# PARALLELIZABLE INTERIOR-POINT ADMM FOR BLOCK-STRUCTURED NONCONVEX PROGRAMS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Physics simulation is an increasingly important computational primitive for AI, and local–global/ADMM splitting exposes its parallel structure but lacks matching finite-time convergence guarantees. We close this gap with two ADMM variants, Interior-Point ADMM (IP-ADMM), for block-structured nonconvex composite programs with smooth nonlinear inequality and affine constraints. A one-sided finite-activation wall keeps each constraint split state strictly feasible while the original-space infeasibility vanishes, giving closed-form constraint updates, one proximal call per objective block, and one equality-constrained quadratic consensus solve per iteration. Ordinary IP-ADMM attains a CQ-free best-iterate weak-KKT rate of $O(K^{-1/3})$ or a banded Fritz–John rate of $O(K^{-1/4})$, which converts to KKT at $O(K^{-1/4})$ under a uniform constraint qualification; these are nearby-point certificates that reduce to conventional KKT for smooth blocks. A checkpoint variant, IP-ADMM-Ckpt, improves the asymptotic KKT exponent to $O(K^{-1/3})$. We also show how the bounds scale with the objective- and constraint-block counts and the coupling norm.

## 1 INTRODUCTION

Physics is becoming a computational foundation for AI. Differentiable simulators place mechanics inside learning and control loops de Avila Belbute-Peres et al. (2018); Qiao et al. (2020); Hu et al. (2020); physics-informed networks and learned mesh simulators inject governing equations or physical structure into predictive models Raissi et al. (2019); Pfaff et al. (2021); and world models and embodied agents learn or exploit interaction dynamics for planning and control Hafner et al. (2020); Savva et al. (2019). In all of these settings, we run simulation many times. Each run must be accurate enough to respect physical structure, and fast enough to supply the experience or gradients that the surrounding AI system needs.

Parallelism is a main way to reach that efficiency. Physical models naturally break into local constitutive, subsystem, and contact computations that are coupled through shared state, and high-throughput robot learning already benefits from running many physics environments at the same time Makoviychuk et al. (2021). Local–global and ADMM solvers are a promising direction along this line of research. These methods recast the simulation as a nonconvex constrained optimization problem, and expose parallelism by splitting the global model into local ones. This strategy has been very successful in practice for modeling deformable bodies, multibody systems, and contact-rich mechanics Narain et al. (2016); Overby et al. (2017); Lee et al. (2023); Zhu et al. (2024); Liu et al. (2026). These methods establish the practical value of decomposition. Unfortunately, existing finite-time convergence analyses still cannot explain their empirical success: standard nonconvex ADMM mainly treats objective blocks under affine coupling, whereas general nonlinear-constraint methods do not preserve this local–global physics architecture Jia & Grimmer (2025).

We ask whether a parallelizable splitting method can provide finite-time convergence guarantees for this nonlinear constrained model. To answer this question, we analyze two variants of the ADMM algorithm. Both introduce one split state $y_i = A_i x$ per local model, where each local model is either an objective term or a nonlinear inequality constraint. Our key idea is to apply a one-sided finite-activation quadratic wall to each inequality block, which keeps each split state strictly inside its feasible region. The original variable need not satisfy $f_i(A_i x) \geq 0$ at every iteration; instead, the consensus error and the original-space infeasibility vanish as the iterations proceed. Because



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.

the split states stay interior, our method is a variant of the interior-point method, which we call IP-ADMM.

Our analysis separates what we can certify with and without a constraint qualification (CQ). Ordinary IP-ADMM follows the idea of IPDS-ADMM Yuan (2025) and increases the penalty parameter at a fixed rate. This gives a CQ-free best-iterate w-KKT bound of $O(K^{-1/3})$, or a b-FJ bound of $O(K^{-1/4})$. Under Uniform Band-MFCQ, a sufficiently accurate b-FJ certificate converts to KKT, also at a rate of $O(K^{-1/4})$. To improve the asymptotic certificate rate under CQ, we consider a variant of IP-ADMM that updates the penalty parameter only at discrete checkpoints. We denote this algorithm IP-ADMM-Ckpt. After its first checkpoint completes at some iteration $K_0$, it gives a latest-checkpoint KKT rate of $O(K^{-1/3})$; no certificate is available before $K_0$, and $K_0$ can be large. Fixed-instance rates alone do not reveal how these guarantees scale with simulation size, because their hidden constants absorb all dependence on the discretization. As Figure 3 illustrates, a physics simulator instead defines a family of problems. As we refine the body or mesh, or add geometric and contact complexity, the number of local objective and inequality blocks grows, and shared states are reused more often Qiao et al. (2020); Liu et al. (2026). We therefore keep the objective-block count $N$, the constraint-block count $M$, and the stacked coupling norm $\|A\|_2$ in optional family-aware certified bounds. These bounds expose the dependence on both the number of local models and the total consensus coupling.

## 2 Related Work

Table 1: Physics regimes for our certificate hierarchy.

| Certificate | Physics regime and precedent |
|---|---|
| w-KKT | Fixed activation band Li et al. (2020) |
| b-FJ | Shrinking penalty layers Harmon et al. (2009) |
| KKT | MFCQ-qualified contact Anitescu & Hart (2004) |

**Physics+AI and parallel physics simulation.** Physics enters AI through several complementary ways. Differentiable engines expose simulation to end-to-end learning and control de Avila Belbute-Peres et al. (2018); Qiao et al. (2020); Hu et al. (2020). Physics-informed networks encode governing-equation residuals, and graph-based learned simulators use physical discretizations with local message passing Raissi et al. (2019); Pfaff et al. (2021). Latent world models and embodied platforms support prediction and interaction at scale Hafner et al. (2020); Savva et al. (2019); Makoviychuk et al. (2021). These developments motivate scalable physics simulators that make good use of parallel processors. Local–global and ADMM formulations are a promising direction along this line. These algorithms separate constitutive, subsystem, or contact handling from global consensus updates in projective dynamics, multibody systems, soft-body contact, and distributed affine-body dynamics Narain et al. (2016); Overby et al. (2017); Lee et al. (2023); Zhu et al. (2024); Liu et al. (2026). This literature motivates our architecture. Our main goal is to establish finite-time original-problem certification for the nonlinear program in Problem 1 under its local-consensus split.

**Contact and admissible trajectories.** Collision avoidance is an important source of nonlinear and often nonconvex constraints in mechanics: feasible motion depends on pose-dependent distances, changing closest features, and self-contact geometry. Robust cloth contact and continuous collision detection seek noninterpenetrating motion over a time interval Bridson et al. (2002); Harmon et al. (2009); Wang et al. (2021). IPC uses a barrier-augmented incremental potential and a CCD-aware line search from a collision-free state; C-IPC and rigid IPC extend this admissible-trajectory philosophy to codimensional and reduced-coordinate rigid contact Li et al. (2020; 2021); Ferguson et al. (2021). BC-ADMM provides a parallelizable algorithm for a class of biconvex functions, but it does not establish finite-time guarantees for certificates of the original problem Pan & Wu (2026).

**Nonconvex splitting and functional constraints.** Nonconvex ADMM commonly allows nonconvex or nonsmooth objective blocks while keeping affine coupling. IPDS-ADMM provides a closely aligned continuation precedent: it combines a polynomially increasing penalty with inverse-penalty smoothing and a square-root-penalty-rescaled multiplier Yuan (2025). Adaptive proximal ADMM instead uses progress tests to gate multiplier updates Maia et al. (2026). For nonlinear functional constraints, Jia & Grimmer (2025) obtain normalized Fritz–John certificates without a CQ, and KKT certificates under a uniform strong-MFCQ condition for the proximal subproblems, while Hinder and Ye analyze a nonconvex log-barrier method from a strictly feasible start Hinder & Ye (2024). ABIP and ACVI combine ADMM and interior-point ideas for linear programs and constrained vari-

ational inequalities, respectively Lin et al. (2021); Yang et al. (2023). However, these algorithms do not exploit the parallelizable structure in local models. Relative to these formulations, we study nonlinear-inequality block walls, one-call or closed-form local updates, original-model certificates, and family-aware bounds in a single local-consensus method.

**Certificates and qualification.** Table 1 maps our increasingly informative certificate hierarchy to three contact regimes. Weak-KKT (w-KKT) gives consensus and force balance without complementarity, paralleling IPC's fixed target distance $\hat{d}$, outside which contact repulsion is zero Li et al. (2020). Normalized banded Fritz–John (b-FJ) additionally supplies complementarity, nontrivial normalized multipliers, and finite near-active support without a CQ; its shrinking support band parallels ACM's nested penalty layers, whose thickness decreases while stiffness increases Harmon et al. (2009). Under Uniform Band-MFCQ, a sufficiently accurate b-FJ certificate converts to KKT, which removes the Fritz–John scaling ambiguity. Anitescu and Hart provide a narrow mechanics precedent: a pointed-friction-cone condition implies MFCQ for their rigid-contact quadratic program Anitescu & Hart (2004). Thus the cited physics papers instantiate the three regimes, while our analysis defines and connects the exact certificates; the CQ-free FJ versus CQ-dependent KKT boundary follows the broader optimization literature Jia & Grimmer (2025); Hinder & Ye (2024).

## 3 Problem Setting and Simulation Instances

We study a block-separable composite program. It has smooth (possibly nonconvex) inequality-constraint functions $f_i$, smooth (possibly nonconvex) objective terms $h_i$, convex nonsmooth proximable terms $\psi_i$, affine equality constraints through $C$, and affine coupling through the maps $A_i$:

$$\operatorname*{argmin}_{x} \sum_{i\in\mathcal{I}_o} \varphi_i(A_i x) \qquad \text{s.t.} \quad f_i(A_i x) \geq 0 \ (i \in \mathcal{I}_c), \quad Cx = d. \tag{1}$$

Here $N, M \geq 1$, $\mathcal{I} = \{1, \ldots, N+M\}$, $\mathcal{I}_o = \{1, \ldots, N\}$, and $\mathcal{I}_c = \mathcal{I} \setminus \mathcal{I}_o$.

**Assumption 3.1** (Constraint regularity)**.** *For every $i \in \mathcal{I}_c$, $f_i$ has an $L_i$-Lipschitz gradient.*

**Assumption 3.2** (Bounded constraint gradients)**.** *For every $i \in \mathcal{I}_c$, $\|\nabla f_i(y)\| \leq M_i$ for all $y$.*

**Assumption 3.3** (Feasibility)**.** *The initial point $x^1$ is strictly feasible: $Cx^1 = d$ and $m := \min_{i\in\mathcal{I}_c} f_i(A_i x^1) > 0$. A point is feasible when $Cx = d$ and $f_i(A_i x) \geq 0$ for every $i \in \mathcal{I}_c$.*

**Assumption 3.4** (Composite objective blocks)**.** *For each $i \in \mathcal{I}_o$, $\varphi_i = h_i + \psi_i$, where $h_i$ is continuously differentiable, possibly nonconvex, globally $D_i$-Lipschitz, and has an $H_i$-Lipschitz gradient. The function $\psi_i$ is proper, closed, convex, finite-valued on $\mathbb{R}^{n_i}$, globally $C_i^{\psi}$-Lipschitz, and $\operatorname{prox}_{t\psi_i}$ is exactly evaluable for every $t > 0$. We also assume the lower bound $\sum_{i\in\mathcal{I}_o} \varphi_i(A_i x) \geq 0$ for every $x$.*

These are the only four standing assumptions. For later use, define $\widehat{C}_i := D_i + C_i^{\psi}$ for $i \in \mathcal{I}_o$. Appendix G records the global estimates implied by these assumptions. Problem 1 overlaps two established problem classes but matches neither exactly. Affine-coupled nonconvex ADMM allows rich composite objective blocks, but it does not directly represent the nonlinear inequalities we use here Yuan (2025); Maia et al. (2026). Functional-constraint methods handle nonlinear inequalities and give FJ/KKT guarantees under their own assumptions Jia & Grimmer (2025); Hinder & Ye (2024), but they do not exploit this local-consensus decomposition. Our method combines the strengths of both.

### Motivating Applications

It is well known that physics simulation problems can be recast as nonconvex constrained optimization Narain et al. (2016); Overby et al. (2017), and we show that our problem setting applies directly to such problems and solves them in parallel. As Figure 1 illustrates, we compute the resting configurations of rigid and deformable objects, where (self-)collision constraints are the main source of nonconvexity. Detailed formulations and checks against the four standing assumptions appear in Appendix A, where we apply a series of constraint mollifications so the problem conforms to our assumptions. These mollifications are standard techniques; see, e.g., Pan & Wu (2026); Li et al. (2020). **Capsule stacking:** We minimize gravitational potential over $n$ planar capsule poses inside a circular container. The $3n$ pose variables form $N = n$ objective blocks; directed capsule–capsule

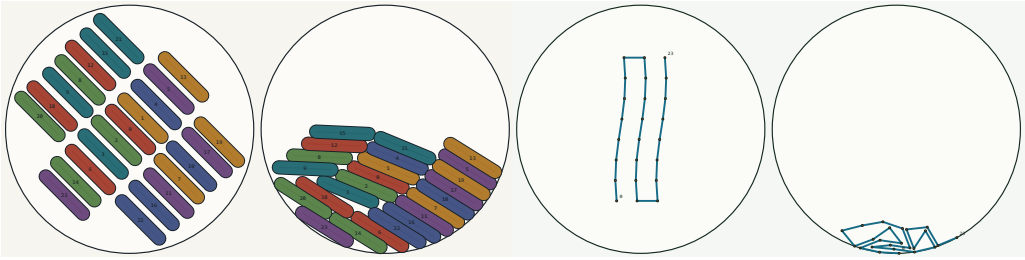

Figure 1: Initial and final configurations for the capsule and cloth resting instances.

Table 2: Convergence rate upper bounds. MFCQ means Uniform Band-MFCQ. Family bounds assume Assumption 4.1, under which $m = \mathcal{O}(1)$ and $M = \mathcal{O}(N^2)$; we omit $m$ for brevity but keep $M$ explicit, since practical instances often have $M \ll N^2$. For IP-ADMM-Ckpt, $K_0$ is the startup budget at which the first checkpoint completes; no certified output exists for $K < K_0$, and the family bound on $K_0$ can be large (Theorem 4.4). The checkpoint family rate carries no separate $M$ outside $\|A\|_2$ only conditional on family-uniform qualification margins (Assumption 4.2). The w-KKT, b-FJ, and KKT columns are nearby-point residuals (Definition H.1).

| Mode | Cert. | CQ | Validity | General | Family |
|---|---|---|---|---|---|
| IP-ADMM | w-KKT | none | $p = 1/3$, fixed $\xi$, $K \geq 3$ | $\mathcal{O}(K^{-1/3})$ | $\mathcal{O}(\|A\|_2 N\sqrt{M}K^{-1/3})$ |
| | b-FJ | none | $p = 1/2$, fixed $\xi$, $K \geq 3$ | $\mathcal{O}(K^{-1/4})$ | $\mathcal{O}(\|A\|_2 N\sqrt{M}K^{-1/4})$ |
| | b-FJ | none | $p = 1/2$, $\xi = \xi_\star$, $\xi_{\text{nom}} < 1$, $K \geq 3$ | $\mathcal{O}(K^{-1/4})$ | $\mathcal{O}\big((\|A\|_2 N + \|A\|_2^2\sqrt{N})K^{-1/2} + \sqrt{\|A\|_2 N}\,K^{-1/4} + K^{-1}\big)$ |
| | KKT | MFCQ | $p = 1/2$, fixed $\xi$, $K \geq K_{\text{cq}}(1/2, \xi)$ | $\mathcal{O}(K^{-1/4})$ | $\mathcal{O}(D_{\text{cq}}\|A\|_2 N\sqrt{M}K^{-1/4})$ |
| IP-ADMM-Ckpt | KKT | MFCQ | $K \geq K_0$ | $\mathcal{O}(K^{-1/3})$ | $\mathcal{O}(\|A\|_2 N^{4/3}K^{-1/3})$ |

collision-clearance constraints and capsule–container boundary collision constraints form $M = 2n^2$ inequality blocks. All blocks use local selectors $A_i$, there are no affine equalities, and initialization is strictly feasible. **Cloth resting:** We minimize gravity plus stretch and two-hop bending energies for $n$ planar cloth nodes, modeling an elastic 2D piece of cloth as described in Choi & Ko (2002). The $2n$ coordinates give $N = 3n - 3$ objective blocks; cloth self-collision constraints over nonincident node–segment pairs and cloth–container boundary collision constraints give $M = (n-1)(n-2) + n$ inequality blocks.

## 4 IP-ADMM(-Ckpt) and Convergence Analysis

We develop two variants of the ADMM algorithm for Problem 1. Both smooth the convex nonsmooth objective components with Moreau envelopes and represent each nonlinear inequality by a one-sided finite-activation wall, yielding independent local updates followed by one equality-constrained quadratic consensus update. Ordinary IP-ADMM uses polynomial continuation, similar to IPDS-ADMM Yuan (2025), and provides CQ-free best-iterate w-KKT and b-FJ guarantees, whereas IP-ADMM-Ckpt promotes its continuation state only at successful checkpoints and, under Uniform Band-MFCQ, returns a latest-checkpoint KKT certificate with an improved asymptotic rate once its first checkpoint completes. We first define their common model and algorithm and then state the mode-specific convergence results; complete constants and proofs are deferred to the appendices. Table 2 compares the summarized guarantees. IP-ADMM rows select the best certified iterate in the stated window, whereas IP-ADMM-Ckpt returns the latest checkpoint.

**Algorithm Design.** We begin by defining our proximal operators. For the convex component in objective, we use the standard Moreau envelopes below

$$(\psi_i)_\mu(y) := \min_z \left\{ \psi_i(z) + \frac{1}{2\mu}\|z - y\|^2 \right\}, \qquad P^{\psi}_{i,\mu}(y) := \text{prox}_{\mu\psi_i}(y). \tag{2}$$

For each mode $\mathsf{m} \in \{\mathsf{IP}, \mathsf{Ckpt}\}$, bind its proximal and wall parameters by

$$(\theta_{1,\mathsf{m}}, \theta_{2,\mathsf{m}}, G_{\mathsf{m}}) := \begin{cases} (\theta_1, \theta_2, G), & \mathsf{m} = \mathsf{IP}, \\ (\theta_1^{\text{cp}}, \theta_2^{\text{cp}}, G^{\text{cp}}), & \mathsf{m} = \mathsf{Ckpt}. \end{cases} \tag{3}$$

The mode-specific values are fixed in Appendix F. For a constraint and $L > 0$, define the active one-sided wall by

$$g_{\mathsf{m}}(s, L) = \begin{cases} G_{\mathsf{m}} - \sqrt{2LG_{\mathsf{m}}}\, s, & s < 0, \\ \frac{L}{2} \max(\sqrt{2G_{\mathsf{m}}/L} - s, 0)^2, & s \geq 0. \end{cases} \tag{4}$$

Write $\beta_{\mathsf{m}}(L) := \sqrt{2G_{\mathsf{m}}/L}$ for the active wall width. In the rest of the algorithm and in each mode-specific proof, $\mathsf{m}$ is fixed by the selected mode; we suppress this fixed mode index on $\bar{f}_i$, $\mathcal{M}_i^k$, and $\mathcal{L}$. We therefore unify objective and constraint blocks as

$$\bar{f}_i(y, L, \mu) := \begin{cases} h_i(y) + (\psi_i)_\mu(y), & i \in \mathcal{I}_{\mathrm{o}}, \\ g_{\mathsf{m}}(f_i(y), L), & i \in \mathcal{I}_{\mathrm{c}}. \end{cases} \tag{5}$$

The calculus of Moreau envelopes is well understood, and the activation wall follows from standard convex analysis. We summarize the useful properties in Appendix G. Introducing split variables $y_i = A_i x$ gives the wall-penalized split formulation used by the algorithm

$$\operatorname*{argmin}_{x,y} \sum_{i \in \mathcal{I}_{\mathrm{o}}} \varphi_i(y_i) + \sum_{i \in \mathcal{I}_{\mathrm{c}}} g_{\mathsf{m}}(f_i(y_i), L) \qquad \text{s.t.} \quad y = Ax, \quad Cx = d. \tag{6}$$

For the algorithm, we smooth the convex objective components and use the smoothed augmented Lagrangian

$$\mathcal{L}(x, y, \lambda, L, \mu, \rho) = \sum_{i \in \mathcal{I}} \bar{f}_i(y_i, L, \mu) + \lambda^T (y - Ax) + \frac{\rho}{2} \|y - Ax\|^2. \tag{7}$$

At iteration $k$, let $d_i^k := y_i^{k+1} - y_i^k$ denote the block step and $e_i^k := \lambda_i^k + \rho^k(y_i^k - A_i x^k)$ the current consensus force; the stacked vectors are $d^k := (d_i^k)_{i \in \mathcal{I}}$ and $e^k := (e_i^k)_{i \in \mathcal{I}}$, obtained by concatenating the blocks in the order $i = 1, \ldots, N + M$, so that $\|d^k\|^2 = \sum_{i \in \mathcal{I}} \|d_i^k\|^2$ and $\|e^k\|^2 = \sum_{i \in \mathcal{I}} \|e_i^k\|^2$. For a constraint block, $\ell_i^k(u) := f_i(y_i^k) + \langle \nabla f_i(y_i^k), u - y_i^k \rangle$ is its affine linearization. On an objective block, $w_i^{k+1} := \nabla(\psi_i)_{\mu^k}(y_i^{k+1})$ is the smoothed convex force; on a constraint block, $\iota_i^{k+1} := g'_{\mathsf{m}}(\ell_i^k(y_i^{k+1}), L^k)$ is the scalar wall force. The model-error curvature $\tau_i^k$, model smoothness $\zeta_i^k$, and resulting block force $q_i^k$ are consolidated as

$$(\tau_i^k, \zeta_i^k, q_i^k) := \begin{cases} (H_i, 1/\mu^k, \nabla h_i(y_i^k) + w_i^{k+1}), & i \in \mathcal{I}_{\mathrm{o}}, \\ (\sqrt{2L^k G_{\mathsf{m}}}\, L_i, L^k M_i^2, \iota_i^{k+1} \nabla f_i(y_i^k)), & i \in \mathcal{I}_{\mathrm{c}}. \end{cases} \tag{8}$$

Set $\mathsf{B}_i^k := \rho^k + \tau_i^k$, the curvature that majorizes the linearized block Lagrangian, and $m_i^k := \theta_{1,\mathsf{m}} \mathsf{B}_i^k$, the strongly convex proximal curvature. The block models are

$$\mathcal{M}_i^k(u) := \begin{cases} (\psi_i)_{\mu^k}(u) + \langle \nabla h_i(y_i^k) + e_i^k, u - y_i^k \rangle + \dfrac{m_i^k}{2} \|u - y_i^k\|^2, & i \in \mathcal{I}_{\mathrm{o}}, \\ g_{\mathsf{m}}(\ell_i^k(u), L^k) + \langle e_i^k, u - y_i^k \rangle + \dfrac{m_i^k}{2} \|u - y_i^k\|^2, & i \in \mathcal{I}_{\mathrm{c}}. \end{cases} \tag{9}$$

Both are strongly convex. Linearizing the smooth block terms makes the first model evaluable with one convex proximal call and gives a closed form for the second model.

For an objective block, set $t_i^k := 1/m_i^k$ and $z_{o,i}^k := y_i^k - t_i^k(\nabla h_i(y_i^k) + e_i^k)$. The one-call interface in Lemma G.3 returns $y_i^{k+1}$ together with the proximal point $p_i^{k+1}$ and force $w_i^{k+1}$; its exact identities are recorded in Equation 41. The constraint model has the three-case closed form of Lemma G.6; hence neither branch requires an iterative inner solve. In Ckpt mode a failed test freezes the stored dual and all continuation parameters but retains the new primal variables; the virtual multiplier is discarded. A successful step records its checkpoint certificate tuple with pre-promotion parameters, then commits the multiplier and promotes the continuation state. Thus the method has one global primal trajectory, and no certified latest-checkpoint output exists before the first success.

Our two modes below use the same primal update with positive continuation parameters $(\rho^k, L^k, \mu^k)$ with the only difference being the scheduling of these parameters. In the ordinary IP-ADMM mode, choose $\sigma \in (0, 2)$, $\rho^0, L^0, \eta > 0$, $\xi \in (0, 1]$, and $p \in [0, 1)$, and set

$$\rho^k = \rho^0(1 + \xi k^p), \qquad L^k = L^0(1 + \xi k^p), \qquad \mu^k = \frac{\eta}{\rho^k}. \tag{10}$$

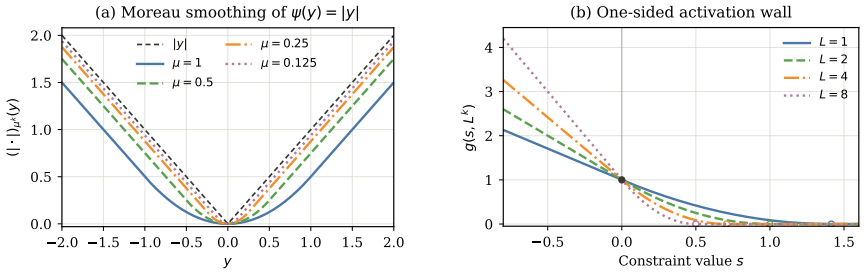


Figure 2: Geometry induced by the continuation schedule in Equation 10, shown under the illustrative normalization $\rho^0 = L^0 = \eta = G = 1$. Panel (a) uses $\mu \in \{1, 1/2, 1/4, 1/8\}$, and panel (b) uses $L \in \{1, 2, 4, 8\}$; these values represent scheduled models rather than iterations for a particular exponent, while $p$ determines how quickly the iterates traverse them. In panel (b), open markers indicate the zero endpoints $\beta = \sqrt{2/L}$.

The common factor in $(\rho^k, L^k)$ keeps $\rho^k/L^k = \rho^0/L^0$ fixed, whereas $\mu^k = \eta/[\rho^0(1 + \xi k^p)]$. Thus, for the model convex term $\psi(y) = |y|$ in the left panel of Figure 2, the quadratic transition region of its Moreau envelope contracts like $1/(1+\xi k^p)$ and the envelope approaches $|y|$. Meanwhile the wall in the right panel has feasible-side activation width

$$\beta^k := \beta_{\mathsf{IP}}(L^k) = \sqrt{2G/L^k} = \sqrt{2G/(L^0(1 + \xi k^p))}. \tag{11}$$

Its boundary height remains $g_{\mathsf{IP}}(0, L^k) = G$, but for each fixed violation $s < 0$, $g_{\mathsf{IP}}(s, L^k) = G - \sqrt{2L^0 G(1 + \xi k^p)}\, s$ grows with the schedule. Hence $p > 0$ moves through increasingly sharp objective and constraint models, while $p = 0$ uses one fixed member of each family.

The increasing-penalty/decreasing-smoothing continuation is inspired by IPDS-ADMM Yuan (2025). We depart from their method by additionally increasing the nonlinear-constraint wall curvature, so its activation band shrinks while consensus and reconstruction errors vanish. The IP-ADMM-Ckpt mode instead freezes $(\rho^k, L^k, \mu^k)$ between successful checkpoints and uses the geometric update shown in Algorithm 1. Its gated dual update is loosely motivated by the progress-tested multiplier updates of Maia et al. (2026); the retained-primal and full continuation-state semantics below are specific to our nonlinear-inequality checkpoint analysis.

**Algorithm 1** IP-ADMM(-Ckpt)

**Input:** $x^1$ satisfying Assumption 3.3; $\mathsf{mode} \in \{\mathsf{IP}, \mathsf{Ckpt}\}$, $\mathsf{m} = \mathsf{mode}$, $\chi > 1$, $\ell_{\mathrm{cp}} > 0$, and $\tau_{\mathrm{cp}} > 0$
  $y_i^1 \leftarrow A_i x^1$, $\lambda^1 \leftarrow 0$; initialize $(\rho^1, L^1, \mu^1)$ for the selected mode
  **for** $k = 1, 2, \ldots$ **do**
    **for** $i = 1, \ldots, N + M$ **do**
      $y_i^{k+1} \leftarrow \operatorname{argmin}_u \mathcal{M}_i^k(u)$
    $x^{k+1} \leftarrow \operatorname{argmin}_{Cx=d} \left\{\mathcal{L}(x, y^{k+1}, \lambda^k, L^k, \mu^k, \rho^k) + (\theta_{2,\mathsf{m}}\rho^k/2)\|A(x - x^k)\|^2\right\}$
    $r^{k+1} \leftarrow y^{k+1} - Ax^{k+1}$
    **if** mode = Ckpt **then**
      $\widehat{\lambda}^{k+1} \leftarrow \lambda^k + \rho^k r^{k+1}$
      $E^{k+1} \leftarrow \rho^k \sqrt{\|y^{k+1} - y^k\|^2 + \|A(x^{k+1} - x^k)\|^2}$
      **if** $E^{k+1} > \tau_{\mathrm{cp}}/\rho^k$ **then**
        $(\lambda^{k+1}, \rho^{k+1}, L^{k+1}, \mu^{k+1}) \leftarrow (\lambda^k, \rho^k, L^k, \mu^k)$
      **else**
        record the step-$k$ tuple with $(\rho^k, L^k, \mu^k)$ as the latest-checkpoint output
        $(\lambda^{k+1}, \rho^{k+1}, L^{k+1}, \mu^{k+1}) \leftarrow (\widehat{\lambda}^{k+1}, \chi\rho^k, \chi^2 L^k, \mu^k/\chi)$
    **else**
      $\lambda^{k+1} \leftarrow \lambda^k + \sigma\rho^k r^{k+1}$ and set $(\rho^{k+1}, L^{k+1}, \mu^{k+1})$ by Equation 10

**Remark 4.1.** *The $x$-update in Algorithm 1 minimizes an equality-constrained quadratic that depends on $x$ only through $Ax$ subject to $Cx = d$, so the minimizer need not be unique when $\ker A \cap \ker C \neq \{0\}$; in that case any minimizing $x^{k+1}$ may be selected, since $Ax^{k+1}$ is determined*

*and every quantity used by the algorithm and its analysis is invariant to the unobserved component of $x$ in* $\ker A \cap \ker C$.

**Family-level block scaling.** Finite-time analysis of a single problem usually hides the block-count dependence in its constants. In physics simulation, however, users change the problem dimension by refining the discretization, which gives a family of problems. We therefore make that dependence explicit through the following convention.

**Assumption 4.1** (Family-level block-scaling convention)**.** *For the optional family-level rate interpretation, consider a sequence of instances with $N \to \infty$ and $M \geq 1$, and fix $p \in (0,1)$. All constants below are uniform in the family member and block index. Every element of $\{\widehat{C}_i, C_i^{\psi}, H_i, \varphi_i(A_i x^1) : i \in \mathcal{I}_o\} \cup \{M_i, L_i : i \in \mathcal{I}_c\} \cup \{m\}$ is $\mathcal{O}(1)$, $\dim(x) = \mathcal{O}(N)$, and $M = \mathcal{O}(N^2)$. Moreover, there are constants $s_A \in \mathbb{N}$ and $0 < c_A \leq C_A < \infty$ such that $\|\operatorname{vec}(A_i)\|_0 \leq s_A$, and $c_A \leq \|\operatorname{vec}(A_i)\|_1 \leq C_A$ for all $i \in \mathcal{I}$. Finally, there are constants $c_\varphi, c_{\widehat{C}}, c_M, c_m > 0$ such that $\varphi_i(A_i x^1) \geq c_\varphi$ and $\widehat{C}_i \geq c_{\widehat{C}}$ for every $i \in \mathcal{I}_o$, $M_i \geq c_M$ for every $i \in \mathcal{I}_c$, and $m \geq c_m$.*

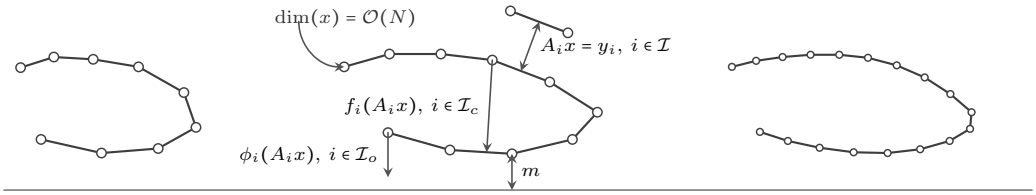


Figure 3: A family of problem instances simulating a 2D cloth falling on the ground.

Figure 3 illustrates Assumption 4.1 on a piece of cloth falling to the ground, discretized into $N$ elements of 2 vertices each. Each element carries a fixed number of objective terms, such as gravitational and internal potential energy, so $\dim(x) = \mathcal{O}(N)$. Pairwise self-collision constraints give $M = \mathcal{O}(N^2)$, each $A_i$ gathers a bounded local stencil so $\|A_i\|_2 = \Theta(1)$, and the other $\mathcal{O}(1)$ bounds encode fixed local energy, force, and curvature scales. The lower bounds on $\widehat{C}_i$ and $M_i$ normalize valid majorants, and that on $\varphi_i(A_i x^1)$ fixes an energy gauge, pinning each block's initial energy to a positive $\Theta(1)$ scale; $m \geq c_m$ is the substantive geometric condition, requiring a uniform initial self-collision and boundary margin. This gauge is a normalization choice: because the initial energy is defined only up to an additive constant, the family bounds use a fixed common per-block energy normalization and are not invariant to arbitrary additive energy shifts, such as the aggregate constant shift used to meet Assumption 3.4 in Appendix A. These conditions are natural for consistently normalized refinements that preserve clearance, but remain conditional under arbitrary mesh refinement Narain et al. (2016); Overby et al. (2017); Liu et al. (2026).

### 4.1 Convergence of IP-ADMM

Progress is measured by the reconstructed original-problem residuals $\mathrm{Crit}_{\text{w-KKT}}$, $\mathrm{Crit}_{\text{b-FJ}}$, and $\mathrm{Crit}_{\text{KKT}}$ defined in Definition H.1–Equation 106. These are nearby-point certificates: the objective and constraint (sub)gradients are evaluated at reconstructed block points $\breve{y}_i$ that are close to, but not identical with, $A_i x$, with the gap controlled by the consensus term $\|Ax - \breve{y}\|$. Evaluating stationarity at such nearby reconstructed points is standard for splitting and proximal methods on nonsmooth composite problems Rockafellar & Wets (2009); Yuan (2025); Jia & Grimmer (2025). Ordinary IP-ADMM returns the best certified candidate in the stated iteration window with the following guarantee.

**Theorem 4.2** (Original-problem rates for IP-ADMM)**.** *Run the ordinary IP-ADMM mode of Algorithm 1 with the parameter choice Equation 28 and polynomial schedule Equation 10. Fix $p \in (0,1)$ and $\xi \in (0,1]$. For the window $j = 3, \ldots, K+1$ and $K \geq 3$,*

$$\begin{aligned} \min_j \mathit{Crit}^j_{\text{w-KKT}} &= \mathcal{O}(K^{(p-1)/2}) + \mathcal{O}(K^{-p}), \\ \min_j \mathit{Crit}^j_{\text{b-FJ}} &= \mathcal{O}(K^{(p-1)/2}) + \mathcal{O}(K^{-p}) + \mathcal{O}(K^{-p/2}) + \mathcal{O}(K^{-1}). \end{aligned} \tag{12}$$

*If, in addition, Assumption 4.1 holds, the fixed-$\xi$ bounds specialize to $\mathcal{O}(\|A\|_2 N\sqrt{M}\, K^{-1/3})$ for w-KKT at $p = 1/3$ and $\mathcal{O}(\|A\|_2 N\sqrt{M}\, K^{-1/4})$ for b-FJ at $p = 1/2$. For the optional tuned $p = 1/2$*

*branch, define*

$$\xi_{\mathrm{nom}} := m/(\|A\|_2 N), \qquad \xi := \min\{1, \xi_{\mathrm{nom}}\}. \tag{13}$$

*On the interior branch* $\xi_{\mathrm{nom}} < 1$*, write* $\xi_\star := \xi_{\mathrm{nom}}$*. Choosing* $\xi = \xi_\star$ *gives*

$$\min_j \mathit{Crit}^j_{b\text{-}FJ} = \mathcal{O}\left(\left[\|A\|_2 N + (\|A\|_2^2\sqrt{N}/m)\right] K^{-1/2} + \sqrt{\|A\|_2 N m}\, K^{-1/4} + K^{-1}\right). \tag{14}$$

*If* $\xi_{\mathrm{nom}} \geq 1$*, the rule clips to* $\xi = 1$ *and the fixed-*$\xi$ *bound applies.*

**Assumption 4.2** (Uniform Band-MFCQ)**.** *There are* $\gamma_{\mathrm{cq}}, \delta_{\mathrm{cq}} > 0$ *such that, whenever* $Cx = d$*,* $f_i(y_i) \geq 0$*, and* $\max_{i \in \mathcal{I}_c} \|A_i x - y_i\| \leq \delta_{\mathrm{cq}}$*, there is a unit* $d_{\mathrm{cq}} \in \ker C$ *satisfying*

$$\nabla f_i(y_i)^T A_i d_{\mathrm{cq}} \geq \gamma_{\mathrm{cq}} \quad \textit{whenever } 0 \leq f_i(y_i) \leq \delta_{\mathrm{cq}}. \tag{15}$$

**Corollary 4.3** (KKT rates for IP-ADMM under Uniform Band-MFCQ)**.** *Run the ordinary IP-ADMM mode of Algorithm 1 with the parameter choice Equation 28,* $p = 1/2$*,* $\xi \in (0, 1]$*, and polynomial schedule Equation 10, and suppose Assumption 4.2 holds. There is a finite, instance-dependent qualification threshold* $K_{\mathrm{cq}}(1/2, \xi) \geq 3$ *such that, for every* $K \geq K_{\mathrm{cq}}(1/2, \xi)$*,* $\min_{j=3,\ldots,K+1} \mathit{Crit}^j_{KKT} = \mathcal{O}(K^{-1/4})$*. If, in addition, Assumption 4.1 holds, define* $D_{\mathrm{cq}} := 2(1 + \|A\|_2\sqrt{\widehat{G}_0}/\gamma_{\mathrm{cq}}) = \Theta(1 + \|A\|_2\sqrt{N}/\gamma_{\mathrm{cq}})$*; then the fixed-*$\xi$ *upper bound is* $\mathcal{O}(D_{\mathrm{cq}}\|A\|_2 N\sqrt{M}\, K^{-1/4})$.

**Proof sketch.** Our schedule increases the penalty and decreases the smoothing, following the continuation idea of IPDS-ADMM Yuan (2025), whose analysis builds a Lyapunov function. Our nonlinear wall, however, raises a new issue: the continuation argument holds only while the split point stays strictly feasible. To handle this, Lemma H.3 lower-bounds the Lyapunov function without assuming the wall-interior condition. Theorem H.6 then combines this bound with the descent of the Lyapunov candidate to prove strict feasibility by induction. The drift-corrected lower bound Lemma H.7 gives the movement budget Corollary H.8. Exact objective-force reconstruction turns this budget into original-problem residuals in Theorem H.12, and Corollary H.13 converts the same residual to KKT under Uniform Band-MFCQ. The family-level and tuned-$\xi$ bounds follow as specializations. Full constants and the dependency chain are in Appendix H.

**Interpretation of the continuation balance.** IPDS-ADMM allows problem-dependent continuation amplitudes $\xi$, including a condition-number-dependent choice in one regime Yuan (2025). Our family-level result instead identifies the explicit nominal balance $\xi_{\mathrm{nom}} = m/(\|A\|_2 N)$ for the nonlinear-inequality split. It changes the leading $K^{-1/4}$ coefficient from the fixed-$\xi$ family scale $\|A\|_2 N\sqrt{M}$ to $\sqrt{\|A\|_2 N m}$, improving the block-scale dependence of that leading high-accuracy term. It does not improve the finite-$K$ bound: the $K^{-1/2}$ terms and the clipping condition remain.

### 4.2 Convergence of IP-ADMM-Ckpt

Under the CQ condition, the checkpoint mechanism yields an improved asymptotic certificate rate after the first completed checkpoint, as stated below. This is a statement about the latest-checkpoint certificate as $K \to \infty$, not a claim of faster progress at a fixed budget: no certified output exists before the first checkpoint completes at iteration $K_0$, and the family bound on $K_0$ can be large.

**Theorem 4.4** (Checkpoint KKT rates for IP-ADMM-Ckpt under Uniform Band-MFCQ)**.** *Run the* Ckpt *mode of Algorithm 1 with checkpoint test* $E^{k+1} \leq \tau_{\mathrm{cp}}/\rho^k$*, the explicit initialization Equation 29, and the geometric checkpoint promotion in Algorithm 1, and suppose Assumption 4.2 holds. Let* $K_0$ *be the iteration budget at which checkpoint* 0 *completes, and for* $K \geq K_0$ *let* $\mathit{Crit}_{KKT^{\mathrm{out}}}(K)$ *denote the residual of the latest-checkpoint output. Then* $\mathit{Crit}_{KKT^{\mathrm{out}}}(K) = \mathcal{O}(K^{-1/3})$ *whenever* $K \geq K_0$*, and no certified output exists for* $K < K_0$*. If, in addition, Assumption 4.1 holds across a family, the Uniform Band-MFCQ constants are uniformly bounded away from zero, and* $\chi$*,* $\ell_{\mathrm{cp}}$*, and* $\tau_{\mathrm{cp}}$ *are fixed, then* $\mathit{Crit}_{KKT^{\mathrm{out}}}(K) = \mathcal{O}\left(\|A\|_2 N^{4/3} K^{-1/3}\right)$ *whenever* $K \geq K_0$*. In this family the startup budget obeys* $K_0 = \mathcal{O}\left(N\left(N + \|A\|_2^2/\gamma_{\mathrm{cq}}^2\right)^3\right)$*, so the improved exponent describes only post-startup behavior and need not reduce the total iteration budget.*

**Proof sketch.** The checkpoint gate is loosely motivated by the progress-tested multiplier updates of adaptive proximal ADMM Maia et al. (2026), but our proof uses a pointwise level induction. We condition each level on two caps: a cap on the stored multiplier and a cap on the entry merit. Under these caps, Lemma I.7 charges every rejection against the frozen-level merit and proves finite arrival with $T_j = \mathcal{O}(\rho_j^3)$. Independently, Lemma I.6 uses split feasibility, support control, and Uniform Band-MFCQ to bound the wall forces, the accepted virtual multiplier, and the consensus by $\mathcal{O}(\rho_j^{-1})$. The promotion charge is then summable, so Lemma I.8 propagates both caps and closes the induction. Each completed level thus carries the pointwise $\mathcal{O}(\rho_j^{-1})$ KKT certificate of Lemma I.9, and Corollary I.10 combines it with geometric level growth to give the latest-checkpoint rate $\mathcal{O}(K^{-1/3})$. Family and multiplier scaling are developed in Appendix I.

**Multiplier bounds.** The two analyses also provide complementary multiplier controls. For ordinary IP-ADMM, the rescaled invariant $\sup_j \|\lambda^j\|^2/\rho^j < \infty$ and $\lambda^j/\rho^j \to 0$ holds for $p > 0$; this does not bound the raw multipliers uniformly, whereas the constant-penalty endpoint $p = 0$ does. For IP-ADMM-Ckpt, Uniform Band-MFCQ bounds the accepted wall forces and virtual consensus multipliers, and hence every stored consensus multiplier; rejected virtual multipliers are not covered. These controls mirror the rescaled-multiplier viewpoint of IPDS-ADMM Yuan (2025); see Lemma H.9, Corollary H.10, and Corollary I.12.

## 5 Numerical Evaluation

We evaluate capsule stacking and cloth resting for $n \in \{8, 16, 24\}$, with one deterministic run per family–size–method case. We compare against two baselines. SLSQP is a sequential quadratic programming (SQP) baseline, and BC-ADMM Pan & Wu (2026) is the closest published optimizer to ours, since it is also a parallel ADMM that starts from a feasible point and keeps every iterate original-space feasible. Every method receives the same strictly feasible initialization and the same analytic derivatives. Across the 21 runs we test the three certificates w-KKT, b-FJ, and KKT. On the small instances ($n = 8$), SLSQP and both IP-ADMM variants reach the KKT threshold of $10^{-3}$. On the larger instances ($n \in \{16, 24\}$), all three still reach the w-KKT threshold and both IP-ADMM variants reach b-FJ, but neither IP-ADMM variant reaches KKT, and SLSQP reaches KKT only on cloth resting, not on capsule stacking. This is consistent with CQ degeneracy or difficulty of KKT certification, but our experiments do not distinguish these explanations. Either way, it agrees with prior works Pan & Wu (2026); Li et al. (2020) that certify only w-KKT or b-FJ solutions. We further observe that the configuration settles early: the normalized state-space $L_2$ distance of each iterate to the final output (Figure 6) falls below $5\%$ of its initial value well before KKT would be certified, so an early stop, or a stopping rule based on the relative change in the physical configuration, is a useful practical heuristic in these experiments. Appendix B gives the protocol and the per-run numbers, and Appendix C connects the theorem-sufficient parameters to the tuned values used here. We further show that our method has applications beyond physics simulation, where Appendix D optimizes a UAV trajectory whose objective carries a sharp $\ell_1$ goal term, which our proximal oracle handles exactly. Finally, Appendix E reports real wall-clock time on the cloth testbed.

## 6 Limitations and Conclusion

This work lays a theoretical foundation for parallel physics optimization: independent nonlinear-inequality updates, one equality-constrained quadratic consensus solve per iteration, a CQ-explicit hierarchy of finite-time certificates, and conditional family-aware bounds in $N$, $M$, and $\|A\|_2$ under a family-scaling convention. Together these give finite-time, original-problem guarantees that respect the local–global splitting used in practice. Several limitations temper these guarantees. First, the KKT results rely on Uniform Band-MFCQ, whose constants $(\gamma_{\mathrm{cq}}, \delta_{\mathrm{cq}})$ we cannot compute or even lower-bound in our applications. Second, the family-aware rates are conditional on the scaling convention and its energy-gauge normalization rather than intrinsic properties of the algorithm. Finally, our main convergence benchmarks remain small, static, single-threaded, and 2D; the additional multicore scaling study provides only limited scalability evidence. These gaps point to concrete next steps. Distributed and GPU implementations would test the parallel structure at scale; and richer benchmarks would stress the family-scaling assumptions. Integrating the solver into learning pipelines and world models is a natural path toward large-scale Physics+AI.

## AI USE STATEMENT

(This section is required by the ICLR 2027 AI Policy for Authors and does not count toward the page limit.)

We used generative AI tools in preparing this manuscript and disclose their use here in accordance with the ICLR 2027 AI Policy for Authors, which covers both recommended- and required-disclosure activities. Concretely, we used generative AI to (i) search for and identify related literature; (ii) edit and condense author-written text for clarity and length; (iii) perform adversarial checking of the mathematical arguments and proofs, flagging gaps, unstated assumptions, and imprecise claims for the authors to resolve. The authors independently derived, checked, and verified all mathematical claims and proofs, and designed, implemented, ran, and verified all experiments, including data generation and analysis. Every AI-assisted contribution—references, edits, arguments, and interpretations—was reviewed and validated by the authors against the primary sources and the underlying mathematics before inclusion. The authors take full responsibility for the final content of this work, including all text, claims, proofs, and artifacts.

## REFERENCES

Mihai Anitescu and Gary D. Hart. A constraint-stabilized time-stepping approach for rigid multibody dynamics with joints, contact and friction. *International Journal for Numerical Methods in Engineering*, 60(14):2335–2371, 2004. doi: 10.1002/nme.1047.

Robert Bridson, Ronald Fedkiw, and John Anderson. Robust treatment of collisions, contact and friction for cloth animation. *ACM Transactions on Graphics*, 21(3):594–603, 2002. doi: 10.1145/566654.566623.

Kwang-Jin Choi and Hyeong-Seok Ko. Stable but responsive cloth. *ACM Trans. Graph.*, 21(3): 604–611, July 2002. ISSN 0730-0301. doi: 10.1145/566654.566624. URL https://doi.org/10.1145/566654.566624.

Filipe de Avila Belbute-Peres, Kevin Smith, Kelsey Allen, Josh Tenenbaum, and J. Zico Kolter. End-to-end differentiable physics for learning and control. In *Advances in Neural Information Processing Systems*, volume 31, pp. 7178–7189, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/842424a1d0595b76ec4fa03c46e8d755-Abstract.html.

Zachary Ferguson, Minchen Li, Teseo Schneider, Francisca Gil-Ureta, Timothy Langlois, Chenfanfu Jiang, Denis Zorin, Danny M. Kaufman, and Daniele Panozzo. Intersection-free rigid body dynamics. *ACM Transactions on Graphics*, 40(4):183:1–183:16, 2021. doi: 10.1145/3450626.3459802.

Danijar Hafner, Timothy P. Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *The Eighth International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1lOTC4tDS.

David Harmon, Etienne Vouga, Breannan Smith, Rasmus Tamstorf, and Eitan Grinspun. Asynchronous contact mechanics. *ACM Transactions on Graphics*, 28(3):87:1–87:12, 2009. doi: 10.1145/1531326.1531393.

Oliver Hinder and Yinyu Ye. Worst-case iteration bounds for log barrier methods on problems with nonconvex constraints. *Mathematics of Operations Research*, 49(4):2402–2424, 2024. doi: 10.1287/moor.2020.0274.

Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. Difftaichi: Differentiable programming for physical simulation. In *The Eighth International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=B1eB5xSFvr.

Zhichao Jia and Benjamin Grimmer. First-order methods for nonsmooth nonconvex functional constrained optimization with or without slater points. *SIAM Journal on Optimization*, 35(2):1300–1329, 2025. doi: 10.1137/23M1569551.

Jeongmin Lee, Minji Lee, and Dongjun Lee. Modular and parallelizable multibody physics simulation via subsystem-based ADMM. In *IEEE International Conference on Robotics and Automation*, pp. 10132–10138, 2023. doi: 10.1109/ICRA48891.2023.10161052.

Minchen Li, Zachary Ferguson, Teseo Schneider, Timothy Langlois, Denis Zorin, Daniele Panozzo, Chenfanfu Jiang, and Danny M. Kaufman. Incremental potential contact: Intersection- and inversion-free large deformation dynamics. *ACM Transactions on Graphics*, 39(4):49:1–49:20, 2020. doi: 10.1145/3386569.3392425.

Minchen Li, Danny M. Kaufman, and Chenfanfu Jiang. Codimensional incremental potential contact. *ACM Transactions on Graphics*, 40(4):170:1–170:24, 2021. doi: 10.1145/3450626.3459767.

Tianyi Lin, Shiqian Ma, Yinyu Ye, and Shuzhong Zhang. An ADMM-based interior-point method for large-scale linear programming. *Optimization Methods and Software*, 36(2–3):389–424, 2021. doi: 10.1080/10556788.2020.1821200.

Jiafeng Liu, Wenhui Zhou, Xinming Pei, Yifan Peng, Huamin Wang, Yin Yang, Lei Lan, and Weiwei Xu. Distributed affine body dynamics with adaptive consensus. In *ACM SIGGRAPH Conference Papers*, 2026. doi: 10.1145/3799902.3811106.

Leandro Farias Maia, David H. Gutman, Renato D. C. Monteiro, and Gilson N. Silva. An adaptive proximal ADMM for nonconvex linearly constrained composite programs. *Mathematical Programming*, pp. 1–56, 2026. doi: 10.1007/s10107-026-02366-w.

Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. Isaac Gym: High performance GPU-based physics simulation for robot learning. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. URL https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/hash/28dd2c7955ce926456240b2ff0100bde-Abstract-round2.html.

Rahul Narain, Matthew Overby, and George E. Brown. ADMM ⊇ projective dynamics: Fast simulation of general constitutive models. In *Eurographics/ACM SIGGRAPH Symposium on Computer Animation*, pp. 21–28, 2016. doi: 10.2312/sca.20161219.

Matthew Overby, George E. Brown, Jie Li, and Rahul Narain. ADMM ⊇ projective dynamics: Fast simulation of hyperelastic models with dynamic constraints. *IEEE Transactions on Visualization and Computer Graphics*, 23(10):2222–2234, 2017. doi: 10.1109/TVCG.2017.2730875.

Zherong Pan and Kui Wu. BC-ADMM: A parallel decoupled non-convex constrained optimizer for robot applications. *The International Journal of Robotics Research*, 2026. doi: 10.1177/02783649261458408.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks. In *The Ninth International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=roNqYL0_XP.

Yi-Ling Qiao, Junbang Liang, Vladlen Koltun, and Ming C. Lin. Scalable differentiable physics for learning and control. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7847–7856, 2020.

Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, 2009.

Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A platform for embodied AI research. In *2019 IEEE/CVF International Conference on Computer Vision*, pp. 9338–9346, 2019. doi: 10.1109/ICCV.2019.00943.

Bolun Wang, Zachary Ferguson, Teseo Schneider, Xin Jiang, Marco Attene, and Daniele Panozzo. A large-scale benchmark and an inclusion-based algorithm for continuous collision detection. *ACM Transactions on Graphics*, 40(5):188:1–188:16, 2021.

Tong Yang, Michael I. Jordan, and Tatjana Chavdarova. Solving constrained variational inequalities via a first-order interior point-based method. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=RQY2AXFMRiu.

Ganzhao Yuan. ADMM for nonconvex optimization under minimal continuity assumption. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=GKAQ92ua3A.

Siyan Zhu, Cheng Fang, Peng Yu, Xiao Zhai, Aimin Hao, and Junjun Pan. Efficient frictional contacts for soft body dynamics via ADMM. *The Visual Computer*, 40:4569–4583, 2024. doi: 10.1007/s00371-024-03438-8.

# A Application Formulations and Assumption Justifications

This appendix records the exact block structure of the two benchmark families. In both cases $C$ and $d$ are empty and $\psi_i \equiv 0$.

**Compact-domain realization.** Suppose exact boundary blocks confine every feasible configuration to a compact set $\Omega$ (modulo exact periodic coordinates), and each remaining raw $f_i$ and $h_i$ is $C^{1,1}$ on an open neighborhood of $A_i\Omega$. Then Assumption 3.1–Assumption 3.2 hold automatically. Indeed, on this compact set each block has bounded value and gradient and a finite gradient-Lipschitz constant. A Whitney extension of this value–gradient jet, followed by a smooth cutoff outside a larger neighborhood, yields a globally defined block that agrees in value and gradient on $A_i\Omega$. The constraint extensions then have globally Lipschitz and bounded gradients, which supply Assumption 3.1–Assumption 3.2. The objective extensions are globally Lipschitz with Lipschitz gradients, and an aggregate constant shift supplies the lower bound in Assumption 3.4 without changing derivatives or minimizers. This shift is a gauge choice: it leaves derivatives and minimizers unchanged, but it does change absolute energy aggregates such as those entering the family-level constants of Assumption 4.1, so those family bounds are stated relative to a fixed common energy normalization. Because the exact boundary blocks force feasible points into $\Omega$, the raw and extended constrained objectives agree on their common feasible envelope. This is the BC-ADMM extension argument of Pan & Wu (2026). We use it to show that our two applications satisfy all the problem assumptions.

**Distance regularization.** To satisfy our problem assumptions, we locally mollify the objective and constraint functions using the helper functions defined below. Similar techniques have been used in Pan & Wu (2026); Li et al. (2020). For the spring scale $\epsilon > 0$, define the pseudo-Huber length modulation

$$\mathscr{M}_\epsilon^{\mathrm{len}}(r) = \sqrt{\|r\|^2 + \epsilon^2} - \epsilon, \tag{16}$$

and for a squared distance $t \geq 0$ define

$$\mathscr{M}_\delta^{\mathrm{rad}}(t) = \delta^2 \left( \sqrt{1 + t/\delta^2} - 1 \right). \tag{17}$$

The latter equals $\delta \mathscr{M}_\delta^{\mathrm{len}}(r)$ when $t = \|r\|^2$. It is smooth and strictly increasing, its residual-space gradient is globally bounded by $\delta$, and that gradient is 1-Lipschitz. Consequently,

$$\mathscr{M}_\delta^{\mathrm{rad}}(D) - \mathscr{M}_\delta^{\mathrm{rad}}(s^2) \geq 0 \quad \Longleftrightarrow \quad D \geq s^2. \tag{18}$$

Thus radial modulation preserves the threshold set of the squared distance supplied to it. For a point $p$ and a segment with endpoints $u, v$, and activation margin constants $\ell_{\mathrm{act}} > 0$, let

$$D_{\mathrm{seg}}(p; u, v) = \min_{t \in [0,1]} \|p - (1-t)u - tv\|^2, \; D_0(p; u, v) = \left\| p - \frac{u+v}{2} \right\|^2, \; \sigma = \|v - u\|^2. \tag{19}$$

With $S(z) = 6z^5 - 15z^4 + 10z^3$, define the blend

$$\mathrm{blend}_{\ell_{\mathrm{act}}}(\sigma) = \begin{cases} 0, & \sigma \leq \ell_{\mathrm{act}}^2/4, \\ S\left( \dfrac{\sigma - \ell_{\mathrm{act}}^2/4}{3\ell_{\mathrm{act}}^2/4} \right), & \ell_{\mathrm{act}}^2/4 < \sigma < \ell_{\mathrm{act}}^2, \\ 1, & \sigma \geq \ell_{\mathrm{act}}^2, \end{cases} \tag{20}$$

and the short-segment modulation

$$\mathscr{M}_{\ell_{\mathrm{act}}}^{\mathrm{seg}}(p; u, v) = (1 - \mathrm{blend}_{\ell_{\mathrm{act}}}(\sigma)) D_0 + \mathrm{blend}_{\ell_{\mathrm{act}}}(\sigma) D_{\mathrm{seg}}. \tag{21}$$

The quintic blend has zero first and second derivatives at both junctions. Since $D_{\mathrm{seg}}$ is jointly $C^{1,1}$ wherever $u \neq v$, including the endpoint/interior projection transitions, $\mathscr{M}_{\ell_{\mathrm{act}}}^{\mathrm{seg}}$ is endpoint-symmetric and $C^{1,1}$ through $u = v$. It equals the standard squared point–segment distance exactly when $\|v - u\| \geq \ell_{\mathrm{act}}$. We define $\mathscr{M}_0^{\mathrm{seg}} = D_{\mathrm{seg}}$, so zero disables this modulation. Because $D_{\mathrm{seg}} \leq D_0$, the positive-margin modulation can overestimate the exact distance inside its short activation band; there it is a local model regularization rather than an exact reparameterization. Thus $\mathscr{M}_{\ell_{\mathrm{act}}}^{\mathrm{seg}}$ provides a globally smooth surrogate collision constraint. The exact distance $D_{\mathrm{seg}}$ is nonsmooth when $\sigma = 0$; the smooth blending keeps $\mathscr{M}_{\ell_{\mathrm{act}}}^{\mathrm{seg}}$ away from this nonsmooth configuration as $\sigma \to 0$.

**Capsule stacking.** Capsule $i$ has pose $z_i = (c_i, \theta_i) \in \mathbb{R}^2 \times \mathbb{R}$, and $x = (z_1, \dots, z_n) \in \mathbb{R}^{3n}$. With planar rotation $R(\theta_i)$ and fixed body-coordinate medial-axis endpoints $a_e$, $e \in \{1, 2\}$, its world endpoint is

$$p_{ie}(z_i) = c_i + R(\theta_i)a_e. \tag{22}$$

There is one gravity block per capsule, $h_i(z_i) = e_2^T c_i$, so $N = n$. For each ordered source endpoint $(i, e)$ and target capsule $j \neq i$, let $D_{iej}(z_i, z_j)$ be its squared distance to capsule $j$'s medial segment. With capsule radius $r_{\mathrm{cap}}$ and clearance $s = 2r_{\mathrm{cap}}$, the capsule–capsule collision block is

$$f_{iej} = \mathscr{M}_\delta^{\mathrm{rad}}(D_{iej}) - \mathscr{M}_\delta^{\mathrm{rad}}(s^2). \tag{23}$$

Each endpoint also gives the capsule–container boundary block

$$f_{ie}^{\mathrm{circ}} = \mathscr{M}_\delta^{\mathrm{rad}}(R_{\mathrm{eff}}^2) - \mathscr{M}_\delta^{\mathrm{rad}}(\|p_{ie} - c_{\mathrm{circ}}\|^2), \tag{24}$$

where $R_{\mathrm{eff}}$ is the container radius minus $r_{\mathrm{cap}}$. Hence the dimension is $3n$, $N = n$, and $M = 2n^2$.

**Assumption justification.** The exact boundary blocks place every medial endpoint in a closed disk and therefore bound every center. Orientation is periodic, so one-period representatives give a compact center–orientation envelope; the raw trigonometric blocks are its periodic pullback. Every target medial segment has fixed positive length, hence $D_{iej}$ is $C^{1,1}$ on a neighborhood of each compact local-pose image, and radial composition preserves this property. The compact-domain realization above therefore supplies Assumption 3.1–Assumption 3.2. Gravity is smooth on the compact center domain, its extension supplies the smooth clauses and lower bound of Assumption 3.4, and $\psi_i \equiv 0$ is closed, convex, globally Lipschitz, and has the identity proximal map. The separated initial poses have strict pairwise and boundary clearance, verifying Assumption 3.3. Thus the extended capsule formulation satisfies all standing assumptions.

**Cloth resting.** Let $x = (q_1, \dots, q_n) \in \mathbb{R}^{2n}$ collect the planar node positions. The gravity energy is $h_i^{\mathrm{grav}}(q_i) = m_g e_2^T q_i$, and a stretch or bending energy on pair $(i, j)$ is defined as

$$h_{ij}^{\mathrm{spring}}(q_i - q_j) = \frac{k_{ij}}{2}\left(\mathscr{M}_\epsilon^{\mathrm{len}}(q_i - q_j) - \ell_{ij}\right)^2. \tag{25}$$

We add one stretch spring energy for each segment, and one bending spring energy for each pair of endpoints of two consecutive segments. Thus $N = n + (n-1) + (n-2) = 3n - 3$. For every node $i$ and nonincident segment $(q_j, q_{j+1})$, the theoretical regularized self-collision block takes $\ell_{\mathrm{act}} > 0$ and uses

$$\begin{aligned} f_{ij}^{\mathrm{self}} &= \mathscr{M}_\delta^{\mathrm{rad}}\left(\mathscr{M}_{\ell_{\mathrm{act}}}^{\mathrm{seg}}(q_i; q_j, q_{j+1})\right) - \mathscr{M}_\delta^{\mathrm{rad}}(s^2), \\ f_i^{\mathrm{circ}} &= \mathscr{M}_\delta^{\mathrm{rad}}(R^2) - \mathscr{M}_\delta^{\mathrm{rad}}(\|q_i - c_{\mathrm{circ}}\|^2). \end{aligned} \tag{26}$$

The ordered nonincident construction has $(n-1)(n-2)$ blocks and the boundary conditions add $n$, so $M = (n-1)(n-2) + n$.

**Assumption justification.** The exact circle blocks confine every node to a closed disk, making the global configuration envelope compact. Positive $\ell_{\mathrm{act}}$ makes every short-segment squared distance $C^{1,1}$ even at collapsed adjacent edges; radial composition and compact-domain realization then supply Assumption 3.1–Assumption 3.2. Gravity and the positive-$\epsilon$ spring blocks are $C^{1,1}$ on the compact objective images, so their extensions supply the smooth clauses and lower bound of Assumption 3.4; again $\psi_i \equiv 0$ has the required convex proximal structure. The initialized strand has strict boundary and nonincident-clearance margins. Moreover, $\mathscr{M}_{\ell_{\mathrm{act}}}^{\mathrm{seg}} \geq D_{\mathrm{seg}}$, so the regularization preserves that strict clearance and verifies Assumption 3.3. Thus the globally extended, positive-margin cloth formulation satisfies all standing assumptions.

**What the collision blocks encode.** The capsule and cloth collision blocks are point-to-segment clearances—endpoint-to-medial-segment for capsules and node-to-nonincident-segment for cloth—and the construction includes both directions of every nonincident pair. These are surrogate contact conditions: a positive point-to-segment clearance does not by itself forbid two segment interiors from crossing, since two finite segments can first contact interior-to-interior while all four endpoint-to-opposite-segment distances remain positive. We therefore treat these blocks as surrogate benchmarks rather than a formal no-crossing guarantee, and certify the absence of crossings on every reported output with an independent exact geometric check that rejects capsule overlap or medial-axis crossing and cloth self-crossing or degenerate segments (Appendix B).

## B Numerical Protocol and Additional Results

The evaluation contains capsule stacking and cloth resting at sizes 8, 16, and 24, with one deterministic SLSQP, BC-ADMM Pan & Wu (2026), IP-ADMM, and IP-ADMM-Ckpt run per case. All methods receive the same strictly feasible initial configuration and analytic first derivatives. Our empirical experimental setup and parameter choices are summarized in Table 3. Our software environment is Python with NumPy, SciPy for SLSQP, and PyTorch; CPU execution with a single numerical thread.

All tables and images are generated from canonical stored run records. Completion means that a solver produced a final output; feasibility evaluates the original nonlinear inequalities. Physical validity applies independent exact-geometry checks for capsule overlap and medial-axis crossing or cloth self-crossing and segment degeneracy. Verification is method-specific. Ordinary IP-ADMM uses its native b-FJ certificate together with split and original-space feasibility. SLSQP and IP-ADMM-Ckpt use a post-hoc same-$x$ KKT diagnostic: constraints with $f_i \le 2 \times 10^{-4}$ are declared active, nonnegative inequality multipliers are fitted by bounded least squares, and stationarity and complementarity are then evaluated. These reconstructed multipliers are not native solver multipliers. BC-ADMM is scored by the residual of the problem it actually solves, as explained next.

BC-ADMM Pan & Wu (2026) is the closest published optimizer to ours, so we add it as a baseline. It is also a parallel ADMM that starts from a strictly feasible point and keeps every iterate original-space feasible, but it gets there a different way. It adds a helper variable to each nonconvex constraint and turns that constraint into a biconvex one. For a cloth node and a segment that have to stay apart, the helper variable is a separating plane: the node has to sit on one side of it and the segment on the other. With the plane held fixed, both sides of the constraint are convex, so every block update is a small convex solve, and the plane itself has a closed form once the blocks are fixed. The springs get the same treatment, with the spring direction as the helper variable, and this reproduces the exact spring energy. The circular container is already convex and needs no helper variable, and gravity is linear, so it stays in the consensus solve. We keep the settings from that paper ($\kappa_y = 2$, $\kappa = 2.1$, $\gamma = 0.95$, $\eta = 1 - 10^{-5}$) and tune only the two penalties and the barrier margin, which is what it prescribes. Unfortunately, BC-ADMM does not apply to capsule stacking. A capsule pose carries an angle, and the world position of a capsule tip is not a convex function of that angle, even with the separating plane held fixed. Therefore, we do not know a BC-ADMM-compatible biconvex reformulation for this capsule formulation, and the BC-ADMM paper lists this reliance on a biconvex relaxation as a limitation of the method. We mark every capsule entry for BC-ADMM as not applicable. BC-ADMM also does not solve our problem. Its barrier keeps a fixed margin that never shrinks, so its answer stops a fixed distance short of the constraint boundary, and its stationarity condition belongs to that relaxed problem rather than to ours. We therefore score it by the residual it actually drives to zero, its own w-KKT residual: the larger of the relaxed stationarity $\|\nabla f + A^T \nabla_y g\|_\infty$ and the consensus gap $\|Ax - y\|_\infty$. It has no b-FJ certificate at all, and its barrier takes the place of multipliers, so it has no complementarity residual for our constraints either. Both of those columns are therefore empty for BC-ADMM.

Table 4 reports, for every run, the earliest iteration at which each certificate first reaches the shared $10^{-3}$ target. All 21 runs complete, and every reported output passes the original-space feasibility and physical-validity checks within the numerical tolerances of Table 3: the 18 runs of the three main methods, plus the 3 BC-ADMM cloth runs. For the two smallest instances ($n = 8$) both IP-ADMM variants establish the w-KKT, b-FJ, and KKT certificates; for the larger instances they still establish w-KKT and b-FJ, but KKT is not reached within the run. Even SLSQP, an SQP baseline, does not robustly certify KKT on capsule stacking at sizes 16 and 24, and on the larger instances reaches KKT only for cloth resting. This is consistent with CQ degeneracy or difficulty of KKT certification, but our experiments do not distinguish these explanations. It suggests that requiring KKT is a fragile stopping criterion in practice. SLSQP has no per-iteration history and no b-FJ certificate, so it reports its final-iterate result; Term. is the terminal execution iteration. BC-ADMM is scored against its own relaxed target, which it reaches only on the smallest cloth instance; at the larger sizes its residual settles above the target and the run stops without a certificate. Its b-FJ and KKT columns are empty for the structural reasons given above. This is the trade we want to show. BC-ADMM assumes the nonconvex constraints are biconvex and exploits that structure, which we never assume, and it only ever has to certify a relaxed problem, one whose constraints are all strictly inactive by construction. The convergence of BC-ADMM is not cheap: on cloth-8 it needs 178k

Table 3: Empirical nondimensional benchmark and solver settings used in the canonical runs; these numerical choices are not claimed to satisfy the theorem-sufficient parameter construction. Values written $a/b/c$ are the per-size settings for $n = 8/16/24$.

| Group | Parameter | Value |
|---|---|---|
| Capsule model | $n$ | 8, 16, 24 |
| | length, diameter, container radius | 1.05, 0.2, 2 |
| | constraint smoothing $\delta$ | 0.25 |
| Cloth model | $n$ | 8, 16, 24 |
| | mass, gravity, container radius | 1, 1, 2 |
| | stretch rest value, stiffness | 0.32, 100 |
| | bending stiffness, spring $\epsilon$ | 4, 0.01 |
| | clearance, constraint smoothing $\delta$ | 0.055, 0.25 |
| SLSQP | `maxiter`, `ftol` | $2 \times 10^4$, $10^{-12}$ |
| Ordinary capsule | $\rho_0, L_0, G, \eta, \sigma, \xi, p$ | 80/80/160, $10^6$, 0.005, $3 \times 10^5$, 1, 0.1/0.04/0.04, 0.5 |
| | $\theta_1, \theta_2$, `maxiter` | 1.01, 0.01, $1.5 \times 10^6$/2/2 |
| Ordinary cloth | $\rho_0, L_0, G, \eta, \sigma, \xi, p$ | 40/40/160, $10^6$, 0.006, 15.5, 1, 0.1, 0.6/0.5/0.5 |
| | $\theta_1, \theta_2$, `maxiter` | 1.01, 0.1, $1.2 \times 10^5$/2/$10^5$ |
| Ckpt capsule | $\rho_0, L_0, G, \eta, \chi, \tau_{cp}, \ell_{cp}$ | 320, $10^6$, 0.005, 1, 1.000003, 1000, 9.765625 |
| | $\theta_1, \theta_2$, `maxiter` | 1.01, 0.01, $8 \times 10^5$/1000/2000 |
| Ckpt cloth | $\rho_0, L_0, G, \eta, \chi, \tau_{cp}, \ell_{cp}$ | 40/40/160, $10^6$/$10^6$/$4 \times 10^6$, 0.006, 1, 1.000003, 2, 625/625/156.25 |
| | $\theta_1, \theta_2$, `maxiter` | 1.01, 0.1, $8 \times 10^5$/$10^4$/$2 \times 10^5$ |
| BC-ADMM cloth | $\beta, \beta_y, \gamma, \eta, \kappa, \kappa_y$ | 100, 100, 0.95, 0.99999, 2.1, 2 |
| | margin $\varepsilon$, $\sigma$, `maxiter` | $10^{-3}$, $10^{-6}$, $3 \times 10^6$ |
| Verification | active, violation, b-FJ, Stat., Comp. | $2 \times 10^{-4}$, $10^{-7}$, $10^{-3}$, $10^{-3}$, $10^{-3}$ |
| Physical gate | capsule, cloth tolerance | $10^{-10}$, $10^{-9}$ |
| Runtime | precision, device, CPU threads | float64, CPU, 1 |

iterations to bring its relaxed residual to $10^{-3}$, while ordinary IP-ADMM certifies b-FJ, a statement about the original constraints, at 60k. BC-ADMM stalls because it refreshes the separating planes only when the merit strictly decreases. Once the iterate reaches the optimum for the current planes, the merit stops decreasing, so the planes are frozen for good and the residual that also asks them to be optimal stays far above the target. In contrast, our algorithm assumes less and certify more: our b-FJ certificate is a statement about the original constraints, and Theorem 4.2 bounds how many iterations it takes to reach a given level.

Figure 4 shows ordinary IP-ADMM snapshots from the start to the final output. Figure 5 plots the w-KKT, b-FJ, and KKT certificates against the total solver iteration for every instance and mode, with the shared $10^{-3}$ target marked; at the larger sizes the KKT curve stalls above the target while the other two drop below it. Figure 6 plots how far each iterate still is from the final output, measured by the $L_2$ distance between the current and final solution vectors and scaled so the start is 1. This distance drops below $5\%$ early and then flattens near zero, long before KKT is reached (among the plotted IP-ADMM methods, KKT is reached only at $n = 8$). The configuration already looks like the final output well before stationarity is certified, so an early stop is a reasonable heuristic in practice. The catch is that this distance can only be measured after the run, because it needs the final solution. On the same axes we therefore also plot a quantity available during the run: the relative step $r^k = \|x^k - x^{k-\Delta}\| / \max\{1, \|x^k\|\}$, where $\Delta$ is the gap between stored samples. It uses only the current and previous iterate. The plot shows that $r^k$ drops at the same time as the distance to the final output, so a simple threshold on $r^k$ is a practical stopping rule, as in earlier physics solvers Narain et al. (2016); Overby et al. (2017); Pan & Wu (2026). In short, a full KKT certificate can take much longer than the moment the motion has effectively stopped, and that moment can be detected online.

Table 4: All 21 canonical runs: the 18 runs of the three main methods on both families, plus 3 BC-ADMM cloth runs. Term. is the terminal execution iteration. The w-KKT, b-FJ, and KKT columns report the earliest iteration (in thousands) at which each solver's own w-KKT residual, b-FJ certificate, and robust common-KKT residual first reach the shared $10^{-3}$ target; N/A means the target was not reached within the run. SLSQP has no per-iteration history and no b-FJ certificate, so it reports its final-iterate result. BC-ADMM Pan & Wu (2026) solves a fixed-margin biconvex relaxation, so its w-KKT entry, marked †, is the stationarity residual of that relaxed problem and is not comparable with the w-KKT of the other rows: at the same point the original-space w-KKT stays $O(1)$, because the fixed margin holds every constraint strictly inactive. BC-ADMM supplies neither a b-FJ certificate nor a complementarity residual for the original problem, and it does not apply to the capsule family, whose poses carry a rotation for which no biconvex relaxation is known. Viol. is the maximum original-space inequality violation.

| Instance | Method | Objective | Term. | Viol. | w-KKT | b-FJ | KKT |
|---|---|---|---|---|---|---|---|
| Capsule 8 | SLSQP | -11.90 | 22 | $5.55\times10^{-16}$ | 22 | N/A | 22 |
| Capsule 8 | IP-ADMM | -11.92 | 1639k | 0 | 139k | 40k | 1329k |
| Capsule 8 | IP-ADMM-Ckpt | -11.92 | 1045k | 0 | 240k | 60k | 1035k |
| Capsule 8 | BC-ADMM | N/A | N/A | N/A | N/A | N/A | N/A |
| Capsule 16 | SLSQP | -19.20 | 32 | $8.88\times10^{-16}$ | 32 | N/A | N/A |
| Capsule 16 | IP-ADMM | -18.74 | 379k | 0 | 379k | 60k | N/A |
| Capsule 16 | IP-ADMM-Ckpt | -18.74 | 710k | 0 | 640k | 120k | N/A |
| Capsule 16 | BC-ADMM | N/A | N/A | N/A | N/A | N/A | N/A |
| Capsule 24 | SLSQP | -22.45 | 25 | $2.90\times10^{-15}$ | 25 | N/A | N/A |
| Capsule 24 | IP-ADMM | -22.40 | 876k | 0 | 800k | 100k | N/A |
| Capsule 24 | IP-ADMM-Ckpt | -22.40 | 1285k | 0 | 1120k | 140k | N/A |
| Capsule 24 | BC-ADMM | N/A | N/A | N/A | N/A | N/A | N/A |
| Cloth 8 | SLSQP | -14.25 | 67 | $4.77\times10^{-15}$ | 67 | N/A | 67 |
| Cloth 8 | IP-ADMM | -14.38 | 291k | 0 | 160k | 60k | 261k |
| Cloth 8 | IP-ADMM-Ckpt | -14.38 | 971k | 0 | 160k | 60k | 931k |
| Cloth 8 | BC-ADMM | -14.35 | 178k | 0 | 178k$^{\dagger}$ | N/A | N/A |
| Cloth 16 | SLSQP | -27.19 | 105 | $4.33\times10^{-14}$ | 105 | N/A | 105 |
| Cloth 16 | IP-ADMM | -27.87 | 585k | 0 | 560k | 320k | N/A |
| Cloth 16 | IP-ADMM-Ckpt | -27.87 | 595k | 0 | 560k | 320k | N/A |
| Cloth 16 | BC-ADMM | -28.09 | 2099k | 0 | N/A | N/A | N/A |
| Cloth 24 | SLSQP | -39.77 | 135 | $5.88\times10^{-15}$ | 135 | N/A | 135 |
| Cloth 24 | IP-ADMM | -39.93 | 1100k | 0 | 960k | 280k | N/A |
| Cloth 24 | IP-ADMM-Ckpt | -39.93 | 2200k | 0 | 1740k | 460k | N/A |
| Cloth 24 | BC-ADMM | -38.80 | 865k | 0 | N/A | N/A | N/A |

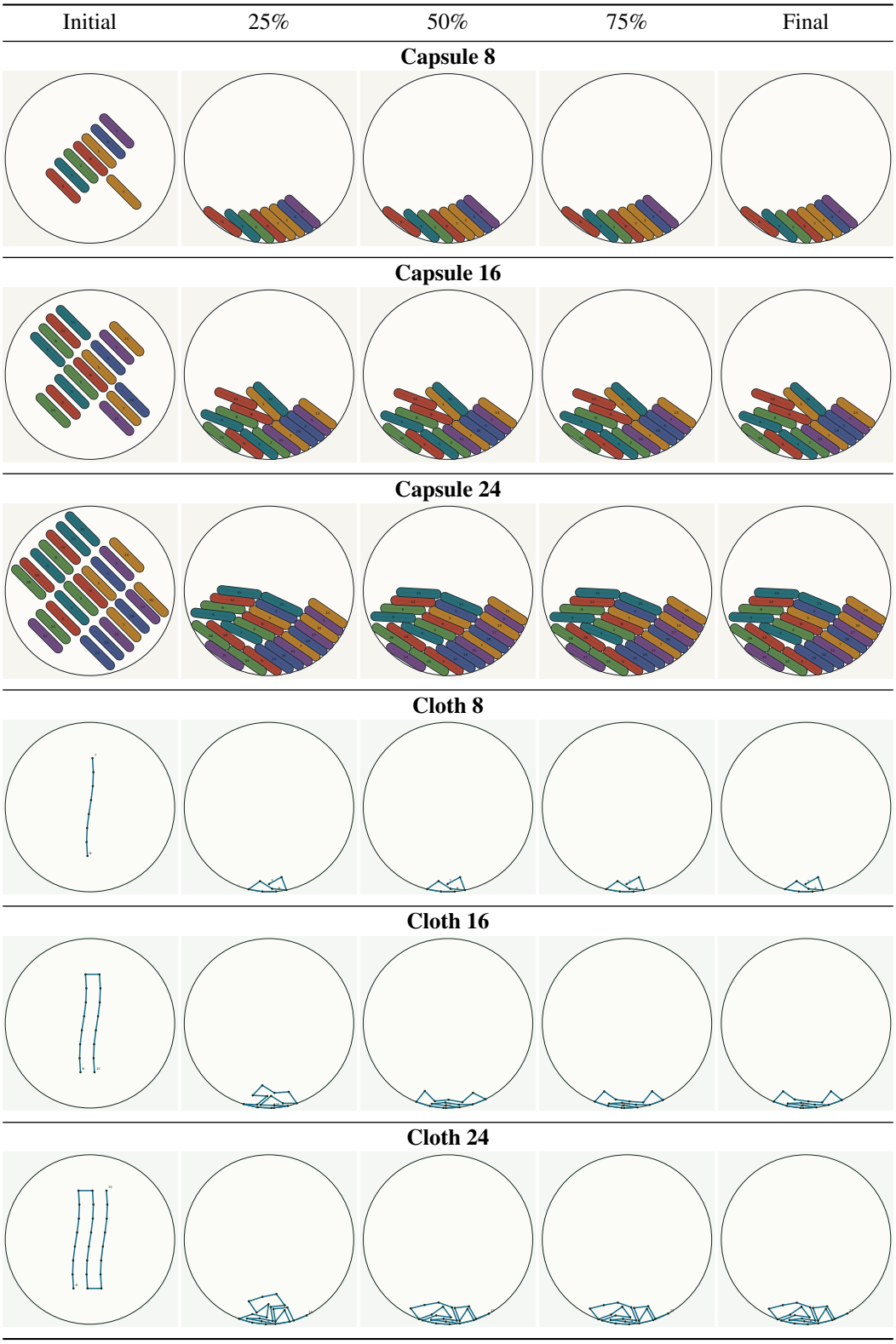


Figure 4: Ordinary IP-ADMM trajectories for all six instances. Intermediate columns are the stored samples nearest 25%, 50%, and 75% of total execution; the last column is the final output.

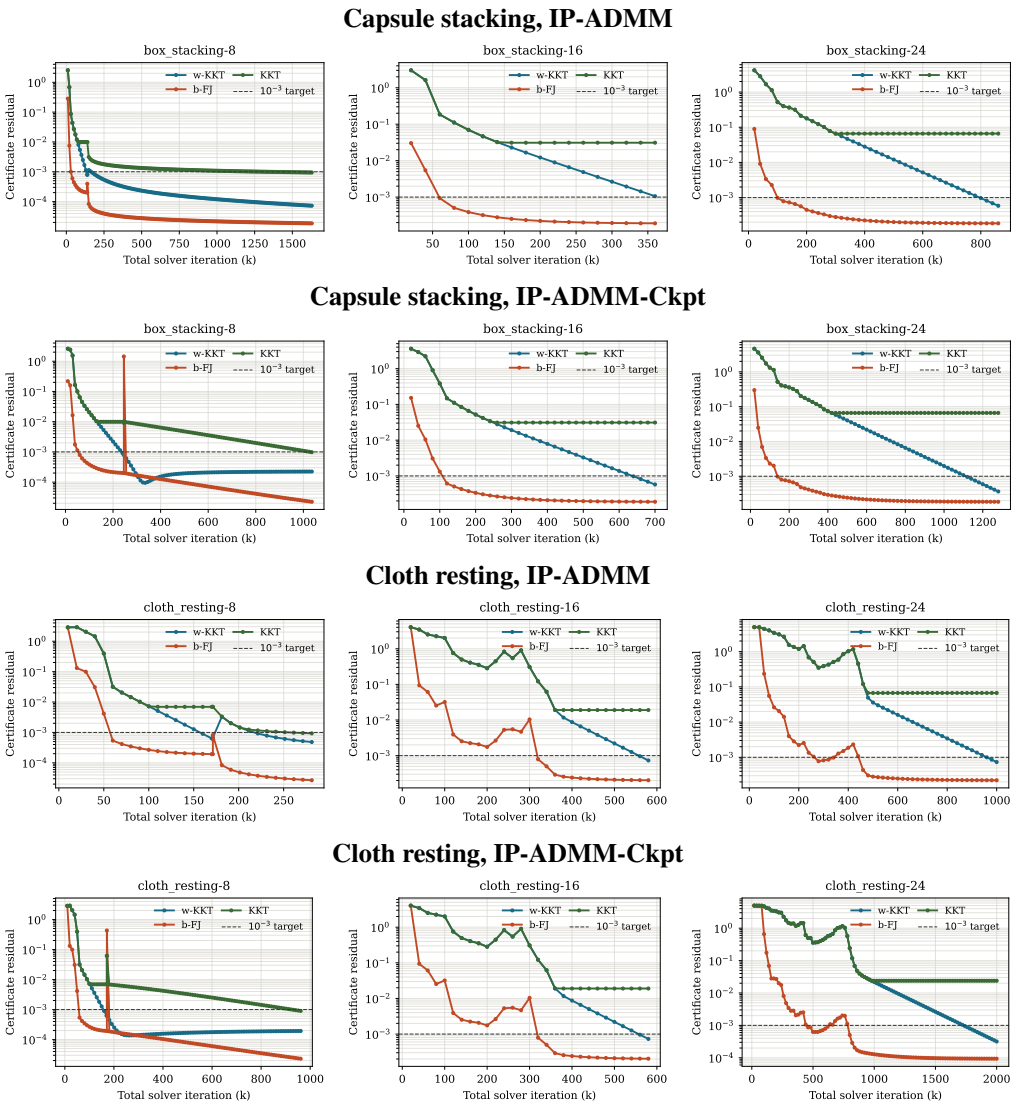


Figure 5: w-KKT, b-FJ, and KKT certificates versus total solver iteration for every instance and ADMM mode. The dashed line is the shared $10^{-3}$ target. KKT reaches the target only at $n = 8$; at larger sizes it stalls above the target while w-KKT and b-FJ still descend below it.

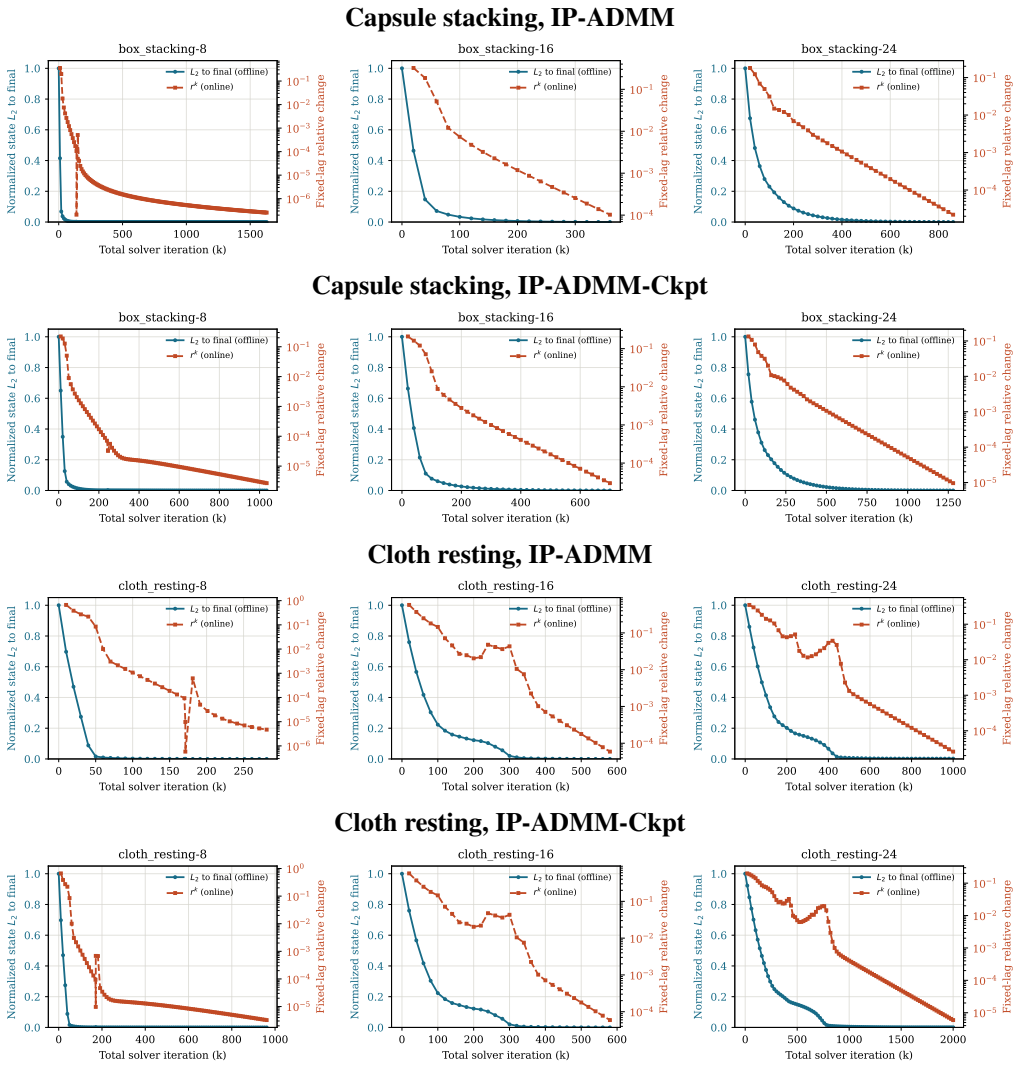


Figure 6: Left axis (blue): normalized state-space $L_2$ distance (Euclidean distance between the current and final solution vectors) of each iterate to the final output, versus total solver iteration (in thousands), for every instance and ADMM mode. Right axis (orange, log): the online relative-change diagnostic $\|x^k - x^{k-\Delta}\| / \max\{1, \|x^k\|\}$ over consecutive samples ($\Delta$ the history sampling interval). Both quantities decay together well before KKT is established (cf. Figure 5), so the online diagnostic flags physically meaningful settling long before stationarity is certified.

# C Parameter Sensitivity: Theorem-Sufficient versus Practical Tuning

Our convergence analysis uses one specific, data-dependent choice of parameters (Equation 28). This choice is enough to prove the finite-time guarantees, but it is deliberately conservative. The main runs in Appendix B instead use values that we tuned aggressively for speed. This appendix connects the two settings on the $n = 8$ instances. We report one run of the exact theorem-sufficient choice, and a sensitivity study around the practical baseline. Each ablation changes only one certified parameter, so we can see how that parameter alone affects the certified schedule. Table 5 lists the certificates, and Figure 7 plots the native b-FJ certificate against the total iteration count in each sweep.

**Theorem-sufficient construction.** Evaluating Equation 28 on cloth-8 gives $\rho^0 \approx 4.7 \times 10^7$, $L^0 \approx 1.2 \times 10^5$, and $G \approx 96$, compared with the practical $\rho^0 = 40$, $L^0 = 10^6$, $G = 6 \times 10^{-3}$; capsule-8 is even larger ($\rho^0 \approx 1.6 \times 10^9$). This run still executes and remains numerically feasible (zero original-space violation, positive split margin), but its native b-FJ certificate levels off near $7 \times 10^{-2}$ (cloth) and $4 \times 10^{-2}$ (capsule). That is about three orders of magnitude looser than the practical baselines at the same budget (Table 5). This agrees with the analysis rather than contradicting it. The certificate the solver reports is exactly $\mathrm{Crit}_{\text{b-FJ}}$ from Definition H.1, and Theorem 4.2 only bounds $\min_{j \le K} \mathrm{Crit}^j_{\text{b-FJ}} \le C\, K^{-1/4}$ for a fixed constant $C$ that does not depend on $K$. At the budget $K \approx 1.2 \times 10^5$ we have $K^{-1/4} \approx 0.054$, and the observed minimum $\approx 6.8 \times 10^{-2}$ gives $C \approx 1.3$. So reaching a tolerance $\epsilon$ needs about $K \approx (C/\epsilon)^4$ iterations. The constant is large here for a structural reason. At the conservative $G$, the wall force on every near-active block saturates at $-\sqrt{2L^kG}$, which is far above the true multipliers, so the normalized stationarity term $\alpha^k \|\Xi^k\|$ stays near a floor set by the geometry. The $G$ sweep in Table 5 shows this directly: shrinking $G$ tightens the native b-FJ (cloth $G$=$2 \times 10^{-3} \to 1.5 \times 10^{-5}$ versus $G$=$1.8 \times 10^{-2} \to 4.8 \times 10^{-5}$). In short, the conservative $G$ inflates the rate constant, but it does not contradict the $\mathcal{O}(K^{-1/4})$ rate.

**Practical sensitivity.** Around the practical baseline the certified phase is robust. The exponent behaves as the theory predicts: the intermediate $p \approx 1/2$ gives the smallest b-FJ, while both $p = 0.33$ and $p = 0.9$ are worse on at least one certificate. The relaxation $\sigma$ has almost no effect, and moderate changes in $\rho^0$, $\theta_2$, and the checkpoint values $\chi$ and $\tau_{cp}$ move the certificates by less than a factor of three. The most sensitive knob is $L^0$: values far from the baseline make b-FJ worse (cloth $L^0 = 10^7 \to 2.0 \times 10^{-4}$). Every completed ablation passes the shared feasibility and physical checks. The capsule family adds one more check: it keeps every split state inside the compact domain that its certified contact-curvature bounds rely on. Perturbations that weaken the barrier or inflate $L$—reducing $G$ to $G/3$, doubling the checkpoint $\ell_{cp}$, or halving $\rho^0$ to 40—push the iterates just outside that domain (margin $\approx -10^{-6}$) and abort. We therefore leave these directions out of the capsule sweep, while the directions that strengthen the barrier or lower $L$ complete normally. Cloth has no such guard and tolerates both directions.

# D Applications in Control and Motion Planning

Although our experiments are mostly focused on physical simulation, our method can be extended to many other applications, e.g. differentiable trajectory optimization, model-predictive control with obstacles. This section illustrates one of such example, where we optimize a trajectory subject to contact or collision constraints. Specifically, we assume a planar UAV follows a piecewise-linear trajectory $x_0, x_1, \ldots, x_n \in \mathbb{R}^2$ with a fixed start $x_0$. We optimize the free nodes $x = (x_1, \ldots, x_n)$ to reach a target $x_{\mathrm{goal}}$ while avoiding circular obstacles. The objective is a small smoothness (squared-acceleration) term plus a sharp goal term,

$$\sum_{j=2}^{n} \frac{w}{2} \left\| x_j - 2x_{j-1} + x_{j-2} \right\|^2 \; + \; \gamma \left\| x_n - x_{\mathrm{goal}} \right\|_1, \tag{27}$$

and each obstacle with center $c$ and radius $r$ contributes, for every segment $(x_{s-1}, x_s)$, the clearance constraint $\mathscr{M}^{\mathrm{rad}}_\delta(D_{\mathrm{seg}}(c; x_{s-1}, x_s)) - \mathscr{M}^{\mathrm{rad}}_\delta(r^2) \ge 0$, i.e. the same radially mollified point-to-segment distance used for capsule stacking (Appendix A). Goal reaching is more important than

Table 5: Size-8 parameter sensitivity. Each ablation changes one certified parameter; the theorem-sufficient row uses the full Equation 28 construction. native b-FJ is the solver's banded Fritz–John certificate; w-KKT and KKT are the reconstructed common residuals; a checkmark marks runs meeting the $10^{-3}$ b-FJ target with valid feasibility and physical checks.

| Parameter | Value | native b-FJ | w-KKT | KKT | OK |
|---|---|---|---|---|---|
| **Cloth resting** ($n$ = 8) | | | | | |
| Ordinary baseline | – | $2.59\times10^{-5}$ | $4.56\times10^{-4}$ | $9.07\times10^{-4}$ | ✓ |
| Ckpt baseline | – | $2.24\times10^{-5}$ | $1.93\times10^{-4}$ | $8.74\times10^{-4}$ | ✓ |
| Theorem-sufficient | – | 0.07 | 0.05 | 0.05 | – |
| $\rho_0$ | 20 | $2.50\times10^{-5}$ | $3.91\times10^{-4}$ | $9.07\times10^{-4}$ | ✓ |
| | 80 | $2.78\times10^{-5}$ | $6.12\times10^{-4}$ | $9.07\times10^{-4}$ | ✓ |
| $L_0$ | $10^5$ | $6.87\times10^{-5}$ | $4.64\times10^{-4}$ | $2.68\times10^{-3}$ | ✓ |
| | $10^7$ | $1.98\times10^{-4}$ | 0.02 | 0.02 | ✓ |
| $G$ | 0.002 | $1.50\times10^{-5}$ | $2.98\times10^{-4}$ | $5.13\times10^{-4}$ | ✓ |
| | 0.018 | $4.75\times10^{-5}$ | $9.40\times10^{-4}$ | $1.59\times10^{-3}$ | ✓ |
| $\xi$ | 0.05 | $3.28\times10^{-5}$ | $3.56\times10^{-4}$ | $1.26\times10^{-3}$ | ✓ |
| | 0.2 | $2.24\times10^{-5}$ | $6.75\times10^{-4}$ | $6.75\times10^{-4}$ | ✓ |
| $p$ | 0.33 | $8.68\times10^{-5}$ | $1.05\times10^{-4}$ | $3.60\times10^{-3}$ | ✓ |
| | 0.66 | $2.24\times10^{-5}$ | $6.70\times10^{-4}$ | $6.70\times10^{-4}$ | ✓ |
| | 0.9 | $4.97\times10^{-5}$ | $4.11\times10^{-3}$ | $4.11\times10^{-3}$ | ✓ |
| $\theta_2$ | 1 | $2.74\times10^{-5}$ | $5.78\times10^{-4}$ | $9.07\times10^{-4}$ | ✓ |
| | 5 | $3.61\times10^{-5}$ | $1.31\times10^{-3}$ | $1.31\times10^{-3}$ | ✓ |
| $\sigma$ | 0.5 | $2.59\times10^{-5}$ | $4.56\times10^{-4}$ | $9.07\times10^{-4}$ | ✓ |
| | 1.5 | $2.59\times10^{-5}$ | $4.56\times10^{-4}$ | $9.07\times10^{-4}$ | ✓ |
| $\chi$ | 1.00001 | $1.03\times10^{-5}$ | $6.70\times10^{-4}$ | $6.70\times10^{-4}$ | ✓ |
| | 1.00003 | $9.51\times10^{-6}$ | $8.65\times10^{-4}$ | $8.65\times10^{-4}$ | ✓ |
| $\tau_{cp}$ | 1 | $2.25\times10^{-5}$ | $1.93\times10^{-4}$ | $8.78\times10^{-4}$ | ✓ |
| | 4 | $2.24\times10^{-5}$ | $1.93\times10^{-4}$ | $8.74\times10^{-4}$ | ✓ |
| $\ell_{cp}$ | 312.5 | $3.02\times10^{-5}$ | $1.13\times10^{-4}$ | $1.00\times10^{-3}$ | – |
| | 1250 | $1.68\times10^{-5}$ | $1.93\times10^{-4}$ | $6.23\times10^{-4}$ | ✓ |
| **Capsule resting** ($n$ = 8) | | | | | |
| Ordinary baseline | – | $1.86\times10^{-5}$ | $7.24\times10^{-5}$ | $9.44\times10^{-4}$ | ✓ |
| Ckpt baseline | – | $2.18\times10^{-5}$ | $2.24\times10^{-4}$ | $9.52\times10^{-4}$ | ✓ |
| Theorem-sufficient | – | 0.04 | $7.86\times10^{-4}$ | 0.01 | – |
| $\rho_0$ | 60 | $2.88\times10^{-5}$ | $2.14\times10^{-4}$ | $1.40\times10^{-3}$ | ✓ |
| | 160 | $3.17\times10^{-5}$ | $4.91\times10^{-4}$ | $1.40\times10^{-3}$ | ✓ |
| $L_0$ | $10^5$ | $9.15\times10^{-5}$ | $8.09\times10^{-4}$ | $4.35\times10^{-3}$ | ✓ |
| | $3\times10^6$ | $1.77\times10^{-5}$ | $2.36\times10^{-4}$ | $8.12\times10^{-4}$ | ✓ |
| $G$ | 0.015 | $5.04\times10^{-5}$ | $3.54\times10^{-4}$ | $2.43\times10^{-3}$ | ✓ |
| $\xi$ | 0.05 | $3.88\times10^{-5}$ | $1.42\times10^{-4}$ | $1.96\times10^{-3}$ | ✓ |
| | 0.2 | $2.42\times10^{-5}$ | $5.40\times10^{-4}$ | $9.98\times10^{-4}$ | ✓ |
| $p$ | 0.33 | $7.28\times10^{-5}$ | $3.78\times10^{-5}$ | $3.79\times10^{-3}$ | ✓ |
| | 0.66 | $2.18\times10^{-5}$ | $1.20\times10^{-3}$ | $1.20\times10^{-3}$ | ✓ |
| | 0.9 | $3.16\times10^{-5}$ | $2.06\times10^{-3}$ | $2.06\times10^{-3}$ | ✓ |
| $\theta_2$ | 1 | $3.16\times10^{-5}$ | $4.86\times10^{-4}$ | $1.40\times10^{-3}$ | ✓ |
| | 5 | $3.81\times10^{-5}$ | $1.06\times10^{-3}$ | $1.40\times10^{-3}$ | ✓ |
| $\sigma$ | 0.5 | $2.94\times10^{-5}$ | $2.67\times10^{-4}$ | $1.40\times10^{-3}$ | ✓ |
| | 1.5 | $2.94\times10^{-5}$ | $2.67\times10^{-4}$ | $1.40\times10^{-3}$ | ✓ |
| $\chi$ | 1.00001 | $2.19\times10^{-5}$ | $5.86\times10^{-4}$ | $5.86\times10^{-4}$ | ✓ |
| | 1.00003 | $2.62\times10^{-5}$ | $1.10\times10^{-3}$ | $1.10\times10^{-3}$ | ✓ |
| $\tau_{cp}$ | 500 | $8.35\times10^{-5}$ | $2.03\times10^{-4}$ | $4.18\times10^{-3}$ | ✓ |
| | 2000 | $8.35\times10^{-5}$ | $2.03\times10^{-4}$ | $4.18\times10^{-3}$ | ✓ |
| $\ell_{cp}$ | 4.8828125 | $1.17\times10^{-4}$ | $2.50\times10^{-4}$ | $5.85\times10^{-3}$ | ✓ |

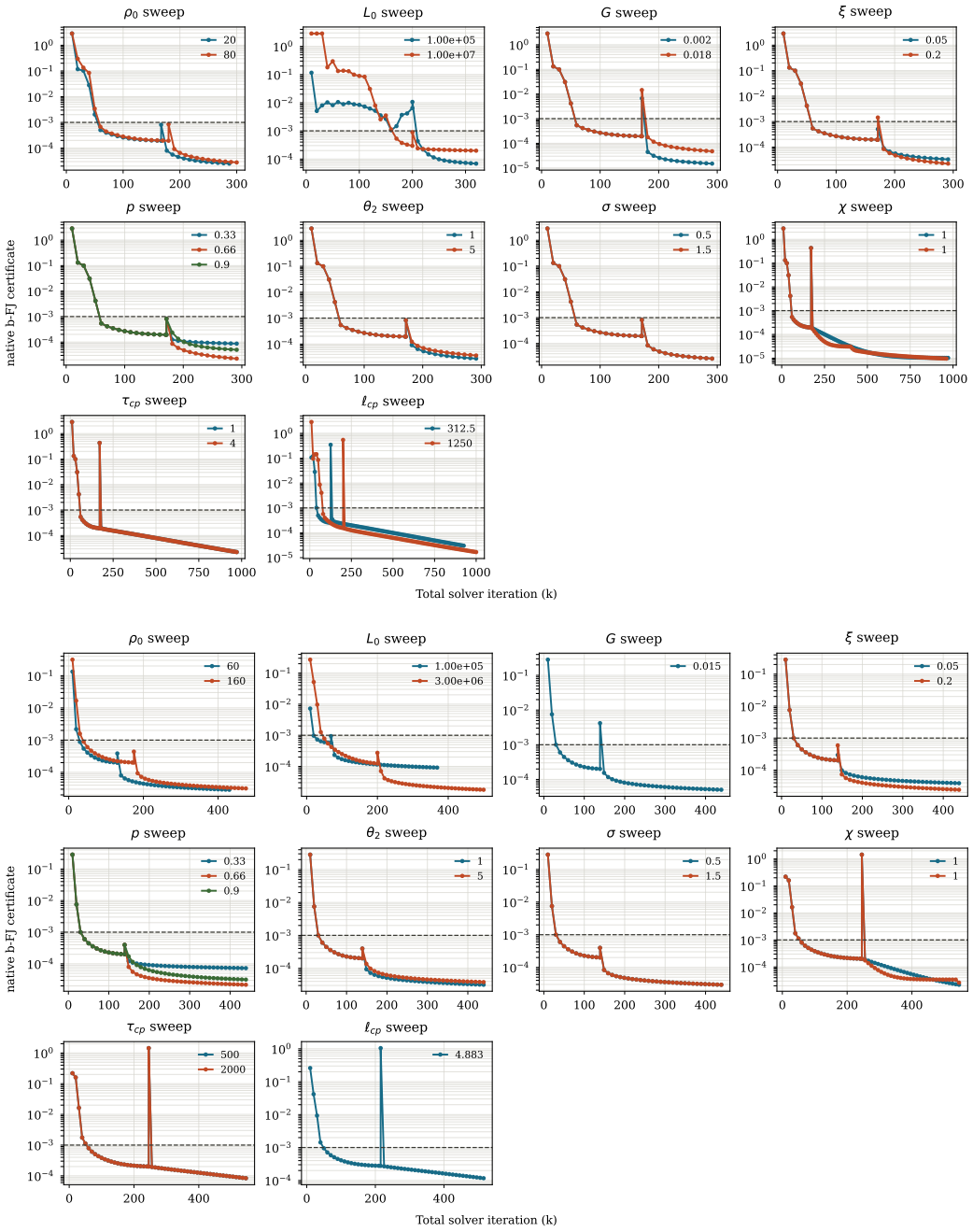


Figure 7: Native b-FJ certificate versus total solver iteration (in thousands) for the cloth-8 (top) and capsule-8 (bottom) parameter sweeps. Each panel overlays the values of one certified parameter around the practical baseline; the dashed line is the shared $10^{-3}$ target.

smoothness, so we make it sharp: the $\ell_1$ term is exactly proximable, and its proximal map is soft-thresholding. This is exactly an instance of Problem 1: the smoothness terms are the smooth blocks $h_i$, the goal term is a convex nonsmooth block $\psi = \gamma\|\cdot - x_{\text{goal}}\|_1$, and the clearance blocks are the nonlinear inequalities. We use $n = 20$ nodes and two obstacles placed on the straight start–goal line, so the smoothness-optimal straight path is infeasible and the obstacle constraints are active at the optimum.

Table 6: UAV trajectory planning ($n$ = 20 nodes, two obstacles). The nonsmooth goal term $\gamma\|x_n - x_{\text{goal}}\|_1$ makes goal reaching exact. native b-FJ is the solver's banded Fritz–John certificate (IP-ADMM variants only); SLSQP and BC-ADMM Pan & Wu (2026) have no native certificate, and the post-hoc KKT reconstruction is unreliable at the $\ell_1$ kink. Clearance is the exact minimum obstacle margin (nonnegative means collision-free). BC-ADMM's clearance is two orders of magnitude larger than the rest because its barrier holds a fixed margin, so it stops short of the active constraint rather than resolving it; that standoff is the reason it certifies nothing here.

| Method | Goal gap | Clearance | Path len. | native b-FJ | Iters |
|---|---|---|---|---|---|
| SLSQP | $2.17\times10^{-14}$ | $2.44\times10^{-15}$ | 10.84 | – | 79 |
| IP-ADMM | $2.13\times10^{-5}$ | $1.82\times10^{-5}$ | 10.84 | $1.15\times10^{-4}$ | 2200k |
| IP-ADMM-Ckpt | $8.08\times10^{-6}$ | $2.63\times10^{-7}$ | 10.84 | $8.35\times10^{-5}$ | 2200k |
| BC-ADMM | $7.32\times10^{-15}$ | $1.99\times10^{-3}$ | 10.84 | – | 118k |

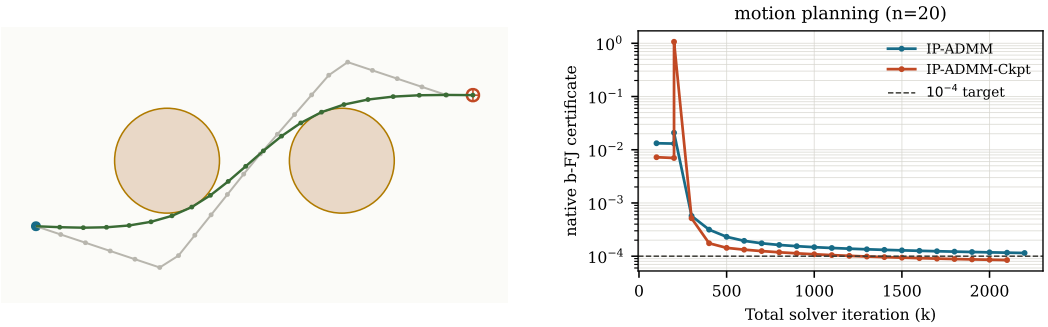


Figure 8: UAV trajectory planning. Left: the initial detour (gray) and the optimized IP-ADMM trajectory (green) with the two circular obstacles; the start is the filled dot and the target is the crossed circle. Right: the native banded Fritz–John certificate versus total solver iteration for the two IP-ADMM variants, with the shared $10^{-4}$ target marked.

Table 6 and Figure 8 summarize one deterministic run of SLSQP, BC-ADMM Pan & Wu (2026), IP-ADMM, and IP-ADMM-Ckpt from the same strictly feasible initial detour. All four converge to essentially the same taut, collision-free path (the IP-ADMM and BC-ADMM trajectories match SLSQP to within the plot line width), and the $\ell_1$ term drives the final node onto the target. The checkpoint variant certifies the solution below the $10^{-4}$ banded Fritz–John level, while the ordinary variant is still approaching it at the same iteration budget. This is qualitatively consistent with the improved post-checkpoint exponent in Table 2. This instance also fits BC-ADMM, since each obstacle-clearance constraint becomes a separating plane between the obstacle disk and one trajectory segment, and the $\ell_1$ goal term is convex with a soft-thresholding proximal map. BC-ADMM lands on the same taut path and reaches the target exactly, but it certifies nothing here due to the non-smooth objective function. Its assumptions ask for a twice-differentiable $g$, which this objective is not, so the sharp term we handle exactly falls outside the class it covers.

## E Parallel Wall-Clock Scaling

The main experiments are small and run on a single thread, so we use this appendix to measure real wall-clock time on the cloth-resting testbed of Appendix B: where each iteration spends its time, and how much faster it runs when we add CPU threads. We do not solve to convergence here—we fix the penalty schedule, run a fixed number of sweeps, and split each sweep into the four steps of Algorithm 1: the local objective update, the constraint update, the global $x$-solve, and the rest (residuals, movement, and dual updates). In the cloth resting problem, we grow the cloth chain over $n \in \{16, 24, 32, 48, 64\}$ nodes ($2n$ variables, $3n-3$ objective blocks, and $(n-1)(n-2)+n$ collision constraints), going well past the $n \le 24$ sizes used for the convergence study. For each size and each thread count in $\{1, 2, 4, 8, 16\}$ we set `torch.set_num_threads`, and run 300 sweeps with `perf_counter`. Each point runs in its own process, and we report the fastest of nine repeats.

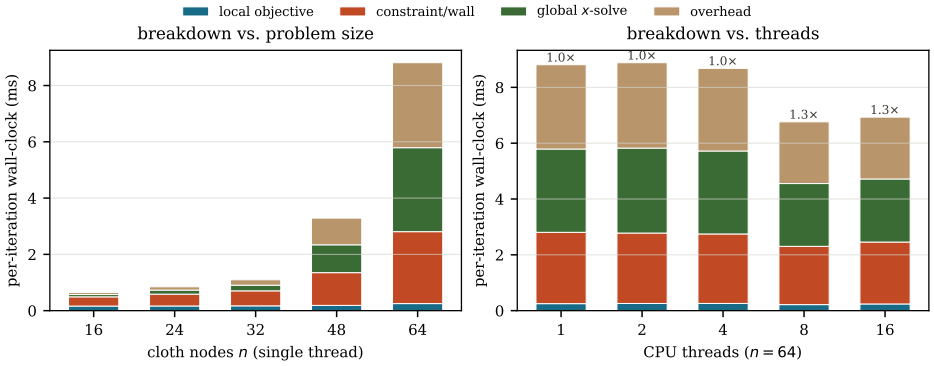


Figure 9: Per-iteration wall-clock time split into the four sweep steps. Left: versus cloth size $n$ on a single thread; the batched local objective update stays small while the constraint update, global $x$-solve, and assembly overhead grow. Right: versus CPU thread count at $n = 64$, where adding threads cuts the per-iteration time by up to $1.3\times$ (annotated).

Figure 9 (left) shows how one iteration divides among the four steps on a single thread. The local objective update is applied to all blocks at once, so it stays cheap: it falls from about a quarter of the time at $n = 16$ to a thin sliver at $n = 64$. As the problem grows, the constraint update and the global $x$-solve—together with the assembly work counted as overhead—take up most of the time. The independent per-block work that the method exposes is already fast; what remains is the shared solve and the matrix assembly around it. Figure 9 (right) shows the same split at $n = 64$ as we add CPU threads. For the small sizes each block touches so little data that the cost of splitting the work across threads outweighs the gain, so extra threads do not help. At $n = 64$ the batched linear algebra is large enough to spread across cores, and the constraint, global, and overhead steps all get faster, cutting the per-iteration time by up to $1.3\times$ (at 8 threads). This is a real speedup on a single machine, and it supports the parallelizability claim with wall-clock numbers rather than structure alone: most of the cost lies in the independent local blocks and one shared solve, and simply adding threads already makes each iteration faster.

## F Explicit Parameter Constructions

This appendix records the sufficient data-dependent choices referenced by the main convergence statements. Empty maxima are read as zero. The data aggregates in the ordinary display are shared by both modes, while the algorithmic tunings and wall-height constructions are mode-specific.

### F.1 Ordinary IP-ADMM

$$
\begin{cases}
H_{\max} = \max\limits_{i\in\mathcal{I}_{\mathrm{o}}} H_i, \quad L_{\max} = \max\limits_{i\in\mathcal{I}_{\mathrm{c}}} L_i, \quad M_{\max} = \max\limits_{i\in\mathcal{I}_{\mathrm{c}}} M_i, \quad \bar{M} = \sum\limits_{i\in\mathcal{I}_{\mathrm{c}}} M_i^2, \\
S_0 = \sum\limits_{i\in\mathcal{I}_{\mathrm{o}}} \varphi_i(A_i x^1), \quad \widehat{G}_0 = \sum\limits_{i\in\mathcal{I}_{\mathrm{o}}} \widehat{C}_i^2, \quad G_\psi = \sum\limits_{i\in\mathcal{I}_{\mathrm{o}}} (C_i^\psi)^2, \\
\sigma = \sigma_1 = \sigma_3 = 1, \quad \sigma_2 = 0, \\
\theta_1 = \dfrac{121909}{121000}, \quad \theta_2 = 25, \quad \eta = 1209, \quad d_0 = \dfrac{1}{2}, \quad \varepsilon = \dfrac{1}{1209}, \\
L^0 = \max\left\{1, \dfrac{1}{m}, \dfrac{12 S_0}{m^2}\right\}, \qquad G = \dfrac{m^2 L^0}{2}, \\
\rho^0 = \max\left\{
\begin{array}{c}
1, 1209 H_{\max}, 1209 m L^0 L_{\max}, 1209 L^0 M_{\max}^2, \\
6\left(\dfrac{\eta}{2} + \dfrac{11}{2}\right)\dfrac{\widehat{G}_0}{G}, \\
6p^2 H_p(\xi)\dfrac{\frac{25}{2} G_\psi + \frac{25}{4} G L^0 \bar{M}}{G}
\end{array}
\right\}.
\end{cases}
\tag{28}
$$

This exact ordinary IP-ADMM construction is noncircular: by Assumption 3.4, $S_0 \geq 0$; moreover, $G > 0$ follows from $m > 0$, and the displayed maximum is finite. There is no weak-convexity or smoothing-smallness clause.

### F.2 IP-ADMM-CKPT

$$
\begin{cases}
\theta_1^{\mathrm{cp}} = \dfrac{101}{100}, \quad \theta_2^{\mathrm{cp}} = \dfrac{1}{100}, \quad \eta^{\mathrm{cp}} = \dfrac{1}{100}, \\
L^{0,\mathrm{cp}} = \max\left\{1, \dfrac{1}{m}, \dfrac{3S_0}{m^2}\right\}, \qquad G^{\mathrm{cp}} = \dfrac{m^2 L^{0,\mathrm{cp}}}{2}, \\
\Delta_G = \dfrac{G^{\mathrm{cp}} - S_0}{4}, \quad \overline{V} = S_0 + 2\Delta_G, \quad a_0 = \dfrac{51}{100}, \\
b_{\mathrm{cp}} = \sqrt{\dfrac{2G^{\mathrm{cp}}}{\ell_{\mathrm{cp}}}}, \quad r_{\mathrm{cp}} = \max\left\{H_{\max}, \sqrt{2G^{\mathrm{cp}}\ell_{\mathrm{cp}}}L_{\max}\right\}, \\
c_{\chi,\mathrm{cp}} = \theta_1^{\mathrm{cp}}(1 + r_{\mathrm{cp}}) - 1, \quad c_{\mathrm{mov}} = \sqrt{(r_{\mathrm{cp}} + c_{\chi,\mathrm{cp}})^2 + (1 + \theta_2^{\mathrm{cp}})^2}, \\
J_H = \left(\sum_{i \in \mathcal{I}_\circ} (H_i C_i^{\psi})^2\right)^{1/2}, \quad M_F = \|A\|_2 \sqrt{\widehat{G}_0}, \\
S_\Xi = \|A\|_2(c_{\mathrm{mov}}\tau_{\mathrm{cp}} + \eta^{\mathrm{cp}} J_H), \quad I_{\mathrm{cp}} = \dfrac{M_F + S_\Xi}{\gamma_{\mathrm{cq}}}, \\
Q_{\mathrm{cp}} = \sqrt{\widehat{G}_0} + M_{\max} I_{\mathrm{cp}} + (c_{\chi,\mathrm{cp}} + 1)\tau_{\mathrm{cp}}, \\
D_{\mathrm{cp}} = \dfrac{(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})^2}{2}, \\
\rho_{\mathrm{cp}}^0 = \max\left\{\begin{array}{c} 1, \dfrac{\eta^{\mathrm{cp}} G_\psi}{2\Delta_G}, \dfrac{D_{\mathrm{cp}}}{\Delta_G}, \dfrac{b_{\mathrm{cp}}}{m}, \dfrac{2b_{\mathrm{cp}}}{\delta_{\mathrm{cq}}}, \dfrac{8\overline{V}}{\delta_{\mathrm{cq}}^2}, \\ \dfrac{2(\chi+1)Q_{\mathrm{cp}}^2}{(1-\chi^{-1})\Delta_G}, \dfrac{4(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})}{\delta_{\mathrm{cq}}}, \left(\dfrac{L_{\max}\tau_{\mathrm{cp}}^2}{\delta_{\mathrm{cq}}}\right)^{1/4} \end{array}\right\}.
\end{cases}
\tag{29}
$$

Indeed, $L^{0,\mathrm{cp}} \geq 3S_0/m^2$ gives $G^{\mathrm{cp}} \geq 3S_0/2 > S_0$ when $S_0 > 0$, while $G^{\mathrm{cp}} > 0$ also when $S_0 = 0$. Thus $\Delta_G = (G^{\mathrm{cp}} - S_0)/4 > 0$. The guide constants through $D_{\mathrm{cp}}$ do not depend on the base checkpoint penalty; in particular, $I_{\mathrm{cp}}$ deliberately uses $S_\Xi$, rather than $S_\Xi/\rho_{\mathrm{cp}}^0$. The nine displayed branches of $\rho_{\mathrm{cp}}^0$ are exactly the mode-specific requirements used by the checkpoint proof; no separate ordinary-derived core floor is required.

**Geometric checkpoint schedule.** If $j \geq 0$ checkpoints have been committed, the level parameters are

$$
\begin{aligned}
\rho_j &:= \rho_{\mathrm{cp}}^0 \chi^j, \qquad L_j := \ell_{\mathrm{cp}} \rho_j^2, \qquad \mu_j := \frac{\eta^{\mathrm{cp}}}{\rho_j}, \qquad \beta_j := \frac{b_{\mathrm{cp}}}{\rho_j}, \\
(\rho^1, L^1, \mu^1) &= (\rho_0, L_0, \mu_0) = \left(\rho_{\mathrm{cp}}^0, \ell_{\mathrm{cp}}(\rho_{\mathrm{cp}}^0)^2, \frac{\eta^{\mathrm{cp}}}{\rho_{\mathrm{cp}}^0}\right).
\end{aligned}
\tag{30}
$$

The subscript $j$ indexes checkpoint levels and the superscript $k$ indexes outer iterations. If iteration $k$ begins after exactly $j$ committed checkpoints, then $(\rho^k, L^k, \mu^k) = (\rho_j, L_j, \mu_j)$; an accepted checkpoint records its tuple at index $k+1$ with these pre-promotion values before the algorithm stores $(\rho^{k+1}, L^{k+1}, \mu^{k+1}) = (\rho_{j+1}, L_{j+1}, \mu_{j+1})$.

## G ALGEBRAIC LEMMAS

We collect the componentwise envelope and wall calculus used below. The objective assumptions immediately give the global estimates

$$
\|\nabla h_i(y)\| \leq D_i, \qquad \|w\| \leq C_i^{\psi} \quad (w \in \partial\psi_i(y)), \qquad |\varphi_i(u) - \varphi_i(v)| \leq \widehat{C}_i \|u - v\|. \tag{31}
$$

Indeed, differentiability and global Lipschitz continuity bound $\nabla h_i$, and every subgradient of the finite convex Lipschitz function $\psi_i$ has norm at most $C_i^{\psi}$. Adding the two Lipschitz bounds gives

the last estimate, with the main-text definition $\widehat{C}_i = D_i + C_i^\psi$. Moreover, $H_i$-smoothness of $h_i$ and convexity of $\psi_i$ make $\varphi_i$ $H_i$-weakly convex. The Moreau gradient bound below also yields

$$\|\nabla \bar{f}_i(y, L, \mu)\| = \|\nabla h_i(y) + \nabla(\psi_i)_\mu(y)\| \le D_i + C_i^\psi = \widehat{C}_i, \qquad i \in \mathcal{I}_o. \tag{32}$$

Thus the objective surrogate is globally $\widehat{C}_i$-Lipschitz, uniformly in $(L, \mu)$.

**Lemma G.1** (Convex Moreau calculus)**.** *For $i \in \mathcal{I}_o$, $\mu > 0$, and every $y$, $P_{i,\mu}^\psi(y)$ is single valued, $(\psi_i)_\mu$ is continuously differentiable, and*

$$\begin{aligned} \nabla(\psi_i)_\mu(y) &= \frac{y - P_{i,\mu}^\psi(y)}{\mu} \\ &\in \partial\psi_i(P_{i,\mu}^\psi(y)). \end{aligned} \tag{33}$$

*Furthermore,*

$$\|\nabla(\psi_i)_\mu(y)\| \le C_i^\psi, \qquad \|y - P_{i,\mu}^\psi(y)\| \le \mu C_i^\psi, \tag{34}$$

*and*

$$0 \le \psi_i(y) - (\psi_i)_\mu(y) \le \frac{\mu}{2}(C_i^\psi)^2. \tag{35}$$

*Proof.* Introduce the minimization variable explicitly and write

$$p = P_{i,\mu}^\psi(y) = \operatorname*{argmin}_r \left\{\psi_i(r) + \frac{1}{2\mu}\|r - y\|^2\right\}.$$

The objective is $1/\mu$-strongly convex and coercive, so $p$ is unique, and its optimality condition gives $(y - p)/\mu \in \partial\psi_i(p)$. If $p_d = P_{i,\mu}^\psi(y + d)$, comparison at the two minimizers gives

$$\left\langle \frac{y + d - p_d}{\mu}, d \right\rangle - \frac{\|d\|^2}{2\mu} \le (\psi_i)_\mu(y + d) - (\psi_i)_\mu(y) \le \left\langle \frac{y - p}{\mu}, d \right\rangle + \frac{\|d\|^2}{2\mu}.$$

Monotonicity of $\partial\psi_i$ applied to the two optimality conditions gives $\|p_d - p\|^2 \le \langle p_d - p, d\rangle$, hence $p_d \to p$ as $d \to 0$. Dividing the comparison inequalities by $\|d\|$ proves differentiability and the gradient formula; the same proximal Lipschitz bound makes the gradient continuous. Every subgradient of a finite convex $C_i^\psi$-Lipschitz function has norm at most $C_i^\psi$, proving the two bounds. Finally, evaluating the proximal objective at $r = y$ gives $(\psi_i)_\mu(y) \le \psi_i(y)$, while $\psi_i(p) \ge \psi_i(y) - C_i^\psi\|p - y\|$. With the scalar variable $r = \|p - y\| \ge 0$, minimizing $-C_i^\psi r + r^2/(2\mu)$ gives the right gap inequality. □

**Lemma G.2** (Moreau gradient and parameter drift)**.** *For $i \in \mathcal{I}_o$, every $\mu > 0$, and all $u, v, y$,*

$$\|\nabla(\psi_i)_\mu(u) - \nabla(\psi_i)_\mu(v)\| \le \frac{1}{\mu}\|u - v\|. \tag{36}$$

*For $0 < \mu_2 \le \mu_1$,*

$$0 \le (\psi_i)_{\mu_2}(y) - (\psi_i)_{\mu_1}(y) \le \frac{(C_i^\psi)^2}{2}(\mu_1 - \mu_2), \tag{37}$$

*and*

$$\|\nabla(\psi_i)_{\mu_2}(y) - \nabla(\psi_i)_{\mu_1}(y)\| \le C_i^\psi(\mu_1/\mu_2 - 1). \tag{38}$$

*Proof.* For $P = P_{i,\mu}^\psi$, monotonicity of $\partial\psi_i$ in the optimality conditions at $u$ and $v$ gives

$$\|P(u) - P(v)\|^2 \le \langle P(u) - P(v), u - v\rangle.$$

Thus $P$ is firmly nonexpansive. Expanding this inequality shows that $I - P$ satisfies the same inequality and is therefore nonexpansive; together with Equation 33, this proves Equation 36. For fixed $y$, the objective defining $(\psi_i)_\mu(y)$ is pointwise nonincreasing in $\mu$, which proves the left inequality in Equation 37. Write $p_\alpha = P_{i,\alpha}^\psi(y)$. For $\nu > \mu$, comparison at $p_\mu$ and $p_\nu$ gives

$$\left(\frac{1}{2\nu} - \frac{1}{2\mu}\right)\|y - p_\nu\|^2 \le (\psi_i)_\nu(y) - (\psi_i)_\mu(y) \le \left(\frac{1}{2\nu} - \frac{1}{2\mu}\right)\|y - p_\mu\|^2.$$

Moreover, monotonicity of $\partial\psi_i$ in the two optimality conditions and Equation 34 give $\|p_\nu - p_\mu\| \le C_i^\psi|\nu-\mu|$. Dividing the comparison inequalities by $\nu-\mu$ and taking either one-sided limit therefore yields

$$\begin{aligned}\frac{d}{d\mu}(\psi_i)_\mu(y) &= -\frac{\|y - P_{i,\mu}^\psi(y)\|^2}{2\mu^2}\\ &= -\frac{1}{2}\|\nabla(\psi_i)_\mu(y)\|^2.\end{aligned}$$

The right side is continuous on $(0,\infty)$. Thus, integrating over $[\mu_2,\mu_1]$ and applying Equation 34 yields

$$\begin{aligned}(\psi_i)_{\mu_2}(y) - (\psi_i)_{\mu_1}(y) &= \int_{\mu_2}^{\mu_1} \frac{1}{2}\|\nabla(\psi_i)_\mu(y)\|^2\, d\mu\\ &\le \frac{(C_i^\psi)^2}{2}(\mu_1-\mu_2).\end{aligned}$$

For the gradient estimate put $g_j = \nabla(\psi_i)_{\mu_j}(y)$ and $p_j = y - \mu_j g_j$. Monotonicity of $\partial\psi_i$ gives $0 \le \langle g_2-g_1, p_2-p_1\rangle$, hence $\mu_2\|g_2-g_1\|^2 \le (\mu_1-\mu_2)\langle g_2-g_1, g_1\rangle$. If $g_2 = g_1$ the claim is immediate; otherwise divide by $\mu_2\|g_2-g_1\|$ and use $\|g_1\| \le C_i^\psi$. □

**Lemma G.3** (One-call envelope prox)**.** *For $i \in \mathcal{I}_o$, $t,\mu > 0$, and every $z$,*

$$prox_{t(\psi_i)_\mu}(z) = \frac{\mu}{\mu+t}z + \frac{t}{\mu+t}prox_{(\mu+t)\psi_i}(z). \tag{39}$$

*If $p = prox_{(\mu+t)\psi_i}(z)$ and the right side is $y^+$, then*

$$\begin{aligned}p &= P_{i,\mu}^\psi(y^+),\\ \nabla(\psi_i)_\mu(y^+) &= \frac{z-p}{\mu+t}\\ &= \frac{y^+-p}{\mu}\\ &= \frac{z-y^+}{t}\\ &\in \partial\psi_i(p).\end{aligned} \tag{40}$$

*In particular, the objective-block quantities from the main text satisfy the exact one-call identities*

$$\begin{aligned}p_i^{k+1} &= prox_{(\mu^k+t_i^k)\psi_i}(z_{o,i}^k),\\ y_i^{k+1} &= \frac{\mu^k}{\mu^k+t_i^k}z_{o,i}^k + \frac{t_i^k}{\mu^k+t_i^k}p_i^{k+1},\\ w_i^{k+1} &= \frac{z_{o,i}^k - p_i^{k+1}}{\mu^k+t_i^k}.\end{aligned} \tag{41}$$

*Proof.* The optimality condition for $p$ is $(z-p)/(\mu+t) \in \partial\psi_i(p)$. The displayed affine combination gives both $y^+-p = \mu(z-p)/(\mu+t)$ and $z-y^+ = t(z-p)/(\mu+t)$. The first identity and $(z-p)/(\mu+t) \in \partial\psi_i(p)$ give $(y^+-p)/\mu \in \partial\psi_i(p)$, hence $p = \mathrm{prox}_{\mu\psi_i}(y^+)$. By Equation 33, the second identity then gives $0 = \nabla(\psi_i)_\mu(y^+) + (y^+-z)/t$. Since the Moreau envelope of the convex function $\psi_i$ is convex, this condition is necessary and sufficient for the unique minimizer $y^+ = \mathrm{prox}_{t(\psi_i)_\mu}(z)$.

For the main-text objective-block subproblem, completing the square and dropping terms independent of $u$ gives

$$(\psi_i)_{\mu^k}(u) + \frac{1}{2t_i^k}\|u - z_{o,i}^k\|^2.$$

Thus its minimizer is $\mathrm{prox}_{t_i^k(\psi_i)_{\mu^k}}(z_{o,i}^k)$, and substituting $(t,\mu,z) = (t_i^k, \mu^k, z_{o,i}^k)$ gives Equation 41. □

**Original-problem reconstruction.** For the original-problem candidate certificate tuple generated by step $k$, let $p_i^{k+1}$ and $w_i^{k+1}$ be the one-call outputs in Equation 41, and define

$$\breve{y}_i^{k+1} := \begin{cases} p_i^{k+1}, & i \in \mathcal{I}_{\mathrm{o}}, \\ y_i^{k+1}, & i \in \mathcal{I}_{\mathrm{c}}, \end{cases} \qquad v_i^{k+1} := \nabla h_i(p_i^{k+1}) + w_i^{k+1} \quad (i \in \mathcal{I}_{\mathrm{o}}). \tag{42}$$

Then

$$\begin{gathered} v_i^{k+1} \in \partial\varphi_i(\breve{y}_i^{k+1}), \\ \|y^{k+1} - \breve{y}^{k+1}\| \le \mu^k C_\psi, \qquad C_\psi := \left(\sum_{i\in\mathcal{I}_{\mathrm{o}}} (C_i^\psi)^2\right)^{1/2}, \\ \|v_i^{k+1} - q_i^k\| \le H_i(\|d_i^k\| + \mu^k C_i^\psi). \end{gathered} \tag{43}$$

Indeed, Equation 40 gives $w_i^{k+1} \in \partial\psi_i(p_i^{k+1})$ and $y_i^{k+1} - p_i^{k+1} = \mu^k w_i^{k+1}$. The exact smooth-plus-convex sum rule gives the subgradient membership, the bound $\|w_i^{k+1}\| \le C_i^\psi$ gives the aggregate reconstruction error, and the $H_i$-Lipschitz continuity of $\nabla h_i$ gives the final force-transfer bound.

**Lemma G.4** (Continuation correction)**.** *For the ordinary IP schedule define* $\mu_{\mathrm{IP}}(L) := \eta L^0/(\rho^0 L)$ *and*

$$R(L) := \frac{\mu_{\mathrm{IP}}(L)}{2} \sum_{i\in\mathcal{I}_o} (C_i^\psi)^2. \tag{44}$$

*Then* $R(L) \ge 0$ *for every* $L > 0$*, and for* $L' \ge L > 0$ *and every* $y = (y_i)_{i\in\mathcal{I}_o}$*,*

$$\sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L', \mu_{\mathrm{IP}}(L')) + R(L') \le \sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L, \mu_{\mathrm{IP}}(L)) + R(L). \tag{45}$$

*Moreover,*

$$\sum_{i\in\mathcal{I}_o} \bar{f}_i(A_i x, L, \mu_{\mathrm{IP}}(L)) + R(L) \ge 0 \qquad \forall x. \tag{46}$$

*Proof.* Nonnegativity follows because $\mu_{\mathrm{IP}}(L) > 0$ and the finite sum of squares in Equation 44 is nonnegative. For $i \in \mathcal{I}_{\mathrm{o}}$, the objective branch $\bar{f}_i(\cdot, L, \mu) = h_i + (\psi_i)_\mu$ is independent of $L$, so the change from $L$ to $L'$ in Equation 45 affects only the smoothing parameter. Apply Equation 37 blockwise with $\mu_{\mathrm{IP}}(L') \le \mu_{\mathrm{IP}}(L)$; the resulting increase is at most $R(L) - R(L')$, proving Equation 45. By Equation 35, $h_i(A_i x) + (\psi_i)_{\mu_{\mathrm{IP}}(L)}(A_i x) + \mu_{\mathrm{IP}}(L)(C_i^\psi)^2/2 \ge \varphi_i(A_i x)$. Summing and using Assumption 3.4 proves Equation 46. □

**Lemma G.5** (Decoupled Moreau continuation correction)**.** *Define*

$$R_\psi(\mu) := \frac{\mu}{2} G_\psi, \qquad G_\psi := \sum_{i\in\mathcal{I}_o} (C_i^\psi)^2. \tag{47}$$

*Then* $R_\psi(\mu) \ge 0$ *for every* $\mu > 0$*. For* $0 < \mu' \le \mu$*, arbitrary* $L, L' > 0$*, and every* $y = (y_i)_{i\in\mathcal{I}_o}$*,*

$$\sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L', \mu') + R_\psi(\mu') \le \sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L, \mu) + R_\psi(\mu). \tag{48}$$

*Moreover, for every* $L, \mu > 0$ *and* $x$*,*

$$\sum_{i\in\mathcal{I}_o} \bar{f}_i(A_i x, L, \mu) + R_\psi(\mu) \ge 0. \tag{49}$$

*Proof.* Nonnegativity follows from $\mu > 0$ and $G_\psi \ge 0$. For $i \in \mathcal{I}_{\mathrm{o}}$, the branch $\bar{f}_i(\cdot, L, \mu) = h_i + (\psi_i)_\mu$ is independent of $L$. Hence the $L$-argument is inert on $\mathcal{I}_{\mathrm{o}}$, the $h_i$ terms cancel, and only the envelopes change. Applying Equation 37 blockwise shows that their total increase when $\mu$ decreases to $\mu'$ is at most $(\mu - \mu')G_\psi/2 = R_\psi(\mu) - R_\psi(\mu')$, proving Equation 48. By Equation 35, for every $i \in \mathcal{I}_{\mathrm{o}}$ and $x$,

$$h_i(A_i x) + (\psi_i)_\mu(A_i x) + \frac{\mu}{2}(C_i^\psi)^2 \ge \varphi_i(A_i x).$$

Sum over $i \in \mathcal{I}_{\mathrm{o}}$ and use the aggregate lower bound in Assumption 3.4 to obtain Equation 49. □

**Lemma G.6** (Wall calculus, model error, and exact affine prox)**.** *Fix an arbitrary wall height $G > 0$ and let $g$ be the corresponding wall formula from the main text. Define the activation width by*

$$\beta(L) := \sqrt{2G/L}. \tag{50}$$

*For every $L > 0$, the wall function $g(\cdot, L)$ is nonnegative, continuously differentiable, convex, and $L$-smooth, with explicit force*

$$g'(s, L) = \max\{\min\{Ls - \sqrt{2LG}, 0\}, -\sqrt{2LG}\}. \tag{51}$$

*In particular, $|g'(s, L)| \le \sqrt{2LG}$ and $g(s, L) > G$ whenever $s < 0$. For all $L, L' > 0$ and $s \in \mathbb{R}$,*

$$|g'(s, L) - g'(s, L')| \le \sqrt{2G}\,|\sqrt{L} - \sqrt{L'}|. \tag{52}$$

*Thus $g'(s, \cdot)$ is $\sqrt{2G}$-Lipschitz as a function of $\sqrt{L}$, uniformly in $s$. For $i \in \mathcal{I}_c$, every iteration index $k$, and every $u$,*

$$|g(f_i(u), L^k) - g(\ell_i^k(u), L^k)| \le \frac{\sqrt{2L^kG}\,L_i}{2}\|u - y_i^k\|^2. \tag{53}$$

*Moreover, consider*

$$\min_u \left\{ g(c + a^T u, L) + \frac{Q}{2}\|u - z\|^2 \right\}, \qquad Q > 0.$$

*Put $s_z = c + a^T z$, $a_L = \sqrt{2LG}$, $\beta_L = \sqrt{2G/L}$, and $\kappa = \|a\|^2/Q$. Its unique minimizer is*

$$u^+ = z - \frac{\iota}{Q}a, \qquad \iota = \begin{cases} -a_L, & s_z < -\kappa a_L, \\ L\dfrac{s_z - \beta_L}{1 + \kappa L}, & -\kappa a_L \le s_z < \beta_L, \\ 0, & s_z \ge \beta_L. \end{cases} \tag{54}$$

*The formulas agree at both boundaries, and $\iota = g'(c + a^T u^+, L)$.*

*Proof.* The definition of $g$, together with $G, L > 0$, directly gives nonnegativity and $g(s, L) > G$ for $s < 0$. At $s = 0$, the negative and middle branches have the same value and one-sided derivative; within the $s \ge 0$ formula, the middle and zero branches likewise match in value and derivative at $s = \beta(L)$. These branchwise derivatives and matching boundary values prove continuous differentiability and give Equation 51. The derivative is nondecreasing and $L$-Lipschitz in $s$, proving convexity and $L$-smoothness; its explicit form also gives $|g'(s, L)| \le \sqrt{2LG}$.

It remains to prove the uniform parameter bound. Put $t = \sqrt{L}$ and $c_G = \sqrt{2G}$. For fixed $s$,

$$g'(s, t^2) = \begin{cases} -c_G t, & s \le 0, \\ t^2 s - c_G t, & s > 0,\ 0 < t < c_G/s, \\ 0, & s > 0,\ t \ge c_G/s. \end{cases}$$

This map is continuous and piecewise $C^1$. On its smooth pieces its derivative is respectively $-c_G$, $2ts - c_G$, and 0; on the middle piece, $0 < ts < c_G$ gives $|2ts - c_G| < c_G$, so every derivative has absolute value at most $c_G$. Integrating piecewise between $\sqrt{L}$ and $\sqrt{L'}$ proves Equation 52; equality holds when $s \le 0$. The Lipschitz-gradient remainder bound gives $|f_i(u) - \ell_i^k(u)| \le L_i\|u - y_i^k\|^2/2$; composing with $|g'| \le \sqrt{2L^kG}$ proves Equation 53.

For the prox, the objective is $Q$-strongly convex, so its minimizer is unique; when $a = 0$, the formulas reduce directly to $u^+ = z$. Optimality has the form $Q(u - z) + \iota a = 0$ with $\iota = g'(c + a^T u, L)$, hence $u = z - \iota a/Q$ and $s = c + a^T u = s_z - \kappa\iota$. On the first branch, $\iota = -a_L$ and $s = s_z + \kappa a_L < 0$ exactly when $s_z < -\kappa a_L$. On the middle branch, using $L\beta_L = a_L$ gives

$$\iota = L(s - \beta_L), \qquad s = \frac{s_z + \kappa a_L}{1 + \kappa L}, \qquad \iota = L\frac{s_z - \beta_L}{1 + \kappa L},$$

and $0 \le s < \beta_L$ is equivalent to $-\kappa a_L \le s_z < \beta_L$. On the last branch, $\iota = 0$ and $s = s_z \ge \beta_L$. At $s_z = -\kappa a_L$ the first two formulas both give $\iota = -a_L$; at $s_z = \beta_L$ the latter two both give zero. Thus Equation 54 satisfies the unique optimality condition and $\iota = g'(c + a^T u^+, L)$. □

For the active mode, instantiate the preceding lemma with $G = G_{\mathsf{m}}$ and $g = g_{\mathsf{m}}$, and write $\beta_{\mathsf{m}}(L)$ for its width. The two true/model discrepancies share the notation $\tau_i^k$ from the main text. If

$$\widetilde{f}_i^k(u) = \begin{cases} h_i(y_i^k) + \langle \nabla h_i(y_i^k), u - y_i^k \rangle + (\psi_i)_{\mu^k}(u), & i \in \mathcal{I}_{\mathrm{o}}, \\ g(\ell_i^k(u), L^k), & i \in \mathcal{I}_{\mathrm{c}}, \end{cases} \tag{55}$$

then, globally,

$$|\bar{f}_i(u, L^k, \mu^k) - \widetilde{f}_i^k(u)| \le \frac{\tau_i^k}{2} \|u - y_i^k\|^2, \tag{56}$$

and $\widetilde{f}_i^k$ is $\zeta_i^k$-smooth. Consequently, adding $\langle e_i^k, u - y_i^k \rangle + \mathsf{B}_i^k \|u - y_i^k\|^2/2$, where $\mathsf{B}_i^k = \rho^k + \tau_i^k$, to $\widetilde{f}_i^k$ yields, up to an additive constant, a majorizer of the true block Lagrangian at $x^k$.

**Schedule consequences.** The ordinary IP schedules in Equation 10 are $\rho^k = \rho^0(1 + \xi k^p)$, $L^k = L^0(1 + \xi k^p)$, and $\mu^k = \eta/\rho^k = \mu_{\mathrm{IP}}(L^k)$. Thus, for every $k \ge 1$,

$$\frac{\rho^k}{L^k} = \frac{\rho^0}{L^0}. \tag{57}$$

They are nondecreasing:

$$\rho^{k+1} \ge \rho^k, \qquad L^{k+1} \ge L^k. \tag{58}$$

The associated smoothing radii consequently satisfy, for $k \ge 2$,

$$\mu^{k+1} \le \mu^k, \qquad \frac{\mu^{k-1}}{\mu^k} = \frac{\rho^k}{\rho^{k-1}}. \tag{59}$$

**Lemma G.7** (IP schedule ratios)**.** *In ordinary IP mode, for $k \ge 1$, $1 \le \rho^{k+1}/\rho^k = L^{k+1}/L^k \le 1+\xi \le 2$.*

*Proof.* The schedules share the factor $1 + \xi k^p$, so

$$\begin{aligned} \frac{\rho^{k+1}}{\rho^k} &= \frac{L^{k+1}}{L^k} \\ &= \frac{1 + \xi(k+1)^p}{1 + \xi k^p}. \end{aligned}$$

For $p \in (0, 1)$, the map $t \mapsto t^p$ is nondecreasing and concave on $(0, \infty)$, so the ratio is at least one and, for $k \ge 1$,

$$(k+1)^p - k^p \le p k^{p-1} \le 1,$$

where the last inequality uses $p \le 1$ and $k^{p-1} \le 1$. Therefore, explicitly using $1 + \xi k^p \ge 1$,

$$\begin{aligned} \frac{1 + \xi(k+1)^p}{1 + \xi k^p} &= 1 + \frac{\xi((k+1)^p - k^p)}{1 + \xi k^p} \\ &\le 1 + \xi((k+1)^p - k^p) \\ &\le 1 + \xi. \end{aligned}$$

When $p = 0$, the ratio is exactly one. Finally, $\xi \le 1$ gives $1 + \xi \le 2$. □

**Lemma G.8.** *Let $\sigma \in (0, 2)$. If $b^+ = \sigma a + (1 - \sigma)b$, then*

$$\frac{1}{\sigma} \|b^+\|^2 \le \sigma_1 \|a\|^2 + \sigma_2(\|b\|^2 - \|b^+\|^2),$$

*where $\sigma_1 = \sigma/(1-|1-\sigma|)^2$ and $\sigma_2 = |1-\sigma|/[\sigma(1-|1-\sigma|)]$. In particular, for $\sigma = 1$, $(\sigma_1, \sigma_2) = (1, 0)$.*

*Proof.* If $\sigma = 1$, then $b^+ = a$ and the claim holds with $(\sigma_1, \sigma_2) = (1, 0)$. Otherwise put $t = |1 - \sigma| \in (0, 1)$, so $1 - t > 0$. The triangle inequality gives $\|b^+\| \le \sigma\|a\| + t\|b\|$. Young's inequality with $\epsilon = t/(1 - t) > 0$ yields

$$\begin{aligned} \|b^+\|^2 &\le (1 + \epsilon)\sigma^2 \|a\|^2 + (1 + 1/\epsilon)t^2 \|b\|^2 \\ &= \frac{\sigma^2}{1 - t} \|a\|^2 + t\|b\|^2. \end{aligned}$$

Since $1/\sigma + \sigma_2 = 1/[\sigma(1-t)]$, division by $\sigma(1-t)$ gives

$$\left(\frac{1}{\sigma} + \sigma_2\right)\|b^+\|^2 \le \frac{\sigma}{(1-t)^2}\|a\|^2 + \frac{t}{\sigma(1-t)}\|b\|^2$$
$$= \sigma_1\|a\|^2 + \sigma_2\|b\|^2.$$

Moving $\sigma_2\|b^+\|^2$ to the right proves the result. □

For $p \in (0,1)$ define

$$H_p(\xi) = \xi^2 \sum_{j=1}^{\infty} \frac{j^{2p-2}}{(1+\xi j^p)^2}, \tag{60}$$

and set $H_0(\xi) = 0$ by convention.

**Lemma G.9.** *Under the ordinary polynomial schedule Equation 10, with $p \in [0,1)$ and $\xi \in (0,1]$, the schedule sums satisfy*

$$\sum_{k=2}^{\infty} \frac{(\sqrt{L^k} - \sqrt{L^{k-1}})^2}{\rho^k} \le \frac{L^0 p^2}{4\rho^0} H_p(\xi), \tag{61}$$

*and*

$$\sum_{k=2}^{\infty} \frac{(\rho^k/\rho^{k-1} - 1)^2}{\rho^k} \le \frac{p^2}{\rho^0} H_p(\xi). \tag{62}$$

*The left sides vanish for $p = 0$.*

*Proof.* Let $p > 0$ and $k \ge 2$, and put $j = k-1 \ge 1$. By the mean value theorem, $k^p - j^p = pc^{p-1}(k-j) = pc^{p-1}$ for some $c \in (j,k)$, since $k - j = 1$. Since $t \mapsto t^p$ is concave, its derivative is nonincreasing, and therefore

$$k^p - j^p \le pj^{p-1}. \tag{63}$$

For the ratio series, the common schedule gives the identity

$$\frac{\rho^k}{\rho^{k-1}} - 1 = \frac{\xi(k^p - j^p)}{1+\xi j^p}$$
$$\le \frac{\xi p j^{p-1}}{1+\xi j^p}.$$

Both sides are nonnegative by schedule monotonicity, so squaring preserves the inequality. Using $\rho^k \ge \rho^0$ and shifting from $k \ge 2$ to $j = k-1 \ge 1$ yield

$$\sum_{k=2}^{\infty} \frac{(\rho^k/\rho^{k-1} - 1)^2}{\rho^k} \le \frac{p^2}{\rho^0}\xi^2 \sum_{j=1}^{\infty} \frac{j^{2p-2}}{(1+\xi j^p)^2} = \frac{p^2}{\rho^0} H_p(\xi).$$

For the square-root series, rationalization gives

$$\sqrt{L^k} - \sqrt{L^{k-1}} = \frac{L^0 \xi(k^p - j^p)}{\sqrt{L^k} + \sqrt{L^{k-1}}}.$$

Moreover,

$$\sqrt{L^k} + \sqrt{L^{k-1}} \ge 2\sqrt{L^{k-1}}$$
$$= 2\sqrt{L^0}\sqrt{1+\xi j^p}.$$

Consequently,

$$\sqrt{L^k} - \sqrt{L^{k-1}} \le \frac{\sqrt{L^0}\,\xi p j^{p-1}}{2\sqrt{1+\xi j^p}}.$$

Both sides are nonnegative by schedule monotonicity, so this inequality may also be squared. Use the sharper denominator bound $\rho^k = \rho^0(1+\xi k^p) \ge \rho^0(1+\xi j^p)$ and again set $j = k-1$ to obtain

$$\sum_{k=2}^{\infty} \frac{(\sqrt{L^k} - \sqrt{L^{k-1}})^2}{\rho^k} \le \frac{L^0 p^2}{4\rho^0}\xi^2 \sum_{j=1}^{\infty} \frac{j^{2p-2}}{(1+\xi j^p)^2} = \frac{L^0 p^2}{4\rho^0} H_p(\xi).$$

The defining series converges because, for fixed $\xi > 0$, its full summand is at most $j^{-2}$. If $p = 0$, then $k^p = 1$ for every positive index, so both $L^k$ and $\rho^k$ are constant and every summand on the two left sides vanishes. □

**Lemma G.10** (Small-$\xi$ regimes of $H_p$)**.** *For fixed $p \in (0,1)$, as $\xi \downarrow 0$,*

$$H_p(\xi) = \begin{cases} \Theta(\xi^2), & 0 < p < 1/2, \\ \Theta(\xi^2 \log(e/\xi)), & p = 1/2, \\ \Theta(\xi^{1/p}), & 1/2 < p < 1. \end{cases} \tag{64}$$

*The hidden constants may depend on the fixed $p$; no uniformity as $p \to 1/2$ is claimed.*

*Proof.* Because the claim is asymptotic as $\xi \downarrow 0$, it suffices to take $0 < \xi \le 1/2$. Let $J = \lceil \xi^{-1/p} \rceil$ and split the series in Equation 60 at $J$. Since $\xi^{-1/p} \ge 1$ and hence $J \le 2\xi^{-1/p}$, for $j \le J$ one has $1 \le 1 + \xi j^p \le 1 + 2^p$. Thus the head denominator is bounded above and below by positive constants depending only on $p$, and the head is comparable to

$$\xi^2 \sum_{j=1}^{J} j^{2p-2}.$$

By comparison with $\int_1^J t^{2p-2}\,dt$, this is respectively $\Theta(\xi^2)$, $\Theta(\xi^2 \log J)$, or $\Theta(\xi^2 J^{2p-1})$ according as $p < 1/2$, $p = 1/2$, or $p > 1/2$. On $j > J$, one has $\xi j^p > 1$, so the full summand in Equation 60, including the $\xi^2$ prefactor, lies between $\frac{1}{4} j^{-2}$ and $j^{-2}$, and the tail is $\Theta(J^{-1}) = \Theta(\xi^{1/p})$. If $p < 1/2$, then $\xi^{1/p} = o(\xi^2)$, so the head dominates; if $p = 1/2$, the $\Theta(\xi^2 \log J)$ head dominates the $\Theta(\xi^2)$ tail; and if $p > 1/2$, the head satisfies $\xi^2 J^{2p-1} = \Theta(\xi^2 \xi^{-(2p-1)/p}) = \Theta(\xi^{1/p})$, matching the tail. Using $\log J = \Theta(\log(e/\xi))$ at $p = 1/2$ proves Equation 64. □

**Lemma G.11** (Geometric checkpoint schedule identities)**.** *Under the checkpoint schedule Equation 30, with the generic wall specialized to $G = G^{\mathrm{cp}}$, $\rho_{\mathrm{cp}}^0 > 0$, $\chi > 1$, $\ell_{\mathrm{cp}} > 0$, and $b_{\mathrm{cp}}$ from Equation 29, the following ratios and the first wall scaling hold for every integer $j \ge 0$; the second wall scaling also holds for every $i \in \mathcal{I}_c$:*

$$\frac{\rho_{j+1}}{\rho_j} = \chi, \qquad \frac{L_{j+1}}{L_j} = \chi^2, \qquad \frac{\mu_{j+1}}{\mu_j} = \chi^{-1}, \tag{65}$$

$$\beta_{\mathsf{Ckpt}}(L_j) = \frac{b_{\mathrm{cp}}}{\rho_j} = \beta_j, \qquad \frac{\sqrt{2L_j G^{\mathrm{cp}}}\, L_i}{\rho_j} = \sqrt{2G^{\mathrm{cp}} \ell_{\mathrm{cp}}}\, L_i. \tag{66}$$

*The geometric sums satisfy*

$$\sum_{j=0}^{\infty} \frac{1}{\rho_j} = \frac{1}{\rho_{\mathrm{cp}}^0 (1 - \chi^{-1})}, \qquad \sum_{j=0}^{J} \rho_j^3 \le \frac{\rho_J^3}{1 - \chi^{-3}} \quad \text{for every integer } J \ge 0. \tag{67}$$

*For every integer $j \ge 0$ and $s \ge 0$, one has $g_{\mathsf{Ckpt}}(s, L_{j+1}) \le g_{\mathsf{Ckpt}}(s, L_j)$.*

*Proof.* The three ratios follow directly from Equation 30. Since $\rho_j > 0$, substituting $L_j = \ell_{\mathrm{cp}} \rho_j^2$ and $b_{\mathrm{cp}} = \sqrt{2G^{\mathrm{cp}}/\ell_{\mathrm{cp}}}$, and $\beta_j = b_{\mathrm{cp}}/\rho_j$ into the definition of $\beta_{\mathsf{Ckpt}}$ gives the first identity in Equation 66; the second follows from $\sqrt{2L_j G^{\mathrm{cp}}} = \rho_j \sqrt{2G^{\mathrm{cp}} \ell_{\mathrm{cp}}}$, using $\rho_j > 0$. Both sums in Equation 67 are geometric; for the finite sum, write $\rho_j^3 = \rho_J^3 \chi^{-3(J-j)}$ and enlarge it to the infinite backward series. Finally, for $s \ge 0$ and $L > 0$, the explicit wall formula gives

$$g_{\mathsf{Ckpt}}(s, L) = \frac{1}{2} \max\{\sqrt{2G^{\mathrm{cp}}} - \sqrt{L}\, s, 0\}^2,$$

because $\sqrt{2G^{\mathrm{cp}}} - s\sqrt{L}$ is nonincreasing in $L$ and $t \mapsto \max\{t, 0\}^2$ is nondecreasing. Thus the display is nonincreasing in $L$; use $L_{j+1} \ge L_j$. □

The force drift constants used below are

$$a_i^k := \zeta_i^k, \qquad b_i^k := \begin{cases} H_i, & i \in \mathcal{I}_{\mathrm{o}}, \\ 2L^k M_i^2 + \sqrt{2L^k G}\, L_i, & i \in \mathcal{I}_{\mathrm{c}}. \end{cases} \tag{68}$$

Set

$$\delta_i^k := \begin{cases} C_i^\psi(\rho^k/\rho^{k-1} - 1), & i \in \mathcal{I}_{\mathrm{o}}, \\ \sqrt{2G}M_i(\sqrt{L^k} - \sqrt{L^{k-1}}), & i \in \mathcal{I}_{\mathrm{c}}. \end{cases} \tag{69}$$

Then

$$\|q_i^k - q_i^{k-1}\| \le a_i^k\|d_i^k\| + b_i^k\|d_i^{k-1}\| + \delta_i^k, \tag{70}$$

and define

$$U^k := \frac{5\sigma_1}{\rho^k}\sum_{i\in\mathcal{I}}(\delta_i^k)^2. \tag{71}$$

**Lemma G.12** (Uniform force drift)**.** *Under the polynomial schedule Equation 10, for every $i \in \mathcal{I}$ and $k \ge 2$, the quantities $a_i^k$, $b_i^k$, and $\delta_i^k$ defined above satisfy Equation 70 at every state, without a feasibility hypothesis. Moreover, $\delta_i^k \ge 0$.*

*Proof.* Schedule monotonicity first gives $\delta_i^k \ge 0$ on both branches. For objective blocks, split the force difference into a state change of the convex envelope at the common parameter $\mu^k$, a parameter change at the common point $y_i^k$, and the change of $\nabla h_i$. Apply Equation 36, $H_i$-smoothness, and Equation 38; the explicit schedule identity $\mu^{k-1}/\mu^k = \rho^k/\rho^{k-1}$ from Equation 59 gives the objective branch of $\delta_i^k$. For constraint blocks put $s_i^k = \ell_i^k(y_i^{k+1})$. Bounded gradients give $|s_i^k - s_i^{k-1}| \le M_i\|d_i^k\| + 2M_i\|d_i^{k-1}\|$. Since $L^{k-1} \le L^k$ by Equation 58, both scalar state-change terms may use the $L^k$ Lipschitz and force bounds. The parameter estimate in Equation 52 contributes $\sqrt{2G}(\sqrt{L^k} - \sqrt{L^{k-1}})$, where the absolute value drops by the same monotonicity. Multiplying by $\nabla f_i(y_i^k)$ and then using smoothness to transfer that gradient from $y_i^{k-1}$ yields the constraint branch. No sign condition on $s_i^k$ is used. □

**Lemma G.13** (Summable force drift)**.** *Under Equation 28, which gives $\sigma_1 = 1$, and the polynomial schedule Equation 10 with $p \in [0, 1)$ and $\xi \in (0, 1]$,*

$$\sum_{k=2}^{\infty} U^k \le \Sigma_U := \frac{p^2 H_p(\xi)}{\rho^0}\left(5G_\psi + \frac{5}{2}GL^0\bar{M}\right), \tag{72}$$

*where $G_\psi$ and $\bar{M}$ are the quantities already defined in Equation 28. At $p = 0$, $\Sigma_U = 0$.*

*Proof.* Because $\mathcal{I}$ is finite and every summand is nonnegative, the block sum in Equation 71 may be interchanged with the infinite $k$-sum. The definition Equation 71 contributes the common factor $5\sigma_1$, with $\sigma_1 = 1$ by Equation 28. For objective blocks, inserting $\delta_i^k = C_i^\psi(\rho^k/\rho^{k-1} - 1)$ into Equation 62 and summing over $i \in \mathcal{I}_{\mathrm{o}}$ gives $p^2 H_p(\xi)G_\psi/\rho^0$ before that factor, hence the objective constant 5. For constraint blocks, $(\delta_i^k)^2 = 2GM_i^2(\sqrt{L^k} - \sqrt{L^{k-1}})^2$; applying Equation 61 and summing over $i \in \mathcal{I}_{\mathrm{c}}$ gives $p^2 H_p(\xi)GL^0\bar{M}/(2\rho^0)$ before the common factor, hence the constraint constant $5/2$. At $p = 0$, the definition $H_0(\xi) = 0$ makes $\Sigma_U = 0$. □

## H Convergence Analysis of IP-ADMM

This appendix proves the ordinary-mode results of Theorem 4.2 and Corollary 4.3. In the IP mode of Algorithm 1, the multiplier is updated at every iteration and $(\rho^k, L^k, \mu^k)$ follow Equation 10. The coupled relation $\mu^k = \mu_{\mathrm{IP}}(L^k)$ from Equation 10 and Equation 57 supplies the objective continuation correction Lemma G.4. The nonlinear wall adds the main difficulty: comparing successive wall parameters requires the new split iterate to be wall-interior, which is not known when the step is first analyzed. We resolve this dependence with an unconditional terminal estimate and explicit parameter margins before using the wall continuation comparison globally. The checkpoint mode uses a different proof and is treated in Appendix I.

**Organization.** The proof is wired as follows. The local model and multiplier identities in Section H.1 give one-step descent and a Lyapunov bridge whose terminal continuation comparison is still conditional. The unconditional estimate Lemma H.3 controls that terminal point before wall-interiority is known, while Lemma H.4 and the explicit margins in Lemma H.5 supply the

finite budget needed by Theorem H.6. That theorem closes the wall-interiority induction without a global movement bound; only afterward is the continuation comparison available at every step. In Section H.4, the drift-corrected lower bound Lemma H.7 yields the movement budget Corollary H.8. Original-problem reconstruction and pointwise residual identities then feed Theorem H.12, and Corollary H.13 converts the same selected candidate under Uniform Band-MFCQ. The scaled-multiplier results branch from the movement budget, whereas the family and tuned-$\xi$ results specialize the exact rate constants; neither branch is required to establish the core residual rates.

### H.1 Lagrangian and Lyapunov Descent

Put $\Delta x^k = x^{k+1} - x^k$, $\Delta\lambda^{k+1} = \lambda^{k+1} - \lambda^k$, and

$$\upsilon_i^k := m_i^k - \rho^k, \quad \Gamma_i^k := |\upsilon_i^k| + a_i^k, \quad \widehat{\Gamma}_i^k := |\upsilon_i^{k-1}| + b_i^k. \tag{73}$$

For $k \geq 2$, define

$$\begin{aligned}
\Theta^k &:= \frac{5\sigma_1}{\rho^k} \sum_{i\in\mathcal{I}} \left[ (\widehat{\Gamma}_i^k)^2 \|d_i^{k-1}\|^2 + (\rho^{k-1})^2 \|A_i \Delta x^{k-1}\|^2 \right] + \frac{\sigma_2}{\rho^k} \|\Delta\lambda^k\|^2, \\
\kappa_\Psi &:= \frac{\xi + 2\sigma + 2\sigma^2\sigma_3}{2\sigma}, \\
\bar{\Psi}^k &:= \mathcal{L}(x^k, y^k, \lambda^k, L^{k-1}, \mu^{k-1}, \rho^k) + R(L^{k-1}) + \kappa_\Psi \Theta^k, \\
\Psi^k &:= \mathcal{L}(x^k, y^k, \lambda^k, L^k, \mu^k, \rho^k) + R(L^k) + \kappa_\Psi \Theta^k.
\end{aligned} \tag{74}$$

For either mode $\mathsf{m}$, the exact optimality condition of the active block model is

$$0 = q_i^k + \lambda_i^k + \rho^k (y_i^k - A_i x^k) + m_i^k d_i^k. \tag{75}$$

For ordinary IP-ADMM, using $\Delta\lambda_i^{k+1}/\sigma = \rho^k(y_i^{k+1} - A_i x^{k+1})$ and $y_i^k - A_i x^k = y_i^{k+1} - A_i x^{k+1} - d_i^k + A_i \Delta x^k$ gives

$$\begin{aligned}
\frac{\Delta\lambda_i^{k+1}}{\sigma} + \lambda_i^k &= -q_i^k - \upsilon_i^k d_i^k - \rho^k A_i \Delta x^k, \\
\upsilon_i^k &= m_i^k - \rho^k.
\end{aligned} \tag{76}$$

The appearance of $m_i^k - \rho^k$, rather than $m_i^k$, is why the consensus quadratic is linearized in the block model.

For either mode $\mathsf{m}$, the $x$-subproblem has a multiplier $\varpi^{k+1}$ satisfying

$$\begin{aligned}
(1 + \theta_{2,\mathsf{m}}) A^T A \Delta x^k + C^T \varpi^{k+1} &= A^T (y^{k+1} - A x^k + \lambda^k/\rho^k), \\
C x^{k+1} &= d.
\end{aligned} \tag{77}$$

From this point through the end of the ordinary appendix, set $\mathsf{m} = \mathsf{IP}$, so $\theta_{1,\mathsf{m}} = \theta_1$, $\theta_{2,\mathsf{m}} = \theta_2$, $G_\mathsf{m} = G$, and $g := g_\mathsf{m} = g_{\mathsf{IP}}$.

**Lemma H.1** (One-step descent)**.** *For $k \geq 1$,*

$$\begin{aligned}
&\mathcal{L}(x^{k+1}, y^{k+1}, \lambda^{k+1}, L^k, \mu^k, \rho^{k+1}) - \mathcal{L}(x^k, y^k, \lambda^k, L^k, \mu^k, \rho^k) \\
&\leq -\frac{2\theta_1 - 1}{2} \sum_{i\in\mathcal{I}} \mathsf{B}_i^k \|d_i^k\|^2 - \frac{1 + 2\theta_2}{2} \rho^k \|A\Delta x^k\|^2 + \frac{\xi + 2\sigma}{2\sigma^2\rho^k} \|\Delta\lambda^{k+1}\|^2.
\end{aligned} \tag{78}$$

*If $f_i(y_i^{k+1}) \geq 0$ for every $i \in \mathcal{I}_c$, then*

$$\mathcal{L}(x^{k+1}, y^{k+1}, \lambda^{k+1}, L^{k+1}, \mu^{k+1}, \rho^{k+1}) + R(L^{k+1}) \leq \mathcal{L}(x^{k+1}, y^{k+1}, \lambda^{k+1}, L^k, \mu^k, \rho^{k+1}) + R(L^k). \tag{79}$$

*Proof.* For a block $i$, only the upper half of Equation 56 is used here, namely $\bar{f}_i(u, L^k, \mu^k) \leq \widetilde{f}_i^k(u) + \tau_i^k \|u - y_i^k\|^2/2$ evaluated at $u = y_i^{k+1}$; at $u = y_i^k$ the two functions agree exactly. Together with exact expansion of the consensus quadratic, this shows that the true Lagrangian change of the

block is at most its model change after removing the proximal term, plus $\mathsf{B}_i^k\|d_i^k\|^2/2$. Since $\mathcal{M}_i^k$ is $m_i^k$-strongly convex and is minimized at $y_i^{k+1}$,

$$\mathcal{M}_i^k(y_i^{k+1}) - \mathcal{M}_i^k(y_i^k) \le -\frac{m_i^k}{2}\|d_i^k\|^2.$$

Removing the proximal term contributes another $-m_i^k\|d_i^k\|^2/2$; therefore the true block change is at most $-(m_i^k - \mathsf{B}_i^k/2)\|d_i^k\|^2 = -(2\theta_1 - 1)\mathsf{B}_i^k\|d_i^k\|^2/2$. Summing gives the first term of the lemma.

Contracting Equation 77 with $\Delta x^k$ and using $C\Delta x^k = 0$ gives the $x$-change $-(1 + 2\theta_2)\rho^k\|A\Delta x^k\|^2/2$. The multiplier and penalty-parameter changes equal $[\rho^{k+1} + (2\sigma - 1)\rho^k]\|y^{k+1} - Ax^{k+1}\|^2/2$. The multiplier update $\Delta\lambda^{k+1} = \sigma\rho^k(y^{k+1} - Ax^{k+1})$ substitutes the residual identity

$$\|y^{k+1} - Ax^{k+1}\|^2 = \frac{\|\Delta\lambda^{k+1}\|^2}{(\sigma\rho^k)^2},$$

while Lemma G.7 gives $\rho^{k+1} + (2\sigma - 1)\rho^k \le (\xi + 2\sigma)\rho^k$ through $\rho^{k+1} \le (1 + \xi)\rho^k$. Multiplying the two yields exactly the last term $(\xi + 2\sigma)\|\Delta\lambda^{k+1}\|^2/(2\sigma^2\rho^k)$ of the first assertion.

For the continuation comparison, the objective aggregate plus $R$ decreases by Equation 45. The hypothesis $f_i(y_i^{k+1}) \ge 0$ places every constraint block on the branch $s \ge 0$ of the piecewise wall, where it admits the equivalent form

$$g(s, L) = \frac{1}{2}\max\{\sqrt{2G} - \sqrt{L}\,s, 0\}^2,$$

which is manifestly nonincreasing in $L$; with $L^{k+1} \ge L^k$ from Equation 58 this gives $g(s, L^{k+1}) \le g(s, L^k)$ without differentiating the only piecewise smooth map $g(s, \cdot)$. The multiplier and penalty terms do not depend on $L$, proving Equation 79. □

Subtracting Equation 76 at consecutive indices yields

$$\Delta\lambda_i^{k+1} = (1 - \sigma)\Delta\lambda_i^k - \sigma z_i^k, \tag{80}$$

where

$$z_i^k = q_i^k - q_i^{k-1} + \upsilon_i^k d_i^k - \upsilon_i^{k-1} d_i^{k-1} + \rho^k A_i\Delta x^k - \rho^{k-1} A_i\Delta x^{k-1}. \tag{81}$$

**Lemma H.2** (Multiplier-difference estimate)**.** *For $k \ge 2$,*

$$\begin{aligned}\frac{\|\Delta\lambda^{k+1}\|^2}{\sigma\rho^k} \le U^k + 5\sigma_1 \sum_{i\in\mathcal{I}}\left[\frac{(\Gamma_i^k)^2}{\rho^k} + \frac{(\widehat{\Gamma}_i^{k+1})^2}{\rho^{k+1}}\right]\|d_i^k\|^2 \\ + 10\sigma_1\rho^k\|A\Delta x^k\|^2 + \Theta^k - \Theta^{k+1}.\end{aligned} \tag{82}$$

*Proof.* Apply Lemma G.8 to Equation 80 with the identification $b^+ = \Delta\lambda_i^{k+1}$, $a = -z_i^k$, and $b = \Delta\lambda_i^k$. By Equation 70, which Lemma G.12 makes valid at an arbitrary, possibly wall-exterior state,

$$\|z_i^k\| \le \Gamma_i^k\|d_i^k\| + \widehat{\Gamma}_i^k\|d_i^{k-1}\| + \rho^k\|A_i\Delta x^k\| + \rho^{k-1}\|A_i\Delta x^{k-1}\| + \delta_i^k.$$

Square with $(a_1 + \cdots + a_5)^2 \le 5\sum_j a_j^2$, sum over blocks, and divide by $\rho^k$. Add and subtract the next-index terms $5\sigma_1(\widehat{\Gamma}_i^{k+1})^2\|d_i^k\|^2/\rho^{k+1}$, $5\sigma_1(\rho^k)^2\|A_i\Delta x^k\|^2/\rho^{k+1}$, and $\sigma_2\|\Delta\lambda^{k+1}\|^2/\rho^{k+1}$. The two terms carrying the current displacement $A\Delta x^k$ are the squared term $5\sigma_1\rho^k\|A_i\Delta x^k\|^2$ produced by the third summand of the bound on $\|z_i^k\|$ and the added next-index term $5\sigma_1(\rho^k)^2\|A_i\Delta x^k\|^2/\rho^{k+1}$; schedule monotonicity Equation 58 bounds the latter by the former, so the two together are at most $10\sigma_1\rho^k\|A\Delta x^k\|^2$. The primal lagged terms give their corresponding part of $\Theta^k - \Theta^{k+1}$ exactly. The dual pair leaves $\sigma_2\|\Delta\lambda^{k+1}\|^2(1/\rho^{k+1} - 1/\rho^k) \le 0$, again by Equation 58; hence the full remaining difference is bounded above by $\Theta^k - \Theta^{k+1}$. The drift squares are exactly $U^k$. □

Define the positive candidate coefficients

$$\begin{aligned}D_{y,i}^k &:= \frac{2\theta_1 - 1}{2}\mathsf{B}_i^k - 5\kappa_\Psi\sigma_1\left[\frac{(\Gamma_i^k)^2}{\rho^k} + \frac{(\widehat{\Gamma}_i^{k+1})^2}{\rho^{k+1}}\right], \\ D_x^k &:= \left(\frac{1 + 2\theta_2}{2} - 10\kappa_\Psi\sigma_1\right)\rho^k.\end{aligned} \tag{83}$$

Combining Lemma H.1 and Lemma H.2, with the coefficient $\kappa_\Psi$ chosen to leave $-\sigma_3/\rho^k$ on the dual increment, gives

$$\begin{aligned}\bar{\Psi}^{k+1} - \Psi^k \le &-\frac{\sigma_3}{\rho^k}\|\Delta\lambda^{k+1}\|^2 \\ &- \sum_{i\in\mathcal{I}} D_{y,i}^k \|d_i^k\|^2 \\ &- D_x^k \|A\Delta x^k\|^2 + \kappa_\Psi U^k.\end{aligned} \tag{84}$$

If $y^{k+1}$ is feasible, Equation 79 gives $\Psi^{k+1} \le \bar{\Psi}^{k+1}$.

### H.2 Split-Feasibility Bounds

For later use define

$$\begin{aligned}C_{d,i}^k &:= \frac{m_i^k}{2} + \frac{3\tau_i^k}{2}, \\ \widetilde{C}_{r,i}^k &:= \begin{cases} \dfrac{m_i^k}{2} + \tau_i^k + \dfrac{\zeta_i^k}{2}, & i \in \mathcal{I}_{\mathrm{o}}, \\ \dfrac{m_i^k}{2} + \dfrac{5\tau_i^k}{2} + 2\zeta_i^k, & i \in \mathcal{I}_{\mathrm{c}}. \end{cases}\end{aligned} \tag{85}$$

**Lemma H.3** (Unconditional terminal lower bound)**.** *For every $k \ge 1$, without assuming feasibility of $y^{k+1}$,*

$$\begin{aligned}&\mathcal{L}(x^{k+1}, y^{k+1}, \lambda^{k+1}, L^k, \mu^k, \rho^{k+1}) \\ &\ge \sum_{i\in\mathcal{I}_o} \bar{f}_i(A_i x^{k+1}, L^k, \mu^k) + \frac{1}{2}\sum_{i\in\mathcal{I}_c} \bar{f}_i(y_i^{k+1}, L^k, \mu^k) \\ &\quad - \sum_{i\in\mathcal{I}} \left[ \widetilde{C}_{r,i}^k \|y_i^{k+1} - A_i x^{k+1}\|^2 + C_{d,i}^k \|d_i^k\|^2 + \frac{\rho^k}{4\sigma}\|y_i^k - A_i x^k\|^2 \right].\end{aligned} \tag{86}$$

*Proof.* Write $r_i^{k+1} = y_i^{k+1} - A_i x^{k+1}$ and let $\widetilde{f}_i^k$ be the model-only function from Equation 56. It is $\zeta_i^k$-smooth and $q_i^k = \nabla \widetilde{f}_i^k(y_i^{k+1})$. Apply Equation 56 at $y_i^{k+1}$ and $A_i x^{k+1}$ and apply the descent lemma to $\widetilde{f}_i^k$ between those points. Since $A_i x^{k+1} - y_i^k = d_i^k - r_i^{k+1}$, expanding and using the Young inequality $\|d_i^k - r_i^{k+1}\|^2 \le 2\|d_i^k\|^2 + 2\|r_i^{k+1}\|^2$ gives

$$\bar{f}_i(y_i^{k+1}, L^k, \mu^k) - \langle q_i^k, r_i^{k+1}\rangle \ge \bar{f}_i(A_i x^{k+1}, L^k, \mu^k) - \frac{3\tau_i^k}{2}\|d_i^k\|^2 - \left(\tau_i^k + \frac{\zeta_i^k}{2}\right)\|r_i^{k+1}\|^2. \tag{87}$$

Model optimality and the multiplier update give

$$\lambda_i^{k+1} = -q_i^k - m_i^k d_i^k - \rho^k(y_i^k - A_i x^k) + \sigma\rho^k r_i^{k+1}.$$

Insert this into the terminal Lagrangian. The inequalities

$$-m_i^k\langle d_i^k, r_i^{k+1}\rangle \ge -\frac{m_i^k}{2}(\|d_i^k\|^2 + \|r_i^{k+1}\|^2)$$

and

$$-\rho^k\langle y_i^k - A_i x^k, r_i^{k+1}\rangle \ge -\sigma\rho^k\|r_i^{k+1}\|^2 - \frac{\rho^k}{4\sigma}\|y_i^k - A_i x^k\|^2$$

cancel the positive $\sigma\rho^k\|r_i^{k+1}\|^2$. Discarding the remaining positive terminal penalty gives the objective-block coefficients above and the pre-wall constraint coefficients.

For $i \in \mathcal{I}_{\mathrm{c}}$ and $F_i^k(y) = g(f_i(y), L^k)$, the chain rule gives $\nabla F_i^k(y) = g'(f_i(y), L^k)\nabla f_i(y)$, so $F_i^k$ is $(\tau_i^k + \zeta_i^k)$-smooth: the product-difference bound contributes $\sqrt{2L^kG}L_i = \tau_i^k$ through $|g'| \le \sqrt{2L^kG}$ and the $L_i$-smoothness of $f_i$, and $L^k M_i^2 = \zeta_i^k$ through the $L^k$-Lipschitz continuity of $g'(\cdot, L^k)$ and $\|\nabla f_i\| \le M_i$. Moreover $|g'(s,L)|^2 \le 2Lg(s,L)$ for every $s$: for $s \ge 0$, Equation 51 gives $|g'(s,L)|^2 = L^2\max(\beta(L) - s, 0)^2 = 2Lg(s,L)$, while for $s < 0$ it gives $|g'(s,L)|^2 = 2LG \le 2Lg(s,L)$ because $g(s,L) > G$ there by Lemma G.6. Hence $\|\nabla F_i^k(y)\|^2 \le |g'(f_i(y), L^k)|^2 M_i^2 \le$

$2L^k M_i^2 F_i^k(y) = 2\zeta_i^k F_i^k(y)$. Combining the descent lemma for the $(\tau_i^k + \zeta_i^k)$-smooth $F_i^k$ with this gradient bound and the Young inequality $\sqrt{2\zeta_i^k F_i^k(y)}\|r\| \le F_i^k(y)/2 + \zeta_i^k\|r\|^2$, we obtain

$$F_i^k(y-r) \ge \frac{1}{2}F_i^k(y) - \left(\frac{\tau_i^k}{2} + \frac{3\zeta_i^k}{2}\right)\|r\|^2.$$

Apply this with $y = y_i^{k+1}$ and $r = r_i^{k+1}$. Adding its loss to the pre-wall residual coefficient gives the sharper coefficient $m_i^k/2+3\tau_i^k/2+2\zeta_i^k$. Since $\tau_i^k \ge 0$, enlarging it to the displayed $m_i^k/2+5\tau_i^k/2+2\zeta_i^k$ weakens the lower bound in the valid direction. The enlargement is deliberate and purely cosmetic: it costs only a larger constant, is used nowhere in this proof, and lets the single constant $\widetilde{C}_{r,i}^k$ declared at the head of this subsection be quoted verbatim in Equation 88. This completes the proof. □

Combining Equation 86 with Equation 84 and $r_i^{k+1} = \Delta\lambda_i^{k+1}/(\sigma\rho^k)$ gives

$$\begin{aligned} &W_{k+1} - \Psi^k \\ &\le \sum_i \left[\frac{\widetilde{C}_{r,i}^k}{\sigma^2(\rho^k)^2} - \frac{\sigma_3}{\rho^k}\right]\|\Delta\lambda_i^{k+1}\|^2 + \sum_i (C_{d,i}^k - D_{y,i}^k)\|d_i^k\|^2 \\ &\quad - D_x^k\|A\Delta x^k\|^2 + \kappa_\Psi U^k + \frac{\rho^k}{4\sigma^3(\rho^{k-1})^2}\|\Delta\lambda^k\|^2, \end{aligned} \tag{88}$$

where

$$W_{k+1} \coloneqq \sum_{i\in\mathcal{I}_o} \bar{f}_i(A_i x^{k+1}, L^k, \mu^k) + R(L^k) + \frac{1}{2}\sum_{i\in\mathcal{I}_c} \bar{f}_i(y_i^{k+1}, L^k, \mu^k). \tag{89}$$

This estimate is valid before $y^{k+1}$ is known feasible.

**Lemma H.4** (Initialization)**.** *Under Equation 28, if $\beta^1 \le m$, then $d_i^1 = 0$ for $i \in \mathcal{I}_c$ and $\|d_i^1\| \le \widehat{C}_i/m_i^1$ for $i \in \mathcal{I}_o$. Moreover, $\|y^2 - Ax^2\| \le \|d^1\|$, and*

$$\begin{aligned} \Psi^2 \le\ & S_0 + R(L^1) + \sum_{i\in\mathcal{I}_o} \frac{\widehat{C}_i^2}{m_i^1} \\ &+ \left(\rho^1 + \frac{\rho^2}{2}\right)\sum_{i\in\mathcal{I}_o} \frac{\widehat{C}_i^2}{(m_i^1)^2} \\ &+ \frac{25}{2\rho^2}\sum_{i\in\mathcal{I}_o}\left[(\widehat{\Gamma}_i^2)^2 + \frac{(\rho^1)^2}{26^2}\right]\frac{\widehat{C}_i^2}{(m_i^1)^2}. \end{aligned} \tag{90}$$

*Proof.* The hypothesis $\beta^1 \le m$ is automatic under Equation 28: since $G = m^2L^0/2$ and $L^1 = L^0(1+\xi)$, Equation 50 gives $\beta^1 = \sqrt{2G/L^1} = m/\sqrt{1+\xi} \le m$. At initialization $e_i^1 = 0$. If $i \in \mathcal{I}_c$, then $f_i(y_i^1) \ge m \ge \beta^1$ and hence $g(\ell_i^1(y_i^1), L^1) = 0$. The constraint model is a sum of two nonnegative terms and has value zero at $y_i^1$; strong convexity makes this its unique minimizer, so $d_i^1 = 0$. For $i \in \mathcal{I}_o$, the block optimality condition Equation 75 at $k = 1$ collapses to $0 = q_i^1 + m_i^1 d_i^1$ with $q_i^1 = \nabla h_i(y_i^1) + w_i^2$, because $\lambda^1 = 0$ and $y_i^1 = A_i x^1$ annihilate its two consensus terms. By Equation 31, $\|\nabla h_i(y_i^1)\| \le D_i$, while $w_i^2 \in \partial\psi_i(p_i^2)$ by Equation 40 gives $\|w_i^2\| \le C_i^\psi$; hence $\|d_i^1\| = \|q_i^1\|/m_i^1 \le (D_i + C_i^\psi)/m_i^1 = \widehat{C}_i/m_i^1$.

Let $Q$ be the orthogonal projector onto $A(\ker C)$. At $k = 1$ the right side of Equation 77 is $A^T d^1$: its multiplier term drops because $\lambda^1 = 0$, and $y^1 = Ax^1$ turns $y^2 - Ax^1$ into $d^1$. Contract Equation 77 with arbitrary directions in $\ker C$. Since $C\Delta x^1 = 0$, this shows that $d^1 - (1+\theta_2)A\Delta x^1$ is orthogonal to $A(\ker C)$, while $A\Delta x^1 \in A(\ker C)$. Applying $Q$ therefore gives $A\Delta x^1 = Qd^1/(1+\theta_2)$. Hence $\|A\Delta x^1\| \le \|d^1\|/(1+\theta_2)$. Moreover, $y^2 - Ax^2 = d^1 - A\Delta x^1 = (I-Q)d^1 + [\theta_2/(1+\theta_2)]Qd^1$; orthogonality of $Qd^1$ and $(I-Q)d^1$ gives $\|y^2 - Ax^2\| \le \|d^1\|$. The corrected objective aggregate at $y^2$ is at most

$$S_0 + R(L^1) + \sum_{i\in\mathcal{I}_o} \widehat{C}_i\|d_i^1\| \le S_0 + R(L^1) + \sum_{i\in\mathcal{I}_o} \widehat{C}_i^2/m_i^1,$$

using Equation 45, the $\widehat{C}_i$-Lipschitz property, and $(\psi_i)_{\mu^1} \le \psi_i$. For $i \in \mathcal{I}_c$, $d_i^1 = 0$, hence $y_i^2 = y_i^1$. Since $\beta^2 \le \beta^1 \le m \le f_i(y_i^2)$, one has $g(f_i(y_i^2), L^2) = 0$. Thus every constraint-wall value

$g(f_i(y_i^2), L^2)$ appearing in $\Psi^2$ is exactly zero. Expanding $\Psi^2$ and inserting the two projector bounds yields Equation 90. Here $\lambda^2 = \rho^1(y^2 - Ax^2)$ is the multiplier update at $k = 1$ read with $\lambda^1 = 0$ and $\sigma = 1$, and in $\kappa_\Psi \Theta^2$ the dual term $\sigma_2 \|\Delta\lambda^2\|^2/\rho^2$ disappears because $\sigma_2 = 0$, while $\sigma_1 = 1$, $\kappa_\Psi \le 5/2$, and $1 + \theta_2 = 26$ bound the surviving primal term by the one displayed in Equation 90; these values are fixed by Equation 28. □

### H.3 Parameter Verification and Split-Feasibility Induction

**Lemma H.5** (Explicit parameter margins)**.** *Under Equation 28, for every block and every admissible index,*

$$-\sigma_3 \le -d_0, \tag{91}$$

$$\frac{\rho^k}{4\sigma^3(\rho^{k-1})^2} - \frac{\sigma_3}{\rho^{k-1}} \le 0, \tag{92}$$

$$\frac{\widetilde{C}_{r,i}^k}{\sigma^2(\rho^k)^2} - \frac{\sigma_3}{\rho^k} \le 0, \tag{93}$$

$$D_x^k \ge d_0 \rho^k, \tag{94}$$

$$D_{y,i}^k \ge d_0 \rho^k, \tag{95}$$

$$C_{d,i}^k - D_{y,i}^k \le 0, \tag{96}$$

$$\sqrt{2G/L^1} \le m. \tag{97}$$

*In addition, the finite-wall budget holds:*

$$\begin{aligned} S_0 + R(L^1) + \sum_{i\in\mathcal{I}_o} \frac{\widehat{C}_i^2}{m_i^1} + \left(\rho^1 + \frac{\rho^2}{2} + \frac{\rho^2}{4}\right) \sum_{i\in\mathcal{I}_o} \frac{\widehat{C}_i^2}{(m_i^1)^2} \\ + \frac{25}{2\rho^2} \sum_{i\in\mathcal{I}_o} \left[(\widehat{\Gamma}_i^2)^2 + \frac{(\rho^1)^2}{26^2}\right] \frac{\widehat{C}_i^2}{(m_i^1)^2} + \kappa_\Psi \Sigma_U \le \frac{G}{2}. \end{aligned} \tag{98}$$

*Proof.* Set $\varepsilon = 1/1209$, $\theta_1 = 121909/121000$, $\theta_2 = 25$, $\eta = 1209$, and $d_0 = 1/2$. The first four clauses in the definition of $\rho^0$, namely the constant 1 together with the three data clauses $1209H_{\max}$, $1209mL^0L_{\max}$, and $1209L^0M_{\max}^2$, together with $\eta = 1209$, imply, for every $i, k$,

$$0 \le \frac{\tau_i^k}{\rho^k} \le \varepsilon, \qquad 0 \le \frac{\zeta_i^k}{\rho^k} \le \varepsilon. \tag{99}$$

Indeed, this is immediate for objective blocks. For constraint blocks use $\sqrt{2L^kG} = mL^0\sqrt{1+\xi k^p}$ and the constant ratio $L^k/\rho^k = L^0/\rho^0$. Since

$$\theta_1 - 1 = \frac{909}{121000}, \qquad \upsilon_i^k = (\theta_1 - 1)\rho^k + \theta_1 \tau_i^k,$$

the displayed bounds give the exact normalized estimates

$$\begin{aligned} \frac{|\upsilon_i^k|}{\rho^k} &\le \theta_1(1+\varepsilon) - 1 = \frac{1009}{120900}, \\ \frac{\Gamma_i^k}{\rho^k} &= \frac{|\upsilon_i^k| + a_i^k}{\rho^k} \le \frac{1009}{120900} + \frac{1}{1209} = \frac{1109}{120900}, \end{aligned} \tag{100}$$

The same clauses bound the lagged drift constant by $b_i^k/\rho^k \le 3\varepsilon = 1/403$: for $i \in \mathcal{I}_o$ it equals $H_i/\rho^k \le \varepsilon$, and for $i \in \mathcal{I}_c$ it equals $(2\zeta_i^k + \tau_i^k)/\rho^k \le 3\varepsilon$. Put

$$c = \frac{1009}{120900}, \qquad e = \frac{1}{403}, \qquad R_k = \frac{\rho^{k+1}}{\rho^k} \in [1, 2].$$

Then $\widehat{\Gamma}_i^{k+1} = |\upsilon_i^k| + b_i^{k+1}$ gives

$$\widehat{\Gamma}_i^{k+1}/\rho^{k+1} \le c/R_k + e.$$

The exact inequality

$$c^2 - 2e^2 = \frac{838081}{14616810000} > 0$$

implies $c^2/R_k \geq c^2/2 > e^2$. Hence the elementary identity

$$(c+e)^2 - R_k(c/R_k + e)^2 = (R_k - 1)(c^2/R_k - e^2) \geq 0$$

gives, with $c + e = 1309/120900$,

$$\frac{(\widehat{\Gamma}_i^{k+1})^2}{\rho^k \rho^{k+1}} \leq \left(\frac{1309}{120900}\right)^2.$$

Since $\kappa_\Psi \leq 5/2$, the definitions of $D_{y,i}^k$ and $D_x^k$ now give

$$\begin{aligned} \frac{D_{y,i}^k}{\rho^k} &\geq \frac{178630370053}{353726802000} > \frac{1}{2}, \\ \frac{D_x^k}{\rho^k} &\geq \frac{1+2\theta_2}{2} - 25 = \frac{1}{2}. \end{aligned} \tag{101}$$

Thus Equation 94–Equation 95 hold with $d_0 = 1/2$. For the terminal residual coefficient,

$$\frac{\widetilde{C}_{r,i}^k}{\rho^k} \leq \frac{\theta_1(1+\varepsilon)}{2} + \frac{9\varepsilon}{2} = \frac{122809}{241800} < 1,$$

which proves Equation 93. To prove Equation 96 without separately rounding $C_{d,i}^k$ and $D_{y,i}^k$, put $s = \tau_i^k/\rho^k \in [0, \varepsilon]$. The exact bounds above yield

$$\frac{D_{y,i}^k - C_{d,i}^k}{\rho^k} \geq \frac{(\theta_1 - 1)(1+s)}{2} - \frac{3s}{2} - \frac{25}{2}\left[\left(\frac{1109}{120900}\right)^2 + \left(\frac{1309}{120900}\right)^2\right].$$

The right side is affine in $s$, and at its two endpoints $s = 0$ and $s = \varepsilon$ it equals, respectively,

$$\frac{27393814}{22107925125} > 0, \qquad \frac{881}{584672400} > 0.$$

It is therefore nonnegative throughout the interval, proving Equation 96. Since $\rho^k/\rho^{k-1} \leq 2$, the left side of Equation 92 is at most $-1/(2\rho^{k-1})$. Conditions Equation 91 and Equation 97 follow directly from $\sigma_3 = 1$, $d_0 = 1/2$, and $G = m^2 L^0/2$, the latter also using $L^1 = L^0(1+\xi) \geq L^0$: indeed $\sqrt{2G/L^1} = m\sqrt{L^0/L^1} \leq m$.

It remains to prove Equation 98. The branch $L^0 \geq 12S_0/m^2$ gives

$$S_0 \leq \frac{m^2 L^0}{12} = \frac{G}{6}.$$

Next, $R(L^1) = \eta G_\psi/(2\rho^1) \leq \eta \widehat{G}_0/(2\rho^0)$, using $\rho^1 \geq \rho^0$ and $G_\psi = \sum_{i \in \mathcal{I}_\circ} (C_i^\psi)^2 \leq \sum_{i \in \mathcal{I}_\circ} \widehat{C}_i^2 = \widehat{G}_0$, which holds because $\widehat{C}_i = D_i + C_i^\psi \geq C_i^\psi$. Since $m_i^1 \geq \theta_1 \rho^1 \geq \theta_1 \rho^0$, $\rho^1 \leq 2\rho^0$, and $\rho^2 \leq 3\rho^0$ (using $\xi \leq 1$ and $2^p \leq 2$), the linear and three quadratic initialization terms are bounded by

$$\left(\frac{1}{\theta_1} + \frac{17}{4\theta_1^2}\right)\frac{\widehat{G}_0}{\rho^0}.$$

Furthermore, $5\kappa_\Psi \leq 25/2$, $(\widehat{\Gamma}_i^2)^2 \leq (1309/120900)^2 \rho^1 \rho^2$, and $1 + \theta_2 = 26$, so the initial $\Theta^2$ term is at most

$$\frac{25}{2\theta_1^2}\left[\left(\frac{1309}{120900}\right)^2 + \frac{1}{26^2}\right]\frac{\widehat{G}_0}{\rho^0}.$$

The exact initialization coefficient satisfies

$$\frac{1}{\theta_1} + \frac{17}{4\theta_1^2} + \frac{25}{2\theta_1^2}\left[\left(\frac{1309}{120900}\right)^2 + \frac{1}{26^2}\right] = \frac{112940321939035250}{21723216943256361} < \frac{11}{2}.$$

Consequently the initialization and smoothing contribution is at most

$$\left(\frac{\eta}{2}+\frac{11}{2}\right)\frac{\widehat{G}_0}{\rho^0}\le\frac{G}{6}$$

by the penultimate clause in the definition of $\rho^0$. Finally, Equation 72 and $\kappa_\Psi \le 5/2$ give the exact bound

$$\kappa_\Psi\Sigma_U\le\frac{p^2H_p(\xi)}{\rho^0}\left(\frac{25}{2}G_\psi+\frac{25}{4}GL^0\bar{M}\right)\le\frac{G}{6}$$

by the last clause in $\rho^0$. Thus the three allocated contributions on the left side of Equation 98 are each at most $G/6$, and their sum is at most $G/2$. □

**Theorem H.6** (Feasibility and one-step Lyapunov decrease)**.** *Under Equation 28, $f_i(y_i^k)\ge 0$ for every $i\in\mathcal{I}_c$ and $k\ge 1$. Moreover, for $k\ge 2$,*

$$\Psi^{k+1}-\Psi^k\le-\frac{d_0}{\rho^k}\|\Delta\lambda^{k+1}\|^2-d_0\rho^k\|d^k\|^2-d_0\rho^k\|A\Delta x^k\|^2+\kappa_\Psi U^k. \tag{102}$$

*Proof.* Throughout this proof, call a split iterate $y^k$ *wall-interior* when $f_i(y_i^k)\ge 0$ for every $i\in\mathcal{I}_c$; this is the condition asserted by the theorem, and it is distinct from feasibility of a point $x$ in the sense of Assumption 3.3, which also requires $Cx=d$. The initial split iterate is wall-interior: $y_i^1=A_ix^1$ and $f_i(A_ix^1)\ge m>0$ by Assumption 3.3. By Equation 97 and Lemma H.4, every constraint block satisfies $y_i^2=y_i^1$, so $y^2$ is wall-interior as well.

For the first unverified point, set $k=2$ in Equation 88. Conditions Equation 93, Equation 94, and Equation 96 discard its current dual, $x$, and $y$ terms. The unmatched lagged term is

$$\frac{\rho^2}{4(\rho^1)^2}\|\Delta\lambda^2\|^2=\frac{\rho^2}{4}\|y^2-Ax^2\|^2\le\frac{\rho^2}{4}\sum_{i\in\mathcal{I}_o}\frac{\widehat{C}_i^2}{(m_i^1)^2},$$

where the first equality uses $\Delta\lambda^2=\lambda^2=\rho^1(y^2-Ax^2)$, i.e. the multiplier update at $k=1$ together with $\lambda^1=0$ and $\sigma=1$. Combine this with Equation 90, $U^2\le\Sigma_U$, and Equation 98; this yields $W_3\le G/2$. By Equation 46, the corrected objective part of $W_3$ is nonnegative, so $\frac{1}{2}\sum_{i\in\mathcal{I}_c}g(f_i(y_i^3),L^2)\le G/2$; since every wall term is nonnegative, each one individually satisfies $g(f_i(y_i^3),L^2)\le G$. Since $g(s,L)>G$ for $s<0$ by Lemma G.6, all $f_i(y_i^3)\ge 0$.

Suppose now that $y^2,\dots,y^K$ are wall-interior, $K\ge 3$. For $k=2,\dots,K-1$, continuation comparison is legal and Equation 84, Equation 94, and Equation 95 give

$$\Psi^{k+1}-\Psi^k\le-\frac{\sigma_3}{\rho^k}\|\Delta\lambda^{k+1}\|^2+\kappa_\Psi U^k.$$

Apply Equation 88 at the unverified terminal index $K$. Conditions Equation 93, Equation 94, and Equation 96 absorb its current terms, while Equation 92 absorbs its lagged dual term $\rho^K\|\Delta\lambda^K\|^2/(4\sigma^3(\rho^{K-1})^2)$ against the matching term $-\sigma_3\|\Delta\lambda^K\|^2/\rho^{K-1}$, which the telescoped decrease above contributes at $k=K-1$. Consequently

$$W_{K+1}\le\Psi^2+\kappa_\Psi\sum_{k=2}^{K}U^k+\frac{\rho^2}{4}\sum_{i\in\mathcal{I}_o}\frac{\widehat{C}_i^2}{(m_i^1)^2}\le G/2$$

by the same initialization budget Equation 98. Here the summand $\frac{\rho^2}{4}\sum_{i\in\mathcal{I}_o}\widehat{C}_i^2/(m_i^1)^2$ is not needed at the terminal index $K\ge 3$, where Equation 92 already absorbs the lagged dual term; it is retained only so that Equation 98 can be invoked verbatim. Applying the argument of the previous paragraph to $W_{K+1}\le G/2$ then proves $f_i(y_i^{K+1})\ge 0$ for every $i\in\mathcal{I}_c$, that is, $y^{K+1}$ is wall-interior. Since $y^1$, $y^2$, and $y^3$ are wall-interior and this step is available for every $K\ge 3$, strong induction gives wall-interiority of $y^k$ for every $k\ge 1$, which is the first assertion of the theorem. The induction closes without using a global movement budget.

Once wall-interiority is known, $\Psi^{k+1}\le\bar{\Psi}^{k+1}$ is always legal. Combining Equation 84 with Equation 91, Equation 94, and Equation 95 gives Equation 102. □

### H.4 Stationarity and Rates

**Original-problem certificates.**

**Definition H.1** (Original-problem residuals)**.** *In all three notions, $\breve{y}_i = y_i$ for every constraint block $i \in \mathcal{I}_c$. These are nearby-point certificates: stationarity is evaluated at the reconstructed block points $\breve{y}_i$ rather than at the common original point $A_i x$, and the two agree only up to the consensus term $\|Ax - \breve{y}\|$. The weak $\epsilon$-KKT and $\epsilon$-KKT conditions are, respectively,*

$$\text{weak } \epsilon\text{-KKT:} \quad \begin{cases} \|Ax - \breve{y}\| \le \epsilon, \\ \left\| \sum_{i\in\mathcal{I}_o} A_i^T v_i + \sum_{i\in\mathcal{I}_c} \iota_i A_i^T \nabla f_i(y_i) + C^T \zeta \right\| \le \epsilon, \\ v_i \in \partial\varphi_i(\breve{y}_i),\ Cx = d,\ f_i(y_i) \ge 0,\ \iota_i \le 0; \end{cases}$$

$$\epsilon\text{-KKT:} \quad \begin{cases} \|Ax - \breve{y}\| \le \epsilon, \\ \left\| \sum_{i\in\mathcal{I}_o} A_i^T v_i + \sum_{i\in\mathcal{I}_c} \iota_i A_i^T \nabla f_i(y_i) + C^T \zeta \right\| \le \epsilon, \\ v_i \in \partial\varphi_i(\breve{y}_i),\ Cx = d,\ f_i(y_i) \ge 0,\ \iota_i \le 0, \\ |\iota_i f_i(y_i)| \le \epsilon \quad (i \in \mathcal{I}_c). \end{cases} \tag{103}$$

*The $\epsilon$-b-FJ conditions are*

$$\begin{cases} \|Ax - \breve{y}\| \le \epsilon, \\ \left\| \alpha \sum_{i\in\mathcal{I}_o} A_i^T v_i + \sum_{i\in\mathcal{I}_c} \bar{\iota}_i A_i^T \nabla f_i(y_i) + C^T \bar{\zeta} \right\| \le \epsilon, \\ v_i \in \partial\varphi_i(\breve{y}_i),\ Cx = d,\ f_i(y_i) \ge 0,\ \alpha > 0,\ \bar{\iota}_i \le 0, \quad \alpha + \|\bar{\iota}\|_1 = 1, \\ |\bar{\iota}_i f_i(y_i)| \le \epsilon, \qquad \bar{\iota}_i \max\{f_i(y_i) - \epsilon, 0\} = 0 \quad (i \in \mathcal{I}_c). \end{cases} \tag{104}$$

For the candidate certificate tuple generated by step $k$, let $\varpi^{k+1}$ be the multiplier of the $x^{k+1}$ subproblem and set

$$\begin{aligned} \zeta^{k+1} &= \rho^k \varpi^{k+1}, & \alpha^{k+1} &= \frac{1}{1 + \|\iota^{k+1}\|_1}, \\ \bar{\iota}^{k+1} &= \alpha^{k+1} \iota^{k+1}, & \bar{\zeta}^{k+1} &= \alpha^{k+1} \zeta^{k+1}, \\ \omega^{k+1} &:= \beta^k + \frac{1}{2} \max_{i\in\mathcal{I}_c} L_i \|d_i^k\|^2. \end{aligned} \tag{105}$$

By its definition, $\alpha^{k+1} \in (0, 1]$. Define

$$\begin{aligned} \Xi^{k+1} &:= \sum_{i\in\mathcal{I}_o} A_i^T v_i^{k+1} + \sum_{i\in\mathcal{I}_c} \iota_i^{k+1} A_i^T \nabla f_i(y_i^{k+1}) + C^T \zeta^{k+1}, \\ \mathrm{Crit}_{\text{w-KKT}}^{k+1} &:= \|Ax^{k+1} - \breve{y}^{k+1}\| + \|\Xi^{k+1}\|, \\ \mathrm{Crit}_{\text{KKT}}^{k+1} &:= \mathrm{Crit}_{\text{w-KKT}}^{k+1} + \sum_{i\in\mathcal{I}_c} |\iota_i^{k+1} f_i(y_i^{k+1})|, \\ \mathrm{Crit}_{\text{b-FJ}}^{k+1} &:= \|Ax^{k+1} - \breve{y}^{k+1}\| + \alpha^{k+1} \|\Xi^{k+1}\| + \sum_{i\in\mathcal{I}_c} |\bar{\iota}_i^{k+1} f_i(y_i^{k+1})| + \omega^{k+1}. \end{aligned} \tag{106}$$

**Lemma H.7** (Lyapunov lower bound)**.** *Under Equation 28 and the ordinary polynomial schedule Equation 10, for every $k \ge 2$,*

$$\Psi^k \ge \underline{\Psi} := -\frac{\widehat{G}_0}{2\rho^0} - \kappa_\Psi \Sigma_U > -\infty. \tag{107}$$

*Proof.* By Equation 46, with $\mu^k = \mu_{\mathrm{IP}}(L^k)$ by Equation 10 and Equation 57, and the $\widehat{C}_i$-Lipschitz property of the objective surrogate,

$$\sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i^k, L^k, \mu^k) + R(L^k) \ge -\sum_{i\in\mathcal{I}_o} \widehat{C}_i \|y_i^k - A_i x^k\|.$$

For each $i \in \mathcal{I}_\circ$, Young's inequality completes the square between $-\widehat{C}_i\|y_i^k - A_i x^k\|$ and the corresponding objective-block quadratic penalty, contributing at worst $-\widehat{C}_i^2/(2\rho^k) \geq -\widehat{C}_i^2/(2\rho^0)$. The constraint-block quadratic penalties are nonnegative and are discarded, as are the nonnegative wall and $\Theta^k$ terms. Using the multiplier update then gives

$$\Psi^k \geq (\lambda^k)^T(y^k - Ax^k) - \frac{\widehat{G}_0}{2\rho^0} = \frac{(\lambda^k)^T\Delta\lambda^k}{\sigma\rho^{k-1}} - \frac{\widehat{G}_0}{2\rho^0}.$$

By Equation 74 and Equation 28, $\kappa_\Psi = (\xi + 4)/2 \in (2, 5/2]$. Moreover, $U^j \geq 0$ by Equation 71, and Lemma G.13 gives the finite tail $0 \leq \sum_{j=k}^{\infty} U^j \leq \Sigma_U$. Define the exact nonnegative drift-tail correction $\widehat{\Psi}^k = \Psi^k + \kappa_\Psi \sum_{j=k}^{\infty} U^j$. By Equation 102, $\widehat{\Psi}^k$ is nonincreasing. If it ever fell below $-\widehat{G}_0/(2\rho^0)$, its tail sum relative to that level would diverge to $-\infty$. On the other hand, the polarization identity $(\lambda^k)^T\Delta\lambda^k = (\|\lambda^k\|^2 - \|\lambda^{k-1}\|^2 + \|\Delta\lambda^k\|^2)/2$ and $\rho^{k-1} \leq \rho^k$ give

$$\widehat{\Psi}^k + \frac{\widehat{G}_0}{2\rho^0} \geq \frac{\|\lambda^k\|^2 - \|\lambda^{k-1}\|^2}{2\sigma\rho^{k-1}} \geq \frac{\|\lambda^k\|^2}{2\sigma\rho^k} - \frac{\|\lambda^{k-1}\|^2}{2\sigma\rho^{k-1}}.$$

The partial sums telescope; discarding the nonnegative terminal term leaves the finite initial endpoint as a uniform lower bound, a contradiction. Hence $\widehat{\Psi}^k \geq -\widehat{G}_0/(2\rho^0)$, and subtracting its nonnegative drift tail, bounded above by $\kappa_\Psi\Sigma_U$, proves the claim. □

**Corollary H.8** (Global movement budget)**.** *Under Equation 28 and the ordinary polynomial schedule Equation 10, define*

$$\mathcal{B} := \frac{\Psi^2 - \underline{\Psi} + \kappa_\Psi\Sigma_U}{d_0}. \tag{108}$$

*The upper bound Equation 90 and the lower bound from Lemma H.7 at $k = 2$ make $\Psi^2$ finite; $\underline{\Psi}$, $\kappa_\Psi$, and $\Sigma_U$ are also finite and $d_0 > 0$, so $\mathcal{B} < \infty$ and*

$$\sum_{k=2}^{\infty}\left[\frac{\|\Delta\lambda^{k+1}\|^2}{\rho^k} + \rho^k\|d^k\|^2 + \rho^k\|A\Delta x^k\|^2\right] \leq \mathcal{B}. \tag{109}$$

*Proof.* Sum Equation 102 from $k = 2$ to $K$, use $\Psi^{K+1} \geq \underline{\Psi}$ from Lemma H.7, and bound $\sum_{k=2}^{K} U^k \leq \Sigma_U$ by Lemma G.13. The result is

$$d_0\sum_{k=2}^{K}\left[\frac{\|\Delta\lambda^{k+1}\|^2}{\rho^k} + \rho^k\|d^k\|^2 + \rho^k\|A\Delta x^k\|^2\right] \leq \Psi^2 - \underline{\Psi} + \kappa_\Psi\Sigma_U.$$

After division by $d_0 > 0$, the partial sums on the left of Equation 109 are nondecreasing because every summand is nonnegative, and they are uniformly bounded above by the finite constant $\mathcal{B}$. They therefore converge to a finite limit; letting $K \to \infty$ proves Equation 109. □

**Lemma H.9** (Bounded rescaled multipliers)**.** *Under Equation 28 and the ordinary polynomial schedule Equation 10, define*

$$\begin{aligned} c_\upsilon &:= \theta_1 - 1 + \frac{\theta_1}{1209} = \frac{1009}{120900}, \\ C_\lambda &:= \frac{\sqrt{\widehat{G}_0}}{\sqrt{\rho^0}} + \sqrt{\frac{2GL^0\bar{M}}{\rho^0}} + (1 + c_\upsilon)\sqrt{\mathcal{B}}. \end{aligned} \tag{110}$$

*Then, for every $k \geq 1$,*

$$\frac{\|\lambda^{k+1}\|}{\sqrt{\rho^k}} \leq C_\lambda. \tag{111}$$

*Consequently, for every $j \geq 1$,*

$$\frac{\|\lambda^j\|}{\sqrt{\rho^j}} \leq C_\lambda, \quad \frac{\|\lambda^j\|^2}{\rho^j} \leq C_\lambda^2, \quad \frac{\|\lambda^j\|}{\rho^j} \leq \frac{C_\lambda}{\sqrt{\rho^j}}. \tag{112}$$

*In particular, $\{\lambda^j/\rho^j\}_{j\geq 1}$ is bounded. If $p > 0$, then*

$$\frac{\|\lambda^j\|}{\rho^j} = \mathcal{O}((\rho^j)^{-1/2}) = \mathcal{O}(j^{-p/2}) \longrightarrow 0. \tag{113}$$

*When $p = 0$, the sequence remains bounded, but no decay is asserted.*

*Proof.* We first bound the stacked force $q^k$. For $i \in \mathcal{I}_o$, the current force formula, the bound $\|\nabla h_i(y_i^k)\| \le D_i$, and Equation 34 give $\|q_i^k\| \le D_i + C_i^\psi = \widehat{C}_i$. For $i \in \mathcal{I}_c$, Equation 51 and $\|\nabla f_i(y_i^k)\| \le M_i$ give $\|q_i^k\| \le \sqrt{2L^k G}\, M_i$. Since $q^k$ is the Euclidean stacking of these blocks,

$$\|q^k\|^2 = \sum_{i\in\mathcal{I}} \|q_i^k\|^2 \le \widehat{G}_0 + 2L^k G\bar{M}.$$

Thus $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for $a, b \ge 0$ gives

$$\|q^k\| \le \sqrt{\widehat{G}_0} + \sqrt{2L^k G\bar{M}}.$$

By definition, $\upsilon_i^k = m_i^k - \rho^k = (\theta_1 - 1)\rho^k + \theta_1\tau_i^k$. The bound $0 \le \tau_i^k/\rho^k \le 1/1209$ established in the proof of Lemma H.5 therefore gives $|\upsilon_i^k|/\rho^k \le \theta_1 - 1 + \theta_1/1209 = c_\upsilon$ for every block. If $\Upsilon^k = \operatorname{diag}(\upsilon_i^k I_{n_i})_{i\in\mathcal{I}}$, then this block-diagonal structure yields

$$\|\Upsilon^k d^k\|^2 = \sum_{i\in\mathcal{I}} |\upsilon_i^k|^2 \|d_i^k\|^2 \le c_\upsilon^2 (\rho^k)^2 \|d^k\|^2.$$

Under Equation 28, $\sigma = 1$, so stacking Equation 76 over the blocks gives

$$\lambda^{k+1} = -q^k - \Upsilon^k d^k - \rho^k A\Delta x^k.$$

For $k \ge 2$, each nonnegative summand in Equation 109 is at most $\mathcal{B}$. Hence $\sqrt{\rho^k}\|d^k\| \le \sqrt{\mathcal{B}}$ and $\sqrt{\rho^k}\|A\Delta x^k\| \le \sqrt{\mathcal{B}}$. Dividing the preceding identity by $\sqrt{\rho^k}$, applying the triangle inequality and the block-diagonal estimate, and then using Equation 57 gives

$$\begin{aligned}
\frac{\|\lambda^{k+1}\|}{\sqrt{\rho^k}} &\le \frac{\|q^k\|}{\sqrt{\rho^k}} + c_\upsilon\sqrt{\rho^k}\|d^k\| + \sqrt{\rho^k}\|A\Delta x^k\| \\
&\le \frac{\sqrt{\widehat{G}_0}}{\sqrt{\rho^k}} + \sqrt{\frac{2GL^k\bar{M}}{\rho^k}} + (1 + c_\upsilon)\sqrt{\mathcal{B}} \\
&\le \frac{\sqrt{\widehat{G}_0}}{\sqrt{\rho^0}} + \sqrt{\frac{2GL^0\bar{M}}{\rho^0}} + (1 + c_\upsilon)\sqrt{\mathcal{B}} = C_\lambda.
\end{aligned}$$

Here the last line also uses $\rho^k \ge \rho^0$.

It remains to treat $k = 1$, because Equation 109 starts at $k = 2$. Condition Equation 97, namely $\beta^1 \le m$, supplies the constraint-block endpoint hypothesis of Lemma H.4; that lemma gives $d_i^1 = 0$ for $i \in \mathcal{I}_c$ and $\|d_i^1\| \le \widehat{C}_i/m_i^1$ for $i \in \mathcal{I}_o$. Since $m_i^1 \ge \theta_1\rho^1$,

$$\|d^1\| \le \frac{\sqrt{\widehat{G}_0}}{\theta_1\rho^1}.$$

The second endpoint estimate $\|y^2 - Ax^2\| \le \|d^1\|$ is also exactly the conclusion of Lemma H.4; the multiplier update together with the initialization $\lambda^1 = 0$ and the parameter choice $\sigma = 1$ then gives $\lambda^2 = \rho^1(y^2 - Ax^2)$. Therefore

$$\frac{\|\lambda^2\|}{\sqrt{\rho^1}} \le \frac{\sqrt{\widehat{G}_0}}{\theta_1\sqrt{\rho^1}} \le \frac{\sqrt{\widehat{G}_0}}{\sqrt{\rho^0}} \le C_\lambda,$$

where the middle inequality uses $\theta_1 > 1$ and $\rho^1 \ge \rho^0$. This proves Equation 111 for all $k \ge 1$.

For $j \ge 2$, apply Equation 111 with $k = j - 1$ and use schedule monotonicity Equation 58:

$$\frac{\|\lambda^j\|}{\sqrt{\rho^j}} \le \frac{\|\lambda^j\|}{\sqrt{\rho^{j-1}}} \le C_\lambda.$$

The same inequality is immediate for $j = 1$ because $\lambda^1 = 0$. Squaring the first of the three same-index bounds in Equation 112 gives $\|\lambda^j\|^2/\rho^j \le C_\lambda^2$; dividing that first bound by $\sqrt{\rho^j}$ gives $\|\lambda^j\|/\rho^j \le C_\lambda/\sqrt{\rho^j}$. Finally, for $p > 0$, the schedule $\rho^j = \rho^0(1 + \xi j^p)$ proves Equation 113. At $p = 0$ it is constant on positive indices, so the proved bound remains uniform without implying decay. □

**Corollary H.10** (Bounded normalized dual shift)**.** *Under Equation 28 and the ordinary polynomial schedule Equation 10,* $e^1 = 0$*,*

$$\frac{\|e^2\|}{\rho^2} \le \frac{2\sqrt{\widehat{G}_0}}{\theta_1 \rho^1}. \tag{114}$$

*For every* $k \ge 3$*,*

$$\frac{\|e^k\|}{\rho^k} \le \frac{C_\lambda + \sqrt{\mathcal{B}}}{\sqrt{\rho^{k-1}}}. \tag{115}$$

*Consequently,* $\{e^k/\rho^k\}_{k\ge1}$ *is uniformly bounded for every* $p \in [0,1)$*, and* $e^k/\rho^k \to 0$ *when* $p > 0$*. Moreover, for every block* $i \in \mathcal{I}$ *and every* $k \ge 1$*,*

$$\frac{\|e_i^k\|}{m_i^k} \le \frac{1}{\theta_1}\frac{\|e^k\|}{\rho^k}. \tag{116}$$

*Proof.* The initialization $y^1 = Ax^1$ and $\lambda^1 = 0$ gives $e^1 = 0$. For $k = 2$, the multiplier update with $\sigma = 1$ gives $\lambda^2 = \rho^1(y^2 - Ax^2)$, and hence

$$\frac{\|e^2\|}{\rho^2} = \left(1 + \frac{\rho^1}{\rho^2}\right)\|y^2 - Ax^2\| \le 2\|d^1\| \le \frac{2\sqrt{\widehat{G}_0}}{\theta_1\rho^1}.$$

Here schedule monotonicity gives $\rho^1 \le \rho^2$, while $\|y^2 - Ax^2\| \le \|d^1\|$ is an endpoint conclusion of Lemma H.4; the stacked estimate $\|d^1\| \le \sqrt{\widehat{G}_0}/(\theta_1\rho^1)$ is the consequence derived in the proof of Lemma H.9 from the blockwise endpoint bounds and Equation 97.

Now let $k \ge 3$. Applying the multiplier update at index $k-1$ gives $y^k - Ax^k = \Delta\lambda^k/\rho^{k-1}$, so the definition of $e^k$ yields

$$\frac{\|e^k\|}{\rho^k} \le \frac{\|\lambda^k\|}{\rho^k} + \frac{\|\Delta\lambda^k\|}{\rho^{k-1}}.$$

By the same-index bound in Lemma H.9 and monotonicity of the schedule, the first term is at most $C_\lambda/\sqrt{\rho^{k-1}}$. The dual summand of Equation 109 at index $k-1 \ge 2$ gives

$$\frac{\|\Delta\lambda^k\|^2}{\rho^{k-1}} \le \mathcal{B}, \qquad \frac{\|\Delta\lambda^k\|}{\rho^{k-1}} \le \frac{\sqrt{\mathcal{B}}}{\sqrt{\rho^{k-1}}},$$

which proves Equation 115.

Since $\rho^{k-1} \ge \rho^0$, the two initial cases and Equation 115 give uniform boundedness for $p \in [0,1)$. If $p > 0$, then $\rho^{k-1} \to \infty$, so the same bound gives $e^k/\rho^k \to 0$. Finally, $m_i^k = \theta_1(\rho^k + \tau_i^k) \ge \theta_1\rho^k$ and $\|e_i^k\| \le \|e^k\|$, which proves Equation 116. □

For the step-$k$ candidate certificate tuple, the objective-block reconstruction and its transfer bounds are given in Equation 42–Equation 43. For constraint blocks,

$$\|\iota_i^{k+1}\nabla f_i(y_i^{k+1}) - q_i^k\| \le \sqrt{2L^kG}\,L_i\|d_i^k\|. \tag{117}$$

Also, $\iota_i^{k+1} \ne 0$ implies $\ell_i^k(y_i^{k+1}) < \beta^k$, and smoothness of $f_i$ gives

$$\iota_i^{k+1} \ne 0 \quad\Longrightarrow\quad 0 \le f_i(y_i^{k+1}) < \beta^k + \frac{L_i}{2}\|d_i^k\|^2 \le \omega^{k+1}. \tag{118}$$

This is the widened affine-force support; the old exact width $\beta^{k+1}$ does not apply.

Let $r_i^{k+1} = \upsilon_i^{k+1}$ for $i \in \mathcal{I}_o$ and $r_i^{k+1} = \iota_i^{k+1}\nabla f_i(y_i^{k+1})$ for $i \in \mathcal{I}_c$, and put $u_i^k = r_i^{k+1} - q_i^k$. Combining Equation 75 and Equation 77 gives the exact identity

$$\Xi^{k+1} = A^T[u^k - (m^k - \rho^k I)d^k] - (1+\theta_2)\rho^k A^T A\Delta x^k. \tag{119}$$

Set

$$r_* := \max\left\{H_{\max}/\rho^0, \sqrt{2GL^0}L_{\max}/\rho^0\right\}, \qquad c_y := \frac{1009}{120900} + \frac{242909}{121000}r_*.$$

The transfer bounds and $c_\upsilon = 1009/120900$ imply

$$\|\Xi^{k+1}\| \le \|A\|_2 \left[c_y \rho^k \|d^k\| + 26\rho^k \|A\Delta x^k\| + \mu^k J_H\right]. \tag{120}$$

The multiplier update and reconstruction similarly give

$$\|Ax^{k+1} - \breve{y}^{k+1}\| \le \frac{\|\Delta\lambda^{k+1}\|}{\rho^k} + \mu^k C_\psi. \tag{121}$$

Hence

$$\mathrm{Crit}^{k+1}_{\text{w-KKT}} \le \frac{\|\Delta\lambda^{k+1}\|}{\rho^k} + \|A\|_2 c_y \rho^k \|d^k\| + 26\|A\|_2 \rho^k \|A\Delta x^k\| + \mu^k Q_\psi. \tag{122}$$

Normalization gives $\iota_i^{k+1}, \bar{\iota}_i^{k+1} \le 0$ and $\alpha^{k+1} + \|\bar{\iota}^{k+1}\|_1 = 1$. From Equation 118 and feasibility,

$$\sum_{i\in\mathcal{I}_c} |\bar{\iota}_i^{k+1} f_i(y_i^{k+1})| \le \omega^{k+1} \|\bar{\iota}^{k+1}\|_1 \le \omega^{k+1}. \tag{123}$$

Thus

$$\mathrm{Crit}^{k+1}_{\text{b-FJ}} \le \mathrm{Crit}^{k+1}_{\text{w-KKT}} + 2\omega^{k+1}. \tag{124}$$

If $\mathrm{Crit}^{k+1}_{\text{b-FJ}} \le \epsilon$, then $\omega^{k+1} \le \epsilon$ and Equation 118 proves the exact final support line of Equation 104. Therefore the residual certifies the unchanged original b-FJ definition.

Set

$$M_F := \|A\|_2 \sqrt{\widehat{G}_0}. \tag{125}$$

**Lemma H.11** (Uniform Band-MFCQ bounds the objective multiplier)**.** *Under Assumption 4.2, let $(x, y, \breve{y}, v, \alpha, \bar{\iota}, \bar{\zeta})$ be an $\epsilon$-b-FJ point in the sense of Definition H.1, with $\epsilon \le \delta_{\mathrm{cq}}$. Then*

$$\alpha \ge \frac{\gamma_{\mathrm{cq}} - \epsilon}{\gamma_{\mathrm{cq}} + M_F}. \tag{126}$$

*If also $\epsilon \le \gamma_{\mathrm{cq}}/2$, then $\alpha \ge \alpha_{\min} := \gamma_{\mathrm{cq}}/[2(\gamma_{\mathrm{cq}} + M_F)]$.*

*Proof.* Let $\lambda_i = -\bar{\iota}_i \ge 0$. Since $\bar{\iota}_i \le 0$, the b-FJ normalization gives $\alpha + \sum_{i\in\mathcal{I}_c} \lambda_i = 1$. Thus $\alpha \in (0, 1]$ and $\sum_{i\in\mathcal{I}_c} \lambda_i = 1 - \alpha$. By Definition H.1, $\breve{y}_i = y_i$ for every $i \in \mathcal{I}_c$, and therefore

$$\max_{i\in\mathcal{I}_c} \|A_i x - y_i\| \le \|Ax - \breve{y}\| \le \epsilon \le \delta_{\mathrm{cq}}.$$

For every index with $\lambda_i > 0$, the sign and support conditions in Definition H.1 also give $0 \le f_i(y_i) \le \epsilon \le \delta_{\mathrm{cq}}$. Hence Assumption 4.2 supplies a single unit $d_{\mathrm{cq}} \in \ker C$ for which Equation 15 holds at every such index. Write the b-FJ stationarity residual as

$$s = \alpha \sum_{i\in\mathcal{I}_o} A_i^T v_i - \sum_{i\in\mathcal{I}_c} \lambda_i A_i^T \nabla f_i(y_i) + C^T \bar{\zeta}, \qquad \|s\| \le \epsilon.$$

For this direction, $Cd_{\mathrm{cq}} = 0$, so the equality-multiplier term vanishes. Moreover, the exact sum rule $\partial\varphi_i = \nabla h_i + \partial\psi_i$ and Equation 31 give $\|v_i\| \le D_i + C_i^\psi = \widehat{C}_i$. Zero-padding the vector $(v_i)_{i\in\mathcal{I}_o}$ on $\mathcal{I}_c$ therefore gives

$$\left\| \sum_{i\in\mathcal{I}_o} A_i^T v_i \right\| \le \|A\|_2 \left( \sum_{i\in\mathcal{I}_o} \|v_i\|^2 \right)^{1/2} \le \|A\|_2 \sqrt{\widehat{G}_0} = M_F.$$

On the other hand, $\nabla f_i(y_i)^T A_i d_{\mathrm{cq}} \ge \gamma_{\mathrm{cq}}$ whenever $\lambda_i > 0$; the terms with $\lambda_i = 0$ vanish. Contracting with $d_{\mathrm{cq}}$ and using $-d_{\mathrm{cq}}^T s \le \epsilon$ therefore yields

$$\gamma_{\mathrm{cq}}(1 - \alpha) \le \sum_{i\in\mathcal{I}_c} \lambda_i \nabla f_i(y_i)^T A_i d_{\mathrm{cq}} = \alpha d_{\mathrm{cq}}^T \sum_{i\in\mathcal{I}_o} A_i^T v_i - d_{\mathrm{cq}}^T s \le \alpha M_F + \epsilon.$$

Rearrangement proves Equation 126; the second assertion follows from $\epsilon \le \gamma_{\mathrm{cq}}/2$. □

**Theorem H.12** (Exact finite-time original-problem rates)**.** *Under Equation 28 and the ordinary polynomial schedule Equation 10, let $\mathcal{B} < \infty$ be the movement constant supplied by Corollary H.8. For every $K \geq 3$, let*

$$\begin{aligned} R_K &= \sum_{k=2}^{K} \rho^k, \qquad I_K = \sum_{k=2}^{K} (\rho^k)^{-1}, \qquad B_K = \sum_{k=2}^{K} \beta^k, \\ J_H &= \left( \sum_{i \in \mathcal{I}_o} (H_i C_i^{\psi})^2 \right)^{1/2}, \quad Q_\psi = C_\psi + \|A\|_2 J_H, \\ \mathcal{Q} &= \frac{\|A\|_2^2 c_y^2}{d_0} + \frac{26^2 \|A\|_2^2}{d_0}. \end{aligned} \tag{127}$$

*Then*

$$\begin{aligned} F_{\mathrm{W}}(K) &:= \frac{\sqrt{\mathcal{B}}}{K-1} \sqrt{I_K + \mathcal{Q} R_K} + \frac{\eta Q_\psi}{K-1} I_K, \\ F_{\mathrm{bFJ}}(K) &:= F_{\mathrm{W}}(K) + \frac{2B_K}{K-1} + \frac{L_{\max} \mathcal{B}}{d_0 \rho^2 (K-1)}, \\ \min_{j=3,\dots,K+1} \mathit{Crit}^j_{\mathit{w\text{-}KKT}} &\leq F_{\mathrm{W}}(K), \\ \min_{j=3,\dots,K+1} \mathit{Crit}^j_{\mathit{b\text{-}FJ}} &\leq F_{\mathrm{bFJ}}(K). \end{aligned} \tag{128}$$

*The second minimum is attained at one selected certificate-bearing iterate whose tuple is an $F_{\mathrm{bFJ}}(K)$-b-FJ point in the sense of Equation 104. For $0 < p < 1$,*

$$\begin{aligned} F_{\mathrm{W}}(K) &= \mathcal{O}(K^{(p-1)/2}) + \mathcal{O}(K^{-p}), \\ F_{\mathrm{bFJ}}(K) &= \mathcal{O}(K^{(p-1)/2}) + \mathcal{O}(K^{-p}) + \mathcal{O}(K^{-p/2}) + \mathcal{O}(K^{-1}). \end{aligned} \tag{129}$$

*Here the hidden constants may depend on the fixed $p \in (0,1)$, $\xi$, and problem and algorithm data, but not on $K$; these estimates are not uniform as $p$ or $\xi$ approaches an endpoint. Thus $p = 1/3$ gives w-KKT rate $\mathcal{O}(K^{-1/3})$, while $p = 1/2$ gives b-FJ rate $\mathcal{O}(K^{-1/4})$. At $p = 0$ the movement contribution is $\mathcal{O}(K^{-1/2})$, while the bounds $F_{\mathrm{W}}(K)$ and $F_{\mathrm{bFJ}}(K)$ have exact limiting floors $\eta Q_\psi / [\rho^0(1+\xi)]$ and $\eta Q_\psi / [\rho^0(1+\xi)] + 2m/\sqrt{1+\xi}$, respectively; hence original-problem convergence is not asserted.*

*Proof.* The factors $1/d_0$ in $\mathcal{Q}$ conservatively enlarge the sharper Cauchy–Schwarz constant. Sum Equation 122 for $k = 2, \dots, K$. For each $k$, write the three movement terms as the inner product of

$$\left( \frac{\|\Delta\lambda^{k+1}\|}{\sqrt{\rho^k}}, \sqrt{\rho^k}\|d^k\|, \sqrt{\rho^k}\|A\Delta x^k\| \right) \quad \text{and} \quad \left( \frac{1}{\sqrt{\rho^k}}, \|A\|_2 c_y \sqrt{\rho^k}, 26\|A\|_2 \sqrt{\rho^k} \right).$$

A single Cauchy–Schwarz inequality over all indices and all three channels, followed by Equation 109, gives the sharper common-factor bound

$$\begin{aligned} &\sum_{k=2}^{K} \left[ \frac{\|\Delta\lambda^{k+1}\|}{\rho^k} + \|A\|_2 c_y \rho^k \|d^k\| + 26\|A\|_2 \rho^k \|A\Delta x^k\| \right] \\ \leq &\sqrt{\mathcal{B}} \sqrt{I_K + \|A\|_2^2 (c_y^2 + 26^2) R_K} \\ \leq &\sqrt{\mathcal{B}} \sqrt{I_K + \mathcal{Q} R_K}, \end{aligned} \tag{130}$$

where the last inequality uses $d_0 = 1/2 < 1$. Thus the common multiplier $\sqrt{\mathcal{B}}$ applies to all three movement channels. Since $\sum_{k=2}^{K} \mu^k = \eta I_K$, averaging yields the bound $F_{\mathrm{W}}(K)$.

Next, $\beta^k \geq 0$ by Equation 50, and $2\omega^{k+1} \leq 2\beta^k + L_{\max}\|d^k\|^2$. Since $\rho^k \geq \rho^2$ for $k \geq 2$, the movement budget gives the sharper bound $\sum_{k=2}^{K} \|d^k\|^2 \leq \mathcal{B}/\rho^2$; since $d_0 = 1/2$, the conservative bound $\mathcal{B}/(d_0\rho^2)$ used in Equation 128 follows. Consequently,

$$\sum_{k=2}^{K} \mathrm{Crit}^{k+1}_{\text{b-FJ}} \leq \sum_{k=2}^{K} \mathrm{Crit}^{k+1}_{\text{w-KKT}} + 2 \sum_{k=2}^{K} \omega^{k+1} \leq (K-1) F_{\mathrm{bFJ}}(K).$$

Choose $j_* \in \{3, \dots, K+1\}$ minimizing the entire b-FJ residual. Then $\mathrm{Crit}^{j_*}_{\text{b-FJ}} \le F_{\mathrm{bFJ}}(K)$ at one common selected certificate-bearing iterate; its consensus, normalized stationarity, complementarity, and nonnegative $\omega^{j_*}$ terms are all bounded there, while Equation 118 supplies the exact support clause.

For $0 < p < 1$,

$$R_K = \mathcal{O}(\rho^0[K + \xi K^{p+1}]), \qquad I_K = \mathcal{O}((\rho^0\xi)^{-1}K^{1-p}), \qquad B_K = \mathcal{O}(m\xi^{-1/2}K^{1-p/2}).$$

Substitution gives Equation 129; balancing its exponents yields $p = 1/3$ for w-KKT and $p = 1/2$ for b-FJ. The widened-support term is $\mathcal{O}(K^{-1})$. If $p = 0$, direct substitution gives $R_K = (K-1)\rho^0(1+\xi)$, $I_K = (K-1)/[\rho^0(1+\xi)]$, and $B_K = (K-1)m/\sqrt{1+\xi}$, proving the stated floors. □

**Corollary H.13** (Exact KKT conversion under Uniform Band-MFCQ)**.** *Under Equation 28, the ordinary polynomial schedule Equation 10, and Assumption 4.2, for each fixed $p \in (0,1)$ and $\xi \in (0,1]$, there is a finite, instance-dependent threshold $K_{\mathrm{cq}}(p,\xi) \ge 3$, independent of $K$, such that for $K \ge K_{\mathrm{cq}}(p,\xi)$,*

$$\min_{j=3,\dots,K+1} \mathit{Crit}^j_{\mathit{KKT}} \le \frac{2(\gamma_{\mathrm{cq}} + M_F)}{\gamma_{\mathrm{cq}}} F_{\mathrm{bFJ}}(K). \tag{131}$$

*In particular, for fixed $\xi$ and $p = 1/2$, the same range $j \in \{3, \dots, K+1\}$ gives $\mathcal{O}(K^{-1/4})$ for every $K \ge K_{\mathrm{cq}}(1/2, \xi)$.*

*Proof.* Pointwise, the definitions and $\bar{\iota}^j = \alpha^j \iota^j$, $\alpha^j \in (0,1]$, and $\omega^j \ge 0$ give the explicit chain

$$\begin{aligned} \mathrm{Crit}^j_{\mathrm{KKT}} &= \|Ax^j - \breve{y}^j\| + \|\Xi^j\| + \sum_{i \in \mathcal{I}_c} |\iota^j_i f_i(y^j_i)| \\ &\le \frac{1}{\alpha^j}\left[\|Ax^j - \breve{y}^j\| + \alpha^j\|\Xi^j\| + \sum_{i \in \mathcal{I}_c} |\bar{\iota}^j_i f_i(y^j_i)| + \omega^j\right] = \frac{\mathrm{Crit}^j_{\text{b-FJ}}}{\alpha^j}. \end{aligned} \tag{132}$$

For fixed $p \in (0,1)$ and $\xi \in (0,1]$, $F_{\mathrm{bFJ}}(K) \to 0$. Choose the stated $K_{\mathrm{cq}} \ge 3$ so that $F_{\mathrm{bFJ}}(K) \le \min\{\delta_{\mathrm{cq}}, \gamma_{\mathrm{cq}}/2\}$ for every $K \ge K_{\mathrm{cq}}(p,\xi)$. Let $j_* \in \{3, \dots, K+1\}$ be the common b-FJ minimizer from Theorem H.12. The two threshold clauses permit Lemma H.11 to be applied at $j_*$ and give $1/\alpha^{j_*} \le 2(\gamma_{\mathrm{cq}} + M_F)/\gamma_{\mathrm{cq}}$. Therefore

$$\min_{j=3,\dots,K+1} \mathrm{Crit}^j_{\mathrm{KKT}} \le \mathrm{Crit}^{j_*}_{\mathrm{KKT}} \le \frac{\mathrm{Crit}^{j_*}_{\text{b-FJ}}}{\alpha^{j_*}} \le \frac{2(\gamma_{\mathrm{cq}} + M_F)}{\gamma_{\mathrm{cq}}} F_{\mathrm{bFJ}}(K),$$

which proves Equation 131; the $p = 1/2$ rate follows from Equation 129, with the $K$-independent factor $2(\gamma_{\mathrm{cq}} + M_F)/\gamma_{\mathrm{cq}}$ absorbed into the $\mathcal{O}(\cdot)$ constant. □

**Same-$x$ KKT residual in the smooth case.** The residuals of Definition H.1 are nearby-point certificates: the objective forces $v_i$ use the proximal points $\breve{y}_i = p_i$ and the constraint gradients use the split points $y_i$, both of which approach $A_i x$ only as the consensus term $\|Ax - \breve{y}\|$ vanishes. When $\psi_i \equiv 0$, the Moreau force vanishes ($w_i = 0$), so $p_i = y_i$ and $\breve{y}_i = y_i$ for every block, and each $\varphi_i = h_i$ is smooth. Replacing every $y_i$ by $A_i x$ then changes the stationarity vector by at most $\big(\|A\|_2 \sum_{i \in \mathcal{I}_o} H_i + \|\iota\|_1 \|A\|_2 \max_{i \in \mathcal{I}_c} L_i\big)\|Ax - y\|$, the feasibility slack by at most $\max_{i \in \mathcal{I}_c} M_i \|Ax - y\|$, and each complementarity term $|\iota_i f_i(A_i x)|$ by at most $|\iota_i| M_i \|Ax - y\|$, using the Lipschitz gradients $H_i, L_i$ and the bounded constraint gradients $M_i$. On the certified iterate, Assumption 4.2 and Lemma H.11 bound $\|\iota\|_1 \le 2(\gamma_{\mathrm{cq}} + M_F)/\gamma_{\mathrm{cq}}$, while $\|Ax - y\| \le \mathrm{Crit}_{\text{w-KKT}} \le \mathrm{Crit}_{\mathrm{KKT}}$. Hence the conventional same-$x$ KKT residual, with all objective and constraint terms evaluated at $A_i x$, is bounded by a fixed multiple of $\mathrm{Crit}_{\mathrm{KKT}}$ and obeys the same $\mathcal{O}(K^{-1/4})$ rate as Corollary H.13. This is the standard transfer from a near-point to a same-point stationarity certificate Rockafellar & Wets (2009).

For the optional family analysis set

$$W_p(\xi) := \max\{1, MH_p(\xi)\}. \tag{133}$$

**Lemma H.14** (Aggregate block scaling)**.** *Under Assumption 4.1, uniformly for $\xi \in (0,1]$,*

$$\begin{aligned} &S_0 = \Theta(N), \quad \widehat{G}_0 = \Theta(N), \quad G_\psi = \mathcal{O}(N), \quad \bar{M} = \Theta(M), \\ &H_{\max}, L_{\max} = \mathcal{O}(1), \quad M_{\max}, m = \Theta(1), \quad L^0, G = \Theta(N), \\ &\rho^0 = \Theta(NW_p(\xi)), \qquad \Sigma_U = \mathcal{O}(N). \end{aligned} \tag{134}$$

*Moreover $H_p(\xi) = \mathcal{O}(1)$ uniformly on $(0,1]$. For fixed $\xi \in (0,1]$, $W_p(\xi) = \Theta(M)$ and $\rho^0 = \Theta(NM)$.*

*Proof.* The per-block and clearance upper bounds in Assumption 4.1 give $S_0, \widehat{G}_0, G_\psi = \mathcal{O}(N)$, $\bar{M} = \mathcal{O}(M)$, and $H_{\max}, L_{\max}, M_{\max}, m = \mathcal{O}(1)$. The clause $\varphi_i(A_i x^1) \geq c_\varphi > 0$ gives $S_0 = \Omega(N)$; the positive lower bounds on $\widehat{C}_i$ and $M_i$ give $\widehat{G}_0 = \Omega(N)$ and $\bar{M} = \Omega(M)$; and those on $M_i$ and $m$ give $M_{\max}, m = \Omega(1)$. Hence $S_0, \widehat{G}_0 = \Theta(N)$, $\bar{M} = \Theta(M)$, and $M_{\max}, m = \Theta(1)$. Therefore Equation 28 gives $L^0, G = \Theta(N)$.

First, Equation 60 implies the uniform bound

$$0 \leq H_p(\xi) \leq \sum_{j=1}^{\infty} j^{-2},$$

because $(1+\xi j^p)^2 \geq \xi^2 j^{2p}$. In the maximum defining $\rho^0$, the term $1209 L^0 M_{\max}^2$ is $\Theta(N)$, while the final term is

$$\mathcal{O}\big(H_p(\xi)[1+NM]\big).$$

The remaining entries in that maximum, namely 1, $1209 H_{\max}$, $1209 m L^0 L_{\max}$, and $6(\eta/2 + 11/2)\widehat{G}_0/G$, are each $\mathcal{O}(N)$. Conversely, the $L^0 \bar{M}$ part of the final entry is $\Theta(NMH_p(\xi))$. Since $N, M \geq 1$, taking the maximum proves $\rho^0 = \Theta(N \max\{1, MH_p(\xi)\})$, with constants allowed to depend on the fixed $p$.

Finally, the definition Equation 72 gives

$$\Sigma_U = \mathcal{O}\left(\frac{H_p(\xi)}{NW_p(\xi)}[N + N^2 M]\right) = \mathcal{O}(N),$$

using $MH_p(\xi)/W_p(\xi) \leq 1$ and the uniform bound on $H_p$. For fixed $\xi > 0$, the positive series in Equation 60 is a fixed constant, so $W_p(\xi) = \Theta(M)$ and hence $\rho^0 = \Theta(NW_p(\xi)) = \Theta(NM)$. □

**Lemma H.15** (Matrix scaling and automatic tuned saturation)**.** *Under Assumption 4.1,*

$$\|A_i\|_2 = \Theta(1) \quad (i \in \mathcal{I}), \qquad \|A\|_F^2 = \Theta(N+M),$$

*and*

$$\|A\|_2^2 = \Omega\left(\frac{N+M}{N}\right), \qquad \|A\|_2^2 = \mathcal{O}(N+M).$$

*For $p = 1/2$, define*

$$\xi_\star := \frac{m}{\|A\|_2 N}.$$

*Then $\xi_\star < 1$ for all sufficiently large family members. Whenever $\xi_\star < 1$,*

$$MH_{1/2}(\xi_\star) = \mathcal{O}\left(\frac{\log(eN)}{N}\right), \qquad W_{1/2}(\xi_\star) = \Theta(1),$$

*uniformly over the family. In fact, $W_{1/2}(\xi_\star) = 1$ for all sufficiently large family members.*

*Proof.* Let $s_A \in \mathbb{N}$ and $0 < c_A \leq C_A < \infty$ be the family-uniform constants from Assumption 4.1. Norm equivalence on an $s_A$-sparse vector gives

$$\frac{c_A}{\sqrt{s_A}} \leq \|A_i\|_F \leq C_A \qquad (i \in \mathcal{I}).$$

The lower bound $c_A > 0$ excludes $A_i = 0$, so $\operatorname{rank}(A_i) \geq 1$. Moreover, $\operatorname{rank}(A_i) \leq s_A$, and therefore

$$\frac{c_A}{s_A} \leq \|A_i\|_2 \leq C_A.$$

Since $A$ vertically stacks the $N + M$ blocks,

$$\|A\|_F^2 = \sum_{i \in \mathcal{I}} \|A_i\|_F^2 = \Theta(N + M).$$

Using $\operatorname{rank}(A) \leq \dim(x) = \mathcal{O}(N)$ yields

$$\|A\|_2^2 \geq \frac{\|A\|_F^2}{\operatorname{rank}(A)} = \Omega\left(\frac{N + M}{N}\right),$$

while $\|A\|_2^2 \leq \|A\|_F^2 = \mathcal{O}(N + M)$ proves the upper bound. In particular, $\|A\|_2 = \Omega(1)$; since $m = \Theta(1)$ by Lemma H.14, this also gives $\xi_\star = \mathcal{O}(N^{-1}) < 1$ for all sufficiently large family members.

By Lemma G.10, there are $\xi_0 \in (0, 1)$ and $C_0 > 0$ such that

$$H_{1/2}(\xi) \leq C_0 \xi^2 \log \frac{e}{\xi} \qquad (0 < \xi \leq \xi_0).$$

For $\xi \in [\xi_0, 1]$, the uniform bound in Lemma H.14 and the positive minimum of $\xi^2 \log(e/\xi)$ on this compact interval give the same inequality after increasing $C_0$. Hence it holds for every $\xi \in (0, 1]$. On the interior branch, substituting $\xi_\star = m/(\|A\|_2 N)$ gives

$$M H_{1/2}(\xi_\star) = \mathcal{O}\left(\frac{M m^2}{\|A\|_2^2 N^2} \log \frac{e \|A\|_2 N}{m}\right).$$

The upper bound $\|A\|_2^2 = \mathcal{O}(N + M) = \mathcal{O}(N^2)$, together with $m = \Theta(1)$, gives

$$\log \frac{e \|A\|_2 N}{m} = \mathcal{O}(\log(eN)).$$

Combining this with $\|A\|_2^2 = \Omega((N + M)/N)$ yields

$$M H_{1/2}(\xi_\star) = \mathcal{O}\left(\frac{M}{N(N + M)} \log(eN)\right) = \mathcal{O}\left(\frac{\log(eN)}{N}\right).$$

The right side is uniformly bounded for $N \geq 1$ and tends to zero. Since $W_{1/2}(\xi) = \max\{1, M H_{1/2}(\xi)\}$ by Equation 133, this proves the uniform relation $W_{1/2}(\xi_\star) = \Theta(1)$ on the interior branch and the eventual equality $W_{1/2}(\xi_\star) = 1$. □

**Lemma H.16** (Block scaling of exact rate constants)**.** *Under Assumption 4.1, uniformly for* $\xi \in (0, 1]$,

$$\begin{aligned} &|\Psi^2|, |\underline{\Psi}|, \mathcal{B} = \mathcal{O}(N), \qquad J_H, C_\psi = \mathcal{O}(\sqrt{N}), \\ &Q_\psi = \mathcal{O}(\|A\|_2 \sqrt{N}), \qquad r_\star = \mathcal{O}(W_p(\xi)^{-1}), \\ &c_y = \Theta(1), \qquad \mathcal{Q} = \Theta(\|A\|_2^2). \end{aligned} \tag{135}$$

*Here* $\Psi^2$ *denotes the iteration-two Lyapunov value, not a square.*

*Proof.* The left side of Equation 98 equals the initialization upper bound in Equation 90 plus two nonnegative extra terms. Thus Equation 98 and Lemma H.4 imply $\Psi^2 \leq G/2 = \mathcal{O}(N)$. The exact formula in Lemma H.7 and Lemma H.14 give $|\underline{\Psi}| = \mathcal{O}(N)$. Since $\Psi^2 \geq \underline{\Psi}$, also $|\Psi^2| = \mathcal{O}(N)$. Substitution in Equation 108 from Corollary H.8 proves $\mathcal{B} = \mathcal{O}(N)$.

The per-block upper bounds give $J_H, C_\psi = \mathcal{O}(\sqrt{N})$. The lower bound $\|A\|_2 = \Omega(1)$ from Lemma H.15 is used here to write $C_\psi = \mathcal{O}(\|A\|_2 \sqrt{N})$, while $\|A\|_2 J_H = \mathcal{O}(\|A\|_2 \sqrt{N})$ follows directly. Thus the definition in Theorem H.12 gives $Q_\psi = \mathcal{O}(\|A\|_2 \sqrt{N})$ without requiring a lower bound on $J_H$.

Moreover,

$$\frac{H_{\max}}{\rho^0} = \mathcal{O}(W_p(\xi)^{-1}), \qquad \frac{\sqrt{2GL^0}L_{\max}}{\rho^0} = \mathcal{O}(W_p(\xi)^{-1}),$$

so $r_* = \mathcal{O}(W_p(\xi)^{-1})$. Since $W_p \geq 1$, the positive floor in $c_y = 1009/120900 + (242909/121000)r_*$ gives $c_y = \Theta(1)$. The exact definition of $\mathcal{Q}$ in Theorem H.12 then gives $\mathcal{Q} = \Theta(\|A\|_2^2)$: its lower bound is supplied by the summand $26^2\|A\|_2^2/d_0$, while the matching upper bound uses $c_y = \Theta(1)$. This comparison does not require a lower bound on $\|A\|_2$ because both sides retain the same factor $\|A\|_2^2$. □

Define the two block-explicit upper-bound shapes

$$\begin{aligned}\mathfrak{F}_{\mathrm{W}}(K) &:= \|A\|_2 N\sqrt{W_p(\xi)}\left[K^{-1/2} + \sqrt{\xi}K^{(p-1)/2}\right] + \frac{\|A\|_2}{\sqrt{N}W_p(\xi)\xi}K^{-p},\\ \mathfrak{F}_{\mathrm{bFJ}}(K) &:= \mathfrak{F}_{\mathrm{W}}(K) + m\xi^{-1/2}K^{-p/2} + \frac{1}{W_p(\xi)K}.\end{aligned} \tag{136}$$

**Corollary H.17** (Block-explicit finite-time rates)**.** *Under Assumption 4.1, there is a constant $C_{\mathrm{bs}} > 0$, possibly depending on the fixed $p$ and the family-uniform constants in Assumption 4.1 but independent of $N, M, \xi, K$, and $\|A\|_2$, such that every $K \geq 3$ satisfies*

$$\begin{aligned}\min_{j=3,\dots,K+1} \mathit{Crit}^j_{\text{w-KKT}} &\leq C_{\mathrm{bs}}\mathfrak{F}_{\mathrm{W}}(K),\\ \min_{j=3,\dots,K+1} \mathit{Crit}^j_{\text{b-FJ}} &\leq C_{\mathrm{bs}}\mathfrak{F}_{\mathrm{bFJ}}(K).\end{aligned} \tag{137}$$

*The second minimum retains the common-iterate certificate of Theorem H.12. For fixed $\xi \in (0, 1]$,*

$$\begin{aligned}p = \frac{1}{3}: &\quad \min_{j=3,\dots,K+1} \mathit{Crit}^j_{\text{w-KKT}} = \mathcal{O}(\|A\|_2 N\sqrt{M}K^{-1/3}),\\ p = \frac{1}{2}: &\quad \min_{j=3,\dots,K+1} \mathit{Crit}^j_{\text{b-FJ}} = \mathcal{O}(\|A\|_2 N\sqrt{M}K^{-1/4}).\end{aligned} \tag{138}$$

*For fixed $\xi$, Lemma H.14 gives $W_p(\xi) = \Theta(M)$. At $p = 1/3$, the movement and reconstruction contributions both scale as $K^{-1/3}$, while the reconstruction coefficient is dominated by $\|A\|_2 N\sqrt{M}$. At $p = 1/2$, the movement and activation-band contributions determine the displayed $K^{-1/4}$ b-FJ rate, and all remaining terms are no larger. Iteration counts for tolerance $\epsilon > 0$ are*

$$\begin{aligned}K_{\mathrm{W}}(\epsilon) &= \mathcal{O}(\|A\|_2^3 N^3 M^{3/2}\epsilon^{-3}),\\ K_{\mathrm{bFJ}}(\epsilon) &= \mathcal{O}(\|A\|_2^4 N^4 M^2\epsilon^{-4}).\end{aligned} \tag{139}$$

*These are certified upper bounds, not matching lower bounds or tight complexity claims.*

*Proof.* For $K \geq 3$, the schedules give

$$R_K = \mathcal{O}(\rho^0[K + \xi K^{p+1}]), \quad I_K = \mathcal{O}((\rho^0\xi)^{-1}K^{1-p}),$$

and $B_K = \mathcal{O}(m\xi^{-1/2}K^{1-p/2})$. By the definition in Theorem H.12, $\mathcal{Q} \geq 26^2\|A\|_2^2/d_0$; the lower bound $\|A\|_2 = \Omega(1)$ from Lemma H.15 therefore makes $\mathcal{Q}$ family-uniformly positive. Since $\rho^k \geq \rho^0 \geq 1$, each $(\rho^k)^{-1} \leq \rho^k$, so $I_K \leq R_K$ and hence $I_K + \mathcal{Q}R_K = \mathcal{O}(\mathcal{Q}R_K)$. Together with $\mathcal{B} = \mathcal{O}(N)$ and $\mathcal{Q} = \Theta(\|A\|_2^2)$ from Lemma H.16, the movement part of Equation 128 is at most

$$\mathcal{O}\left(\|A\|_2 N\sqrt{W_p(\xi)}[K^{-1/2} + \sqrt{\xi}K^{(p-1)/2}]\right).$$

The sharper inverse-schedule estimate, together with $Q_\psi = \mathcal{O}(\|A\|_2\sqrt{N})$ and $\rho^0 = \Theta(NW_p(\xi))$, bounds the reconstruction part by

$$\mathcal{O}\left(\frac{\|A\|_2}{\sqrt{N}W_p(\xi)\xi}K^{-p}\right).$$

The band sum contributes $\mathcal{O}(m\xi^{-1/2}K^{-p/2})$.

For the final term, the superscript in $\rho^2$ is an iteration index:

$$\rho^2 = \rho^0(1 + \xi 2^p) = \Theta(\rho^0) = \Theta(NW_p(\xi)).$$

It is not $(\rho^0)^2$. Thus $L_{\max}\mathcal{B}/[d_0\rho^2(K-1)] = \mathcal{O}([W_p(\xi)K]^{-1})$. Substitution in Equation 128 proves Equation 137, including the common-iterate certificate.

For fixed $\xi$, Lemma H.14 gives $W_p(\xi) = \Theta(M)$. At $p = 1/3$, the movement terms have exponents $K^{-1/2}$ and $K^{-1/3}$, and the reconstruction term has exponent $K^{-1/3}$; moreover, since $N, M \geq 1$, $\|A\|_2/(\sqrt{N}M) \leq \|A\|_2 N\sqrt{M}$. Thus all three are bounded by the displayed w-KKT monomial $\|A\|_2 N\sqrt{M}K^{-1/3}$. At $p = 1/2$, the movement exponents are $K^{-1/2}$ and $K^{-1/4}$, reconstruction is $K^{-1/2}$, the activation band is $K^{-1/4}$, and widened support is $K^{-1}$; the family upper bound $m = \mathcal{O}(1)$ and the lower bound $\|A\|_2 = \Omega(1)$ from Lemma H.15, together with $N, M \geq 1$, let the displayed coefficient $\|A\|_2 N\sqrt{M}$ absorb the remaining coefficients. This proves Equation 138; solving the two certified leading inequalities for $K$ gives Equation 139. □

**Corollary H.18** (Tuned block rate on the interior branch)**.** *Suppose Assumption 4.1 holds with $p = 1/2$ and the nominal value in Equation 13 satisfies $\xi_{\mathrm{nom}} < 1$. Write $\xi_\star := \xi_{\mathrm{nom}}$ and choose $\xi = \xi_\star$. Then, for every $K \geq 3$,*

$$\min_{j=3,\dots,K+1} \mathit{Crit}^j_{b\text{-}FJ} = \mathcal{O}\left(\left[\|A\|_2 N + \frac{\|A\|_2^2\sqrt{N}}{m}\right]K^{-1/2} + \sqrt{\|A\|_2 Nm}\,K^{-1/4} + K^{-1}\right). \tag{140}$$

*The minimum is attained at the same selected certificate-bearing iterate as in Theorem H.12. The two $K^{-1/2}$ contributions are the baseline movement term and the inverse-schedule reconstruction term after substituting $\xi_\star$. Balancing the movement and activation-band coefficients produces the $\sqrt{\|A\|_2 Nm}\,K^{-1/4}$ term, while the widened-support correction contributes $K^{-1}$. This shape applies on the interior tuned branch; when the nominal choice clips to $\xi = 1$, the fixed-$\xi$ rate in Equation 138 applies instead. A sufficient iteration count is*

$$K_{\mathrm{bFJ}}(\epsilon) = \mathcal{O}\left(\|A\|_2^2 N^2 m^2 \epsilon^{-4} + \left[\|A\|_2^2 N^2 + \frac{\|A\|_2^4 N}{m^2}\right]\epsilon^{-2} + \epsilon^{-1}\right). \tag{141}$$

*If*

$$\epsilon \leq c \min\left\{m, \frac{m^2\sqrt{N}}{\|A\|_2}, (\|A\|_2^2 N^2 m^2)^{1/3}\right\} \tag{142}$$

*for a sufficiently small $c > 0$ depending only on the fixed numerical constants and the family-uniform upper and positive lower bounds in Assumption 4.1, but not on $N, M, K, \|A\|_2$, or $\epsilon$, the first term in Equation 141 suffices. If the nominal choice clips to $\xi = 1$, then $W_{1/2}(1) = \Theta(M)$ and the fixed-$\xi$ bound Equation 138, not the interior formula above, applies.*

*Proof.* Lemma H.15 gives $W_{1/2}(\xi_\star) = \Theta(1)$. The two $K^{-1/4}$ coefficients in Equation 136 are then $\|A\|_2 N\sqrt{\xi_\star}$ and $m\xi_\star^{-1/2}$, which are balanced by the exact identity $\xi_\star = m/(\|A\|_2 N)$ in Equation 13. Substitution in Corollary H.17 gives Equation 140; in particular the widened-support term is $\mathcal{O}(K^{-1})$.

Requiring each of the four monomials in Equation 140 to be at most a constant multiple of $\epsilon$ gives Equation 141. Indeed, comparing $\|A\|_2^2 N^2 m^2 \epsilon^{-4}$ with $\|A\|_2^2 N^2 \epsilon^{-2}$ gives $\epsilon \leq c_1 m$; comparison with $\|A\|_2^4 N m^{-2}\epsilon^{-2}$ gives $\epsilon \leq c_2 m^2\sqrt{N}/\|A\|_2$; and comparison with $\epsilon^{-1}$ gives $\epsilon \leq c_3(\|A\|_2^2 N^2 m^2)^{1/3}$. Choosing $c \leq \min\{c_1, c_2, c_3\}$ with the stated family dependence proves Equation 142. If the nominal balance clips to $\xi = 1$, then $H_{1/2}(1)$ is a fixed positive constant and $W_{1/2}(1) = \Theta(M)$, so the fixed-$\xi$ conclusion, rather than the interior tuned substitution, applies. □

**Corollary H.19** (Block scaling under Uniform Band-MFCQ)**.** *Suppose Equation 28, the ordinary polynomial schedule Equation 10, Assumption 4.1, and Assumption 4.2 hold, and set*

$$D_{\mathrm{cq}} := 2\left(1 + \frac{\|A\|_2\sqrt{\widehat{G}_0}}{\gamma_{\mathrm{cq}}}\right) = \Theta\left(1 + \frac{\|A\|_2\sqrt{N}}{\gamma_{\mathrm{cq}}}\right). \tag{143}$$

*For $K \geq K_{\mathrm{cq}}(p,\xi)$, with the instance-dependent threshold from Corollary H.13, the general family bound is*

$$\min_{j=3,\dots,K+1} \mathit{Crit}^{j}_{\mathit{KKT}} \leq C_{\mathrm{bs}} D_{\mathrm{cq}} \mathfrak{F}_{\mathrm{bFJ}}(K). \tag{144}$$

*In particular, for fixed $\xi$ and $p = 1/2$,*

$$\begin{aligned} \min_{j=3,\dots,K+1} \mathit{Crit}^{j}_{\mathit{KKT}} &= \mathcal{O}(D_{\mathrm{cq}} \|A\|_2 N \sqrt{M} K^{-1/4}), \\ K_{\mathrm{KKT}}(\epsilon) &= \mathcal{O}(D_{\mathrm{cq}}^4 \|A\|_2^4 N^4 M^2 \epsilon^{-4}), \end{aligned} \tag{145}$$

*where the separate requirement $K \geq K_{\mathrm{cq}}(1/2,\xi)$ remains in force, and a sufficient iteration count means the maximum of this threshold and the displayed rate-derived count. On the interior tuned branch of Corollary H.18, a sufficient count is*

$$K_{\mathrm{KKT}}(\epsilon) = \mathcal{O}\left( D_{\mathrm{cq}}^4 \|A\|_2^2 N^2 m^2 \epsilon^{-4} + D_{\mathrm{cq}}^2 \left[ \|A\|_2^2 N^2 + \frac{\|A\|_2^4 N}{m^2} \right] \epsilon^{-2} + D_{\mathrm{cq}} \epsilon^{-1} \right), \tag{146}$$

*again together with $K \geq K_{\mathrm{cq}}(1/2,\xi_\star)$, so the actual sufficient count is the maximum of the threshold and the displayed expression.*

*If, additionally, $\gamma_{\mathrm{cq}} = \Theta(1)$ across the family, then Lemma H.15 supplies a uniform positive lower bound on $\|A\|_2$, so the additive one in $D_{\mathrm{cq}}$ can be absorbed in this specialized regime. In that specialized regime, the fixed-$\xi$ sufficient count is*

$$K_{\mathrm{KKT}}(\epsilon) = \mathcal{O}\left( \frac{\|A\|_2^8 N^6 M^2}{\gamma_{\mathrm{cq}}^4} \epsilon^{-4} \right), \tag{147}$$

*and Equation 146 becomes*

$$K_{\mathrm{KKT}}(\epsilon) = \mathcal{O}\left( \frac{\|A\|_2^6 N^4 m^2}{\gamma_{\mathrm{cq}}^4} \epsilon^{-4} + \left[ \frac{\|A\|_2^4 N^3}{\gamma_{\mathrm{cq}}^2} + \frac{\|A\|_2^6 N^2}{m^2 \gamma_{\mathrm{cq}}^2} \right] \epsilon^{-2} + \frac{\|A\|_2 \sqrt{N}}{\gamma_{\mathrm{cq}}} \epsilon^{-1} \right). \tag{148}$$

*The fixed display Equation 147 retains $K \geq K_{\mathrm{cq}}(1/2,\xi)$, while the tuned display Equation 148 requires $K \geq K_{\mathrm{cq}}(1/2,\xi_\star)$. All iteration counts above are sufficient upper bounds, not tightness claims.*

*Proof.* By Equation 125 and Lemma H.14, the exact factor in Equation 131 is

$$\frac{2(\gamma_{\mathrm{cq}} + M_F)}{\gamma_{\mathrm{cq}}} = 2\left( 1 + \frac{\|A\|_2 \sqrt{\widehat{G}_0}}{\gamma_{\mathrm{cq}}} \right) = \Theta\left( 1 + \frac{\|A\|_2 \sqrt{N}}{\gamma_{\mathrm{cq}}} \right).$$

The additive one is retained in the general family bound. For the common b-FJ minimizer $j_\star$ from Theorem H.12, apply the pointwise conversion Equation 132 and the multiplier factor above to Equation 137. This proves Equation 144 at the same selected certificate-bearing iterate, subject to the separate, instance-dependent threshold $K \geq K_{\mathrm{cq}}(p,\xi)$ from Corollary H.13.

The fixed-$\xi$ statement follows from Equation 138; inverting its $K^{-1/4}$ monomial gives Equation 145. On the interior tuned branch, multiply Equation 140 by $D_{\mathrm{cq}}$ and invert its four monomials. This gives Equation 146, including the linear $D_{\mathrm{cq}}\epsilon^{-1}$ contribution.

If $\gamma_{\mathrm{cq}} = \Theta(1)$ across the family, then the lower bound on $\|A\|_2$ from Lemma H.15 makes $\|A\|_2\sqrt{N}/\gamma_{\mathrm{cq}}$ uniformly positive, so the additive one may be absorbed under precisely this additional family condition. Substitution into Equation 145 and Equation 146 yields Equation 147 and Equation 148. □

**Corollary H.20** (Block scaling of the scaled multiplier)**.** *Under Equation 28, the ordinary polynomial schedule Equation 10, and Assumption 4.1, uniformly for $\xi \in (0,1]$, recall from Equation 133 that $W_p(\xi) = \max\{1, M H_p(\xi)\}$. The exact constant in Lemma H.9 satisfies*

$$C_\lambda = \mathcal{O}\left( \sqrt{N} + \sqrt{\frac{NM}{W_p(\xi)}} \right). \tag{149}$$

*Consequently, for every $k \geq 1$,*

$$\frac{\|\lambda^{k+1}\|}{\rho^k} = \mathcal{O}\left(\left[\frac{1}{\sqrt{W_p(\xi)}} + \frac{\sqrt{M}}{W_p(\xi)}\right](1+\xi k^p)^{-1/2}\right). \tag{150}$$

*For fixed $\xi \in (0,1]$, one has $W_p(\xi) = \Theta(M)$ and hence*

$$\frac{\|\lambda^{k+1}\|}{\rho^k} = \mathcal{O}\left(M^{-1/2}(1+\xi k^p)^{-1/2}\right). \tag{151}$$

*On the interior tuned branch of Corollary H.18, Lemma H.15 gives $W_{1/2}(\xi_\star) = \Theta(1)$ and*

$$\frac{\|\lambda^{k+1}\|}{\rho^k} = \mathcal{O}\left(\sqrt{M}\,(1+\xi_\star k^{1/2})^{-1/2}\right). \tag{152}$$

*Uniformly over the family, the coefficient in square brackets in Equation 150 is at most $\mathcal{O}(\sqrt{N+M})$; this is only a sufficient upper bound, not a claim of tightness. No explicit $\|A\|_2$ factor is introduced in these bounds because Equation 76 and the movement budget control $A\Delta x^k$ directly. This does not assert independence from $A$.*

*Proof.* The definition of $C_\lambda$ in Lemma H.9, together with Lemma H.14 and Lemma H.16, gives

$$\frac{\sqrt{\widehat{G}_0}}{\sqrt{\rho^0}} = \mathcal{O}(W_p(\xi)^{-1/2}),$$
$$\sqrt{\frac{2GL^0\bar{M}}{\rho^0}} = \mathcal{O}\left(\sqrt{\frac{NM}{W_p(\xi)}}\right),$$
$$(1+c_\upsilon)\sqrt{\mathcal{B}} = \mathcal{O}(\sqrt{N}).$$

Since $N \geq 1$ and $W_p(\xi) \geq 1$, the first of the three bounds above is $\mathcal{O}(\sqrt{N})$, proving Equation 149.

By Equation 111,

$$\frac{\|\lambda^{k+1}\|}{\rho^k} \leq \frac{C_\lambda}{\sqrt{\rho^k}}.$$

The schedule and Lemma H.14 give $\rho^k = \Theta(NW_p(\xi)(1+\xi k^p))$. Dividing the two terms in Equation 149 by this square root yields, respectively, $W_p(\xi)^{-1/2}$ and $\sqrt{M}/W_p(\xi)$, proving Equation 150.

For fixed $\xi$, Lemma H.14 gives $W_p(\xi) = \Theta(M)$, which proves Equation 151. On the interior tuned branch, Lemma H.15 gives $W_{1/2}(\xi_\star) = \Theta(1)$; substituting this relation and $\xi = \xi_\star$ into Equation 150 proves Equation 152. Finally, $W_p(\xi) \geq 1$ and $M \geq 1$ imply

$$\frac{1}{\sqrt{W_p(\xi)}} + \frac{\sqrt{M}}{W_p(\xi)} \leq 1 + \sqrt{M} \leq 2\sqrt{M} = \mathcal{O}(\sqrt{N+M}),$$

which is the stated sufficient family-wide coefficient bound. □

*Proof of Theorem 4.2.* The instance-wise asymptotic bounds are the corresponding specializations of Theorem H.12. The fixed-$\xi$ family statements follow from Corollary H.17, and the interior tuned statement follows from Corollary H.18 and Lemma H.15 with the exact rule Equation 13. These results give every assertion of Theorem 4.2. □

*Proof of Corollary 4.3.* Apply Corollary H.13 with $p = 1/2$. Its fixed-family specialization is Corollary H.19, which gives the stated $D_{\mathrm{cq}}$ dependence. □

# I Convergence Analysis of IP-ADMM-Ckpt

This appendix proves Theorem 4.4 for the Ckpt mode of Algorithm 1. Throughout this appendix, set $\mathsf{m} = \mathsf{Ckpt}$ and abbreviate $\theta_1 := \theta_{1,\mathsf{m}} = \theta_1^{\mathrm{cp}}$, $\theta_2 := \theta_{2,\mathsf{m}} = \theta_2^{\mathrm{cp}}$, $\eta := \eta^{\mathrm{cp}}$, $G := G_{\mathsf{m}} = G^{\mathrm{cp}}$, and $g := g_{\mathsf{m}}$. These are local aliases for the checkpoint quantities in Equation 29, not shared parameters with ordinary IP-ADMM. The mode uses the unit virtual-multiplier update shown in the algorithm, the checkpoint test $E^{k+1} \le \tau_{\mathrm{cp}}/\rho^k$, the explicit parameters in Equation 29, and the geometric schedule Equation 30. Its proof conditions each level on two caps, derives both finite arrival and accepted-step control under those caps, and then propagates the caps across checkpoint transitions. This produces pointwise certificates for checkpoints. The ordinary IP mode instead uses a global continuation argument and is analyzed in Appendix H.

**Levels, frozen state, and checkpoint outputs.** Level $j \ge 0$ begins after exactly $j$ checkpoints have been committed. While that level is active, the level-defining state $(\lambda^{[j]}, \rho_j, L_j, \mu_j)$ is fixed, but every new primal pair is retained. Put $k_{-1} := 0$. If level $j$ completes, let $k_j$ denote its accepting iteration; then level $j$ consists of $k \in \{k_{j-1}+1, \dots, k_j\}$ and

$$(\rho^k, L^k, \mu^k) = (\rho_j, L_j, \mu_j), \quad \lambda^k = \lambda^{[j]}, \quad T_j := k_j - k_{j-1}, \quad K_j := \sum_{\ell=0}^{j} T_\ell = k_j, \tag{153}$$

for every $k$ in that level. Here $\lambda^{[0]} := \lambda^1 = 0$. The finite-arrival result Lemma I.7 later proves that every level completes, so all of these indices are finite.

Three multiplier roles are distinct. The *stored* multiplier $\lambda^k = \lambda^{[j]}$ enters every subproblem in level $j$. Each iteration forms the *virtual* multiplier $\widehat{\lambda}^{k+1} = \lambda^k + \rho^k r^{k+1}$. Rejection discards that virtual multiplier, whereas acceptance promotes $\widehat{\lambda}^{k_j+1}$ to the next stored multiplier $\lambda^{[j+1]}$. The raw multiplier bounds below cover stored multipliers and accepted virtual multipliers, not rejected virtual multipliers.

At acceptance, the method first records the pre-promotion checkpoint certificate tuple

$$(x^{k_j+1}, y^{k_j+1}, \breve{y}^{k_j+1}, v^{k_j+1}, \iota^{k_j+1}, \zeta^{k_j+1}). \tag{154}$$

It is formed from $(\rho_j, L_j, \mu_j)$ and $\lambda^{[j]}$ by Equation 42 and the definitions preceding Equation 106, with $\zeta^{k_j+1} = \rho_j \varpi^{k_j+1}$. Recording before promotion prevents the tuple from mixing two levels. Before checkpoint 0 completes, no certified output exists. A rejection preserves the level-defining state but retains $(x^{k+1}, y^{k+1})$; an acceptance records the tuple and only then promotes the level-defining state. Thus the method follows one primal trajectory and never restarts.

**Frozen-level merit function.** The checkpoint proof measures primal progress during level $j$ with

$$\mathcal{P}_{\rho,L,\mu}(x,y;\lambda) := \mathcal{L}(x,y,\lambda,L,\mu,\rho) + R_\psi(\mu), \quad \mathcal{P}_j(x,y) := \mathcal{P}_{\rho_j,L_j,\mu_j}(x,y;\lambda^{[j]}), \tag{155}$$

where $\mathcal{L}$ is defined in Equation 7 and $R_\psi$ in Lemma G.5. The level-defining state is fixed, so every primal sweep is measured against the same function $\mathcal{P}_j$; no polynomial drift correction or stored-multiplier update appears within the level. The primals still change at every iteration. Only promotion changes the definition of the potential, and Lemma I.8 charges that one-time change at the retained primal pair.

**Derived constants.** The following constants package the support-width, certificate, merit-budget, and work estimates. They are determined by the explicit choice Equation 29 and introduce no additional assumptions:

$$\begin{aligned} S_\omega &:= b_{\mathrm{cp}} + \frac{L_{\max}\tau_{\mathrm{cp}}^2}{2(\rho_{\mathrm{cp}}^0)^3}, \quad C_{\mathrm{cert}} := 2Q_{\mathrm{cp}} + \eta C_\psi + S_\Xi + I_{\mathrm{cp}} S_\omega, \\ \Delta_* &:= \overline{V} + \Delta_G, \quad C_{\mathrm{work}} := \frac{\Delta_*}{a_0 \tau_{\mathrm{cp}}^2}. \end{aligned} \tag{156}$$

**What is and is not reused.** The proof reuses the shared local calculus for Moreau envelopes, nonlinear walls, model error, block and consensus optimality, reconstruction, and original-problem certificates; it also uses the checkpoint-specific geometric identities in Lemma G.11. It does not import the ordinary mode's polynomial drift estimates, paired Lyapunov functions, global movement budget, or window-averaging argument. Frozen-level descent, a two-cap level induction, accepted-step estimates, finite-arrival accounting, and promotion charging replace that machinery. Consequently, ordinary IP-ADMM selects a best-window candidate, whereas IP-ADMM-Ckpt certifies the latest checkpoint pointwise. At an accepted pre-promotion tuple, Uniform Band-MFCQ first bounds the wall forces; the virtual-dual identity then bounds the accepted virtual multiplier, which becomes the next stored multiplier. No bound is claimed for rejected virtual multipliers.

**Organization.** The dependency chain is as follows. Section I.1 establishes frozen-level descent, virtual-multiplier and stationarity transfer identities, and wall support. Section I.2 supplies the conditional merit lower bound, initialization, and within-level split feasibility. Under the stored-multiplier and entry-merit caps, Section I.3 then develops two complementary branches: Lemma I.6 bounds accepted wall forces, the accepted virtual multiplier, and sharp consensus, while Lemma I.7 proves finite arrival with the per-level work bound. Lemma I.8 combines finite arrival, accepted-step control, and the summable promotion charge to propagate both caps, closing the induction. With those global invariants in place, Section I.4 proves the pointwise checkpoint certificate and inverts cumulative work to obtain the completed-checkpoint and anytime rates. The block-scaling and stored/accepted-multiplier results are final consequences.

### I.1 Frozen-Level Descent and Certificate-Tuple Identities

Throughout this subsection $k$ is an iteration of level $j$, so that $(\rho^k, L^k, \mu^k) = (\rho_j, L_j, \mu_j)$ and $\lambda^k = \lambda^{[j]}$. As in Section H.1, $\Delta x^k = x^{k+1} - x^k$, $d_i^k = y_i^{k+1} - y_i^k$, $r^{k+1} = y^{k+1} - Ax^{k+1}$, and $v_i^k = m_i^k - \rho^k$, which at level $j$ reads $v_i^k = m_i^k - \rho_j = (\theta_1 - 1)\rho_j + \theta_1 \tau_i^k$.

**Lemma I.1** (Frozen-level one-step descent)**.** *Let $k$ be any iteration of level $j$. Then*

$$\begin{aligned}\mathcal{P}_j(x^k, y^k) - \mathcal{P}_j(x^{k+1}, y^{k+1}) &\geq \frac{2\theta_1 - 1}{2}\rho_j \|d^k\|^2 + \frac{1 + 2\theta_2}{2}\rho_j \|A\Delta x^k\|^2 \\ &\geq a_0 \rho_j \left(\|d^k\|^2 + \|A\Delta x^k\|^2\right),\end{aligned} \tag{157}$$

*with $a_0 = (2\theta_1 - 1)/2 = (1 + 2\theta_2)/2 = 51/100 > 0$ from Equation 29. In particular $\mathcal{P}_j$ is nonincreasing along level $j$, and no wall interiority, no multiplier update, and no parameter change is used: the estimate holds whether or not $f_i(y_i^k) > 0$ for $i \in \mathcal{I}_c$. The affine feasibility $Cx^k = Cx^{k+1} = d$ is used, in the $x$ step only.*

*Proof.* Since $\mu^k = \mu_j$ is frozen, $R_\psi(\mu_j)$ is the same constant at $(x^k, y^k)$ and at $(x^{k+1}, y^{k+1})$; it therefore cancels, and it suffices to bound the decrease of $\mathcal{L}(\cdot, \cdot, \lambda^{[j]}, L_j, \mu_j, \rho_j)$. Because the stored multiplier and the penalty are frozen, the same function is decreased by both sweeps.

*Block sweep.* For fixed $x^k$ the Lagrangian is separable: $\mathcal{L}(x^k, y, \lambda^{[j]}, L_j, \mu_j, \rho_j) = \sum_{i \in \mathcal{I}} \Phi_i(y_i)$ with

$$\Phi_i(u) := \bar{f}_i(u, L_j, \mu_j) + \langle \lambda_i^{[j]}, u - A_i x^k \rangle + \frac{\rho_j}{2}\|u - A_i x^k\|^2.$$

Expanding the quadratic about $y_i^k$ and using $e_i^k = \lambda_i^{[j]} + \rho_j(y_i^k - A_i x^k)$ gives the exact identity

$$\Phi_i(u) - \Phi_i(y_i^k) = \bar{f}_i(u, L_j, \mu_j) - \bar{f}_i(y_i^k, L_j, \mu_j) + \langle e_i^k, u - y_i^k \rangle + \frac{\rho_j}{2}\|u - y_i^k\|^2.$$

Let $\widetilde{f}_i^k$ be the model-only function of Equation 56; by construction $\widetilde{f}_i^k(y_i^k) = \bar{f}_i(y_i^k, L_j, \mu_j)$, and $\widetilde{f}_i^k(u) + \langle e_i^k, u - y_i^k \rangle + \frac{m_i^k}{2}\|u - y_i^k\|^2$ equals $\mathcal{M}_i^k(u)$ up to an additive constant independent of $u$. Applying Equation 56 at $u = y_i^{k+1}$,

$$\Phi_i(y_i^{k+1}) - \Phi_i(y_i^k) \leq \left[\mathcal{M}_i^k(y_i^{k+1}) - \mathcal{M}_i^k(y_i^k)\right] - \frac{m_i^k}{2}\|d_i^k\|^2 + \frac{\tau_i^k + \rho_j}{2}\|d_i^k\|^2.$$

The model $\mathcal{M}_i^k$ is $m_i^k$-strongly convex by Assumption 3.4 on objective blocks and by the convexity of $g(\cdot, L_j)$ in Lemma G.6 on constraint blocks. The block update in Algorithm 1 minimizes this model at $y_i^{k+1}$. Hence its bracket is at most $-\frac{m_i^k}{2}\|d_i^k\|^2$. Since $\tau_i^k + \rho_j = \mathsf{B}_i^k$ and $m_i^k = \theta_1 \mathsf{B}_i^k$,

$$\Phi_i(y_i^{k+1}) - \Phi_i(y_i^k) \le -\left(m_i^k - \frac{\mathsf{B}_i^k}{2}\right)\|d_i^k\|^2 = -\frac{2\theta_1 - 1}{2}\mathsf{B}_i^k\|d_i^k\|^2 \le -\frac{2\theta_1 - 1}{2}\rho_j\|d_i^k\|^2,$$

using $\mathsf{B}_i^k = \rho_j + \tau_i^k \ge \rho_j$, since $\tau_i^k \ge 0$, and $2\theta_1 - 1 > 0$. Summing over $i \in \mathcal{I}$ gives the first displayed term.

*$x$ step.* Write $\phi(x) := \mathcal{L}(x, y^{k+1}, \lambda^{[j]}, L_j, \mu_j, \rho_j) + \frac{\theta_2\rho_j}{2}\|A(x - x^k)\|^2$, the exact objective of the $x$ subproblem. It is a convex quadratic in $x$ with Hessian $(1+\theta_2)\rho_j A^T A$, so its second-order Taylor expansion about $x^{k+1}$ is an exact identity at every point; evaluated at $x^k$ it reads

$$\phi(x^k) = \phi(x^{k+1}) + \langle \nabla\phi(x^{k+1}), x^k - x^{k+1}\rangle + \frac{(1+\theta_2)\rho_j}{2}\|A\Delta x^k\|^2.$$

By Equation 77 the gradient $\nabla\phi(x^{k+1})$ lies in the range of $C^T$, while $C\Delta x^k = 0$ because $Cx^k = d$ by induction from Assumption 3.3 and $Cx^{k+1} = d$ by Equation 77; hence the inner product vanishes and $\phi(x^k) - \phi(x^{k+1}) = \frac{(1+\theta_2)\rho_j}{2}\|A\Delta x^k\|^2$. The proximal term of $\phi$ vanishes at $x^k$ and equals $\frac{\theta_2\rho_j}{2}\|A\Delta x^k\|^2$ at $x^{k+1}$, so the Lagrangian decrease exceeds the decrease of $\phi$ by exactly that amount; adding it gives

$$\mathcal{L}(x^k, y^{k+1}, \lambda^{[j]}, L_j, \mu_j, \rho_j) - \mathcal{L}(x^{k+1}, y^{k+1}, \lambda^{[j]}, L_j, \mu_j, \rho_j) = \frac{1 + 2\theta_2}{2}\rho_j\|A\Delta x^k\|^2.$$

Adding the two sweeps and restoring the cancelled constant $R_\psi(\mu_j)$ proves the first inequality in Equation 157. The second follows from the definition of $a_0$ in Equation 29. □

**Lemma I.2** (Virtual multiplier and stationarity transfer)**.** *Let $k$ be any iteration of level $j$. Then the virtual multiplier obeys the exact identity*

$$\widehat{\lambda}^{k+1} = \lambda^{[j]} + \rho_j r^{k+1} = -q^k - \Upsilon^k d^k - \rho_j A\Delta x^k, \qquad \Upsilon^k := \operatorname{diag}(\upsilon_i^k I_{n_i})_{i\in\mathcal{I}}, \tag{158}$$

*and, with $r_{\mathrm{cp}}$ and $c_{\chi,\mathrm{cp}}$ from Equation 29,*

$$0 \le \frac{\tau_i^k}{\rho_j} \le r_{\mathrm{cp}}, \qquad 0 < \frac{\upsilon_i^k}{\rho_j} \le c_{\chi,\mathrm{cp}} \qquad (i \in \mathcal{I}). \tag{159}$$

*Let $\delta_i^k := v_i^{k+1} - q_i^k$ for $i \in \mathcal{I}_o$ and $\delta_i^k := \iota_i^{k+1}\nabla f_i(y_i^{k+1}) - q_i^k$ for $i \in \mathcal{I}_c$. This checkpoint $\delta_i^k$ is the single-index reconstruction gap and is not the* IP*-mode force-drift term $\delta_i^k$ of Equation 69, which measures the parameter change between indices $k-1$ and $k$ and is not used in this appendix. Let $\delta^k := (\delta_i^k)_{i\in\mathcal{I}}$ denote the stacked reconstruction gap. Then*

$$\|\delta^k\| \le r_{\mathrm{cp}}\rho_j\|d^k\| + \mu_j J_H, \tag{160}$$

*and the stationarity residual $\Xi^{k+1}$ of Equation 106 satisfies the exact identity and bound*

$$\Xi^{k+1} = A^T\left[\delta^k - \Upsilon^k d^k\right] - (1+\theta_2)\rho_j A^T A\Delta x^k, \tag{161}$$

$$\|\Xi^{k+1}\| \le \|A\|_2\left[(r_{\mathrm{cp}} + c_{\chi,\mathrm{cp}})\rho_j\|d^k\| + (1+\theta_2)\rho_j\|A\Delta x^k\| + \mu_j J_H\right]. \tag{162}$$

*If in addition the step-$k$ test is accepted, that is $E^{k+1} \le \tau_{\mathrm{cp}}/\rho_j$, then*

$$\|\Xi^{k+1}\| \le \frac{S_\Xi}{\rho_j}, \qquad S_\Xi = \|A\|_2(c_{\mathrm{mov}}\tau_{\mathrm{cp}} + \eta J_H). \tag{163}$$

*Proof. Identity Equation 158.* Substituting $y_i^k - A_i x^k = r_i^{k+1} - d_i^k + A_i\Delta x^k$ into Equation 75 with $\lambda_i^k = \lambda_i^{[j]}$ and $\rho^k = \rho_j$ gives

$$0 = q_i^k + \lambda_i^{[j]} + \rho_j r_i^{k+1} + (m_i^k - \rho_j)d_i^k + \rho_j A_i\Delta x^k,$$

which is exactly $\widehat{\lambda}_i^{k+1} = -q_i^k - \upsilon_i^k d_i^k - \rho_j A_i \Delta x^k$. Stacking over $i$ proves Equation 158. It is the frozen-stored-multiplier analogue of Equation 76; no stored-multiplier update is performed and the checkpoint mode has no $\sigma$ parameter.

*Bound Equation 159.* For $i \in \mathcal{I}_o$, $\tau_i^k = H_i \le H_{\max}$, and $\rho_j \ge \rho_{\mathrm{cp}}^0 \ge 1$ by Equation 30 and Equation 29, so $\tau_i^k/\rho_j \le H_{\max} \le r_{\mathrm{cp}}$. For $i \in \mathcal{I}_c$, $\tau_i^k = \sqrt{2L_jG}\,L_i$, and the second identity of Equation 66 gives $\tau_i^k/\rho_j = \sqrt{2G\ell_{\mathrm{cp}}}\,L_i \le \sqrt{2G\ell_{\mathrm{cp}}}\,L_{\max} \le r_{\mathrm{cp}}$. Nonnegativity is clear. Consequently

$$\frac{\upsilon_i^k}{\rho_j} = \frac{\theta_1(\rho_j + \tau_i^k) - \rho_j}{\rho_j} = \theta_1 - 1 + \theta_1 \frac{\tau_i^k}{\rho_j} \le \theta_1(1 + r_{\mathrm{cp}}) - 1 = c_{\chi,\mathrm{cp}},$$

and $\upsilon_i^k \ge (\theta_1 - 1)\rho_j > 0$ because $\theta_1 = \theta_1^{\mathrm{cp}} = 101/100$.

*Bound Equation 160.* For $i \in \mathcal{I}_o$, the third line of Equation 43 gives $\|\delta_i^k\| \le H_i\|d_i^k\| + \mu_j H_i C_i^\psi$. For $i \in \mathcal{I}_c$, the wall lemma Lemma G.6 instantiated with $G = G^{\mathrm{cp}}$ gives $\|\delta_i^k\| \le \sqrt{2L_jG}\,L_i\|d_i^k\|$; that estimate is a consequence of the uniform force bound $|\iota_i^{k+1}| \le \sqrt{2L_jG}$ stated in Lemma G.6 together with the $L_i$-Lipschitz gradient of Assumption 3.1, and uses no schedule property. Set $\kappa_i := H_i$ for $i \in \mathcal{I}_o$ and $\kappa_i := \sqrt{2L_jG}\,L_i$ for $i \in \mathcal{I}_c$; the previous paragraph gives $\kappa_i \le r_{\mathrm{cp}}\rho_j$. Since the stacked vector $\delta^k$ has blockwise bound $\|\delta_i^k\| \le \kappa_i\|d_i^k\| + \mu_j H_i C_i^\psi \mathbb{1}_{\{i\in\mathcal{I}_o\}}$, the blockwise estimates and Minkowski's inequality give

$$\|\delta^k\| \le \Big(\sum_{i\in\mathcal{I}} \kappa_i^2\|d_i^k\|^2\Big)^{1/2} + \mu_j\Big(\sum_{i\in\mathcal{I}_o}(H_iC_i^\psi)^2\Big)^{1/2} \le r_{\mathrm{cp}}\rho_j\|d^k\| + \mu_j J_H,$$

with $J_H$ from Equation 29.

*Identity Equation 161.* By the definition preceding Equation 106, $\Xi^{k+1} = A^T(q^k + \delta^k) + C^T\zeta^{k+1}$ with $\zeta^{k+1} = \rho_j\varpi^{k+1}$. Specializing Equation 77 to $\mathsf{m} = \mathsf{Ckpt}$ and multiplying it by $\rho_j$ and using $y^{k+1} - Ax^k = r^{k+1} + A\Delta x^k$ gives

$$C^T\zeta^{k+1} = A^T\left(\lambda^{[j]} + \rho_j r^{k+1} + \rho_j A\Delta x^k\right) - (1+\theta_2)\rho_j A^T A\Delta x^k = A^T\widehat{\lambda}^{k+1} - \theta_2\rho_j A^T A\Delta x^k,$$

where the last equality uses Equation 158. Also by Equation 158, $q^k = -\widehat{\lambda}^{k+1} - \Upsilon^k d^k - \rho_j A\Delta x^k$. Adding the two displays, the terms $\pm A^T\widehat{\lambda}^{k+1}$ cancel and

$$\Xi^{k+1} = A^T\left[\delta^k - \Upsilon^k d^k - \rho_j A\Delta x^k\right] - \theta_2\rho_j A^T A\Delta x^k = A^T\left[\delta^k - \Upsilon^k d^k\right] - (1+\theta_2)\rho_j A^T A\Delta x^k.$$

*Bounds Equation 162 and Equation 163.* Because $\Upsilon^k$ is block diagonal, $\|\Upsilon^k d^k\|^2 = \sum_i |\upsilon_i^k|^2\|d_i^k\|^2 \le c_{\chi,\mathrm{cp}}^2\rho_j^2\|d^k\|^2$ by Equation 159. Combining this with Equation 160 and $\|A^Tu\| \le \|A\|_2\|u\|$ in Equation 161 proves Equation 162. If the test is accepted, then $E^{k+1} \le \tau_{\mathrm{cp}}/\rho_j$ and, by the Cauchy–Schwarz inequality applied to the two movement channels,

$$(r_{\mathrm{cp}} + c_{\chi,\mathrm{cp}})\rho_j\|d^k\| + (1+\theta_2)\rho_j\|A\Delta x^k\| \le c_{\mathrm{mov}}\,\rho_j\sqrt{\|d^k\|^2 + \|A\Delta x^k\|^2} = c_{\mathrm{mov}}E^{k+1} \le \frac{c_{\mathrm{mov}}\tau_{\mathrm{cp}}}{\rho_j},$$

with $c_{\mathrm{mov}}$ from Equation 29. Since $\mu_j = \eta/\rho_j$ by Equation 30, substitution in Equation 162 gives $\|\Xi^{k+1}\| \le \|A\|_2(c_{\mathrm{mov}}\tau_{\mathrm{cp}} + \eta J_H)/\rho_j = S_\Xi/\rho_j$. □

**Lemma I.3** (Checkpoint support width)**.** *Let $k$ be an iteration of level $j$ whose test is accepted, and suppose $f_i(y_i^{k+1}) \ge 0$ for every $i \in \mathcal{I}_c$. Then, with $S_\omega$ from Equation 156,*

$$\iota_i^{k+1} \ne 0 \quad\Longrightarrow\quad 0 \le f_i(y_i^{k+1}) \le \beta_j + \frac{L_i}{2}\|d_i^k\|^2 \le \frac{S_\omega}{\rho_j}. \tag{164}$$

*Consequently*

$$\sum_{i\in\mathcal{I}_c}\left|\iota_i^{k+1} f_i(y_i^{k+1})\right| \le \frac{S_\omega}{\rho_j}\|\iota^{k+1}\|_1. \tag{165}$$

*Proof.* By definition, $\iota_i^{k+1} = g'(\ell_i^k(y_i^{k+1}), L_j)$, and the explicit force Equation 51 shows that $g'(\cdot, L_j)$ vanishes exactly on $[\beta(L_j), \infty)$. Hence $\iota_i^{k+1} \ne 0$ forces $\ell_i^k(y_i^{k+1}) < \beta(L_j) = \beta_j$, the last equality by Equation 66. Since $f_i$ has an $L_i$-Lipschitz gradient by Assumption 3.1,

$$f_i(y_i^{k+1}) \le \ell_i^k(y_i^{k+1}) + \frac{L_i}{2}\|d_i^k\|^2 < \beta_j + \frac{L_i}{2}\|d_i^k\|^2.$$

This is the widened affine-force support; the exact width $\beta_j$ alone does not apply, exactly as in Equation 118. The lower bound is the hypothesis. Acceptance gives $\rho_j\|d_i^k\| \le \rho_j\|d^k\| \le E^{k+1} \le \tau_{\mathrm{cp}}/\rho_j$, where the first step is the blockwise bound $\|d_i^k\| \le \|d^k\|$ and the second discards the nonnegative $\|A\Delta x^k\|$ channel of $E^{k+1}$; hence $\|d_i^k\|^2 \le \tau_{\mathrm{cp}}^2/\rho_j^4$, and $L_i \le L_{\max}$, so

$$\beta_j + \frac{L_i}{2}\|d_i^k\|^2 \le \frac{b_{\mathrm{cp}}}{\rho_j} + \frac{L_{\max}\tau_{\mathrm{cp}}^2}{2\rho_j^4} = \frac{1}{\rho_j}\left[b_{\mathrm{cp}} + \frac{L_{\max}\tau_{\mathrm{cp}}^2}{2\rho_j^3}\right] \le \frac{1}{\rho_j}\left[b_{\mathrm{cp}} + \frac{L_{\max}\tau_{\mathrm{cp}}^2}{2(\rho_{\mathrm{cp}}^0)^3}\right] = \frac{S_\omega}{\rho_j},$$

using $\rho_j \ge \rho_{\mathrm{cp}}^0$. Finally, $\iota_i^{k+1} f_i(y_i^{k+1})$ vanishes for every index outside the support, so Equation 165 follows by bounding each surviving factor $f_i(y_i^{k+1})$ by $S_\omega/\rho_j$ and summing $|\iota_i^{k+1}|$. □

### I.2 Merit Bounds under the Stored-Multiplier Cap and Split Feasibility

The two results of this subsection are stated for an arbitrary point of the frozen level, not only for iterates, and are conditional on the single hypothesis $\|\lambda^{[j]}\| \le Q_{\mathrm{cp}}$. That hypothesis is supplied by the induction of Lemma I.8, which uses the present results only at levels where it has already been verified; there is therefore no circularity.

**Lemma I.4** (Merit lower bound under the stored-multiplier cap and coarse consensus)**.** *Fix $j \ge 0$ and suppose $\|\lambda^{[j]}\| \le Q_{\mathrm{cp}}$. Let $x$ and $y$ be arbitrary; no feasibility of $x$ is required. Then, with $D_{\mathrm{cp}}$ from Equation 29,*

$$\mathcal{P}_j(x,y) \ge \sum_{i\in\mathcal{I}_c} g(f_i(y_i), L_j) - \frac{D_{\mathrm{cp}}}{\rho_j} \ge -\frac{D_{\mathrm{cp}}}{\rho_j} \ge -\Delta_G. \tag{166}$$

*If moreover $\mathcal{P}_j(x,y) \le \overline{V}$, then the coarse consensus bound*

$$\|y - Ax\| \le \frac{2(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})}{\rho_j} + \sqrt{\frac{2\overline{V}}{\rho_j}} \le \delta_{\mathrm{cq}} \tag{167}$$

*holds, so that in particular $\max_{i\in\mathcal{I}_c}\|A_i x - y_i\| \le \delta_{\mathrm{cq}}$. Here $\delta_{\mathrm{cq}} > 0$ is the band radius supplied by Assumption 4.2, which enters the base penalty $\rho_{\mathrm{cp}}^0$ of Equation 29; only that constant is used, and the qualification conclusion Equation 15 is not invoked here.*

*Proof.* Write $s := \|y - Ax\|$ and $s_i := \|y_i - A_i x\|$, so that $s^2 = \sum_{i\in\mathcal{I}} s_i^2$. Split Equation 155 as

$$\mathcal{P}_j(x,y) = \underbrace{\left[\sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L_j, \mu_j) + R_\psi(\mu_j)\right]}_{\text{(I)}} + \underbrace{\sum_{i\in\mathcal{I}_c} g(f_i(y_i), L_j)}_{\text{(II)}} + \underbrace{\langle \lambda^{[j]}, y - Ax\rangle + \frac{\rho_j}{2}s^2}_{\text{(III)}}.$$

*Term (I).* For $i \in \mathcal{I}_o$, $\bar{f}_i(\cdot, L_j, \mu_j) = h_i + (\psi_i)_{\mu_j}$ is globally $\widehat{C}_i$-Lipschitz: $h_i$ is $D_i$-Lipschitz by Assumption 3.4, and $\|\nabla(\psi_i)_{\mu_j}\| \le C_i^\psi$ by Equation 34, with $\widehat{C}_i = D_i + C_i^\psi$. Hence

$$\sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L_j, \mu_j) \ge \sum_{i\in\mathcal{I}_o} \bar{f}_i(A_i x, L_j, \mu_j) - \sum_{i\in\mathcal{I}_o} \widehat{C}_i s_i \ge \sum_{i\in\mathcal{I}_o} \bar{f}_i(A_i x, L_j, \mu_j) - \sqrt{\widehat{G}_0}\, s,$$

the last step by the Cauchy–Schwarz inequality with $\widehat{G}_0 = \sum_{i\in\mathcal{I}_o} \widehat{C}_i^2$, which gives $\sum_{i\in\mathcal{I}_o} \widehat{C}_i s_i \le \sqrt{\widehat{G}_0}\big(\sum_{i\in\mathcal{I}_o} s_i^2\big)^{1/2}$, followed by $\sum_{i\in\mathcal{I}_o} s_i^2 \le \sum_{i\in\mathcal{I}} s_i^2 = s^2$ because the omitted squares for $i \in \mathcal{I}_c$ are nonnegative. Since Equation 49 holds at every $x$, it gives $\sum_{i\in\mathcal{I}_o} \bar{f}_i(A_i x, L_j, \mu_j) + R_\psi(\mu_j) \ge 0$. Therefore $\text{(I)} \ge -\sqrt{\widehat{G}_0}\, s$. This is the only step of the present proof that invokes Equation 49; the aggregate objective lower bound of Assumption 3.4 enters through that lemma, and no boundedness of the iterates or feasibility of $x$ is needed.

*Term (II).* The wall sum is retained unchanged and is nonnegative.

*Term (III).* By Cauchy–Schwarz and the hypothesis, $\text{(III)} \ge -Q_{\mathrm{cp}} s + \frac{\rho_j}{2}s^2$.

*Lower bound.* Adding the three terms,

$$\mathcal{P}_j(x,y) \geq \sum_{i\in\mathcal{I}_c} g(f_i(y_i), L_j) - \left(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0}\right) s + \frac{\rho_j}{2} s^2. \tag{168}$$

The quantity $s$ is by definition nonnegative, so minimizing the scalar quadratic $-(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})s + \frac{\rho_j}{2}s^2$ over all $s \in \mathbb{R}$ is a relaxation of the minimization over $s \geq 0$ and therefore still yields a valid lower bound; the relaxation is in fact tight, because its unconstrained minimizer $(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})/\rho_j$ is itself nonnegative, and its value is $-(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})^2/(2\rho_j) = -D_{\mathrm{cp}}/\rho_j$, which proves the first inequality of Equation 166. The wall $g$ is nonnegative on both of its branches, so the second inequality follows. The schedule Equation 30 gives $\rho_j \geq \rho^0_{\mathrm{cp}}$, while Equation 29 gives $\rho^0_{\mathrm{cp}} \geq D_{\mathrm{cp}}/\Delta_G$, proving the third.

*Coarse consensus.* Discarding the nonnegative wall sum in Equation 168 and using $\mathcal{P}_j(x,y) \leq \overline{V}$,

$$\frac{\rho_j}{2} s^2 - \left(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0}\right) s - \overline{V} \leq 0.$$

The quadratic formula and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ give

$$s \leq \frac{(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0}) + \sqrt{(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})^2 + 2\rho_j \overline{V}}}{\rho_j} \leq \frac{2(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})}{\rho_j} + \sqrt{\frac{2\overline{V}}{\rho_j}}.$$

By Equation 30 and Equation 29, $\rho_j \geq \rho^0_{\mathrm{cp}} \geq 4(Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})/\delta_{\mathrm{cq}}$ bounds the first term by $\delta_{\mathrm{cq}}/2$, and $\rho_j \geq \rho^0_{\mathrm{cp}} \geq 8\overline{V}/\delta^2_{\mathrm{cq}}$ gives $2\overline{V}/\rho_j \leq \delta^2_{\mathrm{cq}}/4$ and hence bounds the second term by $\delta_{\mathrm{cq}}/2$. Finally $\max_{i\in\mathcal{I}_c} \|A_i x - y_i\| \leq \|y - Ax\| = s$. □

**Lemma I.5** (Initialization and within-level strict split feasibility)**.** *The following hold.*

*(i)* (Initialization.) *The starting point satisfies*

$$\mathcal{P}_0(x^1, y^1) \leq S_0 + \Delta_G = \overline{V} - \Delta_G. \tag{169}$$

*(ii)* (Strict split feasibility from the cap.) *Fix $j \geq 0$ with $\|\lambda^{[j]}\| \leq Q_{\mathrm{cp}}$, and let $(x,y)$ satisfy $\mathcal{P}_j(x,y) \leq \overline{V}$; as in Lemma I.4, no feasibility of $x$ is required. Then*

$$f_i(y_i) > 0 \qquad \text{for every } i \in \mathcal{I}_c. \tag{170}$$

*(iii)* (Propagation along a level.) *Fix $j \geq 0$ with $\|\lambda^{[j]}\| \leq Q_{\mathrm{cp}}$ and suppose the first iterate of level $j$ satisfies $\mathcal{P}_j(x^{k_{j-1}+1}, y^{k_{j-1}+1}) \leq \overline{V}$. Then every generated iterate $(x^k, y^k)$ with $k \geq k_{j-1}+1$ that is produced by an iteration while the level state $(\lambda^{[j]}, \rho_j, L_j, \mu_j)$ is in force satisfies $Cx^k = d$, $\mathcal{P}_j(x^k, y^k) \leq \overline{V}$, and $f_i(y^k_i) > 0$ for every $i \in \mathcal{I}_c$, and Equation 167 holds at each such iterate. No termination of the level is presupposed; if the level does terminate at $k_j < \infty$, the range of admissible indices is $k_{j-1}+1 \leq k \leq k_j+1$ and includes the checkpoint certificate index $k_j+1$.*

*Proof.* (i) By construction $y^1 = Ax^1$ and $\lambda^{[0]} = \lambda^1 = 0$, so the linear and quadratic consensus terms of Equation 155 both vanish and

$$\mathcal{P}_0(x^1, y^1) = \sum_{i\in\mathcal{I}_o} \left[h_i(A_i x^1) + (\psi_i)_{\mu_0}(A_i x^1)\right] + \sum_{i\in\mathcal{I}_c} g(f_i(A_i x^1), L_0) + R_\psi(\mu_0).$$

For $i \in \mathcal{I}_o$, the left inequality of Equation 35 gives $(\psi_i)_{\mu_0} \leq \psi_i$, whence the first sum is at most $\sum_{i\in\mathcal{I}_o} \varphi_i(A_i x^1) = S_0$. For $i \in \mathcal{I}_c$, Assumption 3.3 gives $f_i(A_i x^1) \geq m > 0$, while Equation 29 contains the clause $\rho^0_{\mathrm{cp}} \geq b_{\mathrm{cp}}/m$, so by Equation 66 together with $\rho_0 = \rho^0_{\mathrm{cp}}$ from Equation 30

$$\beta(L_0) = \beta_0 = \frac{b_{\mathrm{cp}}}{\rho^0_{\mathrm{cp}}} \leq m \leq f_i(A_i x^1).$$

The feasible branch of $g$ then gives $g(f_i(A_i x^1), L_0) = \frac{L_0}{2}\max\{\beta_0 - f_i(A_i x^1), 0\}^2 = 0$, so the whole wall sum vanishes: the interior-point term is switched off at the strictly feasible start. Finally $R_\psi(\mu_0) = \eta^{\mathrm{cp}} G_\psi/(2\rho_{\mathrm{cp}}^0) \le \Delta_G$ by the clause $\rho_{\mathrm{cp}}^0 \ge \eta^{\mathrm{cp}} G_\psi/(2\Delta_G)$ in Equation 29. Adding the three estimates and using $\overline{V} = S_0 + 2\Delta_G$ proves Equation 169.

(ii) Suppose, for contradiction, that $f_{i_0}(y_{i_0}) \le 0$ for some $i_0 \in \mathcal{I}_{\mathrm{c}}$. Both branches of $g$ give $g(s, L) \ge G$ for $s \le 0$: for $s < 0$ the value is $G - \sqrt{2LG}\, s > G$, and for $s = 0$ it is $\frac{L}{2}(\sqrt{2G/L})^2 = G$. Since $g \ge 0$ on the remaining constraint blocks, $\sum_{i \in \mathcal{I}_{\mathrm{c}}} g(f_i(y_i), L_j) \ge G$, and Equation 166 gives

$$\mathcal{P}_j(x, y) \ge G - \frac{D_{\mathrm{cp}}}{\rho_j} \ge G - \Delta_G.$$

By the definition $\Delta_G = (G - S_0)/4$ in Equation 29, $G = S_0 + 4\Delta_G$, so $G - \Delta_G = S_0 + 3\Delta_G > S_0 + 2\Delta_G = \overline{V}$, where the strict inequality uses $\Delta_G > 0$. This contradicts $\mathcal{P}_j(x, y) \le \overline{V}$. Thus no index $i_0 \in \mathcal{I}_{\mathrm{c}}$ can satisfy $f_{i_0}(y_{i_0}) \le 0$, so $f_i(y_i) > 0$ for every $i \in \mathcal{I}_{\mathrm{c}}$. The conclusion is strict, with the quantitative slack $\Delta_G$ in the merit value.

(iii) Equality $Cx^k = d$ holds for $k = 1$ by Assumption 3.3 and for every $k \ge 2$ by the second line of Equation 77; the $x$ subproblem is solved over the affine set $\{Cx = d\}$ at every iteration, in both branches of the checkpoint test. Within level $j$ all four parameters $(\lambda^{[j]}, \rho_j, L_j, \mu_j)$ are frozen, so Lemma I.1 applies at every iteration $k \ge k_{j-1} + 1$ of the level and yields $\mathcal{P}_j(x^{k+1}, y^{k+1}) \le \mathcal{P}_j(x^k, y^k)$. Chaining from the first iterate of the level gives $\mathcal{P}_j(x^k, y^k) \le \overline{V}$ at every index reached while the level state is in force, and, when the level terminates, up to and including $k_j + 1$, the index of the pre-promotion checkpoint certificate tuple. Parts (ii) and Lemma I.4 then apply at each such iterate. □

**What split feasibility does and does not assert.** Two distinct feasibility statements must be kept apart, and only the first is proved above. *Split feasibility* is the pair

$$Cx^k = d, \qquad f_i(y_i^k) > 0 \quad (i \in \mathcal{I}_{\mathrm{c}}),$$

which Lemma I.5 establishes for every generated trial iterate, accepted or rejected, at every level, conditionally on the two level hypotheses. Those hypotheses are discharged by the induction of Lemma I.8, whose conclusion Equation 182 is the unconditional form of the pair. The linear constraint is satisfied by the true variable $x^k$, whereas the inequality constraints are satisfied *strictly* by the split variable $y^k$, which is the argument the wall $g$ actually sees.

*Original feasibility*, namely $f_i(A_i x^k) \ge 0$, is *not* asserted at finite $k$. The split and the original argument differ by the consensus residual, which is only $\mathcal{O}(1/\rho_j)$: at an accepted checkpoint, Lemma I.6 gives $\|y^{k_j+1} - Ax^{k_j+1}\| \le 2Q_{\mathrm{cp}}/\rho_j$, so Assumption 3.2 yields only

$$f_i(A_i x^{k_j+1}) \ge f_i(y_i^{k_j+1}) - M_i \|A_i x^{k_j+1} - y_i^{k_j+1}\| > -\frac{2M_{\max} Q_{\mathrm{cp}}}{\rho_j},$$

an asymptotic, not exact, original-feasibility statement. This is why the checkpoint certificate in Section I.4 measures consensus, stationarity, complementarity, and near-active support of the pair $(x^{k_j+1}, \breve{y}^{k_j+1})$ rather than claiming exact original feasibility. Strictness of $f_i(y_i^k) > 0$ keeps the true-constraint wall in the checkpoint merit on its feasible branch and the merit below $G$; the affine model wall used in a block update may still evaluate a negative linearized argument.

Finally, note the asymmetry in the two hypotheses of Lemma I.5(ii): the merit cap $\overline{V}$ is a *level* quantity and must be re-established after each promotion, since $\mathcal{P}_{j+1}$ uses a different multiplier, penalty, curvature, and smoothing parameter than $\mathcal{P}_j$. That re-establishment is the transition charge of Lemma I.8, and it is the only step of the analysis in which the merit value can increase.

### I.3 Accepted Checkpoints, Global Invariants, and Finite Arrival

For $j \ge 0$ consider the two *level hypotheses*

$$(\mathrm{I}_j) \quad \|\lambda^{[j]}\| \le Q_{\mathrm{cp}}, \qquad (\mathrm{II}_j) \quad \mathcal{P}_j\left(x^{k_{j-1}+1}, y^{k_{j-1}+1}\right) \le \overline{V}. \tag{171}$$

Lemma I.6 and Lemma I.7 are proved under $(\mathrm{I}_j)$ and $(\mathrm{II}_j)$ alone, and Lemma I.8 closes the induction that verifies both for every $j$.

**Lemma I.6** (Accepted-checkpoint multiplier and consensus bounds)**.** *Suppose Assumption 4.2 holds. Fix $j \ge 0$, assume $(\mathrm{I}_j)$ and $(\mathrm{II}_j)$, and let $k$ be an iteration of level $j$ whose test is accepted, so that $k = k_j$. Write $r := r^{k+1}$. Then, in this order:*

*(a)* (Coarse consensus and strict feasibility.)

$$Cx^{k+1} = d, \qquad f_i(y_i^{k+1}) > 0 \ \ (i \in \mathcal{I}_c), \qquad \|r\| \le \delta_{\mathrm{cq}}. \tag{172}$$

*(b)* (Support inside the qualification band.)

$$\iota_i^{k+1} \ne 0 \quad \Longrightarrow \quad 0 \le f_i(y_i^{k+1}) \le \frac{S_\omega}{\rho_j} \le \delta_{\mathrm{cq}}. \tag{173}$$

*(c)* (Uniform Band-MFCQ inequality-multiplier bound.) *The wall forces satisfy $\iota_i^{k+1} \le 0$, the associated nonnegative inequality multipliers are $-\iota_i^{k+1} \ge 0$, and*

$$\|\iota^{k+1}\|_1 = \sum_{i \in \mathcal{I}_c} \left(-\iota_i^{k+1}\right) \le \frac{M_F + S_\Xi}{\gamma_{\mathrm{cq}}} = I_{\mathrm{cp}}. \tag{174}$$

*(d)* (Raw accepted virtual multiplier.)

$$\|\widehat{\lambda}^{k+1}\| \le \sqrt{\widehat{G}_0} + M_{\max} I_{\mathrm{cp}} + (c_{\chi,\mathrm{cp}} + 1)\tau_{\mathrm{cp}} = Q_{\mathrm{cp}}. \tag{175}$$

*(e)* (Sharp consensus.)

$$\|r\| \le \frac{2Q_{\mathrm{cp}}}{\rho_j}. \tag{176}$$

*The bound in (d) is asserted only for an accepted virtual multiplier. No bound is claimed for a rejected $\widehat{\lambda}^{k+1}$, which is discarded and never stored.*

*Proof.* The order of the five parts is forced: the qualification Assumption 4.2 is available only inside a $\delta_{\mathrm{cq}}$ consensus band, and its conclusion applies only to indices whose constraint value lies in $[0, \delta_{\mathrm{cq}}]$. Both facts must therefore be established before any multiplier bound, and the sharp consensus estimate is a consequence, not a hypothesis, of the multiplier bound.

*(a).* By $(\mathrm{I}_j)$, $(\mathrm{II}_j)$, and Lemma I.5(iii) applied at the checkpoint certificate index $k+1 = k_j + 1$, we have $Cx^{k+1} = d$, $\mathcal{P}_j(x^{k+1}, y^{k+1}) \le \overline{V}$, and $f_i(y_i^{k+1}) > 0$ for every $i \in \mathcal{I}_c$. Lemma I.4 then gives $\|r\| = \|y^{k+1} - Ax^{k+1}\| \le \delta_{\mathrm{cq}}$.

*(b).* Part (a) supplies the hypothesis $f_i(y_i^{k+1}) \ge 0$ of Lemma I.3, whose conclusion Equation 164 gives $0 \le f_i(y_i^{k+1}) \le S_\omega/\rho_j$ on the support. It remains to check $S_\omega/\rho_j \le \delta_{\mathrm{cq}}$. By Equation 156 and $\rho_j \ge \rho_{\mathrm{cp}}^0$,

$$\frac{S_\omega}{\rho_j} \le \frac{b_{\mathrm{cp}}}{\rho_{\mathrm{cp}}^0} + \frac{L_{\max}\tau_{\mathrm{cp}}^2}{2(\rho_{\mathrm{cp}}^0)^4}.$$

The clause $\rho_{\mathrm{cp}}^0 \ge 2b_{\mathrm{cp}}/\delta_{\mathrm{cq}}$ of Equation 29 bounds the first summand by $\delta_{\mathrm{cq}}/2$, and the clause $\rho_{\mathrm{cp}}^0 \ge (L_{\max}\tau_{\mathrm{cp}}^2/\delta_{\mathrm{cq}})^{1/4}$ gives $(\rho_{\mathrm{cp}}^0)^4 \ge L_{\max}\tau_{\mathrm{cp}}^2/\delta_{\mathrm{cq}}$ and hence bounds the second summand by $\delta_{\mathrm{cq}}/2$.

*(c).* The explicit force Equation 51 gives $\iota_i^{k+1} = g'(\ell_i^k(y_i^{k+1}), L_j) \le 0$ and $-\iota_i^{k+1} \ge 0$; consequently $\|\iota^{k+1}\|_1 = \sum_{i \in \mathcal{I}_c}(-\iota_i^{k+1})$. By part (a) the pair $(x^{k+1}, y^{k+1})$ satisfies every hypothesis of Assumption 4.2: $Cx^{k+1} = d$, $f_i(y_i^{k+1}) \ge 0$, and $\max_{i \in \mathcal{I}_c} \|A_i x^{k+1} - y_i^{k+1}\| \le \|r\| \le \delta_{\mathrm{cq}}$. Let $d_{\mathrm{cq}} \in \ker C$ be the resulting unit vector of Equation 15.

By the definition of $\Xi^{k+1}$ preceding Equation 106 and by the definition of $\delta^k$ in Lemma I.2, $q_i^k + \delta_i^k = v_i^{k+1}$ for $i \in \mathcal{I}_o$ and $q_i^k + \delta_i^k = \iota_i^{k+1} \nabla f_i(y_i^{k+1})$ for $i \in \mathcal{I}_c$, so

$$\Xi^{k+1} = \sum_{i \in \mathcal{I}_o} A_i^T v_i^{k+1} + \sum_{i \in \mathcal{I}_c} \iota_i^{k+1} A_i^T \nabla f_i(y_i^{k+1}) + C^T \zeta^{k+1}.$$

Contract with $d_{\mathrm{cq}}$. Since $Cd_{\mathrm{cq}} = 0$, the last term drops out and

$$\langle \Xi^{k+1}, d_{\mathrm{cq}} \rangle = \sum_{i \in \mathcal{I}_{\mathrm{o}}} \langle v_i^{k+1}, A_i d_{\mathrm{cq}} \rangle - \sum_{i \in \mathcal{I}_{\mathrm{c}}} \left( -\iota_i^{k+1} \right) \nabla f_i(y_i^{k+1})^T A_i d_{\mathrm{cq}}.$$

For $i \in \mathcal{I}_{\mathrm{o}}$, the reconstruction Equation 42 defines $v_i^{k+1} = \nabla h_i(p_i^{k+1}) + w_i^{k+1}$, so $\|v_i^{k+1}\| \le D_i + C_i^{\psi} = \widehat{C}_i$, since $\|\nabla h_i\| \le D_i$ by Equation 31 and $\|w_i^{k+1}\| \le C_i^{\psi}$ by Equation 34. Hence, by the Cauchy–Schwarz inequality in the stacked space and $\|d_{\mathrm{cq}}\| = 1$,

$$\left| \sum_{i \in \mathcal{I}_{\mathrm{o}}} \langle v_i^{k+1}, A_i d_{\mathrm{cq}} \rangle \right| \le \Big( \sum_{i \in \mathcal{I}_{\mathrm{o}}} \widehat{C}_i^2 \Big)^{1/2} \|A d_{\mathrm{cq}}\| \le \|A\|_2 \sqrt{\widehat{G}_0} = M_F,$$

which is exactly the constant Equation 125, recalled in Equation 29. For $i \in \mathcal{I}_{\mathrm{c}}$ with $\iota_i^{k+1} \neq 0$, part (b) puts $f_i(y_i^{k+1})$ in $[0, \delta_{\mathrm{cq}}]$, so Equation 15 gives $\nabla f_i(y_i^{k+1})^T A_i d_{\mathrm{cq}} \ge \gamma_{\mathrm{cq}}$; the indices with $\iota_i^{k+1} = 0$ contribute nothing. Therefore

$$\gamma_{\mathrm{cq}} \|\iota^{k+1}\|_1 \le \sum_{i \in \mathcal{I}_{\mathrm{c}}} \left( -\iota_i^{k+1} \right) \nabla f_i(y_i^{k+1})^T A_i d_{\mathrm{cq}} = \sum_{i \in \mathcal{I}_{\mathrm{o}}} \langle v_i^{k+1}, A_i d_{\mathrm{cq}} \rangle - \langle \Xi^{k+1}, d_{\mathrm{cq}} \rangle \le M_F + \|\Xi^{k+1}\|.$$

Because the step is accepted, Equation 163 gives $\|\Xi^{k+1}\| \le S_\Xi/\rho_j \le S_\Xi$, using $\rho_j \ge \rho_{\mathrm{cp}}^0 \ge 1$. Dividing by $\gamma_{\mathrm{cq}} > 0$ proves Equation 174. This is a *raw* bound: it is uniform in $j$ and carries no factor of $\rho_j$ or $\sqrt{\rho_j}$, which is precisely why $I_{\mathrm{cp}}$ in Equation 29 is defined with $S_\Xi$ rather than $S_\Xi/\rho_{\mathrm{cp}}^0$.

*(d).* Bound the three terms of the exact identity Equation 158. For the force, $\|q_i^k\| \le \|\nabla h_i(y_i^k)\| + \|w_i^{k+1}\| \le D_i + C_i^{\psi} = \widehat{C}_i$ for $i \in \mathcal{I}_{\mathrm{o}}$, using $\|\nabla h_i\| \le D_i$ from Equation 31 and $\|w_i^{k+1}\| \le C_i^{\psi}$ from Equation 34, while $\|q_i^k\| = |\iota_i^{k+1}| \|\nabla f_i(y_i^k)\| \le M_{\max} |\iota_i^{k+1}|$ for $i \in \mathcal{I}_{\mathrm{c}}$ by Assumption 3.2. Hence, using $\|\cdot\|_2 \le \|\cdot\|_1$ on the constraint block and part (c),

$$\|q^k\| \le \Big( \sum_{i \in \mathcal{I}_{\mathrm{o}}} \widehat{C}_i^2 \Big)^{1/2} + M_{\max} \|\iota^{k+1}\|_2 \le \sqrt{\widehat{G}_0} + M_{\max} I_{\mathrm{cp}}.$$

For the movement terms, acceptance gives $\rho_j \|d^k\| \le E^{k+1} \le \tau_{\mathrm{cp}}/\rho_j \le \tau_{\mathrm{cp}}$ and likewise $\rho_j \|A\Delta x^k\| \le \tau_{\mathrm{cp}}$, so Equation 159 gives $\|\Upsilon^k d^k\| \le c_{\chi,\mathrm{cp}} \rho_j \|d^k\| \le c_{\chi,\mathrm{cp}} \tau_{\mathrm{cp}}$ and $\|\rho_j A \Delta x^k\| \le \tau_{\mathrm{cp}}$. Adding the three bounds and comparing with the definition of $Q_{\mathrm{cp}}$ in Equation 29 proves Equation 175.

*(e).* By definition $\rho_j r = \widehat{\lambda}^{k+1} - \lambda^{[j]}$, so $(\mathrm{I}_j)$ and part (d) give $\rho_j \|r\| \le 2Q_{\mathrm{cp}}$. This sharpens the coarse estimate of part (a), which is only the constant $\delta_{\mathrm{cq}}$, to a bound of order $1/\rho_j$. Measured against the underlying coarse bound Equation 167, whose dominant term is $\sqrt{2\overline{V}/\rho_j}$, the effective gain is a factor of order $\rho_j^{1/2}$ rather than $\rho_j$. The sharp estimate could not have been used to obtain (c), because (c) is what produced the bound on $\widehat{\lambda}^{k+1}$. □

**Lemma I.7** (Failure cost and finite arrival)**.** *Fix $j \ge 0$ and assume $(\mathrm{I}_j)$ and $(\mathrm{II}_j)$; the two level hypotheses are used only in parts (b) and (c), while part (a) holds for any level $j$ without them. Then:*

*(a)* (Failure cost.) *If $k$ is an iteration of level $j$ whose test is rejected, then*

$$\mathcal{P}_j(x^k, y^k) - \mathcal{P}_j(x^{k+1}, y^{k+1}) > \frac{a_0 \tau_{\mathrm{cp}}^2}{\rho_j^3}. \tag{177}$$

*(b)* (Level budget.) *Along level $j$ the total available merit decrease is at most*

$$\overline{V} + \frac{D_{\mathrm{cp}}}{\rho_j} \le \overline{V} + \Delta_G = \Delta_*. \tag{178}$$

*(c)* (Finite arrival and per-level work.) *Level $j$ terminates after finitely many iterations, and*

$$T_j < 1 + C_{\mathrm{work}} \rho_j^3, \qquad C_{\mathrm{work}} = \frac{\Delta_*}{a_0 \tau_{\mathrm{cp}}^2}. \tag{179}$$

*Proof.* (a) Rejection at step $k$ means $E^{k+1} > \tau_{\mathrm{cp}}/\rho^k = \tau_{\mathrm{cp}}/\rho_j$, that is,

$$\rho_j\sqrt{\|d^k\|^2 + \|A\Delta x^k\|^2} > \frac{\tau_{\mathrm{cp}}}{\rho_j}, \qquad \text{hence} \qquad \|d^k\|^2 + \|A\Delta x^k\|^2 > \frac{\tau_{\mathrm{cp}}^2}{\rho_j^4}.$$

The state $(\lambda^{[j]}, \rho_j, L_j, \mu_j)$ is unchanged by a rejection, so Lemma I.1 applies verbatim at step $k$ and yields a decrease of at least $a_0\rho_j(\|d^k\|^2 + \|A\Delta x^k\|^2) > a_0\tau_{\mathrm{cp}}^2/\rho_j^3$. The cost is charged in the *same* merit function on both sides, which is possible only because the level state is frozen; this is the reason the checkpoint argument needs no drift correction.

(b) By $(\mathrm{II}_j)$ the level starts at merit value at most $\overline{V}$. The hypothesis $(\mathrm{I}_j)$ supplies the multiplier cap required by Lemma I.4, whose conclusion Equation 166 gives $\mathcal{P}_j(x^k, y^k) \geq -D_{\mathrm{cp}}/\rho_j \geq -\Delta_G$ throughout the level. Since Lemma I.1 makes $\mathcal{P}_j$ nonincreasing, the accumulated decrease over the whole level cannot exceed $\overline{V} - (-\Delta_G) = \Delta_*$.

(c) Suppose first that no test of level $j$ is ever accepted. Then level $j$ never terminates, so it is the infinite tail $\{k_{j-1}+1, k_{j-1}+2, \dots\}$ on all of which the level state is frozen and every test is rejected. Hence, for every $T \geq 1$, the first $T$ iterations of the level are all rejections, and summing Equation 177 telescopically over them gives an accumulated decrease exceeding $Ta_0\tau_{\mathrm{cp}}^2/\rho_j^3$, which is unbounded in $T$ and contradicts (b). Hence some test is accepted and $k_j < \infty$. If $T_j = 1$, the claimed bound follows immediately from $\Delta_* > 0$. If $T_j \geq 2$, the last iteration of the level, $k = k_j$, is the accepted one and the preceding $T_j - 1$ iterations are rejections, so

$$(T_j - 1)\frac{a_0\tau_{\mathrm{cp}}^2}{\rho_j^3} < \Delta_*, \qquad \text{that is,} \qquad T_j < 1 + \frac{\Delta_*}{a_0\tau_{\mathrm{cp}}^2}\rho_j^3 = 1 + C_{\mathrm{work}}\rho_j^3,$$

which gives Equation 179. The cubic exponent is exactly the product of the two $\rho_j^{-1}$ factors in the test threshold $E^{k+1} \leq \tau_{\mathrm{cp}}/\rho_j$, squared into the movement, against the single factor $\rho_j$ in the descent coefficient: the per-failure cost $a_0\tau_{\mathrm{cp}}^2/\rho_j^3$ is $a_0\rho_j \cdot (\tau_{\mathrm{cp}}/\rho_j^2)^2$. □

**Lemma I.8** (Transition charge and global invariants)**.** *Suppose Assumption 4.2 holds. Then:*

*(a)* (One-step charge.) *Fix $j \geq 0$, assume $(\mathrm{I}_j)$ and $(\mathrm{II}_j)$, and let $k_j < \infty$ be the accepting iteration of level $j$. Then the promotion at $k_j$ raises the merit value by at most*

$$\mathcal{P}_{j+1}\left(x^{k_j+1}, y^{k_j+1}\right) - \mathcal{P}_j\left(x^{k_j+1}, y^{k_j+1}\right) \leq \frac{\chi+1}{2}\rho_j\|r^{k_j+1}\|^2 \leq \frac{2(\chi+1)Q_{\mathrm{cp}}^2}{\rho_j} := c_j, \tag{180}$$

*and the charges are summable with*

$$\sum_{j=0}^{\infty} c_j = \frac{2(\chi+1)Q_{\mathrm{cp}}^2}{\rho_{\mathrm{cp}}^0(1-\chi^{-1})} \leq \Delta_G. \tag{181}$$

*(b)* (Global invariants.) *$(\mathrm{I}_j)$ and $(\mathrm{II}_j)$ hold and $k_j < \infty$ for every $j \geq 0$. Consequently every generated trial iterate satisfies*

$$Cx^k = d, \qquad f_i(y_i^k) > 0 \ \ (i \in \mathcal{I}_c), \qquad \|\lambda^k\| \leq Q_{\mathrm{cp}}, \tag{182}$$

*every checkpoint level is reached after finitely many iterations, and the conclusions of Lemma I.4, Lemma I.5, Lemma I.6, and Lemma I.7 hold at every level without separately assuming $(\mathrm{I}_j)$ or $(\mathrm{II}_j)$.*

*The summability bound Equation 181 is a standalone arithmetic consequence of the closed form $c_j = 2(\chi+1)Q_{\mathrm{cp}}^2/\rho_j$ and the base-penalty clauses of Equation 29: it uses neither the level hypotheses under which the charge bound of (a) is derived nor the induction of (b), so invoking it inside (b) is noncircular.*

*Proof.* (a) Write $(x, y) := (x^{k_j+1}, y^{k_j+1})$ and $r := r^{k_j+1} = y - Ax$. Recording precedes promotion, so this primal pair is simultaneously the primal component of the level-$j$ checkpoint output and the first iterate of level $j+1$; only the state changes, from $(\lambda^{[j]}, \rho_j, L_j, \mu_j)$ to $(\lambda^{[j+1]}, \rho_{j+1}, L_{j+1}, \mu_{j+1}) =$

$(\widehat{\lambda}^{k_j+1}, \chi\rho_j, \chi^2 L_j, \mu_j/\chi)$. Grouping Equation 155 into its objective, wall, and consensus parts, the difference splits exactly as

$$\mathcal{P}_{j+1}(x,y) - \mathcal{P}_j(x,y) = \underbrace{\Big[\textstyle\sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L_{j+1}, \mu_{j+1}) + R_\psi(\mu_{j+1}) - \sum_{i\in\mathcal{I}_o} \bar{f}_i(y_i, L_j, \mu_j) - R_\psi(\mu_j)\Big]}_{\text{(I)}}$$

$$+ \underbrace{\sum_{i\in\mathcal{I}_c} [g(f_i(y_i), L_{j+1}) - g(f_i(y_i), L_j)]}_{\text{(II)}} + \underbrace{\langle \lambda^{[j+1]} - \lambda^{[j]}, r\rangle + \frac{\rho_{j+1}-\rho_j}{2}\|r\|^2}_{\text{(III)}}.$$

Term (I) is at most zero: $\mu_{j+1} = \mu_j/\chi \le \mu_j$, so Equation 48 of Lemma G.5 applies with $(\mu, \mu', L, L') = (\mu_j, \mu_{j+1}, L_j, L_{j+1})$. This is exactly why the *decoupled* correction $R_\psi$, and not the ordinary correction $R$ of Lemma G.4, is the right potential here: the checkpoint schedule does not satisfy $\mu = \mu_{\mathrm{IP}}(L)$.

Term (II) is at most zero as well. By $(\mathrm{I}_j)$, $(\mathrm{II}_j)$, and Lemma I.5(iii), $f_i(y_i) > 0$ for every $i \in \mathcal{I}_c$, so the final claim of Lemma G.11, namely $g(s, L_{j+1}) \le g(s, L_j)$ for $s \ge 0$, applies blockwise. Sharpening the penalty at a split point with $f_i(y_i) \ge 0$ never increases the wall value, so nonnegativity of the split constraint values is all that this step requires; the strict inequality proved in Lemma I.5 supplies it a fortiori.

Term (III) is the entire charge. Since $\lambda^{[j+1]} - \lambda^{[j]} = \widehat{\lambda}^{k_j+1} - \lambda^{[j]} = \rho_j r$ and $\rho_{j+1} - \rho_j = (\chi-1)\rho_j$,

$$\text{(III)} = \rho_j\|r\|^2 + \frac{(\chi-1)\rho_j}{2}\|r\|^2 = \frac{\chi+1}{2}\rho_j\|r\|^2.$$

Lemma I.6(e) gives $\|r\| \le 2Q_{\mathrm{cp}}/\rho_j$, so (III) $\le \frac{\chi+1}{2}\rho_j \cdot 4Q_{\mathrm{cp}}^2/\rho_j^2 = c_j$, proving Equation 180. Summing $c_j = 2(\chi+1)Q_{\mathrm{cp}}^2/\rho_j$ with the first identity of Equation 67 gives the equality in Equation 181, and the clause $\rho_{\mathrm{cp}}^0 \ge 2(\chi+1)Q_{\mathrm{cp}}^2/[(1-\chi^{-1})\Delta_G]$ of Equation 29 gives the inequality. Note that $c_j$ is a *geometric* charge, summable because the sharp consensus bound of Lemma I.6(e) is $\mathcal{O}(1/\rho_j)$; the coarse bound Equation 167 alone would give a charge of order 1 and would not be summable.

(b) We prove by induction on $j$ the joint statement

$$\mathcal{S}_j: \qquad (\mathrm{I}_j) \text{ holds}, \qquad \mathcal{P}_j\left(x^{k_{j-1}+1}, y^{k_{j-1}+1}\right) \le \overline{V} - \Delta_G + \sum_{\ell=0}^{j-1} c_\ell, \qquad k_{j-1} < \infty.$$

Because $\sum_{\ell\ge0} c_\ell \le \Delta_G$ by Equation 181, $\mathcal{S}_j$ implies $(\mathrm{II}_j)$.

*Base* $j = 0$. Here $k_{-1} = 0 < \infty$ and $\lambda^{[0]} = \lambda^1 = 0$, so $(\mathrm{I}_0)$ holds because $Q_{\mathrm{cp}} > 0$, its definition in Equation 29 being a sum of nonnegative terms that contains the strictly positive term $(c_{\chi,\mathrm{cp}} + 1)\tau_{\mathrm{cp}}$. The middle clause with the empty sum is exactly Equation 169.

*Step.* Assume $\mathcal{S}_j$; then $(\mathrm{I}_j)$ and $(\mathrm{II}_j)$ hold. Lemma I.7(c) applies and gives $k_j < \infty$, so level $j$ has an accepting iteration. Lemma I.6(d) then gives $\|\lambda^{[j+1]}\| = \|\widehat{\lambda}^{k_j+1}\| \le Q_{\mathrm{cp}}$, which is $(\mathrm{I}_{j+1})$. Lemma I.1 makes $\mathcal{P}_j$ nonincreasing along level $j$, so

$$\mathcal{P}_j\left(x^{k_j+1}, y^{k_j+1}\right) \le \mathcal{P}_j\left(x^{k_{j-1}+1}, y^{k_{j-1}+1}\right) \le \overline{V} - \Delta_G + \sum_{\ell=0}^{j-1} c_\ell,$$

and part (a) adds at most $c_j$ to obtain the middle clause of $\mathcal{S}_{j+1}$. Together with $k_j < \infty$ this proves $\mathcal{S}_{j+1}$.

The induction is noncircular: at each step only the level-$j$ hypotheses are used, and Lemma I.6 and Lemma I.7 were both stated and proved conditionally on exactly those hypotheses. In particular the merit cap is never assumed at level $j+1$ in order to establish it.

Finally, $(\mathrm{I}_j)$ and $(\mathrm{II}_j)$ for all $j$ give Equation 182: the equality $Cx^k = d$ holds for $k = 1$ by Assumption 3.3 and for every $k \ge 2$ by Equation 77, the strict inequalities hold at $k = 1$ by Assumption 3.3 and $y^1 = Ax^1$, and for $k \ge 2$ by Lemma I.5(iii), and the stored multiplier at any $k$ of level $j$ is $\lambda^k = \lambda^{[j]}$, bounded by $(\mathrm{I}_j)$. Every level terminates by Lemma I.7(c), and its accepting iteration $k_j$ belongs to it, so $1 \le T_j < \infty$ for every $j$; hence $k_j = \sum_{\ell=0}^{j} T_\ell$ is finite and strictly increasing in $j$, and every checkpoint index $j$ is reached after finitely many iterations. □

### I.4 Checkpoint Certificate, Complexity, and Scaling

By Lemma I.8(b) the level hypotheses Equation 171 hold for every $j \geq 0$, so all results of this subsection are unconditional. Throughout, Assumption 4.2 is in force and the checkpoint certificate is the pre-promotion tuple Equation 154 at index $k_j + 1$.

**Lemma I.9** (Pointwise KKT certificate at a checkpoint)**.** *Suppose Assumption 4.2 holds. For every $j \geq 0$, the checkpoint certificate tuple Equation 154, namely $(x^{k_j+1}, y^{k_j+1}, \breve{y}^{k_j+1}, v^{k_j+1}, \iota^{k_j+1}, \zeta^{k_j+1})$, satisfies every exact side condition of the $\epsilon$-KKT system in Definition H.1, namely*

$$\begin{aligned} v_i^{k_j+1} \in \partial\varphi_i(\breve{y}_i^{k_j+1}) \ (i \in \mathcal{I}_o), \qquad \breve{y}_i^{k_j+1} = y_i^{k_j+1} \ (i \in \mathcal{I}_c), \\ Cx^{k_j+1} = d, \ f_i(y_i^{k_j+1}) > 0 \ (i \in \mathcal{I}_c), \ \iota_i^{k_j+1} \leq 0 \ (i \in \mathcal{I}_c), \end{aligned} \tag{183}$$

*together with the quantitative bounds*

$$\left\|Ax^{k_j+1} - \breve{y}^{k_j+1}\right\| \leq \frac{2Q_{\mathrm{cp}} + \eta C_\psi}{\rho_j}, \qquad \left\|\Xi^{k_j+1}\right\| \leq \frac{S_\Xi}{\rho_j}, \tag{184}$$

$$\sum_{i\in\mathcal{I}_c} \left|\iota_i^{k_j+1} f_i(y_i^{k_j+1})\right| \leq \frac{I_{\mathrm{cp}} S_\omega}{\rho_j}, \qquad \mathit{Crit}_{KKT}^{k_j+1} \leq \frac{C_{\mathrm{cert}}}{\rho_j}, \tag{185}$$

*consisting of the three component bounds on consensus, stationarity, and complementarity together with the aggregate residual bound, and the explicit near-active support band*

$$\iota_i^{k_j+1} \neq 0 \quad \Longrightarrow \quad 0 \leq f_i(y_i^{k_j+1}) \leq \frac{S_\omega}{\rho_j}, \tag{186}$$

*and the original-space inequality violation obeys*

$$\max_{i\in\mathcal{I}_c} [-f_i(A_i x^{k_j+1})]_+ \leq \frac{2M_{\max} Q_{\mathrm{cp}}}{\rho_j}. \tag{187}$$

*Proof. Side conditions.* The subgradient membership is the first line of Equation 43, while $\breve{y}_i^{k_j+1} = y_i^{k_j+1}$ for $i \in \mathcal{I}_c$ is the definition in Equation 42; both hold for the one-call outputs Equation 41 at any parameter values and is therefore schedule independent. The equality $Cx^{k_j+1} = d$ and the strict inequalities $f_i(y_i^{k_j+1}) > 0$ are Equation 182, and $\iota_i^{k_j+1} \leq 0$ was shown in Lemma I.6(c). The checkpoint certificate tuple's inequality multipliers are the raw wall forces, and the corresponding nonnegative multipliers of the original problem are $-\iota_i^{k_j+1}$. No normalization is applied: unlike the b-FJ certificate of the ordinary mode, the checkpoint tuple is the unnormalized nearby-point KKT certificate of Definition H.1, which is possible because Lemma I.6(c) bounds $\|\iota^{k_j+1}\|_1$ by the constant $I_{\mathrm{cp}}$.

*Consensus.* Since $\breve{y}_i^{k_j+1} = y_i^{k_j+1}$ for $i \in \mathcal{I}_c$, the second line of Equation 43 gives $\|y^{k_j+1} - \breve{y}^{k_j+1}\| \leq \mu_j C_\psi = \eta C_\psi / \rho_j$. Combining with the sharp consensus estimate Equation 176,

$$\left\|Ax^{k_j+1} - \breve{y}^{k_j+1}\right\| \leq \left\|Ax^{k_j+1} - y^{k_j+1}\right\| + \left\|y^{k_j+1} - \breve{y}^{k_j+1}\right\| \leq \frac{2Q_{\mathrm{cp}}}{\rho_j} + \frac{\eta C_\psi}{\rho_j}.$$

*Stationarity.* This is Equation 163 of Lemma I.2, applicable because the step-$k_j$ test is accepted.

*Complementarity, support, and original-space violation.* Lemma I.6(b) gives the band Equation 186, and Lemma I.6(c) gives $\|\iota^{k_j+1}\|_1 \leq I_{\mathrm{cp}}$. Substituting both into the complementarity estimate Equation 165 of Lemma I.3, whose hypotheses hold because the step-$k_j$ test is accepted and Equation 182 gives $f_i(y_i^{k_j+1}) > 0$ for every $i \in \mathcal{I}_c$, yields the complementarity bound. Moreover, $f_i$ is $M_i$-Lipschitz by Assumption 3.2, strict split feasibility gives $f_i(y_i^{k_j+1}) > 0$, and Equation 176 gives

$$[-f_i(A_i x^{k_j+1})]_+ \leq M_i \|A_i x^{k_j+1} - y_i^{k_j+1}\| \leq \frac{2M_{\max} Q_{\mathrm{cp}}}{\rho_j},$$

which proves Equation 187; the second step uses $M_i \leq M_{\max}$ together with the blockwise domination $\|A_i x^{k_j+1} - y_i^{k_j+1}\| \leq \|r^{k_j+1}\|$ of the stacked consensus residual.

*Aggregate.* By the definition of $\mathrm{Crit}_{\mathrm{KKT}}$ in Equation 106, the three component bounds add to the aggregate bound:

$$\mathrm{Crit}_{\mathrm{KKT}}^{k_j+1} \le \frac{2Q_{\mathrm{cp}} + \eta C_\psi}{\rho_j} + \frac{S_\Xi}{\rho_j} + \frac{I_{\mathrm{cp}} S_\omega}{\rho_j} = \frac{C_{\mathrm{cert}}}{\rho_j},$$

with $C_{\mathrm{cert}}$ exactly as in Equation 156. Every term is $\mathcal{O}(1/\rho_j)$; this is a consequence of the $\chi^2$ factor in the curvature promotion, which forces $\beta(L_j) = \beta_j = b_{\mathrm{cp}}/\rho_j$ and $\sqrt{2L_j G}L_i/\rho_j = \sqrt{2G\ell_{\mathrm{cp}}}L_i$ by Equation 66 of Lemma G.11, so that the wall band, the model curvature ratio, and the smoothing radius $\mu_j = \eta/\rho_j$ all scale together. The estimate is pointwise at a single checkpoint certificate index; no averaging over a window is used. □

**Corollary I.10** (Completed-checkpoint and anytime complexity)**.** *Define*

$$C_{\mathrm{any}} := \frac{C_{\mathrm{work}} + (\rho_{\mathrm{cp}}^0)^{-3}}{1 - \chi^{-3}}. \tag{188}$$

*Then, for every $j \ge 0$,*

$$K_j \le j + 1 + \frac{C_{\mathrm{work}}}{1 - \chi^{-3}} \rho_j^3 \le C_{\mathrm{any}} \rho_j^3. \tag{189}$$

*Let the integer $K \ge K_0$ be a stopping budget, where $K_0$ is the cumulative count of Equation 153 at $j = 0$, that is, the startup budget $T_0$ of checkpoint 0, and not the qualification threshold $K_{\mathrm{cq}}(p, \xi)$ of the ordinary mode. Let $J(K) := \max\{j \ge 0 : K_j \le K\}$ be the latest checkpoint, and let $\mathit{Crit}_{\mathit{KKT}\mathrm{out}}(K) := \mathit{Crit}_{\mathit{KKT}}^{k_{J(K)}+1}$ be the residual of the latest-checkpoint output. Then $J(K)$ is well defined and finite, and*

$$\mathit{Crit}_{\mathit{KKT}\mathrm{out}}(K) \le \frac{C_{\mathrm{cert}}}{\rho_{J(K)}} \le \chi C_{\mathrm{cert}} C_{\mathrm{any}}^{1/3} K^{-1/3} = \chi C_{\mathrm{cert}} \left( \frac{C_{\mathrm{work}} + (\rho_{\mathrm{cp}}^0)^{-3}}{1 - \chi^{-3}} \right)^{1/3} K^{-1/3}. \tag{190}$$

*For $\epsilon > 0$ let $j_\epsilon := \max\{0, \lceil \log_\chi(C_{\mathrm{cert}}/(\rho_{\mathrm{cp}}^0 \epsilon)) \rceil\}$. Then $\mathit{Crit}_{\mathit{KKT}}^{k_{j_\epsilon}+1} \le \epsilon$ and*

$$K_{\mathrm{KKT}}(\epsilon) := K_{j_\epsilon} \le j_\epsilon + 1 + \frac{C_{\mathrm{work}}}{1 - \chi^{-3}} (\rho_{\mathrm{cp}}^0)^3 \chi^{3j_\epsilon} = \mathcal{O}(\epsilon^{-3}) \qquad (\epsilon \downarrow 0). \tag{191}$$

*In the nontrivial regime $\epsilon < C_{\mathrm{cert}}/\rho_{\mathrm{cp}}^0$ the explicit form $K_{\mathrm{KKT}}(\epsilon) \le j_\epsilon + 1 + \frac{\chi^3 C_{\mathrm{work}}}{1-\chi^{-3}} (C_{\mathrm{cert}}/\epsilon)^3$ holds. No bound of any kind is claimed for $K < K_0$, since no certified checkpoint output exists before checkpoint 0 completes.*

*Proof. Cumulative bound.* Lemma I.7(c) holds at every level; relaxing its strict bound Equation 179 to $T_\ell \le 1 + C_{\mathrm{work}} \rho_\ell^3$ before summing gives $K_j = \sum_{\ell=0}^j T_\ell \le (j+1) + C_{\mathrm{work}} \sum_{\ell=0}^j \rho_\ell^3$, and the finite geometric sum in Equation 67 bounds the last sum by $\rho_j^3/(1 - \chi^{-3})$. For the second inequality of Equation 189 it suffices to show

$$j + 1 \le \frac{(\rho_{\mathrm{cp}}^0)^{-3}}{1 - \chi^{-3}} \rho_j^3 = \frac{\chi^{3j}}{1 - \chi^{-3}}.$$

Put $t := \chi^3 > 1$. The claim is $(j+1)(1 - t^{-1}) \le t^j$, equivalently $(j+1)(t-1) \le t^{j+1}$. This holds because $t^{j+1} - 1 = (t-1) \sum_{i=0}^j t^i \ge (t-1)(j+1)$, each of the $j+1$ terms being at least 1.

*Anytime bound.* The set $\{j \ge 0 : K_j \le K\}$ contains 0 by $K \ge K_0$ and is finite because every level executes at least one iteration, $T_\ell \ge 1$, so that $K_j = \sum_{\ell=0}^j T_\ell \ge j + 1 \to \infty$; hence $J := J(K)$ is well defined. By maximality $K < K_{J+1}$, and Equation 189 at level $J+1$ with $\rho_{J+1} = \chi \rho_J$ gives $K < C_{\mathrm{any}} \chi^3 \rho_J^3$, that is, $1/\rho_J < \chi (C_{\mathrm{any}}/K)^{1/3}$. Combining with Lemma I.9 at level $J$ proves Equation 190. The returned object is a checkpoint, so it carries the full certificate Equation 183–Equation 187, not only the residual value.

*Complexity.* If $\log_\chi(C_{\mathrm{cert}}/(\rho_{\mathrm{cp}}^0 \epsilon)) \le 0$ then $j_\epsilon = 0$ and $\rho_{j_\epsilon} = \rho_{\mathrm{cp}}^0 \ge C_{\mathrm{cert}}/\epsilon$; otherwise $\chi^{j_\epsilon} \ge C_{\mathrm{cert}}/(\rho_{\mathrm{cp}}^0 \epsilon)$ and again $\rho_{j_\epsilon} = \rho_{\mathrm{cp}}^0 \chi^{j_\epsilon} \ge C_{\mathrm{cert}}/\epsilon$. Lemma I.9 then gives $\mathrm{Crit}_{\mathrm{KKT}}^{k_{j_\epsilon}+1} \le C_{\mathrm{cert}}/\rho_{j_\epsilon} \le \epsilon$, and the first inequality of Equation 189 at $j = j_\epsilon$ gives the displayed bound. If $\epsilon < C_{\mathrm{cert}}/\rho_{\mathrm{cp}}^0$ then $j_\epsilon \ge 1$ and the ceiling overshoots by less than one, so $\rho_{j_\epsilon} \le \chi C_{\mathrm{cert}}/\epsilon$, giving the explicit form. Finally, in

that same nontrivial regime $\epsilon < C_{\mathrm{cert}}/\rho^0_{\mathrm{cp}}$, which is the only one that matters as $\epsilon \downarrow 0$, one has $j_\epsilon \le 1 + \log_\chi\left(C_{\mathrm{cert}}/(\rho^0_{\mathrm{cp}}\epsilon)\right) = \mathcal{O}(\log(1/\epsilon))$ and $\rho^3_{j_\epsilon} = \mathcal{O}(\epsilon^{-3})$, so $K_{\mathrm{KKT}}(\epsilon) = \mathcal{O}(\epsilon^{-3})$. These last two statements are asymptotic in $\epsilon$ alone: the logarithm carries the multiplicative constant $1/\log\chi$ and the additive offset $\log_\chi(C_{\mathrm{cert}}/\rho^0_{\mathrm{cp}})$, and the coefficient of the cubic term is $\chi^3 C_{\mathrm{work}} C^3_{\mathrm{cert}}/(1-\chi^{-3})$; a valid constant for the full $\mathcal{O}(\epsilon^{-3})$ bound also absorbs the lower-order logarithmic term. All of these depend on the problem data through $\chi$, $C_{\mathrm{cert}}$, $C_{\mathrm{work}}$, and $\rho^0_{\mathrm{cp}}$, which are held fixed as $\epsilon \downarrow 0$; no uniformity over a family is claimed here. □

**Same-$x$ KKT residual in the smooth case.** As in the ordinary mode, the checkpoint certificate of Lemma I.9 is a nearby-point certificate: the constraint gradients use the split points $y_i$, and the objective forces use the proximal points $\breve{y}_i = p_i$. When $\psi_i \equiv 0$, the Moreau force vanishes ($w_i = 0$), so $\breve{y}_i = y_i$ for every block and each $\varphi_i = h_i$ is smooth. Replacing every $y_i$ by $A_i x$ then changes the stationarity vector by at most $\left(\|A\|_2 \sum_{i\in\mathcal{I}_\circ} H_i + \|\iota\|_1 \|A\|_2 \max_{i\in\mathcal{I}_c} L_i\right)\|Ax-y\|$, the feasibility slack by at most $\max_{i\in\mathcal{I}_c} M_i \,\|Ax-y\|$, and each complementarity term $|\iota_i f_i(A_i x)|$ by at most $|\iota_i| M_i \|Ax-y\|$, using the Lipschitz gradients $H_i, L_i$ and the bounded constraint gradients $M_i$. At an accepted checkpoint, Lemma I.6(c) bounds $\|\iota^{k_j+1}\|_1 \le I_{\mathrm{cp}}$ and Lemma I.9 gives $\|Ax^{k_j+1} - \breve{y}^{k_j+1}\| = \mathcal{O}(\rho_j^{-1})$. Hence the conventional same-$x$ KKT residual, with all objective and constraint terms evaluated at $A_i x$, is $\mathcal{O}(\rho_j^{-1})$, the same order as the nearby-point certificate; combined with Corollary I.10 it obeys the same latest-checkpoint rate $\mathcal{O}(K^{-1/3})$ as Theorem 4.4. This is the standard transfer from a near-point to a same-point stationarity certificate Rockafellar & Wets (2009).

**Lemma I.11** (Block scaling of the checkpoint constants)**.** *Suppose Assumption 4.1 holds, $\chi$, $\ell_{\mathrm{cp}}$, and $\tau_{\mathrm{cp}}$ are fixed, and the family satisfies $\inf_{\mathcal{F}} \gamma_{\mathrm{cq}} > 0$ and $\inf_{\mathcal{F}} \delta_{\mathrm{cq}} > 0$. Then*

$$S_0,\ \widehat{G}_0,\ L^{0,\mathrm{cp}},\ G^{\mathrm{cp}},\ \Delta_G,\ \overline{V},\ \Delta_* = \Theta(N), \qquad G_\psi = \mathcal{O}(N),$$
$$C_{\mathrm{work}} = \Theta(N), \qquad a_0 = \Theta(1), \tag{192}$$
$$C_\psi,\ J_H,\ S_\omega = \mathcal{O}(\sqrt{N}), \qquad b_{\mathrm{cp}} = \Theta(\sqrt{N}),$$
$$r_{\mathrm{cp}},\ c_{\chi,\mathrm{cp}},\ c_{\mathrm{mov}} = \mathcal{O}(\sqrt{N}), \tag{193}$$
$$M_F = \Theta(\|A\|_2\sqrt{N}), \qquad S_\Xi = \mathcal{O}(\|A\|_2\sqrt{N}),$$
$$I_{\mathrm{cp}} = \mathcal{O}(D_{\mathrm{cq}}), \tag{194}$$
$$Q_{\mathrm{cp}} = \mathcal{O}(\sqrt{N} + D_{\mathrm{cq}}), \qquad D_{\mathrm{cp}} = \mathcal{O}(N + D^2_{\mathrm{cq}}),$$
$$C_{\mathrm{cert}} = \mathcal{O}\left(\sqrt{N}(D_{\mathrm{cq}} + \|A\|_2 + 1)\right). \tag{195}$$

*with $D_{\mathrm{cq}} = 2(1 + \|A\|_2\sqrt{\widehat{G}_0}/\gamma_{\mathrm{cq}}) = \Theta(1 + \|A\|_2\sqrt{N}/\gamma_{\mathrm{cq}})$. Moreover*

$$\rho^0_{\mathrm{cp}} = \Theta\left(N + \frac{\|A\|_2^2}{\gamma^2_{\mathrm{cq}}}\right),$$
$$C_{\mathrm{any}} = \Theta(N), \quad K_0 = \mathcal{O}\left(N\left(N + \frac{\|A\|_2^2}{\gamma^2_{\mathrm{cq}}}\right)^3\right). \tag{196}$$

$$\mathit{Crit}_{\mathit{KKT}^{\mathrm{out}}}(K) = \mathcal{O}\left((D_{\mathrm{cq}} + \|A\|_2 + 1)N^{5/6}K^{-1/3}\right) \qquad (K \ge K_0), \tag{197}$$

*and, writing $\mathcal{N} := N + M \ge N$ for the total block count, replacing each $\Theta(N)$ aggregate by the conservative total-block envelope $\mathcal{N}$ replaces $N^{5/6}$ by $\mathcal{N}^{5/6}$ in Equation 197. In the target family, where the assumed lower bound $\inf_{\mathcal{F}} \gamma_{\mathrm{cq}} > 0$ already gives $\gamma_{\mathrm{cq}} = \Omega(1)$,*

$$C_{\mathrm{cert}} = \mathcal{O}(\|A\|_2 N), \quad \mathit{Crit}_{\mathit{KKT}^{\mathrm{out}}}(K) = \mathcal{O}\left(\|A\|_2 N^{4/3}K^{-1/3}\right),$$
$$K_{\mathrm{KKT}}(\epsilon) = \mathcal{O}\left(\|A\|_2^3 N^4 \epsilon^{-3}\right) \quad (\epsilon \downarrow 0 \textit{ for each fixed instance}). \tag{198}$$

*Finally, one outer iteration performs exactly $N$ objective proximal updates and $M$ closed-form constraint updates, followed by one $x$ solve, so the serial block-update count is $W_{\mathrm{serial}} = \mathcal{N}K$ and*

$$W_{\mathrm{serial,KKT}}(\epsilon) = \mathcal{O}\left((N+M)\|A\|_2^3 N^4 \epsilon^{-3}\right) \quad (\epsilon \downarrow 0 \textit{ for each fixed instance}). \tag{199}$$

*Proof.* All statements follow by substituting Assumption 4.1 into the exact formulas Equation 29 and Equation 156; no new estimate is required.

*Equation 192.* $S_0 = \sum_{i\in\mathcal{I}_\circ} \varphi_i(A_i x^1)$ is a sum of $N$ terms, each bounded above and below away from zero, so $S_0 = \Theta(N)$, and likewise $\widehat{G}_0 = \sum_{i\in\mathcal{I}_\circ} \widehat{C}_i^2 = \Theta(N)$ and $G_\psi = \sum_{i\in\mathcal{I}_\circ} (C_i^\psi)^2 = \mathcal{O}(N)$. Since $m = \Theta(1)$, $L^{0,\mathrm{cp}} = \max\{1, 1/m, 3S_0/m^2\} = \Theta(N)$ and $G^{\mathrm{cp}} = m^2 L^{0,\mathrm{cp}}/2 = \Theta(N)$. The objective branch gives $G^{\mathrm{cp}} \ge 3S_0/2$, hence $S_0 \le 2G^{\mathrm{cp}}/3$ and

$$\frac{G^{\mathrm{cp}}}{12} \le \Delta_G = \frac{G^{\mathrm{cp}} - S_0}{4} \le \frac{G^{\mathrm{cp}}}{4}.$$

Thus $\Delta_G = \Theta(N)$, whence $\overline{V} = S_0 + 2\Delta_G = \Theta(N)$ and $\Delta_* = \overline{V} + \Delta_G = \Theta(N)$. With $\tau_{\mathrm{cp}}$ fixed and $a_0 = (2\theta_1^{\mathrm{cp}} - 1)/2 = (1 + 2\theta_2^{\mathrm{cp}})/2 = 51/100$, $C_{\mathrm{work}} = \Delta_*/(a_0 \tau_{\mathrm{cp}}^2) = \Theta(N)$.

*Equation 193.* $C_\psi = \sqrt{G_\psi} = \mathcal{O}(\sqrt{N})$ and $J_H = (\sum_{i\in\mathcal{I}_\circ} (H_i C_i^\psi)^2)^{1/2} \le H_{\max} C_\psi = \mathcal{O}(\sqrt{N})$. Since $\ell_{\mathrm{cp}}$ is fixed and $G^{\mathrm{cp}} = \Theta(N)$, $b_{\mathrm{cp}} = \sqrt{2G^{\mathrm{cp}}/\ell_{\mathrm{cp}}} = \Theta(\sqrt{N})$ and $r_{\mathrm{cp}} = \max\{H_{\max}, \sqrt{2G^{\mathrm{cp}}\ell_{\mathrm{cp}}} L_{\max}\} = \mathcal{O}(\sqrt{N})$. Hence $c_{\chi,\mathrm{cp}} = \theta_1(1 + r_{\mathrm{cp}}) - 1$ and $c_{\mathrm{mov}}$ are $\mathcal{O}(\sqrt{N})$. As $\rho_{\mathrm{cp}}^0 \ge 1$, $S_\omega \le b_{\mathrm{cp}} + L_{\max}\tau_{\mathrm{cp}}^2/2$, and the right-hand side is $\mathcal{O}(\sqrt{N})$; hence $S_\omega = \mathcal{O}(\sqrt{N})$.

*Equation 194.* $M_F = \|A\|_2 \sqrt{\widehat{G}_0} = \Theta(\|A\|_2 \sqrt{N})$; note that the sparse-support and two-sided $\ell_1$ clauses of Assumption 4.1 imply the per-block bounds in Lemma H.15, but do not bound the stacked $\|A\|_2$ by a family-uniform constant, so its growth is carried explicitly. Then $S_\Xi = \|A\|_2(c_{\mathrm{mov}}\tau_{\mathrm{cp}} + \eta J_H) = \mathcal{O}(\|A\|_2\sqrt{N})$ and

$$I_{\mathrm{cp}} = \frac{M_F + S_\Xi}{\gamma_{\mathrm{cq}}} = \mathcal{O}\left(\frac{\|A\|_2\sqrt{N}}{\gamma_{\mathrm{cq}}}\right) = \mathcal{O}(D_{\mathrm{cq}}),$$

the last step by the stated form of $D_{\mathrm{cq}}$ and $\widehat{G}_0 = \Theta(N)$.

*Equation 195.* $Q_{\mathrm{cp}} = \sqrt{\widehat{G}_0} + M_{\max} I_{\mathrm{cp}} + (c_{\chi,\mathrm{cp}} + 1)\tau_{\mathrm{cp}} = \mathcal{O}(\sqrt{N} + D_{\mathrm{cq}})$ since $M_{\max} = \mathcal{O}(1)$ and $c_{\chi,\mathrm{cp}} = \mathcal{O}(\sqrt{N})$; consequently $D_{\mathrm{cp}} = (Q_{\mathrm{cp}} + \sqrt{\widehat{G}_0})^2/2 = \mathcal{O}(N + D_{\mathrm{cq}}^2)$. The matching lower half, used in the base-penalty step below, follows by discarding the last, nonnegative summand of $Q_{\mathrm{cp}}$: Assumption 4.1 bounds each $\widehat{C}_i$ and each $M_i$ away from zero, so $\sqrt{\widehat{G}_0} = \Theta(\sqrt{N})$ and $M_{\max} = \Theta(1)$, while $S_\Xi \ge 0$ gives $I_{\mathrm{cp}} \ge M_F/\gamma_{\mathrm{cq}} = \Theta(\|A\|_2\sqrt{N}/\gamma_{\mathrm{cq}})$. Hence

$$Q_{\mathrm{cp}} = \Theta\left(\sqrt{N} + \frac{\|A\|_2\sqrt{N}}{\gamma_{\mathrm{cq}}}\right),$$

which is the two-sided form of the upper bound recorded in Equation 195, since $D_{\mathrm{cq}} = \Theta(1 + \|A\|_2\sqrt{N}/\gamma_{\mathrm{cq}})$. For the certificate constant,

$$\begin{aligned} C_{\mathrm{cert}} &= 2Q_{\mathrm{cp}} + \eta C_\psi + S_\Xi + I_{\mathrm{cp}} S_\omega \\ &= \mathcal{O}\left(\sqrt{N} + D_{\mathrm{cq}} + \|A\|_2\sqrt{N} + D_{\mathrm{cq}}\sqrt{N}\right) \\ &= \mathcal{O}\left(\sqrt{N}(D_{\mathrm{cq}} + \|A\|_2 + 1)\right). \end{aligned}$$

*Base penalty, startup, and $C_{\mathrm{any}}$.* Analyze the nine displayed branches of $\rho_{\mathrm{cp}}^0$ directly. The floor is constant; $\eta^{\mathrm{cp}} G_\psi/(2\Delta_G) = \mathcal{O}(1)$; $D_{\mathrm{cp}}/\Delta_G = \mathcal{O}(1 + \|A\|_2^2/\gamma_{\mathrm{cq}}^2)$; the two $b_{\mathrm{cp}}$ branches are $\mathcal{O}(\sqrt{N})$; and $8\overline{V}/\delta_{\mathrm{cq}}^2 = \Theta(N)$ because the family lower bound on $\delta_{\mathrm{cq}}$ is positive and the normalized choice after Assumption 4.2 gives $\delta_{\mathrm{cq}} \le 1$. The transition and coarse-linear branches are bounded by $\mathcal{O}(N + \|A\|_2^2/\gamma_{\mathrm{cq}}^2)$ using the two-sided scaling of $Q_{\mathrm{cp}}$ above and Young's inequality, while the fourth-root branch is $\mathcal{O}(1)$. Conversely, the coarse square-root branch supplies the $\Omega(N)$ part and the merit branch supplies the qualification-dependent part. Hence

$$\rho_{\mathrm{cp}}^0 = \Theta\left(N + \frac{\|A\|_2^2}{\gamma_{\mathrm{cq}}^2}\right).$$

No ordinary drift clause is present: $\bar{M} = \Theta(M)$ and $H_p(\xi)$ occur only in the ordinary construction, and no aggregate constraint-side factor occurs outside the stacked norm $\|A\|_2$. Since $\rho^0_{\mathrm{cp}} \geq 1$, Equation 188 gives $C_{\mathrm{any}} \leq (C_{\mathrm{work}}+1)/(1-\chi^{-3}) = \mathcal{O}(N)$, and $C_{\mathrm{any}} \geq C_{\mathrm{work}} = \Omega(N)$, so $C_{\mathrm{any}} = \Theta(N)$. Finally $K_0 \leq 1 + C_{\mathrm{work}}(\rho^0_{\mathrm{cp}})^3$ by Equation 179, which proves the displayed startup bound.

*Rates.* By Equation 190, the anytime constant is $\chi C_{\mathrm{cert}} C_{\mathrm{any}}^{1/3} = \mathcal{O}(\sqrt{N}(D_{\mathrm{cq}} + \|A\|_2 + 1) \cdot N^{1/3})$, which is Equation 197. In the target family $\gamma_{\mathrm{cq}} = \Omega(1)$ gives $D_{\mathrm{cq}} = \mathcal{O}(1 + \|A\|_2\sqrt{N})$, and Lemma H.15 gives $\|A\|_2 = \Omega(1)$, so $D_{\mathrm{cq}} + \|A\|_2 + 1 = \mathcal{O}(\|A\|_2\sqrt{N})$; hence $C_{\mathrm{cert}} = \mathcal{O}(\|A\|_2 N)$ and the anytime constant is $\mathcal{O}(\|A\|_2 N \cdot N^{1/3}) = \mathcal{O}(\|A\|_2 N^{4/3})$. For the $\epsilon$ bound, use the explicit nontrivial-regime form of Corollary I.10:

$$K_{\mathrm{KKT}}(\epsilon) \leq j_\epsilon + 1 + \frac{\chi^3 C_{\mathrm{work}}}{1-\chi^{-3}} \left(\frac{C_{\mathrm{cert}}}{\epsilon}\right)^3 = \mathcal{O}\left(\log \frac{1}{\epsilon}\right) + \mathcal{O}\left(N \cdot \|A\|_2^3 N^3 \epsilon^{-3}\right) = \mathcal{O}\left(\|A\|_2^3 N^4 \epsilon^{-3}\right).$$

*Absence of a separate $M$ factor and the serial count.* Every aggregate appearing in a checkpoint constant, namely $S_0$, $\widehat{G}_0$, $G_\psi$, and $J_H$, is a sum over $\mathcal{I}_{\mathrm{o}}$ only; the constraint blocks enter solely through the per-block maxima $M_{\max}$, $L_{\max}$, which are $\mathcal{O}(1)$, and through the stacked norm $\|A\|_2$, which is carried explicitly and ranges over all $N + M$ blocks. Hence, conditional on the family-uniform qualification margins $\inf_{\mathcal{F}} \gamma_{\mathrm{cq}} > 0$ and $\inf_{\mathcal{F}} \delta_{\mathrm{cq}} > 0$, no additional factor $\bar{M} = \Theta(M)$ or $\sqrt{M}$ appears outside $\|A\|_2$ and the CQ constants in the certificate constant or post-checkpoint rate; the startup bound retains the displayed dependence on $\rho^0_{\mathrm{cp}}$ and therefore on $\|A\|_2$; this differs from the ordinary IP-ADMM family rates of Theorem 4.2, which carry $\sqrt{M}$ through the drift clause of $\rho^0$. This absence is conditional on fixed positive margins: $\gamma_{\mathrm{cq}}$ and $\delta_{\mathrm{cq}}$ are difficult to compute, and they may shrink as more constraints become simultaneously near-active, a count governed by $M$; how they scale with family size is not known in general, so the absence of a separate $M$ factor should be read relative to family-uniform margins rather than as unconditional independence of $M$. Bounding each objective aggregate by the total-block envelope $\mathcal{N} = N + M \geq N$ is therefore conservative and replaces $N^{5/6}$ by $\mathcal{N}^{5/6}$ in Equation 197, which is the envelope claim of the statement. Finally, the inner loop of Algorithm 1 runs $i = 1, \ldots, N + M$, so one outer iteration performs $N$ objective block updates, each a single proximal call by Equation 41, and $M$ constraint block updates, each a closed-form evaluation by Equation 54, plus one equality-constrained quadratic solve. Multiplying the outer count by $\mathcal{N}$ gives $W_{\mathrm{serial}} = \mathcal{N}K$ and Equation 199. □

**Corollary I.12** (Raw stored and accepted multiplier scaling)**.** *Suppose Assumption 4.2 holds. For every $k \geq 1$ the stored multiplier satisfies $\|\lambda^k\| \leq Q_{\mathrm{cp}}$, and for every $j \geq 0$ the accepted virtual multiplier satisfies $\|\widehat{\lambda}^{k_j+1}\| \leq Q_{\mathrm{cp}}$. Define $j(k) := \#\{j \geq 0 : k_j < k\}$, the number of checkpoints committed before iteration $k$. Then $\rho^k = \rho_{j(k)}$ and*

$$\frac{\|\lambda^k\|}{\rho^k} \leq \frac{Q_{\mathrm{cp}}}{\rho^0_{\mathrm{cp}}} \chi^{-j(k)} \xrightarrow[j(k)\to\infty]{} 0. \tag{200}$$

*Under the hypotheses of Lemma I.11,*

$$\|\lambda^k\| = \mathcal{O}\left(\sqrt{N} + D_{\mathrm{cq}}\right), \qquad \frac{\|\lambda^k\|}{\rho^k} = \mathcal{O}\left(\frac{(\sqrt{N} + D_{\mathrm{cq}})\chi^{-j(k)}}{N}\right). \tag{201}$$

*In the target family the normalized bound is $\mathcal{O}(\|A\|_2 N^{-1/2} \chi^{-j(k)})$. No bound is asserted for a rejected virtual multiplier.*

*Proof.* The stored bound is Equation 182 and the accepted bound is Equation 175, both unconditional by Lemma I.8(b). With the convention $k_{-1} = 0$, one has $k_{j(k)-1} < k \leq k_{j(k)}$, that is, $k$ lies in level $j(k)$; then $\rho^k = \rho_{j(k)} = \rho^0_{\mathrm{cp}} \chi^{j(k)}$ by Equation 30, with $\chi$ the scalar promotion factor, and $\lambda^k = \lambda^{[j(k)]}$, which gives Equation 200. The scaling statements follow from $Q_{\mathrm{cp}} = \mathcal{O}(\sqrt{N} + D_{\mathrm{cq}})$ and $\rho^0_{\mathrm{cp}} = \Omega(N)$ in Lemma I.11; in the target family $D_{\mathrm{cq}} = \mathcal{O}(1 + \|A\|_2\sqrt{N})$ and Lemma H.15 gives $\|A\|_2 = \Omega(1)$, so $(\sqrt{N} + D_{\mathrm{cq}})/N = \mathcal{O}(\|A\|_2/\sqrt{N})$.

A rejected virtual multiplier is formed from the raw movement through Equation 158, whose terms $\Upsilon^k d^k$, with $\Upsilon^k$ the block-diagonal curvature matrix of iteration $k$ rather than the scalar promotion

factor $\chi$, and $\rho_j A\Delta x^k$ are controlled only when the movement test passes; it is discarded immediately and never enters a subproblem, so no bound is needed or claimed. □

**Comparison with the ordinary mode.** Under the checkpoint mode's additional Uniform Band-MFCQ and initialization assumptions, the bound of Corollary I.12 is stronger in form than its IP-mode counterpart. Lemma H.9 and Corollary H.20 prove only $\sup_j \|\lambda^j\|^2/\rho^j < \infty$, that is, $\|\lambda^j\| = \mathcal{O}(\sqrt{\rho^j})$, which permits the multiplier itself to grow without bound; the decay $\lambda^j/\rho^j \to 0$ there is obtained from the summable global movement budget and requires $p > 0$. In Ckpt mode the multiplier is *uniformly bounded* by the single constant $Q_{\mathrm{cp}}$, and the normalized multiplier decays geometrically in the checkpoint index at the explicit rate $\chi^{-j(k)}$, with no schedule exponent and no movement budget. The mechanism is different: the bound comes from the raw Uniform Band-MFCQ estimate Equation 174 applied at the accepted step, not from averaging.

**Proof of Theorem 4.4.** By Lemma I.7(c) and Lemma I.8(b), every checkpoint level terminates after finitely many outer iterations and $1 \le T_j < 1 + C_{\mathrm{work}}\rho_j^3$ for every $j \ge 0$. In particular, $K_0 = T_0 = k_0 < \infty$, matching the startup budget in Corollary I.10. The exact anytime estimate Equation 190 gives the $K^{-1/3}$ residual decay for the latest checkpoint whenever $K \ge K_0$. Under the stated block-scaling and family-uniform qualification assumptions, Equation 197 gives the claimed block-dependent rate, and its target-family specialization is exactly Equation 198. These estimates prove the shortened headline theorem. All other exact certificate, feasibility, multiplier, startup, and work-complexity results remain appendix details.