisms and record the radius hypotheses used by each construction.

Theorem I.1 (Two-channel affine-oracle lower bound)**.** *For every $N \ge 4$, $\kappa_f \ge 4$, and sufficiently small universal $\varepsilon > 0$, there is a smooth strongly convex one-agent CC-MAC instance, normalized so that $\|w^0 - w^\star\| = 1$, with two individually whitened channels and $\chi_\times = \cot(\pi/(2N)) = \Theta(N)$, requiring $\Omega(\sqrt{\kappa_f}\chi_\times \log(1/\varepsilon))$ composite normalized affine calls and $\Omega(\sqrt{\kappa_f}\log(1/\varepsilon))$ gradients in the stated linear-span oracle.*

We now prove Theorem I.1. The construction is a two-coloring of the edges of a path. It is useful because each color class is a matching and therefore an isometric affine channel, whereas the union retains the full path spectrum.

Lemma I.2 (Alternating-matchings path geometry)**.** *Let $E_N \in \mathbb{R}^{(N-1)\times N}$ be the oriented incidence matrix of the path P_N, with row i equal to $e_i^\top - e_{i+1}^\top$. Let E_{odd} and E_{even} contain respectively the odd and even rows of E_N, and define*

$$M_{\mathrm{odd}} = 2^{-1/2}E_{\mathrm{odd}}, \qquad M_{\mathrm{even}} = 2^{-1/2}E_{\mathrm{even}}, \qquad M = \begin{bmatrix} M_{\mathrm{odd}} \\ M_{\mathrm{even}} \end{bmatrix}.$$

Then every positive singular value of each M_{odd} and M_{even} equals one, while

$$M^\top M = \frac{1}{2} L_{P_N}, \qquad \kappa(M) = \frac{1+\cos(\pi/N)}{1-\cos(\pi/N)} = \cot^2 \frac{\pi}{2N}. \tag{49}$$

Consequently the Friedrichs cosine of the two row spaces equals $\cos(\pi/N)$ and their cross factor is $\cot(\pi/(2N))$.

Proof. Within either parity class no two path edges share a vertex. Hence the nonzero rows of E_{odd} (and of E_{even}) are mutually orthogonal and each has squared norm 2. After multiplication by $2^{-1/2}$ the rows are orthonormal, so $M_j M_j^\top = \mathrm{Id}$ on its row space and every positive singular value is one.

Stacking the two parity classes merely permutes the rows of $2^{-1/2} E_N$, hence $M^\top M = E_N^\top E_N/2 = L_{P_N}/2$. The path Laplacian has eigenvalues

$$2 - 2\cos\frac{k\pi}{N}, \qquad k = 0, 1, \ldots, N-1.$$

Therefore the smallest positive and largest eigenvalues of $M^\top M$ are $1-\cos(\pi/N)$ and $1+\cos(\pi/N)$, respectively, which proves (49). Since the two channels are individually whitened, the exact-whitening identity in Theorem D.1 gives

$$\frac{1+c_F}{1-c_F} = \kappa(M) = \frac{1+\cos(\pi/N)}{1-\cos(\pi/N)}.$$

The map $c \mapsto (1+c)/(1-c)$ is injective on $[0,1)$, so $c_F = \cos(\pi/N)$ and $\chi_\times = \cot(\pi/(2N))$. □

I.1 Proof of Theorem I.1

Fix $N \geq 4$. Use the smooth strongly convex hard functions on a linear graph from the proof of Theorem 2 of Scaman et al. (2017), with local condition number κ_f. Their coordinate-tail argument (before the final conversion to function error) gives a product objective F_N for which every black-box procedure using the path gossip operator

$$W_N := \frac{1}{2} L_{P_N}$$

requires

$$\Omega\left(\sqrt{\kappa_f}\,\frac{1}{\sqrt{\gamma_N}}\log\frac{1}{\varepsilon}\right) \quad \text{applications of } W_N, \qquad \gamma_N := \frac{\lambda_{\min}^+(W_N)}{\lambda_{\max}(W_N)}, \tag{50}$$

and $\Omega(\sqrt{\kappa_f}\log(1/\varepsilon))$ gradient computations. For the path spectrum above,

$$\gamma_N = \frac{1-\cos(\pi/N)}{1+\cos(\pi/N)} = \chi_\times^{-2}.$$

Embed this problem into a one-agent CC-MAC instance. Introduce a dummy scalar local variable x_1 with the decoupled quadratic $\mu_f x_1^2/2$, set $A_1 = C_1 = 0$, and let the shared variable z equal the product variable of F_N. Take $B_1 = M_{\mathrm{odd}} \otimes \mathrm{Id}$, $D_1 = M_{\mathrm{even}} \otimes \mathrm{Id}$, and zero right-hand sides. The dummy block does not change the condition number or the oracle lower bound. (A finite hard-coordinate dimension large enough for the requested accuracy is sufficient for the standard resisting-oracle construction.) Thus the two nonzero affine channels are precisely

$$K_1 = M_{\mathrm{odd}} \otimes \mathrm{Id}, \qquad K_2 = M_{\mathrm{even}} \otimes \mathrm{Id},$$

and Lemma I.2 gives their individual condition numbers and cross factor.

It remains to match the oracle classes. Let $K = [K_1; K_2]$ and $b = 0$. In the black-box first-order (BBFO) linear-span model of Salim et al. (2022), every primal vector that can be generated using K and $K^\top$ belongs to spans containing terms of the form

$$K^\top K x = W_N x, \qquad K^\top K \nabla F_N(x) = W_N \nabla F_N(x),$$

(up to the identity on the hard coordinate). More explicitly,

$$K^\top \operatorname{Span}\{Kx^s, K\nabla F_N(x^s) : s \leq t\} = \operatorname{Span}\{W_N x^s, W_N \nabla F_N(x^s) : s \leq t\}.$$

Thus the exact span-reduction used in the proof of Theorem 1 of Salim et al. (2022) applies verbatim: a BBFO procedure with T composite $K/K^\top$ oracle calls induces a black-box procedure using at most a constant multiple of T applications of W_N. Equation (50) therefore implies

$$T = \Omega\left(\sqrt{\kappa_f}\,\chi_\times \log\frac{1}{\varepsilon}\right).$$

Because a composite oracle is at least as informative per charged call as querying either channel separately, the same lower bound holds for the total number of separately charged channel matrix–vector calls. The gradient lower bound is inherited from the same Scaman construction. The unit-radius normalization in the theorem is without loss: for any quadratic hard instance with solution $w^\star$, the transformation

$$F_c(w) := c^2 F(w/c), \qquad v_c := cv$$

leaves L_f, μ_f, all channel row spaces, and all matrix/network condition numbers unchanged while scaling the optimizer and every oracle-generated primal span by c. Choosing $c = 1/\|w^0 - w^\star\|$ therefore gives unit initial radius and replaces absolute accuracy by the corresponding relative accuracy. This proves Theorem I.1. □

Theorem I.3 (Radius-explicit decentralized communication lower bound)**.** *For the path family constructed below, let $t > 0$, $r = \exp[-\operatorname{arcosh}(1 + 3/(\kappa_f t^2))]$, $E_I = (1-r)(3+r)/2$, and let the three physical groups have size m_{grp}. After normalizing the full execution-space radius to one, every legal linear-span method satisfies*

$$N_{\mathrm{comm}} \geq (m_{\mathrm{grp}} + 1)\left[\frac{1}{-\log r}\log\left(\frac{\sqrt{3/(3+t^2E_I)}}{\varepsilon}\right) - 1\right]_+. \tag{51}$$

This construction displays the cross-angle/network propagation mechanism, but the prefactor depends on t and is not asserted as a uniform $\Omega(\chi_\times\sqrt{\kappa_f\kappa_W}\log(1/\varepsilon))$ bound at fixed absolute accuracy.

I.2 Radius-explicit decentralized communication lower bound

We now prove Theorem I.3. The construction adapts the alternating-coordinate propagation argument of Yarmoshik et al. (2025, Appendix C), but the bad affine parameter is generated by the angle between a local cross-link channel and the shared-variable consensus channel. In particular, the local cross-link channel itself has condition number one.

Alternating partial isometries. Work first on the hard coordinate space $\ell_2(\mathbb{N})$. Define R_1, R_2 by

$$(R_1u)_1 = u_1, \qquad (R_1u)_{k+1} = \frac{u_{2k} - u_{2k+1}}{\sqrt{2}}, \quad k \geq 1, \tag{52}$$

$$(R_2u)_k = \frac{u_{2k-1} - u_{2k}}{\sqrt{2}}, \qquad k \geq 1. \tag{53}$$

The rows within each operator have disjoint supports and unit norm, hence

$$R_1R_1^\top = \mathrm{Id}, \qquad R_2R_2^\top = \mathrm{Id}. \tag{54}$$

Let $P_j = R_j^\top R_j$. For $k \geq 2$,

$$[(P_1 + P_2)u]_k = u_k - \frac{1}{2}u_{k-1} - \frac{1}{2}u_{k+1}, \tag{55}$$

while the first coordinate is $\frac{3}{2}u_1 - \frac{1}{2}u_2$. Thus P_1 and P_2 are complementary matching projectors whose union is a one-dimensional zero chain.

A legal cross-coupled instance. Let the physical graph be the path on $n = 3m_{\mathrm{grp}}$ agents and partition its consecutive vertices into V_1, V_2, V_3, each of size m_{grp}. For $i \in V_1$ introduce a local variable x_i and the affine relation

$$x_i = tR_1u, \tag{56}$$

while for $i \in V_3$ use $x_i = tR_2u$. Agents in V_2 have no hard local variable (equivalently, one may add a decoupled dummy scalar if strictly positive local dimensions are required). These relations are a

special case of the original cross-coupled equality: take the coupled output space to be the direct sum of one copy of the row space for every active agent, let J_i inject into the ith summand, and set

$$A_i = J_i, \qquad B_i = -tJ_iR_1 \quad (i \in V_1), \qquad A_i = J_i, \qquad B_i = -tJ_iR_2 \quad (i \in V_3), \tag{57}$$

with zero blocks on V_2. Because the ranges of the J_i are mutually orthogonal, the aggregate equality is equivalent to the local links (56) and their V_3 analogues. Hence this is a core CC-MAC instance with $C_i = D_i = 0$ and nonzero cross blocks B_i. For this subclass the aggregate output is already a direct sum of locally owned blocks, so the generic auxiliary localization variable used elsewhere in the paper is unnecessary; adding it would only enlarge the oracle available to the algorithm and is therefore omitted in the lower-bound representation.

After introducing local copies u_i of the shared variable, the first affine channel is

$$\mathcal{K}_1 : (x, \boldsymbol{u}) \mapsto \mathrm{col}\big(x_i - tR_{g(i)}u_i\big)_{i \in V_1 \cup V_3}, \qquad g(i) = \begin{cases} 1, & i \in V_1, \\ 2, & i \in V_3, \end{cases} \tag{58}$$

and the second channel is the consensus operator $\mathcal{K}_2 = (W \otimes \mathrm{Id})$ acting on $\boldsymbol{u}$. Equation (54) gives

$$\mathcal{K}_1\mathcal{K}_1^\top = (1+t^2)\mathrm{Id}, \qquad \kappa(\mathcal{K}_1) = 1. \tag{59}$$

Let $U = \mathrm{Range}(\mathcal{K}_1^\top)$ and let V be the row space of the consensus channel; for a connected graph, V is precisely the block-disagreement subspace in the u coordinates. Every a in the row-coefficient space of $\mathcal{K}_1$ satisfies

$$\|\mathcal{K}_1^\top a\|^2 = (1+t^2)\|a\|^2, \qquad \|P_V\mathcal{K}_1^\top a\| \le t\|a\|.$$

Equality is attained by choosing two agents in the same outer group with opposite row coefficients. Therefore

$$c_F(U,V) = \frac{t}{\sqrt{1+t^2}}, \qquad \chi_\times = \sqrt{\frac{1+c_F(U,V)}{1-c_F(U,V)}} = \sqrt{1+t^2} + t. \tag{60}$$

Separate left preconditioning of either channel does not change these row spaces, by Lemma F.1.

The hard objective and its geometric tail. If $t = 0$, then $\chi_\times = 1$ and the standard consensus hard block used below already proves the claimed order. Hence it remains to treat $t > 0$. On the cross block let the active agents use

$$f_i^\times(x_i, u) = \begin{cases} \frac{L_f}{2}\|x_i - \xi e_1\|^2 + \frac{\mu_f}{2}\|u\|^2, & i \in V_1, \\ \frac{L_f}{2}\|x_i\|^2 + \frac{\mu_f}{2}\|u\|^2, & i \in V_3, \end{cases} \tag{61}$$

and let agents in V_2 contribute only $\mu_f\|u\|^2/2$. These functions are L_f-smooth and μ_f-strongly convex in their existing variables. On the feasible set, after consensus and elimination of the x_i, the cross block equals, up to the positive factor m_{grp} and an irrelevant constant,

$$\phi_\times(u) = \frac{3\mu_f}{2}\|u\|^2 + \frac{L_f t^2}{2}\langle u, (P_1 + P_2)u\rangle - L_f t\xi u_1. \tag{62}$$

Set

$$\delta := \frac{3\mu_f}{L_f t^2} = \frac{3}{\kappa_f t^2}, \qquad \alpha := \mathrm{arcosh}(1+\delta), \qquad r := e^{-\alpha} \in (0,1). \tag{63}$$

For $c_r := \sqrt{1-r^2}$ choose

$$\xi = \frac{c_r}{L_f t}\left(3\mu_f + \frac{L_f t^2}{2}(3-r)\right). \tag{64}$$

Then the unique minimizer of (62) is

$$u^\star = c_r(1, r, r^2, \ldots), \qquad \|u^\star\| = 1. \tag{65}$$

Indeed, the first optimality equation is exactly enforced by (64); for every $k \ge 2$, (55) reduces the optimality condition to

$$(1+\delta)u_k - \frac{1}{2}u_{k-1} - \frac{1}{2}u_{k+1} = 0,$$

which holds because $(r + r^{-1})/2 = \cosh\alpha = 1 + \delta$. The linked local variables contribute to the full execution-space radius. Direct evaluation gives

$$E_I := \langle u^\star, (P_1 + P_2)u^\star \rangle = \frac{(1-r)(3+r)}{2}, \qquad R^2_{\text{full}} = m_{\text{grp}}(3 + t^2 E_I).$$

After scaling the complete execution vector to unit radius, any vector supported on the first s hard coordinates has

$$\|w - w^\star\| \geq \sqrt{\frac{3}{3 + t^2 E_I}}\, r^s. \tag{66}$$

Propagation through the physical path. We may only strengthen the algorithm when proving a lower bound. Thus, between communication rounds, give every node in V_1 free access to the projector P_1, every node in V_3 free access to P_2, and every node arbitrary local linear algebra and gradients. This oracle contains all hard-coordinate information obtainable from the original local maps $R_j, R_j^\top$ and the quadratic gradients: returning from the local x_i space to the shared coordinate can only produce $R_j^\top R_j = P_j$, together with the fixed seed e_1 in V_1.

Starting from the zero span (the standard resisting-oracle reduction handles arbitrary fixed initial vectors), let $s(N_{\text{round}})$ be the largest hard-coordinate index that can occur at any node after N_{round} neighbor communication rounds. The matching structure gives

$$P_1 \operatorname{span}\{e_1, \ldots, e_s\} \subseteq \begin{cases} \operatorname{span}\{e_1, \ldots, e_{s+1}\}, & s \text{ even}, \\ \operatorname{span}\{e_1, \ldots, e_s\}, & s \text{ odd}, \end{cases} \tag{67}$$

and the two cases are reversed for P_2. Hence creating coordinate 2 requires the seed from V_1 to reach V_3; creating coordinate 3 then requires the new information to return to V_1, and so on. We give the propagation induction explicitly.

For $s \geq 1$, let T_s be the first communication round after which some node can contain a vector with a nonzero sth hard coordinate. The linear term in (61) and one free local return through $R_1^\top$ give $T_1 = 0$ in V_1. Suppose coordinate s is first exposed. If s is odd, then the boundary pair $(s, s+1)$ belongs to the P_2 matching, whereas P_1 preserves $\operatorname{span}\{e_1, \ldots, e_s\}$; if s is even the roles are reversed. Thus coordinate $s+1$ can be created only in the outer group opposite to the one whose local matching created coordinate s. Before the information containing coordinate s reaches that opposite group, arbitrary local linear combinations, gradients, and applications of the available projector remain inside $\operatorname{span}\{e_1, \ldots, e_s\}$ by (67). The closest vertices of V_1 and V_3 are $m_{\text{grp}} + 1$ physical edges apart, and one communication round moves information across at most one edge. Consequently

$$T_{s+1} \geq T_s + (m_{\text{grp}} + 1).$$

Induction gives $T_s \geq (s-1)(m_{\text{grp}} + 1)$. Therefore, after N_{round} neighbor communication rounds, every locally available hard-coordinate vector is supported on the first

$$s(N_{\text{round}}) \leq 1 + \left\lfloor \frac{N_{\text{round}}}{m_{\text{grp}} + 1} \right\rfloor \tag{68}$$

coordinates. This is the same alternating-frontier mechanism as the propagation lemma of Yarmoshik et al. (2025, Appendix C), written here for the cross-link/consensus pair.

Combining (66) and (68), an ε-accurate output must satisfy

$$N_{\text{round}} \geq (m_{\text{grp}} + 1) \left[\frac{1}{-\log r} \log\left(\frac{\sqrt{3/(3 + t^2 E_I)}}{\varepsilon} \right) - 1 \right]_+ . \tag{69}$$

Since $-\log r = \operatorname{arcosh}(1 + \delta)$ and $\cosh x \geq 1 + x^2/2$,

$$-\log r \leq \sqrt{2\delta} = \sqrt{\frac{6}{\kappa_f t^2}}. \tag{70}$$

Thus the propagation rate contains the factor $m_{\text{grp}} t \sqrt{\kappa_f}$, but the logarithmic term retains the full-radius prefactor in (69); it is not uniform in t at fixed absolute accuracy. For the unweighted physical path,

$$\kappa_W = \frac{1 + \cos(\pi/n)}{1 - \cos(\pi/n)} = \cot^2 \frac{\pi}{2n}, \qquad m_{\text{grp}} = \Theta(\sqrt{\kappa_W}). \tag{71}$$

Hence the path contributes $m_{\mathrm{grp}} = \Theta(\sqrt{\kappa_W})$ to the radius-explicit lower bound in (69).

To cover the regime $t < 1$ without changing the prescribed angle, append an independent standard consensus hard block to the shared variable and to each local objective, as in Scaman et al. (2017). The cross-link channel acts only on the first block, while the consensus channel acts on both, so (60) is unchanged. The direct-sum objective remains L_f-smooth and μ_f-strongly convex, and the standard block requires

$$\Omega\left(\sqrt{\kappa_f \kappa_W}\log\frac{1}{\varepsilon}\right) \tag{72}$$

communications. Every accurate output must solve both blocks. The relation

$$\chi_\times = \sqrt{1+t^2} + t \le (1+\sqrt{2})\max\{1,t\}, \tag{73}$$

shows how t controls the cross factor, but it does not remove the t-dependent full-radius prefactor. Equation (69) is precisely (51) and proves Theorem I.3. Finite-dimensional instances are obtained by truncating the hard coordinate after more coordinates than any method can expose before the claimed lower-bound horizon, exactly as in standard zero-chain lower bounds. □

What these lower bounds establish. The affine construction and the ordinary unconstrained subclass establish the two lower orders recorded in Theorem 4.2. The communication family establishes alternating physical propagation together with the radius-explicit dependence in Theorem I.3. A uniform product lower bound over a class controlling only the primal radius would additionally require a hard family whose complete execution-space radius remains uniformly controlled as t varies; that stronger conclusion is not used here.

J Smooth-Convex Log-Free Continuation and Matching Lower Bounds

Accuracy class and multiplier scale. Assume F is convex with L_f-Lipschitz gradient, $\|w^0 - w^\star\| \le R$, and choose the minimum-norm multiplier $\lambda^\star \in \mathrm{Range}(\bar{K})$. Define

$$\Lambda_{\mathrm{cvx}} := \frac{\sigma^+_{\min}(\bar{K})\|\lambda^\star\|}{L_f R}, \qquad F(\bar{w}) - F^\star \le \varepsilon, \quad \frac{L_f R}{\sigma^+_{\min}(\bar{K})}\|\bar{K}\bar{w} - \bar{v}\| \le \varepsilon. \tag{74}$$

When $\bar{K} = 0$ the feasibility condition is omitted and $\Lambda_{\mathrm{cvx}} = 0$. The class fixes L_f, R, the geometric budgets $\chi_\times \le \Xi$, $\kappa_W \le \Omega$, and $\Lambda_{\mathrm{cvx}} \le \bar{\Lambda}$. Set $T_{\mathrm{cvx}} = \sqrt{L_f R^2/\varepsilon}$.

The continuation construction below avoids a multiplicative target-accuracy logarithm in the affine and communication ledgers. Throughout, $\bar{K}$ denotes the fixed separately normalized joint equality operator, $\mathcal{A} := \{w : \bar{K}w = \bar{v}\} \ne \varnothing$, $\mathcal{N} := \mathrm{Range}(\bar{K}^\top)$, $Q := \bar{K}^\top\bar{K}$, and

$$0 < \ell \le \lambda^+_{\min}(Q), \qquad \lambda_{\max}(Q) \le \lambda_{\mathrm{hi}}, \qquad \widehat{\chi} := \sqrt{\lambda_{\mathrm{hi}}/\ell}. \tag{75}$$

The supplied normalization certificates of the main text give $\widehat{\chi} = O(\chi_\times)$ instancewise and $\widehat{\chi} = O(\Xi)$ on the class $\chi_\times \le \Xi$. Let $L := L_f$, $\sigma := \sigma^+_{\min}(\bar{K})$, and assume $0 < \varepsilon \le LR^2$ and $\Lambda_{\mathrm{cvx}} \le \bar{\Lambda}$.

J.1 Executable filters and one regularized stage

Let T_j, U_j be Chebyshev polynomials of the first and second kind and put $d_{\mathrm{RF}} = d_{\mathrm{UF}} := \lceil\widehat{\chi}\rceil$. For $\lambda_{\mathrm{hi}} > \ell$ define

$$r_{d_{\mathrm{RF}}}(t) := \frac{T_{d_{\mathrm{RF}}}((\lambda_{\mathrm{hi}} + \ell - 2t)/(\lambda_{\mathrm{hi}} - \ell))}{T_{d_{\mathrm{RF}}}((\lambda_{\mathrm{hi}} + \ell)/(\lambda_{\mathrm{hi}} - \ell))}, \qquad s_{d_{\mathrm{UF}}}(t) := \frac{U_{d_{\mathrm{UF}}}(1 - t/\lambda_{\mathrm{hi}})}{d_{\mathrm{UF}} + 1}, \tag{76}$$

and for $\lambda_{\mathrm{hi}} = \ell$ use $r_1(t) = 1 - t/\lambda_{\mathrm{hi}}$. Standard Chebyshev identities give

$$r_{d_{\mathrm{RF}}}(0) = s_{d_{\mathrm{UF}}}(0) = 1, \quad |r_{d_{\mathrm{RF}}}(t)| < \tfrac{1}{2}\ (t \in [\ell, \lambda_{\mathrm{hi}}]), \quad |s_{d_{\mathrm{UF}}}(t)| \le 1, \quad \sqrt{t}\,|s_{d_{\mathrm{UF}}}(t)| \le \sqrt{\ell}. \tag{77}$$

Indeed, $\operatorname{arcosh}((\lambda_{\mathrm{hi}} + \ell)/(\lambda_{\mathrm{hi}} - \ell)) \ge 2/\widehat{\chi}$, so the denominator defining $r_{d_{\mathrm{RF}}}$ is at least $\cosh 2$; for $s_{d_{\mathrm{UF}}}$, write $\cos\theta = 1 - t/\lambda_{\mathrm{hi}}$ and use $U_{d_{\mathrm{UF}}}(\cos\theta) = \sin((d_{\mathrm{UF}} + 1)\theta)/\sin\theta$. Consequently, with

$$P := \mathrm{Id} - r_{d_{\mathrm{RF}}}(Q), \qquad \tfrac{1}{2}\Pi_{\mathcal{N}} \preceq P \preceq \tfrac{3}{2}\Pi_{\mathcal{N}}, \qquad \mathrm{Ker}(P) = \mathrm{Ker}(\bar{K}). \tag{78}$$

Define the residual $\mathcal{R}_{\bar{v}}(x) := \bar{K}^\top(\bar{K}x - \bar{v})$.

Algorithm 3 First-kind residual filter RF

Require: $x, \bar{K}, \bar{v}, \ell, \lambda_{\mathrm{hi}}, d_{\mathrm{RF}} = \lceil \sqrt{\lambda_{\mathrm{hi}}/\ell} \rceil$
if $\lambda_{\mathrm{hi}} = \ell$ **then**
  return $x - \mathcal{R}_{\bar{v}}(x)/\lambda_{\mathrm{hi}}$
end if
$c \leftarrow (\lambda_{\mathrm{hi}} + \ell)/2$, $h \leftarrow (\lambda_{\mathrm{hi}} - \ell)/2$, $a \leftarrow c/h$
$x_- \leftarrow x$, $x_c \leftarrow x - \mathcal{R}_{\bar{v}}(x)/c$, $\rho \leftarrow 1/a$
for $j = 1, \ldots, d_{\mathrm{RF}} - 1$ **do**
  $\rho_+ \leftarrow (2a - \rho)^{-1}$
  $x_+ \leftarrow x_c + \rho_+\rho(x_c - x_-) - (2\rho_+/h)\mathcal{R}_{\bar{v}}(x_c)$
  $(x_-, x_c, \rho) \leftarrow (x_c, x_+, \rho_+)$
end for
return x_c

Algorithm 4 Second-kind weighted filter UF

Require: $x, \bar{K}, \bar{v}, \lambda_{\mathrm{hi}}$, integer $d_{\mathrm{UF}} \geq 1$
$x_- \leftarrow x$, $x_c \leftarrow x - \mathcal{R}_{\bar{v}}(x)/\lambda_{\mathrm{hi}}$
for $j = 1, \ldots, d_{\mathrm{UF}} - 1$ **do**
  $x_+ \leftarrow \dfrac{2(j+1)}{j+2}(x_c - \mathcal{R}_{\bar{v}}(x_c)/\lambda_{\mathrm{hi}}) - \dfrac{j}{j+2}x_-$
  $(x_-, x_c) \leftarrow (x_c, x_+)$
end for
return x_c

For every analytical feasible reference $w_{\mathrm{ref}} \in \mathcal{A}$, Algorithms 3–4 return

$$\mathrm{RF}(x) = w_{\mathrm{ref}} + r_{d_{\mathrm{RF}}}(Q)(x - w_{\mathrm{ref}}), \qquad \mathrm{UF}(x) = w_{\mathrm{ref}} + s_{d_{\mathrm{UF}}}(Q)(x - w_{\mathrm{ref}}), \tag{79}$$

although w_{ref} is never supplied or computed. This follows by dividing the three-term recurrences for T_j and U_j by their normalizing factors. Each filter evaluation uses its polynomial degree many residual evaluations; each residual evaluation is one forward and one adjoint affine call.

Define, only for analysis,

$$\begin{aligned} \mathcal{H}(w) &:= F(w) + \frac{L}{2}\langle w - w_{\mathrm{ref}}, P(w - w_{\mathrm{ref}})\rangle, \\ \phi_\nu(w) &:= \mathcal{H}(w) + \frac{\nu}{2}\|w - w^0\|^2, \qquad 0 < \nu \leq L. \end{aligned} \tag{80}$$

The augmentation and its gradient vanish on $\mathcal{A}$, while its executable gradient is

$$\nabla\phi_\nu(w) = \nabla F(w) + L\big(w - \mathrm{RF}(w)\big) + \nu(w - w^0). \tag{81}$$

Let a_ν minimize ϕ_ν on $\mathcal{A}$, set $p_\nu = -\nabla\phi_\nu(a_\nu) \in \mathcal{N}$, and let $\bar{K}^\top\lambda_\nu = p_\nu$ with minimum-norm $\lambda_\nu \in \mathrm{Range}(\bar{K})$. Comparison with $w^\star$ and Lipschitzness of ∇F give

$$\|a_\nu - w^0\| \leq R, \quad \|a_\nu - w^\star\| \leq 2R, \quad F(a_\nu) - F^\star \leq \frac{\nu}{2}R^2, \quad \|\lambda_\nu\| \leq (\bar{\Lambda} + 3)LR/\sigma. \tag{82}$$

Set $\tau = \frac{1}{2}\sqrt{\nu/L}$, $\eta = (12\sqrt{L\nu})^{-1}$, and $\theta = 4\sqrt{L\nu}$.

Algorithm 5 One augmented Cross-APAPC continuation step

Require: $x, f, p, w^0, F, \bar{K}, \bar{v}, L, \nu, \ell, \lambda_{\mathrm{hi}}, d_{\mathrm{RF}}$, with $p \in \mathcal{N}$
1: $g \leftarrow \tau x + (1-\tau) f$
2: $c_g \leftarrow g - \mathrm{RF}(g; \bar{K}, \bar{v}, \ell, \lambda_{\mathrm{hi}}, d_{\mathrm{RF}})$
3: $g_\phi \leftarrow \nabla F(g) + L c_g + \nu(g - w^0)$
4: $x_{1/2} \leftarrow \big(x - \eta(g_\phi - \nu g + p)\big)/(1+\eta\nu)$
5: $c_{1/2} \leftarrow x_{1/2} - \mathrm{RF}(x_{1/2}; \bar{K}, \bar{v}, \ell, \lambda_{\mathrm{hi}}, d_{\mathrm{RF}})$
6: $\Delta p \leftarrow \theta c_{1/2}$, $p^+ \leftarrow p + \Delta p$
7: $x^+ \leftarrow x_{1/2} - \eta \Delta p/(1+\eta\nu)$
8: $f^+ \leftarrow g + \dfrac{2\tau}{2-\tau}(x^+ - x)$
9: **return** (x^+, f^+, p^+)

For the analytical energy

$$E_\nu(x,f,p) := 12\|x - a_\nu\|^2 + \frac{1}{12L\nu}\left\langle p - p_\nu, \left(3P^\dagger - \frac{\mathrm{Id}}{1+\eta\nu}\right)(p - p_\nu)\right\rangle + \frac{4(1-\tau)}{\nu} D_{\phi_\nu}(f, a_\nu), \tag{83}$$

one has

$$E_\nu(x^+, f^+, p^+) \le \left(1 + \frac{1}{24}\sqrt{\nu/L}\right)^{-1} E_\nu(x,f,p). \tag{84}$$

To verify this, take the proof-only equality operator $A_P := P^{1/2}$ and write $p = A_P^\top y$, $p_\nu = A_P^\top y_\nu$. Algorithm 5 is exactly the APAPC step of Salim et al. (2022) applied to ϕ_ν with valid certificates $L_{\mathrm{in}} = 6L$, $\mu_{\mathrm{in}} = \nu$, $\lambda_1 = 3$, $\lambda_2 = 1/2$. On $\mathcal{N}$, $\mathrm{Id} \preceq 3P^\dagger - (1+\eta\nu)^{-1}\mathrm{Id} \preceq 6\mathrm{Id}$, so their metric energy is precisely (83) up to the common scalar $\sqrt{L\nu}$. Their one-step inequality is valid for arbitrary $y - y_\nu \in \mathrm{Range}(A_P)$; no zero-dual initialization is used. Hence the currently retained dual state is admissible, which is the point needed by the continuation argument.

J.2 Warm starts, complete algorithm, and exact ledger

Let $g_0 = \nabla F(w^0)$ and initialize

$$p^0 := \mathrm{UF}(g_0; \bar{K}, 0, \lambda_{\mathrm{hi}}, d_{\mathrm{UF}}) - g_0. \tag{85}$$

The weighted bound in (77), applied to $-\bar{K}^\top \lambda^\star$, yields

$$E_L(w^0, w^0, p^0) \le \left[19 + \frac{(\bar{\Lambda}+5)^2}{2}\right] R^2. \tag{86}$$

Moreover, if $E_\nu(x,f,p) \le R^2$ and $\nu' = \nu/2$, then retaining p and resetting $x \leftarrow f$ gives

$$E_{\nu'}(f,f,p) \le C_{\mathrm{ws}} R^2 < 178R^2 < 200R^2. \tag{87}$$

For completeness, the estimates behind (87) are

$$\|p - p_\nu\| \le \sqrt{12L\nu}R, \quad D_{\phi_\nu}(f, a_\nu) \le \nu R^2/2, \quad \|f - \Pi_{\mathcal{A}} f\| \le \sqrt{2\nu/L}R, \tag{88}$$

and, for $a = a_\nu$, $a' = a_{\nu/2}$,

$$\|\nabla F(a) - \nabla F(a')\| \le \sqrt{L\nu}R, \qquad \|p_\nu - p_{\nu/2}\| \le \tfrac{5}{2}\sqrt{L\nu}R. \tag{89}$$

The first line follows from (83), strong convexity, and (78). Pairing the two regularized KKT systems with $a - a' \in \mathrm{Ker}(\bar{K})$ and using cocoercivity proves (89). Expanding the two Bregman divergences gives the exact transfer identity

$$D_{\phi_{\nu/2}}(f,a') - D_{\phi_\nu}(f,a) = -\frac{\nu}{4}\|f - w^0\|^2 + \phi_\nu(a) - \phi_\nu(a') + \frac{\nu}{4}\|a' - w^0\|^2 + \langle p_{\nu/2} - p_\nu, f - \Pi_{\mathcal{A}} f\rangle. \tag{90}$$

which with (88)–(89) gives $C_{\mathrm{ws}} = 108 + (\sqrt{12} + 5/2)^2 + 8(3/4 + (5/2)\sqrt{2}) = 177.854779\ldots$.

Set

$$B_{\mathrm{cont}} := \max\{200, 20 + (\bar{\Lambda} + 6)^2\}, \qquad S_{\mathrm{cont}} := \left\lceil \log_2 \frac{64LR^2}{\varepsilon} \right\rceil, \qquad \nu_s := L2^{-s},$$
$$k_s := \left\lceil \frac{\log B_{\mathrm{cont}}}{\log(1 + 2^{-s/2}/24)} \right\rceil, \qquad k := \max\left\{0, \left\lceil \log_2 \frac{16\sqrt{2}(\bar{\Lambda} + 6)LR^2}{\varepsilon} \right\rceil\right\}. \tag{91}$$

Algorithm 6 Log-free smooth-convex continuation with affine-only output repair

Require: $F, w^0, L, R, \varepsilon, \bar{\Lambda}$; $\bar{K}, \bar{v}$ and, if $\bar{K} \neq 0$, certificates $\ell, \lambda_{\mathrm{hi}}$
if $\bar{K}$ is absent **then**
  return Algorithm 7$(F, w^0, L, R, \varepsilon)$
end if
$d_{\mathrm{RF}} \leftarrow \lceil \sqrt{\lambda_{\mathrm{hi}}/\ell} \rceil$, $d_{\mathrm{UF}} \leftarrow d_{\mathrm{RF}}$; set $B_{\mathrm{cont}}, S_{\mathrm{cont}}, \nu_s, k_s, k$ by (91)
$g_0 \leftarrow \nabla F(w^0)$; $p \leftarrow \mathrm{UF}(g_0; \bar{K}, 0, \lambda_{\mathrm{hi}}, d_{\mathrm{UF}}) - g_0$; $x \leftarrow w^0$; $f \leftarrow w^0$
for $s = 0, \ldots, S_{\mathrm{cont}}$ **do**
  if $s > 0$ **then**
    $x \leftarrow f$ ▷ retain f and the current dual p
  end if
  for $j = 1, \ldots, k_s$ **do**
    $(x, f, p) \leftarrow$ Algorithm 5$(x, f, p, w^0, F, \bar{K}, \bar{v}, L, \nu_s, \ell, \lambda_{\mathrm{hi}}, d_{\mathrm{RF}})$
  end for
end for
$y \leftarrow f$
for $j = 1, \ldots, k$ **do**
  $y \leftarrow \mathrm{RF}(y; \bar{K}, \bar{v}, \ell, \lambda_{\mathrm{hi}}, d_{\mathrm{RF}})$
end for
$\bar{w} \leftarrow \mathrm{UF}(y; \bar{K}, \bar{v}, \lambda_{\mathrm{hi}}, d_{\mathrm{UF}})$
return $\bar{w}$

Algorithm 7 Accelerated-gradient branch when the affine system is absent

Require: $F, w^0, L, R, \varepsilon$
$x \leftarrow w^0$, $y \leftarrow w^0$, $t \leftarrow 1$, $N_{\mathrm{AGD}} \leftarrow \lceil \sqrt{2LR^2/\varepsilon} \rceil$
for $j = 1, \ldots, N_{\mathrm{AGD}}$ **do**
  $x_+ \leftarrow y - L^{-1}\nabla F(y)$; $t_+ \leftarrow (1 + \sqrt{1 + 4t^2})/2$
  $y \leftarrow x_+ + (t - 1)(x_+ - x)/t_+$; $(x, t) \leftarrow (x_+, t_+)$
end for
return x

Induction using (84) and (87) gives $E_{\nu_s}(x, f, p) \le R^2$ at every completed stage and $\varepsilon/(128R^2) < \nu_{S_{\mathrm{cont}}} \le \varepsilon/(64R^2)$. At the last stage write $z = \Pi_{\mathcal{A}} f$ and $n = f - z$. Then

$$F(z) - F^\star < 8\nu_{S_{\mathrm{cont}}} R^2 \le \varepsilon/8, \qquad \|n\| \le \sqrt{2\nu_{S_{\mathrm{cont}}}/L}\, R. \tag{92}$$

The executed return is

$$\bar{w} = z + e, \qquad e = s_{d_{\mathrm{UF}}}(Q) r_{d_{\mathrm{RF}}}(Q)^k n \in \mathcal{N}, \qquad \|e\| \le \frac{\|\bar{K}e\|}{\sigma} \le \frac{\varepsilon}{16(\bar{\Lambda} + 6)LR}. \tag{93}$$

The key step is the weighted second-kind estimate $\|\bar{K} s_{d_{\mathrm{UF}}}(Q) q\| \le \sqrt{\ell}\|q\|$, which avoids multiplying the repair by $\|\bar{K}\|/\sigma$. Using (82), smoothness, and the regularized KKT relation,

$$F(\bar{w}) - F(z) \le \|\lambda_{\nu_{S_{\mathrm{cont}}}}\|\, \|\bar{K}e\| + (\nu_{S_{\mathrm{cont}}} R + L\|z - a_{\nu_{S_{\mathrm{cont}}}}\|)\|e\| + \tfrac{L}{2}\|e\|^2, \tag{94}$$

so (92)–(93) imply exactly (74).

Let $A_{\mathrm{tot}} := \sum_{s=0}^{S_{\mathrm{cont}}} k_s$. Since $\log(1 + a) \ge a/(1 + a)$,

$$A_{\mathrm{tot}} < \frac{\sqrt{128}(1 + 25\log B_{\mathrm{cont}})}{1 - 1/\sqrt{2}}\, T_{\mathrm{cvx}}, \qquad N_{\nabla f} = 1 + A_{\mathrm{tot}}, \qquad N_{\mathrm{aff}} = d_{\mathrm{RF}}(4A_{\mathrm{tot}} + 2k + 4). \tag{95}$$

Here $d_{\mathrm{RF}} \le 2\widehat{\chi}$ and $k \le 1 + \log_2(16\sqrt{2}(\bar{\Lambda}+6)) + 2\log_2 T_{\mathrm{cvx}}$. Hence the final repair contributes only the additive term $O_{\bar{\Lambda}}(\widehat{\chi}\log T_{\mathrm{cvx}})$, dominated by $O_{\bar{\Lambda}}(\widehat{\chi}T_{\mathrm{cvx}})$. Thus

$$N_{\nabla f} = O_{\bar{\Lambda}}(T_{\mathrm{cvx}}), \qquad N_{\mathrm{aff}} = O_{\bar{\Lambda}}(\widehat{\chi}T_{\mathrm{cvx}}), \qquad N_{\mathrm{comm}} = O_{\bar{\Lambda}}(\widehat{\chi}\sqrt{\kappa_W}\,T_{\mathrm{cvx}}) \tag{96}$$

under the supplied normalized-channel realization. Thus the affine and communication bounds have no multiplicative target-accuracy logarithm; no raw $A/B/C/D/W$ theorem is inferred beyond the primitive conversion already stated in Appendix L.

J.3 Radius-uniform lower bounds

We give a finite construction that attains the matching affine and communication orders while controlling the *full execution radius* uniformly in the cross-angle parameter.

For even N_c, let R_1, R_2 be the alternating normalized edge matchings

$$(R_1u)_j = (u_{2j} - u_{2j+1})/\sqrt{2}, \qquad (R_2u)_j = (u_{2j-1} - u_{2j})/\sqrt{2}, \tag{97}$$

with the natural index ranges. Writing $P_j = R_j^\top R_j$ gives $P_1 + P_2 = \frac{1}{2}L_{\mathrm{path},N_c} =: P_{\mathrm{path}}$. Let $c, \mu > 0$, $\delta = 3\mu/c$, $\alpha = \operatorname{arcosh}(1+\delta) \le 1/4$, $r = e^{-\alpha}$,

$$N_c = 2\lceil 10/\alpha \rceil, \quad \rho_c = r^{2N_c}, \quad a = 1 - r, \quad v_j = r^{j-1} + r^{2N_c - j}, \quad Z_c = (\sum_j v_j^2)^{-1/2}. \tag{98}$$

Then $u^\star = Z_c v$ is the unique minimizer of

$$\frac{3\mu}{2}\|u\|^2 + \frac{c}{2}\langle u, P_{\mathrm{path}}u\rangle - \beta u_1, \qquad \beta = \frac{cZ_c a(1-\rho_c)}{2r}, \tag{99}$$

and $\|u^\star\| = 1$. The reflected term makes the last-coordinate KKT equation exact: $v_{N_c+1} = v_{N_c}$; the first-coordinate equation is $(\delta + 1/2)v_1 - v_2/2 = a(1-\rho_c)/(2r)$. Hence there is no infinite-chain truncation residue. Moreover

$$E_c := \langle u^\star, P_{\mathrm{path}}u^\star\rangle \le \frac{a^2}{2(1-\rho_c)}, \qquad \sum_{j>s}(u_j^\star)^2 \ge 0.998/e, \quad s = \lfloor (2\alpha)^{-1} \rfloor. \tag{100}$$

The energy bound follows from $0 \le v_j - v_{j+1} \le ar^{j-1}$; the tail bound follows from the exact normalization $\|v\|^2 = (1-r^{4N_c})/(1-r^2) + 2N_c r^{2N_c - 1}$ and $20 \le \alpha N_c < 20.5$.

For communication, use $3M$ agents in three consecutive groups, put $t \ge 1$, $\kappa = L/\mu \ge 100$, and take (99) with $c = Lt^2$. The outer groups have constraints $x_i = tR_1u_i$ and $x_i = tR_2u_i$, respectively, while all u_i are in consensus; the objective is

$$f_i = \begin{cases} \frac{L}{2}\|x_i\|^2 + \frac{\mu}{2}\|u_i\|^2 - \beta u_{i,1}, & i \in V_1, \\ \frac{\mu}{2}\|u_i\|^2, & i \in V_2, \\ \frac{L}{2}\|x_i\|^2 + \frac{\mu}{2}\|u_i\|^2, & i \in V_3. \end{cases} \tag{101}$$

Its unique feasible optimizer has $u_i = u^\star$ and $x_i = tR_ju^\star$. Crucially, its *full* squared radius from zero is

$$R_0^2 = M(3 + t^2E_c), \qquad 3M \le R_0^2 \le M(3 + 4/\kappa), \tag{102}$$

because $t^2E_c \le 4/\kappa$. Thus the hard shared tail remains a fixed fraction of the full radius even as the channels become nearly parallel.

Normalize the link channel by $h = \sqrt{1+t^2}$ and the consensus channel by the disagreement projector. Then their row spaces are disjoint and

$$c_F = t/h, \qquad \chi_\times = h + t, \qquad \sigma^2 = 1 - t/h = 1/[h(h+t)]. \tag{103}$$

A valid KKT multiplier is explicit. For the normalized link blocks take $\lambda_{1,i} = -Lhx_i^\star$ and for the consensus block take

$$d_i = \begin{cases} \beta e_1 - \mu u^\star - Lt^2P_1u^\star, & i \in V_1, \\ -\mu u^\star, & i \in V_2, \\ -\mu u^\star - Lt^2P_2u^\star, & i \in V_3. \end{cases}$$

The finite-chain stationarity equation gives $\sum_i d_i = 0$, while the link stationarity equations are $Lx_i^\star + \lambda_{1,i}/h = 0$. Hence (λ_1, d) is a genuine multiplier. Using $t\sqrt{E_c} \le 2/\sqrt{\kappa}$, $ta \le \sqrt{6/\kappa}$, and $R_0 \ge \sqrt{3M}$ yields

$$\frac{\sigma\|\lambda^\star\|}{LR_0} \le \frac{7/\sqrt{\kappa} + 2/\kappa}{\sqrt{3}} < 1. \tag{104}$$

Only V_1 initially contains the seed e_1; local closure in V_1 uses P_1 and in V_3 uses P_2. Since the two matchings alternately expose successive coordinates, coordinate j cannot appear before $(j-1)\Delta$ neighbor steps when the outer-group distance is Δ. Combining this support invariant with (100)–(102) gives $\Omega(Mt\sqrt{\kappa})$ communication rounds on the path. Here $\kappa_W = \cot^2(\pi/(6M)) = \Theta(M^2)$. For bounded network budget the same construction on the complete six-node graph has $\kappa_W = 1$ and gives $\Omega(t\sqrt{\kappa})$. For bounded angle use only consensus equalities and the same reflected chain with $c = L - \mu$ and local objectives

$$f_i(u_i) = \begin{cases} \frac{\mu}{2}\|u_i\|^2 + \frac{L-\mu}{2}\langle u_i, P_1 u_i\rangle - \beta u_{i,1}, & i \in V_1, \\ \frac{\mu}{2}\|u_i\|^2, & i \in V_2, \\ \frac{\mu}{2}\|u_i\|^2 + \frac{L-\mu}{2}\langle u_i, P_2 u_i\rangle, & i \in V_3. \end{cases}$$

Then $R_0^2 = 3M$, $\chi_\times = 1$, and the minimum-norm consensus multiplier obeys

$$\frac{\sigma\|\lambda^\star\|}{LR_0} \le \frac{2\sqrt{6/(\kappa-1)} + \sqrt{3}/\kappa}{\sqrt{3}} < 1 \qquad (\kappa \ge 100),$$

because the stacked gradient at the feasible optimizer is normal to consensus and its seed, scalar, and matching components satisfy the displayed bound. The identical alternating-exposure argument gives $\Omega(M\sqrt{\kappa})$ rounds on the path and $\Omega(\sqrt{\kappa})$ on the complete six-node graph.

For the affine resource, use one physical agent and a virtual path with $N_{\mathrm{virt}} = 3M$ blocks. Split its incidence rows into odd and even matchings and set

$$\widehat{K}_1 = 2^{-1/2} E_{\mathrm{odd}} \otimes I_{N_c}, \qquad \widehat{K}_2 = 2^{-1/2} E_{\mathrm{even}} \otimes I_{N_c}. \tag{105}$$

Let $\bar{K} := \mathrm{col}(\widehat{K}_1, \widehat{K}_2)$. Both channels are individually normalized, while

$$\chi_\times = \cot\frac{\pi}{2N_{\mathrm{virt}}} = \Theta(N_{\mathrm{virt}}), \qquad \kappa_W = 1. \tag{106}$$

A $\bar{K}$ or $\bar{K}^\top$ action crosses at most one virtual edge, so the same alternating-frontier argument and (100) require $\Omega(N_{\mathrm{virt}}\sqrt{\kappa})$ affine calls even if all gradient calls are free. A six-block one-channel member handles bounded Ξ and still requires $\Omega(\sqrt{\kappa})$ affine calls.

Finally rescale amplitudes so the full initial radius is exactly R and set

$$\mu := 24\varepsilon/R^2, \qquad \kappa = L/\mu = T_{\mathrm{cvx}}^2/24. \tag{107}$$

All row spaces, support frontiers, and the scale $\sigma\|\lambda^\star\|/(LR)$ are invariant under this amplitude rescaling. Whenever (74) holds and the multiplier scale is at most one, strong convexity and KKT give

$$F(w) - F^\star \ge -\varepsilon + \frac{\mu}{2}\|w - w^\star\|^2, \tag{108}$$

so an accurate point must lie within $R/2$ of the optimizer. For $T_{\mathrm{cvx}} \ge 50$, $\kappa \ge 100$, and the preceding fixed-contraction obstructions therefore yield

$$\mathfrak{C}_{\mathrm{aff}}^{\mathrm{cvx}}(\varepsilon) = \Omega(\Xi T_{\mathrm{cvx}}), \qquad \mathfrak{C}_{\mathrm{comm}}^{\mathrm{cvx}}(\varepsilon) = \Omega(\Xi\sqrt{\Omega}\, T_{\mathrm{cvx}}). \tag{109}$$

For $\Xi \ge 3$ take $t = (\Xi - \Xi^{-1})/2$; for $\Omega \ge 16$ take $M = \lfloor\sqrt{\Omega}/2\rfloor$. The bounded-budget constructions above lose only universal constants. The separate gradient lower bound is the standard smooth convex zero-chain with all affine constraints absent, giving $\Omega(T_{\mathrm{cvx}})$. Together with (96), this proves Theorem 5.1 for every fixed $\bar{\Lambda} \ge 1$ and threshold $T_{\mathrm{cvx}} \ge 50$.

Scope and relation to prior algorithms. The continuation wrapper is not claimed as the first smooth-convex APAPC acceleration. Condat et al. (2026), Theorem 3.11 and Remark 3.12, already give smooth-convex primal–dual acceleration under a linear equality and explicitly discuss quadratic equality augmentation. The contribution here is the executable CC-MAC wrapper: fixed normal augmentation, weighted range-valued initialization, retained-dual warm starts, the weighted final repair, and the radius-uniform resource-separated lower constructions that together yield the exact class-level orders above. The result assumes fixed supplied spectral certificates and exact arithmetic; certificate acquisition and arbitrary time-varying inexact products are outside the theorem.

K Nonsmooth Extension: Cross-Sliding Upper Bounds

Assume F is finite and convex on the ambient Euclidean space, the affine system is feasible, $\|w^0 - w^\star\| \le R$, and

$$M_{3R} := \sup_{w \in B(w^\star, 3R)} \sup_{g \in \partial F(w)} \|g\| < \infty. \tag{110}$$

For nonzero $\bar{K}$, with $\sigma = \sigma^+_{\min}(\bar{K})$, the target criterion is

$$F(\bar{w}) - F^\star \le \varepsilon, \qquad (M_{3R}/\sigma)\|\bar{K}\bar{w} - \bar{v}\| \le \varepsilon. \tag{111}$$

If $\bar{K} = 0$ the second inequality is omitted. Set $T_{\rm ns} = M_{3R}R/\varepsilon$ and $T_{\rm nsc} = M_{3R}/\sqrt{\mu_f \varepsilon}$.

We prove Theorem 5.2. Throughout this appendix, $\bar{K}$ denotes the separately normalized joint affine operator, $\sigma := \sigma^+_{\min}(\bar{K})$, and $L_{\bar{K}} := \|\bar{K}\|$. The equality is scale-invariant: replacing $(\bar{K}, \bar{v})$ by $(a\bar{K}, a\bar{v})$ for $a > 0$ leaves the feasible set, $\kappa(\bar{K})$, and the criterion (111) unchanged.

K.1 Dual radius and a linear-operator sliding primitive

Lemma K.1 (Minimum-norm multiplier bound under local regularity)**.** *Let $F : \mathbb{R}^d \to \mathbb{R}$ be finite and convex, let $w^\star$ solve $\min\{F(w) : \bar{K}w = \bar{v}\}$, and choose a KKT multiplier $\lambda^\star \in \mathrm{Range}(\bar{K})$ of minimum norm. Under* (110)*,*

$$\|\lambda^\star\| \le \frac{M_{3R}}{\sigma}. \tag{112}$$

Proof. Because F is finite on the ambient space, the standard affine-constraint qualification gives $0 \in \partial F(w^\star) + \bar{K}^\top \lambda^\star$; redundant multiplier components may be projected away, so $\lambda^\star \in \mathrm{Range}(\bar{K})$. Hence for some $g^\star \in \partial F(w^\star)$, $\bar{K}^\top \lambda^\star = -g^\star$. The restriction of $\bar{K}^\top$ to $\mathrm{Range}(\bar{K})$ has smallest singular value σ, and therefore $\sigma\|\lambda^\star\| \le \|\bar{K}^\top \lambda^\star\| = \|g^\star\| \le M_{3R}$. □

Set $R_\lambda := 2M_{3R}/\sigma$ and $\mathcal{Y} := \{\lambda : \|\lambda\| \le R_\lambda\}$. Define

$$\mathrm{Gap}(w) := F(w) - F^\star + \max_{\lambda \in \mathcal{Y}} \langle \lambda, \bar{K}w - \bar{v}\rangle = F(w) - F^\star + R_\lambda \|\bar{K}w - \bar{v}\|. \tag{113}$$

The KKT inequality $F(w) - F^\star \ge -\|\lambda^\star\|\|\bar{K}w - \bar{v}\|$ gives

$$\mathrm{Gap}(w) \ge F(w) - F^\star, \qquad \mathrm{Gap}(w) \ge \frac{R_\lambda}{2}\|\bar{K}w - \bar{v}\|. \tag{114}$$

Thus $\mathrm{Gap}(w) \le \varepsilon$ implies (111) up to a universal factor two.

We use the bilinear recursive-sliding theorem of Nguyen and Gasnikov (2026). Suppose a phase center w^0 satisfies $\|w^0 - w^\star\| \le R_s \le R$ and restrict the phase to the Euclidean ball $\mathcal{X}_s := B(w^0, 2R_s)$. Every queried primal point then lies in $B(w^\star, 3R)$, so (110) controls every subgradient used in the phase. Rescale

$$\bar{K}_a := a_s \bar{K}, \qquad \bar{v}_a := a_s \bar{v}, \qquad a_s := \frac{R_\lambda}{R_s}, \tag{115}$$

and write $\lambda = a_s y$. The saddle problem is

$$\min_{w \in \mathcal{X}_s} \max_{\|y\| \le R_s} F(w) + \langle y, \bar{K}_a w - \bar{v}_a\rangle. \tag{116}$$

Since $\|\lambda^\star/a_s\| \le R_s/2$, the restricted dual set contains an optimal multiplier. In the notation of Nguyen and Gasnikov (2026, Section 4, degenerate case), the primal component has Hölder exponent 0 and constant $O(M_{3R})$, while the bilinear constant is $\|\bar{K}_a\|$. Their separated componentwise complexity therefore specializes to

$$N_{\bar{K}} = O\left(\frac{L_{\bar{K}} R_\lambda R_s}{\eta}\right), \qquad N_{\partial f} = O\left(\frac{M_{3R}^2 R_s^2}{\eta^2}\right) \tag{117}$$

for restricted gap target η. These are calls to the normalized $\bar{K}, \bar{K}^\top$ interface and subgradient calls, respectively; primitive realization is accounted for in Appendix L.

Here RS is the recursive-sliding method of Nguyen and Gasnikov (2026, Section 4, degenerate bilinear case) with the displayed primal and dual balls, the supplied subgradient and linear-operator constants, and termination budget η. The theorem-level count includes calls to its projection oracles. When vectors are distributed by blocks, projection onto a Euclidean ball requires a global norm reduction; that communication is not hidden inside a normalized affine call.

Algorithm 8 CROSS-SLIDING: one balanced nonsmooth phase

Require: F, $(\bar{K}, \bar{v})$, M_{3R}, center w^0, radius R_s, target restricted gap η, spectral norm/gap bounds, exact projections onto the phase balls

1: $a_s \leftarrow R_\lambda / R_s$, $\bar{K}_a \leftarrow a_s \bar{K}$, $\bar{v}_a \leftarrow a_s \bar{v}$
2: $\mathcal{X}_s \leftarrow B(w^0, 2R_s)$, $\mathcal{Y}_s \leftarrow \{y : \|y\| \le R_s\}$
3: $(\bar{w}, \bar{y}) \leftarrow \mathrm{RS}(F, \bar{K}_a, \bar{v}_a, \mathcal{X}_s, \mathcal{Y}_s, \eta)$
4: **return** $\bar{w}$

K.2 PROOF OF THEOREM 5.2

For the convex case take $R_s = R$ and $\eta = \varepsilon$ in (117). Lemma K.1 gives $R_\lambda \le 2M_{3R}/\sigma$. Fixed-accuracy separate normalization gives $L_{\bar{K}}/\sigma = \sqrt{\kappa(\bar{K})} = O(\chi_\times)$, hence

$$N_{\bar{K}} = O\left(\chi_\times \frac{M_{3R} R}{\varepsilon}\right) = O(\chi_\times T_{\mathrm{ns}}), \qquad N_{\partial f} = O(T_{\mathrm{ns}}^2).$$

Equation (114) gives (111).

For the strongly convex case restart this phase. Let $R_0 = R$ and at stage s request $\eta_s := \mu_f R_s^2/8$. If $\mathrm{Gap}(w_s) \le \eta_s$, strong convexity and the KKT inequality imply

$$\mathrm{Gap}(w_s) \ge \frac{\mu_f}{2}\|w_s - w^\star\|^2 + (R_\lambda - \|\lambda^\star\|)\|\bar{K} w_s - \bar{v}\|,$$

so $\|w_s - w^\star\| \le R_s/2$. Thus $R_{s+1} = R_s/2$ is valid. Equation (117) yields

$$N_{\bar{K}}^{(s)} = O\left(\frac{L_{\bar{K}} R_\lambda}{\mu_f R_s}\right), \qquad N_{\partial f}^{(s)} = O\left(\frac{M_{3R}^2}{\mu_f^2 R_s^2}\right).$$

The geometric sums are dominated by the final stage. Stopping at $R_S^2 \asymp \varepsilon/\mu_f$ gives

$$N_{\bar{K}} = O\left(\frac{L_{\bar{K}} R_\lambda}{\sqrt{\mu_f \varepsilon}}\right) = O(\chi_\times T_{\mathrm{nsc}}), \qquad N_{\partial f} = O(T_{\mathrm{nsc}}^2).$$

For any fixed J, the same normalized-affine proof replaces $\chi_\times$ by $\chi_{\times,J}$. A physical-round statement would additionally have to implement and count the global ball projections and scalar reductions, as well as the normalized channels; the certified smooth-SC primitive lift in Appendix L is not reused here without that accounting and a separate lift-radius argument. □

K.3 WHAT IS AND IS NOT PROVED ON THE NONSMOOTH LOWER-BOUND SIDE

The standard centralized oracle constructions give the first-order lower bounds quoted in Theorem 5.2; they are obtained on unconstrained (hence admissible) subclasses and do not involve the affine oracle. A matching resource-separated lower bound $N_{\mathrm{aff}} = \Omega(\chi_\times T_{\mathrm{ns}})$ (and its communication analogue) requires a hard family in which *both* the subgradient oracle and the affine oracle respect the same cross-angle information frontier. A naive coordinate-dilation argument is insufficient, because a subgradient oracle may expose a terminal micro-coordinate before the affine path has propagated to it. We therefore leave the matching $\chi_\times$-dependent nonsmooth affine/communication lower bound open rather than assuming such a frontier without proof.

L PRIMITIVE RESOURCE ACCOUNTING FOR THE SMOOTH-SC SOLVER

Charged-operation ledger.

Resource	One charged operation	Count
First-order oracle	One parallel local gradient (or subgradient in the nonsmooth extension).	$N_{\nabla f}$ or $N_{\partial f}$
Normalized affine oracle	Apply the separately normalized joint stack or its adjoint.	N_{aff}
Communication	One synchronous neighbor round / sparse gossip multiplication.	N_{comm}
Primitive implementation	One block-parallel original A, B, C, D family product or adjoint.	$N_A, \ldots, N_D$

This appendix concerns the executable smooth strongly convex solver of Theorem 4.1. The convex and nonsmooth extension theorems use their own normalized-channel ledgers; no unchanged raw-lift ledger is claimed for those regimes. All logarithmic factors below are $\log_+(R_E/\varepsilon)$ for the supplied initial primal–dual energy certificate. Fix $0 < \delta_{\mathrm{wh}} < 1$ independently of ε and use the certified construction in Appendix G. We count block-parallel products by the original matrices A, B, C, D and their transposes, and raw physical neighbor rounds. For one normalized forward *or* adjoint call, the upper bounds are

	$A/A^\top$	$B/B^\top$	$C/C^\top$	$D/D^\top$	rounds
$\widehat{L}_1, \widehat{L}_1^\top$	$O(\sqrt{\bar{k}_1})$	$O(\sqrt{\bar{k}_1})$	$O(\sqrt{\bar{k}_1}\sqrt{\bar{\kappa}_C})$	0	$O(\sqrt{\bar{k}_1}\sqrt{\bar{\kappa}_W})$
$\widehat{L}_2, \widehat{L}_2^\top$	0	0	0	$O(\sqrt{\bar{k}_2})$	$O(\sqrt{\bar{k}_2}\sqrt{\bar{\kappa}_W})$

Since $Q_j v = \mathcal{L}_j(\mathcal{L}_j^\top v)$, degree-$N_j^{\mathrm{wh}}$ normalization uses $2N_j^{\mathrm{wh}} + 1 = O(\sqrt{\bar{k}_j})$ channel actions. The extra factors $\sqrt{\bar{\kappa}_C}$ and $\sqrt{\bar{\kappa}_W}$ arise only from $C_\star$ and $\widetilde{W}$, respectively. With equal unit cost for the four matrix families, set

$$\begin{aligned} s_{\mathrm{CL}} &= O\Big(\sqrt{\bar{k}_1}(1 + \sqrt{\bar{\kappa}_C})\Big), \quad s_{\mathrm{SH}} = O(\sqrt{\bar{k}_2}), \\ \Lambda_{\mathrm{in}} &:= s_{\mathrm{CL}} + s_{\mathrm{SH}}, \qquad\qquad \Gamma_{\mathrm{in}} := \sqrt{\bar{k}_1} + \sqrt{\bar{k}_2}. \end{aligned} \tag{118}$$

The budgets absorb universal implementation constants; Λ_{in} counts raw affine work and Γ_{in} communication-bearing channel work. Absent matrix families contribute zero, and for $n = 1$ all communication counts vanish. For the supplied $\Xi \geq \chi_\times$, put $T_{\Xi,E} := \sqrt{\kappa_f}\,\Xi \log_+(R_E/\varepsilon)$. The complete smooth strongly convex solve upper bounds are

$$\begin{aligned} N_{\nabla f} &= O\big(\sqrt{\kappa_f}\log_+(R_E/\varepsilon)\big), \\ N_A,\ N_B &= O(T_{\Xi,E}\sqrt{\bar{k}_1}), \\ N_C &= O(T_{\Xi,E}\sqrt{\bar{k}_1}\sqrt{\bar{\kappa}_C}), \\ N_D &= O(T_{\Xi,E}\sqrt{\bar{k}_2}), \\ N_{\mathrm{aff}}^{\mathrm{prim}} &= O(T_{\Xi,E}\Lambda_{\mathrm{in}}), \\ N_{\mathrm{comm}} &= O(T_{\Xi,E}\Gamma_{\mathrm{in}}\sqrt{\bar{\kappa}_W}). \end{aligned} \tag{119}$$

Penalty-gradient products and one-time transformed right-hand sides are absorbed above; certificate acquisition is a separate setup cost under Assumption 1. Constant-factor-tight certificates, including $\Xi = O(\chi_\times)$, replace barred ratios by $k_1, k_2, \kappa_C, \kappa_W$ and $T_{\Xi,E}$ by $O(\sqrt{\kappa_f}\chi_\times \log_+(R_E/\varepsilon))$. Fixed δ_{wh} adds no target-accuracy logarithm.