WHEN BATCH SCHEDULES CANNOT BE CERTIFIED: CAPACITY OF GLOBALLY VALID MONOMIAL CERTIFICATES ON INTEGER LOG-HULLS

Anonymous authors
Paper under double-blind review

ABSTRACT

A discrete training schedule can be nearly optimal and still be impossible to certify with one reusable global witness. For $J(m_0, m_1) = m_0^{-1} + m_1^{-1}$ and budget 3, the integer optimum is $3/2$, whereas every globally valid single monomial certifies at most $\sqrt{2}$. We call the best reusable proof endpoint *single-monomial certificate capacity*; throughout, this term refers only to a monomial lower bound valid on the entire positive orthant, not to arbitrary certificates or true model risk. The capacity gap is the log-ratio to the best integer schedule. For positive reciprocal risk interfaces we show that the strongest reusable monomial is obtained by minimizing $\phi(z) = \log J(e^z)$ over the convex hull of logarithmic integer schedules. This gives an exact plan-level obstruction, a finite-support bound in terms of covering radius, and a contractive-SGD realization on a strongly convex quadratic loss of the predicted gap, which is $N\gamma\theta(1-\theta)/(2b^3)$ to first order for the log-ratio. Support expansion is therefore a measurable proof resource: it admits a covering-radius upper bound on the remaining finite-domain certificate gap. A product frontier and a controlled noncommuting moment-family audit reproduce the predicted support obstruction without relying on a particular optimizer or data set.

1 INTRODUCTION

A finite training budget fixes the number of stochastic gradient evaluations, but its allocation changes how noise propagates. The resulting optimization problem has two different optima: the best feasible schedule and the strongest lower certificate that can be reused across all schedules. The distinction is not a numerical artifact. With $J(m_0, m_1) = m_0^{-1} + m_1^{-1}$ and $m_0 + m_1 \le 3$, the integer optimum is $3/2$, while the best globally valid single monomial has value $\sqrt{2}$. No improvement to the candidate can remove the remaining 6.066% proof gap; a richer certificate family is required.

This paper asks one question: *how much can one reusable global certificate prove on an integer schedule domain?* We answer it for the positive reciprocal interfaces produced by fixed stochastic dynamics. The answer is a geometric capacity, determined by the log integer hull rather than by the usual continuous budget relaxation. The capacity is exact, gives a decision-level impossibility test, and predicts how many supporting tangents are needed to close the gap.

The capacity thesis. Our main theorem characterizes reusable certificate capacity. We derive three consequences in one chain:

1. **Exact capacity.** If D is the integer schedule domain, $P = \mathrm{conv}(\log D)$, and $\phi(z) = \log J(e^z)$, then the best globally valid monomial has endpoint $L_{\mathrm{fam}} = \exp(\min_{z \in P} \phi(z))$.
2. **Capacity obstruction.** Writing $a(m) = J(m)/J^*_{\mathrm{int}}$ and $c = J^*_{\mathrm{int}}/L_{\mathrm{fam}}$, a plan is certifiable at tolerance ε exactly when $a(m)c \le 1 + \varepsilon$.
3. **Support-resource bound.** A K-support envelope admits a covering-radius upper bound on the remaining finite-domain gap according to projected covering radius. We then give a deliberately constructed contractive-SGD realization on a strongly convex quadratic loss, showing that this obstruction occurs in a stochastic learning process rather than only in an abstract polynomial.

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.

Farthest-point traversal, integer recovery, interval arithmetic, and the nonlinear and Digits audits are supporting tools for these statements; they are not additional scientific claims.

Why this matters. Reusable certificates arise when a scheduler, verifier, or deployment pipeline must compare many integer allocations without rebuilding a proof for each one. The theorem identifies when more search is futile, when more supports are the right resource, and when a new certificate family is necessary. The result is broader than a single allocation instance because it applies to any fixed positive reciprocal interface, including noncommuting moment dynamics.

Scope. We assume a precommitted fixed process, sampling with replacement, unit component costs, and globally valid monomial lower bounds. The main theorem is about an exact positive reciprocal interface. Nonlinear upper-model transfer, weighted costs, and finite-precision recovery are reported as limitation and reproducibility material in Appendix D. The experiments are synthetic audits of the mathematical claims; the SGD construction is an existence witness, not a benchmark model or training-policy evaluation.

Relation to prior work. Batch scaling and continuous scheduling study useful allocation structures (Smith et al., 2018; Perko, 2023; Wang et al., 2026; Meterez et al., 2026); operator analyses and exact moment recurrences provide stochastic risk interfaces (Jain et al., 2018; Bordelon & Pehlevan, 2022; Mulayoff & Michaeli, 2024); geometric programming supplies log convexity (Boyd et al., 2007; Agrawal et al., 2019). Our additional object is the attainable lower endpoint of a *reusable family* on the integer log hull. Its exactness and obstruction are different from either continuous allocation or a finite-domain oracle.

The convex ingredients are standard. Geometric programming supplies continuous log-convex relaxations and pointwise tangents; our theorem instead optimizes the minimum of one monomial valid on all of $\mathbb{R}^N_{++}$ over a changing finite integer domain. The resulting max–min identity, exact obstruction c, and integer-anchor support bound are the added reusable discrete capacity statements.

2 Exact reciprocal interface and integer geometry

Fix N updates, lower limits $\ell_i \geq 1$, and a unit-cost budget B. The feasible schedules and their optimum are

$$D = \{m \in \mathbb{Z}^N : m_i \geq \ell_i,\ \sum_i m_i \leq B\},\ J^*_{\mathrm{int}} = \min_{m \in D} J(m). \tag{2.1}$$

The risk interface is a nonzero nonnegative reciprocal polynomial

$$J(m) = \sum_{S \subseteq [N]} c_S \prod_{i \in S} m_i^{-1},\ c_S \geq 0, \quad J(m) > 0. \tag{2.2}$$

Terms with $|S| \geq 2$ encode interactions between noise injected at different updates. They are exactly the terms lost by a first-degree allocation model.

For affine stochastic processes, positive moment propagation gives $M_{k+1} = \mathcal{A}_k(M_k) + [\mathcal{G}_k(M_k) + W_k]/m_k$ and $J = \langle Q_{\mathrm{obs}}, M_N \rangle$. Independence within and across batches and sampling with replacement produce the reciprocal factors in (2.2); a forward pass and an adjoint compute J and its gradient without expanding all 2^N coefficients.

Write

$$P = \mathrm{conv}\{\log m : m \in D\},\ \phi(z) = \log J(e^z). \tag{2.3}$$

For unit costs, let $u_i = \ell_i + B - \sum_j \ell_j$ and let p_i interpolate $p_i(\log k) = k$ between consecutive integers. Then

$$P = \{z : \log \ell_i \leq z_i \leq \log u_i,\ \sum_i p_i(z_i) \leq B\}. \tag{2.4}$$

The slopes of each p_i increase, so this set is convex. The unit-cost assumption is used only for this explicit representation; the abstract capacity identity remains valid for any finite integer domain.

Table 1: Three relaxation levels for the same reciprocal interface.

object	optimization	reusable?	role
continuous allocation	$\min_{\sum_i e^{z_i} \le B} \phi(z)$	no	classical relaxation
reusable monomial	$\sup_{H \le J \text{ on } \mathbb{R}^N_{++}} \min_{m \in D} H(m)$	yes	capacity L_{fam}
finite-domain oracle	$\sup_{H \le J \text{ on } D} \min_{m \in D} H(m)$	current D	equals J^*_{int}

3 THE CAPACITY OF ONE REUSABLE CERTIFICATE

Define the monomial family

$$\mathcal{M} = \left\{ m \mapsto a \prod_i m_i^{\beta_i} : a > 0,\ \beta \in \mathbb{R}^N \right\}.$$

Certificate capacity is the best reusable proof endpoint, rather than the best schedule value:

$$L_{\mathrm{fam}} = \sup_{H \in \mathcal{H}_{\mathrm{reuse}}(J)} \min_{m \in D} H(m), \quad \mathcal{H}_{\mathrm{reuse}}(J) = \{H \in \mathcal{M} : H \le J \text{ on } \mathbb{R}^N_{++}\}. \tag{3.1}$$

Operational semantics. The finite-domain family is $\mathcal{H}_D(J) = \{H \in \mathcal{M} : H(m) \le J(m)\ \forall m \in D\}$. Thus $L_{\mathrm{fam}} = \sup_{H \in \mathcal{H}_{\mathrm{reuse}}} \min_D H$ and $J^*_{\mathrm{int}} = \sup_{H \in \mathcal{H}_D} \min_D H$; the constant monomial $H \equiv J^*_{\mathrm{int}}$ attains the second quantity.

Theorem 3.1 (Exact reusable capacity)**.** *For the interface in* (2.2)*,*

$$L_{\mathrm{fam}} = \exp\left(\min_{z \in P} \phi(z) \right). \tag{3.2}$$

The continuous problem is a relaxation, so its optimum is a lower bound; the reusable family is optimized over the smaller set e^P, and its endpoint L_{fam} can therefore be strictly larger. In all cases $L_{\mathrm{fam}} \le J^*_{\mathrm{int}}$; either inequality can be strict.

Proof. For a valid monomial, $h(z) = \log H(e^z)$ is affine and $h \le \phi$ on $\mathbb{R}^N$. Therefore $\min_{\log D} h \le h(z_*) \le \phi(z_*)$ for a minimizer z_* of ϕ on P. Convexity of ϕ makes the tangent $h_*(z) = \phi(z_*) + \nabla\phi(z_*)^\top (z - z_*)$ globally valid. First-order optimality on P gives $\min_{x \in X} h_*(x) = h_*(z_*) = \phi(z_*)$. Convexity gives $h_*(x) \le \phi(x)$ for all $x \in \mathbb{R}^N$. Thus its exponential attains the upper bound. Finally, if $z = \sum_j \alpha_j \log m^{(j)}$ with $\alpha_j \ge 0$ and $\sum_j \alpha_j = 1$, then $e^{z_i} \le \sum_j \alpha_j m_i^{(j)}$ and hence $\sum_i e^{z_i} \le \sum_j \alpha_j \sum_i m_i^{(j)} \le B$; this is the Jensen direction proving the continuous relaxation ordering. □

For a candidate m, define $a(m) = J(m)/J^*_{\mathrm{int}}$ and $c = J^*_{\mathrm{int}}/L_{\mathrm{fam}}$. The exact decision-level consequence is

$$\boxed{\exists H \in \mathcal{H}_{\mathrm{reuse}}(J) : J(m) \le (1+\varepsilon) \min_{d \in D} H(d) \iff a(m)c \le 1 + \varepsilon.} \tag{3.3}$$

The factor $a(m)$ is candidate quality; c is the irreducible capacity tax. When $c > 1 + \varepsilon$, no candidate can be certified by this family, even if the integer optimizer is known exactly.

Table 1 summarizes the distinction between the continuous relaxation, the reusable monomial family, and the finite-domain oracle. The ordering is $L_{\mathrm{cont}} \le L_{\mathrm{fam}} \le J^*_{\mathrm{int}}$: only the middle quantity is a certificate that can be reused on the whole positive orthant.

Equality geometry. For active terms let $\pi_S(z) = c_S e^{-\mathbf{1}_S^\top z}/J(e^z)$. At a minimizer z_*, the Bregman identity is

$$\phi(x) - \phi(z_*) = \nabla\phi(z_*)^\top (x - z_*) + \mathrm{KL}(\pi(z_*) \| \pi(x)). \tag{3.4}$$

Both terms are nonnegative on P. Thus capacity equals the integer optimum exactly when an integer point lies on the supporting face and preserves the interaction distribution:

$$\boxed{L_{\mathrm{fam}} = J^*_{\mathrm{int}} \iff \exists x \in X : \nabla\phi(z_*)^\top (x - z_*) = 0, \quad (\mathbf{1}_S - \mathbf{1}_T)^\top (x - z_*) = 0\ \forall S, T \in \mathcal{S}.} \tag{3.5}$$

The first condition is the supporting-face test; the second is the KL equality test. This condition also handles rank-deficient interfaces; the full proof is in Appendix A.

Reuse boundary. The capacity identity is about a witness valid on the whole positive orthant. If validity is required only on the finite set D, the constant affine function at $\log J^*_{\mathrm{int}}$ closes the gap immediately. That finite-domain oracle must be rebuilt when D changes and therefore measures a different resource. The reusable family pays a one-time cross-domain validity cost; this is the quantity c in (3.3).

4 SUPPORT BUDGETS CLOSE THE FINITE-DOMAIN CERTIFICATE GAP

Before stating the support theorem, define the projected interaction geometry explicitly. Let $\mathcal{S}$ be the active monomial supports and

$$V = \operatorname{span}\{\mathbf{1}_S - \mathbf{1}_T : S, T \in \mathcal{S}\},\ X = \log D, \tag{4.1}$$

with Π_V the orthogonal projection. For $Z \subseteq X$, set

$$\eta_V(Z) = \max_{x \in X} \min_{z \in Z} \|\Pi_V(x - z)\|_2,\ \eta^*_{V,K} = \min_{Z \subseteq X,\ 1 \le |Z| \le \min(K, |X|)} \eta_V(Z). \tag{4.2}$$

The centers are restricted to integer anchors X, matching the executable support selection. For a finite support set $Z \subseteq X$ of at most K distinct integer anchors, let h_z be the tangent at z and define

$$L(Z) = \min_{x \in X} \exp\left(\max_{z \in Z} h_z(x)\right). \tag{4.3}$$

For every nonempty support set $Z \subseteq X$ of at most K distinct integer anchors,

Theorem 4.1 (Finite-support certificate bound)**.** *Assume the relevant-hull smoothness bound* $\sup_{z \in \operatorname{conv}(\Pi_V X)} \|\nabla^2 \phi(z)|_V\|_{\mathrm{op}} \le \mathsf{L}_V$. *Then*

$$0 < L(Z) \le J^*_{\mathrm{int}},\ 0 \le \log \frac{J^*_{\mathrm{int}}}{L(Z)} \le \frac{\mathsf{L}_V}{2} \eta_V(Z)^2. \tag{4.4}$$

A geometry-only farthest-point traversal satisfies $\eta_V(Z_K) \le 2\eta^*_{V,K}$ *and hence*

$$\log \frac{J^*_{\mathrm{int}}}{L(Z_K)} \le 2\mathsf{L}_V (\eta^*_{V,K})^2. \tag{4.5}$$

Corollary 4.2 (Support complexity)**.** *For tolerance* $\varepsilon > 0$*, define*

$$K^{\mathrm{opt}}_\varepsilon(D, J) = \min\left\{k \in \{1, \dots, |X|\} : \exists Z \subseteq X, |Z| = k,\ \log(J^*_{\mathrm{int}}/L(Z)) \le \log(1 + \varepsilon)\right\}.$$

Then

$$2\mathsf{L}_V (\eta^*_{V,K})^2 \le \log(1 + \varepsilon) \quad \Longrightarrow \quad K^{\mathrm{opt}}_\varepsilon(D, J) \le K.$$

The theorem gives the sufficient budget

$$K^{\mathrm{suff}}_\varepsilon(D, J) = \min\{K : 2\mathsf{L}_V (\eta^*_{V,K})^2 \le \log(1 + \varepsilon)\}.$$

The optimal k*-center radius* $\eta^*_{V,K}$ *may require a combinatorial solve. For an executed geometry-only path we report the observable* $K^{\mathrm{run}}_\varepsilon = \min\{K : \log(J^*_{\mathrm{int}}/L(Z_K)) \le \log(1 + \varepsilon)\}$*; if* $\mathsf{L}_V = 0$*, set* $K^{\mathrm{suff}}_\varepsilon = 1$.

For arbitrary Z, the theorem bounds the finite-domain gap to J^*_{int}; it does not assert $L_{\mathrm{fam}} \le L(Z)$. A nested sequence $Z \subseteq Z\prime$ is monotone because its tangent envelope only increases. The two-dimensional example in Appendix D shows why an arbitrary single anchor can lie below L_{fam}.

For each $x \in X$, smoothness gives $\phi(x) - h_z(x) \le (\mathsf{L}_V/2)\|\Pi_V(x - z)\|_2^2$; taking the closest $z \in Z$ and then the minimum over x yields the displayed bound. The farthest-point guarantee is the usual 2-approximation for the discrete k-center problem on X. The full proof is in Appendix B. The result makes support count a proof resource rather than a second learning objective. The online first-access cost is K support evaluations plus one candidate evaluation; all reference scans are reported separately.

The exact two-dimensional example already exhibits the ladder. One tangent has value $\sqrt{2}$; the two tangents at $(1, 2)$ and $(2, 1)$ attain $3/2$. Thus a support budget of two closes the reusable capacity gap on that domain. For larger domains the bound is intentionally conservative: it predicts how geometric coverage limits the proof, while the exact restricted oracle measures what the selected integer anchors actually achieve.

Algorithm 1 Geometry-only support ladder

Require: Integer log-domain X, support budget K
1: Choose the fixed anchor nearest the projected domain centroid
2: **for** $k = 2, \dots, K$ **do**
3: Add $x \in X$ maximizing its distance to the current projected support
4: **end for**
5: Evaluate the tangents only after the geometry-only support is fixed
6: Return $L(Z_K)$ and the separate candidate/reference ledger

Projected geometry. For reciprocal interfaces, $\nabla^2\phi(z) = \mathrm{Cov}_{\pi(z)}(\mathbf{1}_S)$, so Popoviciu's inequality gives $\mathsf{L}_V \le N/4$ in the ambient Euclidean metric. Let V be the span of active exponent differences and let $d_X = \dim \mathrm{aff}(\Pi_V \log D)$. The effective bound depends on d_X and the projected diameter, not on inactive coordinates. The computation records these dimensions for both product and noncommuting interfaces.

The algorithm is deliberately agnostic to labels, SGD seeds, and policy states. It measures how much geometric support is required to make a global certificate tight.

5 A STOCHASTIC REALIZATION OF THE OBSTRUCTION

The capacity gap is present in a stable learning process, not only in an abstract polynomial. Fix $\gamma > 0$, $0 < \rho < 1$, $\ell_i = 1$, $x_0 \neq 0$, and $t = \rho/(1+\sqrt{\gamma})$. Draw curvature $H = 1 - t \pm t\sqrt{\gamma}$ with equal probability and use unit-step minibatch SGD on $\ell_H(x) = Hx^2/2$. Every sample has curvature at least $1-\rho$; the minibatch update is mean-square contractive with $\mathbb{E}[(1-\bar{H})^2] \le \rho^2$ for every $m \ge 1$. Direct moment propagation gives

$$J(m) = A_N \prod_{i=0}^{N-1} (1 + \gamma/m_i), \; A_N = \frac{(1-t)x_0^2}{2} t^{2N}. \tag{5.1}$$

Theorem 5.1 (Stable-SGD density law)**.** *For $B = Nb + r$, $b \ge 1$, $0 < r < N$, $\theta = r/N$, and $g = b^{1-\theta}(b+1)^{\theta}$, balanced batches are integer-optimal and*

$$J^*_{\mathrm{int}} = A_N(1+\gamma/b)^{N-r}(1+\gamma/(b+1))^{r}, \tag{5.2}$$

$$L_{\mathrm{fam}} = A_N(1+\gamma/g)^{N}, \tag{5.3}$$

$$\Delta := \log \frac{J^*_{\mathrm{int}}}{L_{\mathrm{fam}}} = \frac{N\gamma\theta(1-\theta)}{2b^3}\big(1 + O_\gamma(b^{-1})\big). \tag{5.4}$$

The exact expression holds for every stated integer budget; integer mean batch ($r = 0$) has zero capacity gap.

Exchange proves balance, symmetry gives the constant log anchor, and the remainder is the Jensen gap of $f(s) = \log(1 + \gamma e^{-s})$ on one integer cell. The complete finite-b bounds and the bounded-residual-noise variant are in Appendix C.

Finite-budget check. With $h = \log(1 + 1/b)$ and $\kappa_- \le f''(s) \le \kappa_+$ on the cell,

$$\frac{N\theta(1-\theta)h^2\kappa_-}{2} \le \Delta \le \frac{N\theta(1-\theta)h^2\kappa_+}{2}. \tag{5.5}$$

This inequality is the numerical target used in the frontier checks. It separates the exact integer-versus-hull gap from the asymptotic notation. The construction remains contractive for bounded residual noise; the additive log perturbation is bounded on every fixed batch box.

Product interfaces and budget domain. The product rows use the explicit positive reciprocal interface

$$J_\beta(m) = \prod_{i=1}^{N} \left(\frac{1}{4} + \frac{\beta}{m_i}\right), \qquad \beta \in \left\{\frac{1}{8}, \frac{1}{2}\right\}, \qquad D_{N,B} = \left\{ m \in \mathbb{Z}^N_{\ge 1} : \sum_i m_i \le B \right\}. \tag{5.6}$$

Here $\beta = 1/8$ is low and $\beta = 1/2$ is high; the frontier uses $(N, B) = (4, 6), (4, 9)$ and the scaling audit uses $(8, 12), (8, 14)$, with exact enumeration in every domain.

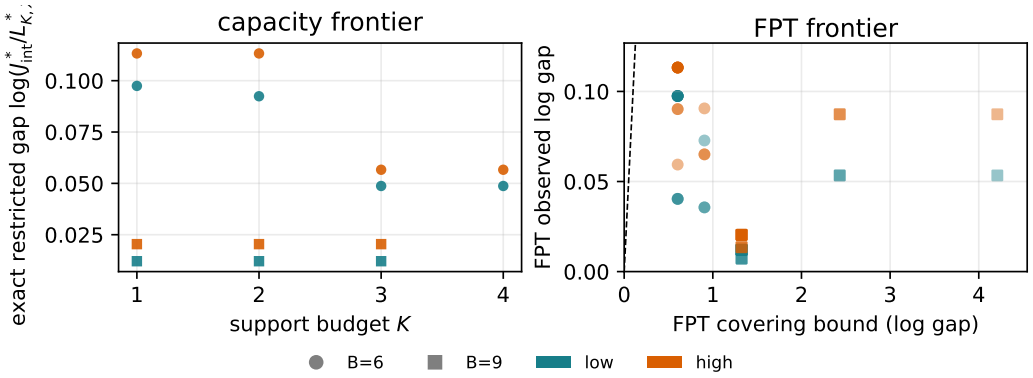

Figure 1: Capacity evidence on product and noncommuting interfaces. Left: exact restricted support gaps decrease with support budget. Right: geometry-only farthest-point traversal (FPT) observed gaps against their covering bounds for both interface families. Circles and squares denote $B = 6$ and $B = 9$; teal and orange denote low and high interaction levels.

Beyond commuting products.

Noncommuting moment family. For $M_k \in \mathbb{S}^2_+$ we use $M_{k+1} = AM_kA^\top + \lambda(W + G_1M_kG_1^\top + G_2M_kG_2^\top)/m_k$, $J(m) = \langle C, M_N\rangle$, with $\lambda \in \{1/4, 1\}$ and

$$A = \begin{pmatrix} 1/2 & 1/8 \\ 0 & 3/4 \end{pmatrix}, \quad G_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 0 & 1/8 \\ 1/4 & 0 \end{pmatrix},$$

$W = \mathrm{diag}(1/16, 1/32)$, $C = \mathrm{diag}(1/2, 1)$, and $M_0 = (1, -1)^\top(1, -1)$. The fixed matrices satisfy $AG_2 \neq G_2A$, so the maps do not commute. We use the deterministic $N \in \{4, 6\}$, $B \in \{6, 9, 8, 11\}$ grid; low/high denote $\lambda = 1/4, 1$. This is an abstract noncommuting moment-mechanism audit, not a claim that these matrices are the update operators of a constructed SGD instance.

Here ρ is reserved for the SGD contraction parameter above; λ denotes only the noncommuting interaction level. Positivity follows by induction:

$$M_k = \sum_{S\subseteq[k]} M_{k,S} \prod_{i\in S} m_i^{-1}, \; M_{k,S} \succeq 0.$$

The homogeneous term is $AM_{k,S}A^\top$, while the new reciprocal term is $\lambda(W\mathbf{1}_{S=\varnothing} + G_1M_{k,S}G_1^\top + G_2M_{k,S}G_2^\top) \succeq 0$; hence $c_S = \langle C, M_{N,S}\rangle \geq 0$. The recurrence therefore expands into a positive reciprocal polynomial. Hence the log-sum-exp convexity, capacity theorem, and support bound apply unchanged; this family is a controlled abstract mechanism check rather than a separate SGD-realization claim.

6 Capacity Frontier Experiments

We evaluate the four quantities that directly test the thesis: exact integer optimum, finite-support value, covering bound, and support count. The main frontier uses two product interfaces ($N = 4$, budgets $B \in \{6, 9\}$) at two interaction levels and the matched noncommuting family defined above. All domains are enumerated exactly; support locations are integer anchors, and the accompanying farthest-point implementation is geometry-only. Figure 1 reports the observable run gap against the covering bound. The four domain sizes for $(N, B) = (4, 6), (4, 9), (6, 8), (6, 11)$ are $15, 126, 28, 462$; the added $N = 8$ stress domains have sizes 495 and 3003.

The left panel of Figure 1 is the exhaustive integer-anchor oracle $L^*_{K,X}$, which is a restricted lower bound on the idealized L^*_K that allows arbitrary hull support locations; the right panel is the observable geometry-only FPT run. The FPT path uses only projected integer geometry to choose anchors; risk values are evaluated after selection and are not used to choose the next anchor. A single support can leave a visible gap, while a small support pool closes much of it. On the $N = 4, B = 6$ product

Table 2: Representative exact restricted capacity gaps.

interface	$K = 1$	$K = 2$	$K = 3$	domain size
product $N = 4$, $B = 6$, low	.0975	.0924	.0487	15
product $N = 4$, $B = 6$, high	.1133	.1133	.0566	15
product $N = 4$, $B = 9$, low	.0121	.0121	.0121	126
product $N = 4$, $B = 9$, high	.0204	.0204	.0204	126

Table 3: Noncommuting audit at the declared endpoint tolerance.

family	cases	family-empty	optimum-certifiable	unresolved
product	8	8	0	0
noncommuting	8	3	5	0

domain, the exact restricted log gap falls from 0.0975 to 0.0487 (low interaction) and from 0.1133 to 0.0566 (high interaction) by $K = 3$. On $B = 9$, the restricted oracle remains within 0.0121 (low) and 0.0204 (high) at $K \leq 3$. The noncommuting family shows the same monotone support effect; its values and projected dimensions are provided in the reproducibility artifact.

The $N = 8$ stress audit covers 495 and 3003 schedules; its observed gaps 0.185–0.227 remain below bounds 2.59–6.73. Across the $N = 8$ cases, the covering-bound/log-gap ratio is 11.4–36.4 (19.3–24.1 after the $B = 14$ two-support step). The theorem should therefore be read as a certified worst-case upper bound, not a calibrated predictor of the minimal K.

Table 2 exposes the exact values behind the frontier. The same calculation supplies the restricted oracle, the support envelope, and the covering certificate, so the empirical test does not depend on a chosen optimizer or on a fitted surrogate. Table 3 reports the same endpoint counts. At the fixed 0.3% endpoint tolerance, the independent noncommuting audit has three family-empty and five optimum-certifiable cases across the $N \in \{4, 6\}$ test grid. These counts establish that the obstruction survives noncommuting propagation; the product frontier remains the primary quantitative evidence.

These are capacity measurements, not claims about a particular policy's wall time or about nonlinear true-risk transfer. The theorem is exact for fixed positive reciprocal interfaces; data-dependent schedules, without-replacement sampling, weighted costs, and nonlinear transfer require separate certificate families. The Digits and interval audits are implementation checks listed in Appendix D, not additional contributions.

Limitations. The theorem does not cover without modification data-dependent schedules, without-replacement sampling, weighted component costs, or a nonlinear risk transfer from an upper model to true risk. Those extensions require a new certificate family or a separate calibration theorem. The current result is strong precisely because its capacity statement is exact under the declared interface.

Reproducibility. The accompanying source contains the deterministic frontier generators, compact numerical records, and tested Python and TeX versions needed to reproduce the reported values.

REPRODUCIBILITY STATEMENT

The supplementary archive provides the code and compact numerical records needed to reproduce the reported values. Raw training streams and public-data files are not redistributed; all reported checks are rerunnable from the included code.

DATA AVAILABILITY AND ETHICS

The core experiments use synthetic product and noncommuting moment interfaces. Optional WDBC and scikit-learn Digits audits are retained only as reproducibility checks; their provenance and attribution are given in the manuscript. Raw data, split files, and raw training streams are not redistributed. The originating UCI datasets are released under CC BY 4.0 (Wolberg et al., 1993;

Alpaydin & Kaynak, 1998). No new participant recruitment or clinical intervention is involved; the classification tasks are methodological allocation tests, not clinical deployment studies.

AI ASSISTANCE DISCLOSURE

Generative AI tools were used only to aid writing and editing. The mathematical results, proofs, and computational checks were produced and verified by the authors, who take full responsibility for the content.

REFERENCES

Akshay Agrawal, Steven Diamond, and Stephen Boyd. Disciplined geometric programming. *Optimization Letters*, 13(5):961–976, 2019. doi: 10.1007/s11590-019-01422-z. URL https://web.stanford.edu/~boyd/papers/dgp.html.

E. Alpaydin and C. Kaynak. Optical recognition of handwritten digits. UCI Machine Learning Repository, dataset, 1998. URL https://archive.ics.uci.edu/dataset/80/optical+recognition+of+handwritten+digits. CC BY 4.0; scikit-learn 1797-row subset used; accessed 20 September 2026.

Blake Bordelon and Cengiz Pehlevan. Learning curves for SGD on structured features. In *International Conference on Learning Representations*, 2022. doi: 10.48550/arXiv.2106.02713. URL https://openreview.net/forum?id=WPI2vbkAl3Q. Camera-ready preprint: arXiv:2106.02713v5.

Stephen Boyd, Seung-Jean Kim, Lieven Vandenberghe, and Arash Hassibi. A tutorial on geometric programming. *Optimization and Engineering*, 8(1):67–127, 2007. doi: 10.1007/s11081-007-9001-7. URL https://web.stanford.edu/~boyd/papers/gp_tutorial.html.

Prateek Jain, Sham M. Kakade, Rahul Kidambi, Praneeth Netrapalli, and Aaron Sidford. Parallelizing stochastic gradient descent for least squares regression: Mini-batching, averaging, and model misspecification. *Journal of Machine Learning Research*, 18(223):1–42, 2018. URL https://jmlr.org/papers/v18/16-595.html.

Alexandru Meterez, Depen Morwani, Jingfeng Wu, Costin-Andrei Oncescu, Cengiz Pehlevan, and Sham Kakade. Seesaw: Accelerating training by balancing learning rate and batch size scheduling. In *International Conference on Learning Representations*, 2026. doi: 10.48550/arXiv.2510.14717. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/187f9ba4cd5a59477305ac712282a1cd-Paper-Conference.pdf. Revised preprint: arXiv:2510.14717v2.

Rotem Mulayoff and Tomer Michaeli. Exact mean square linear stability analysis for SGD. In *Proceedings of the Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 3915–3969, 2024. URL https://proceedings.mlr.press/v247/mulayoff24a.html.

Stefan Perko. Unlocking optimal batch size schedules using continuous-time control and perturbation theory. arXiv:2312.01898v1, 2023. URL https://arxiv.org/abs/2312.01898v1. Preprint.

Samuel L. Smith, Pieter-Jan Kindermans, Chris Ying, and Quoc V. Le. Don't decay the learning rate, increase the batch size. In *International Conference on Learning Representations*, 2018. URL https://research.google/pubs/dont-decay-the-learning-rate-increase-the-batch-size/.

Jinbo Wang, Binghui Li, Zhanpeng Zhou, Mingze Wang, Yuxuan Sun, Jiaqi Zhang, Xunliang Cai, and Lei Wu. Fast catch-up, late switching: Optimal batch size scheduling via functional scaling laws. In *International Conference on Learning Representations*, 2026. doi: 10.48550/arXiv.2602.14208. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/0c7aaeb2745951f01224b3f95c75b528-Paper-Conference.pdf. Preprint version: arXiv:2602.14208v2.

William Wolberg, Olvi Mangasarian, Nick Street, and W. Street. Breast cancer wisconsin (diagnostic). UCI Machine Learning Repository, dataset, 1993. URL https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic. CC BY 4.0; accessed 20 September 2026.

A CORE PROOFS AND EQUALITY CONDITIONS

This supplement records the proof details and numerical checks needed to interpret the three main claims.

Log-hull representation. In a grid cell write $z_i = (1-a_i)\log k_i + a_i \log(k_i+1)$ with $0 \le a_i \le 1$. The interpolated budget is $\sum_i k_i + \sum_i a_i \le B$. Since the right-hand side after subtracting the integer base is integral, a vertex has no two fractional coordinates (opposite perturbations preserve the budget) and cannot have one fractional coordinate at a tight, perturbable boundary. The vertices are therefore exactly the integer log-plans, proving (2.4). This argument is the only place where unit costs are needed.

Capacity theorem. For $H(m) = K\prod_i m_i^{\beta_i}$, $h(z) = \log H(e^z)$ is affine. If $H \le J$ on the positive orthant, then $h \le \phi$ and $\min_{\log D} h \le \min_P \phi$. At a minimizer z_* of the convex function ϕ over P, first-order optimality gives $\nabla\phi(z_*)^\top(x - z_*) \ge 0$ for all $x \in P$. The global tangent $h_*(x) = \phi(z_*) + \nabla\phi(z_*)^\top(x - z_*)$ is valid and has minimum $\phi(z_*)$ on P, proving Theorem 3.1.

Supporting-face equality. The reciprocal log-sum-exp has gradient $-\sum_S \pi_S(z)\mathbf{1}_S$ and Hessian $\mathrm{Cov}_{\pi(z)}(\mathbf{1}_S)$. The standard log-sum-exp decomposition gives

$$\phi(x) - \phi(z_*) = \nabla\phi(z_*)^\top(x - z_*) + \mathrm{KL}(\pi(z_*)\|\pi(x)). \tag{A.1}$$

Thus both the supporting-face term and the interaction-distribution term must vanish for an integer point to attain capacity. If the active exponent differences span a proper subspace, the same statement holds after projection onto that subspace; no full-rank assumption is hidden in the theorem.

A.1 GENERAL FINITE-DOMAIN FORMULATION, EXACTNESS, AND DENSITY LAW

All vector logarithms and exponentials are coordinatewise. We distinguish the declared objective J, its integer optimum J^*_{int}, and the best value L_{fam} obtainable from one global monomial lower bound. The objective may be an exact risk R_N or a proved upper interface $\mathcal{C}_N$; the results below concern whichever objective is fixed.

A.2 THE OPTIMAL FAMILY VALUE IS A LOG-HULL MINIMUM

Let $D \subset \mathbb{R}^N_{++}$ be finite and nonempty, and define

$$X = \log D, \qquad P = \mathrm{conv}\, X, \qquad \phi(z) = \log J(e^z), \qquad J^*_{\mathrm{int}} = \min_{m\in D} J(m). \tag{A.2}$$

Assume that ϕ is finite, differentiable, and convex on $\mathbb{R}^N$. Write $\mathcal{H}(J)$ for all monomials $H(m) = a\prod_i m_i^{b_i}$ with $a > 0$, $b_i \in \mathbb{R}$, that satisfy $H(m) \le J(m)$ on the entire positive orthant, and set $L_{\mathrm{fam}} = \sup_{H\in\mathcal{H}(J)} \min_{m\in D} H(m)$.

Theorem A.1 (Best global monomial)**.** *The family optimum is attained and equals*

$$L_{\mathrm{fam}} = \exp \min_{z\in P} \phi(z). \tag{A.3}$$

For any minimizer z_*, an attaining monomial is $H_*(m) = \exp\{\phi(z_*) + \nabla\phi(z_*)^T(\log m - z_*)\}$.

Proof. Compactness gives a minimizer z_*. For any admissible H, $h(z) = \log H(e^z)$ is affine and satisfies $h \le \phi$. Hence $\min_X h = \min_P h \le h(z_*) \le \phi(z_*)$. Conversely, convexity makes the tangent at z_* a global affine minorant. First-order optimality gives $\nabla\phi(z_*)^T(z - z_*) \ge 0$ for every $z \in P$, so that tangent has minimum $\phi(z_*)$ over P, and also over X. Exponentiation proves the result. □

The tangent construction is classical (Boyd et al., 2007); the theorem identifies the relaxation represented by optimizing this entire family. The order of optimization matters: $\sup_H \min_D H = L_{\mathrm{fam}}$, whereas $\min_D \sup_H H = J^*_{\mathrm{int}}$, because a tangent attains J at each positive point.

Corollary A.2 (Preserving a constant floor)**.** *If $J = c_0 + K$, $c_0 \geq 0$, $K > 0$, and $\log K(e^z)$ is finite, differentiable, and convex, then*

$$\sup_{H_K \in \mathcal{H}(K)} \min_D (c_0 + H_K) = c_0 + \min_P K(e^z) = L_{\text{fam}}.$$

Proof. The logarithm of $c_0 + K(e^z)$ is convex by the log-sum-exp rule (or directly when $c_0 = 0$). Apply Theorem A.1 to K and J; adding a constant preserves their minimizers. □

This equality optimizes the anchor. A floor-preserving support can still be strictly stronger at a fixed anchor. A constant objective is certified exactly without normalized sensitivities.

A.3 An explicit hull for a unit-cost integer budget

Let $\ell, u \in \mathbb{Z}^N$, $1 \leq \ell_i \leq u_i$, and $B \in \mathbb{Z}$ with $B \geq \sum_i \ell_i$. Define

$$D = \{m \in \mathbb{Z}^N : \ell \leq m \leq u,\ \textstyle\sum_i m_i \leq B\}. \tag{A.4}$$

Caps may always be replaced by $\min\{u_i, B - \sum_{j \neq i} \ell_j\}$; without external caps use the second argument. For nonfixed coordinates, let p_i interpolate $(\log k, k)$ for integers $k = \ell_i, \ldots, u_i$:

$$p_i(z_i) = k + \frac{z_i - \log k}{\log(1 + 1/k)}, \qquad \log k \leq z_i \leq \log(k+1). \tag{A.5}$$

For $u_i = \ell_i$, set $p_i(\log \ell_i) = \ell_i$ on that singleton.

Theorem A.3 (Explicit log integer hull)**.** *For* (A.4)*,*

$$P = \{z : \log \ell \leq z \leq \log u,\ \textstyle\sum_i p_i(z_i) \leq B\}. \tag{A.6}$$

Proof. The slopes $1/\log(1 + 1/k)$ increase with k. Thus the displayed set is convex and contains X, proving one inclusion. Conversely, in each nonfixed coordinate write a point in that set as $z_i = (1 - a_i)\log k_i + a_i \log(k_i + 1)$ with $a_i \in [0, 1]$. For a fixed coordinate put $k_i = \ell_i$ and $a_i = 0$. Then $K = B - \sum_i k_i$ is a nonnegative integer and $a \in Q = \{v \in [0,1]^N : \sum_i v_i \leq K,\ v_i = 0 \text{ on fixed coordinates}\}$. Every vertex of Q is binary: two fractional coordinates admit opposite small perturbations; one fractional coordinate admits a two-sided perturbation when the inequality is slack, and contradicts the integer right-hand side when it is tight. Therefore a is a convex combination of binary vertices v^t of Q. Each $k + v^t$ belongs to D, and the same combination of their logs equals z. This proves the reverse inclusion, including singleton domains. □

Each p_i is the maximum of its segment lines on its domain, so (A.6) gives a lifted convex program with auxiliary $t_i \geq p_i(z_i)$ and $\sum_i t_i \leq B$. Its size is $O(N + \sum_i (u_i - \ell_i))$, pseudopolynomial in the numerical budget; its logarithmic coefficients still require numerical protection. The ordinary real budget domain is larger because $e^{z_i} \leq p_i(z_i)$. Neither arbitrary weighted budgets nor replacing $\sum p_i \leq B$ by equality follows from this theorem. For example, the log midpoint of $(1, 3)$ and $(3, 1)$ belongs to the hull of full-budget plans for $B = 4$, yet $\sum_i p_i = 2 + \log 3/\log 2 < 4$.

A.4 Exactness, exponent rank, and curvature

For the rest of this appendix take the nonzero reciprocal multi-affine objective

$$J(m) = \sum_{S \in \mathcal{S}} c_S \prod_{i \in S} m_i^{-1}, \qquad c_S > 0, \quad a_S = \mathbf{1}_S, \quad \pi_S(z) = \frac{c_S e^{-a_S^T z}}{J(e^z)}. \tag{A.7}$$

Only positive coefficients are included in $\mathcal{S}$; the empty set is allowed. The probabilities π_S describe objective contributions, not data-label probabilities. They satisfy

$$\nabla\phi(z) = -\mathbb{E}_{\pi(z)} a_S, \qquad \nabla^2\phi(z) = \text{Cov}_{\pi(z)}(a_S), \qquad 0 \preceq \nabla^2\phi(z) \preceq (N/4)I. \tag{A.8}$$

The last bound follows from the trace bound for binary coordinates.

Theorem A.4 (Exactness criterion)**.** *For all $x, z \in \mathbb{R}^N$,*

$$\phi(x) - \phi(z) = \nabla\phi(z)^T(x - z) + D_{\mathrm{KL}}(\pi(z)\|\pi(x)). \tag{A.9}$$

Fix $z_* \in \arg\min_P \phi$. The family is exact, $L_{\mathrm{fam}} = J^*_{\mathrm{int}}$, if and only if some $x \in X$ satisfies

$$\nabla\phi(z_*)^T(x - z_*) = 0, \qquad (a_S - a_T)^T(x - z_*) = 0 \quad (S, T \in \mathcal{S}). \tag{A.10}$$

Equivalently, some $x \in X$ satisfies $\max_{y\in X} \nabla\phi(x)^T(x - y) = 0$. If the exponent differences span $\mathbb{R}^N$, the hull minimizer is unique and exactness is equivalent to $e^{z_*} \in D$.

Proof. Differentiation proves (A.8). For $d = x - z$, $\log[\pi_S(z)/\pi_S(x)] = a_S^T d + \phi(x) - \phi(z)$; averaging proves (A.9). At z_* the linear term is nonnegative on P by optimality. KL is nonnegative by $-\log t \geq 1 - t$, with equality precisely when its two positive probability vectors coincide. Comparing pairwise probability ratios shows that equality is equivalent to the exponent conditions in (A.10). Minimizing over the finite set X proves the first criterion. A zero oracle gap is exactly the first-order condition for minimizing the convex ϕ on P, proving the second. Finally, $v^T\nabla^2\phi(z)v$ vanishes precisely when all $v^T a_S$ coincide. Full exponent-difference rank excludes this for nonzero v, hence makes ϕ strictly convex and z_* unique. □

Without full rank, a noninteger minimizing anchor need not imply a gap: a single monomial has affine ϕ and is itself an exact certificate. A positive constant and all positive singleton coefficients suffice for full rank. A small floating-point oracle gap does not prove that the exact gap is zero.

Corollary A.5 (Distance and perturbation bounds)**.** *Let $\mathcal{F}_* = \{z \in P : \nabla\phi(z_*)^T(z - z_*) = 0\}$, $\delta_X = \min_{x\in X} \|x - z_*\|_2$ and $\delta_F = \min_{x\in X\cap\mathcal{F}_*} \|x - z_*\|_2$. If $\mu I \preceq \nabla^2\phi \preceq LI$ on P, then*

$$\frac{\mu}{2}\delta_X^2 \leq \log(J^*_{\mathrm{int}}/L_{\mathrm{fam}}) \leq \frac{L}{2}\delta_F^2.$$

One may take $L = N/4$; under full rank a positive μ exists. If another differentiable log-convex objective has $\sup_P |\widetilde{\phi} - \phi| \leq \eta$, its log family gap differs from the original by at most 2η.

Proof. A convex representation of z_* by points of X has zero weighted first-order cost, so every positive-weight point belongs to $\mathcal{F}_*$. Taylor's integral formula and the Hessian bounds give the two distance inequalities, dropping the nonnegative first-order term for the lower bound and choosing a closest face point for the upper. Compactness and positive definiteness give $\mu > 0$ under full rank. Finally, uniform perturbation changes each of $\min_X \phi$ and $\min_P \phi$ by at most η. □

A.5 A SHARP DENSITY LAW REALIZED BY STABLE SGD

Theorem A.6 (Exact values and two-sided density bounds)**.** *Let $J(m) = A\prod_{i=1}^N (1 + \gamma/m_i)$, with $A, \gamma > 0$, $m_i \geq 1$, and $\sum_i m_i \leq B$ an integer unit budget. Write $B = Nb + r$, $b \geq 1$, $0 \leq r < N$, $\theta = r/N$, and $g = b^{1-\theta}(b + 1)^\theta$. Then*

$$J^*_{\mathrm{int}} = A(1 + \gamma/b)^{N-r}(1 + \gamma/(b + 1))^r, \tag{A.11}$$

$$L_{\mathrm{fam}} = A(1 + \gamma/g)^N. \tag{A.12}$$

For $h = \log(1 + 1/b)$, let $\kappa_- = \min_{v\in[b,b+1]} \gamma v/(v + \gamma)^2$ and $\kappa_+ = \max_{v\in[b,b+1]} \gamma v/(v + \gamma)^2$. The log gap $G = \log(J^*_{\mathrm{int}}/L_{\mathrm{fam}})$ satisfies

$$\frac{N\theta(1 - \theta)h^2}{2}\kappa_- \leq G \leq \frac{N\theta(1 - \theta)h^2}{2}\kappa_+ \leq \min\left\{\frac{N}{32b^2}, \frac{N\gamma}{8b^3}\right\}. \tag{A.13}$$

The gap is zero for $r = 0$ and positive otherwise. Every $\gamma > 0$ is realized by SGD on two uniformly sampled convex quadratic losses with a strongly convex mean and strict mean-square contraction even at batch one.

Proof. The function $u \mapsto \log(1 + \gamma/u)$ is strictly decreasing and has second derivative $\gamma(2u + \gamma)/[u^2(u + \gamma)^2] > 0$. An optimum spends the budget. Transferring one unit from a coordinate at least two larger than another strictly decreases the log objective. Thus its coordinates are b and $b + 1$, with r of the latter, proving (A.11).

Average the logs of all permutations of a balanced plan. The result is $z_* = (\log g)\mathbf{1} \in P$. For any global monomial, the geometric mean of its values over these permutations equals $H(g\mathbf{1})$, giving $\min_D H \le J(g\mathbf{1})$. Conversely the tangent there is $H_g(m) = J(g\mathbf{1}) \prod_i (g/m_i)^\alpha$ with $\alpha = \gamma/(g+\gamma) > 0$. Its minimum maximizes $\prod_i m_i$. The same balancing exchange maximizes this product at the balanced plans, where it equals g^N. Thus $\min_D H_g = J(g\mathbf{1})$.

Set $f(t) = \log(1 + \gamma e^{-t})$ and $t = \log b$. Substitution gives

$$G = N\{(1-\theta)f(t) + \theta f(t+h) - f(t+\theta h)\}. \tag{A.14}$$

Here $f''(t) = \gamma e^t/(e^t+\gamma)^2 > 0$. Apply Jensen's inequality to $f - \kappa_- t^2/2$ and $\kappa_+ t^2/2 - f$ on this interval to obtain both bounds. The minimum curvature is the smaller endpoint value; the maximum is $1/4$ when $\gamma \in [b, b+1]$ and the larger endpoint otherwise, as the derivative in v is $\gamma(\gamma - v)/(v+\gamma)^3$. Therefore $\kappa_+ \le \min\{1/4, \gamma/b\}$; using $h \le 1/b$ and $\theta(1-\theta) \le 1/4$ finishes (A.13).

Alternative non-strict realization. The following older construction is retained only as an algebraic cross-check: it permits a zero-curvature component. The main theorem and all headline claims use the strictly $(1-\rho)$-strongly convex realization in the stable-SGD construction in the next section; this alternative is not used to support those claims. For the realization put $a = \sqrt{\gamma}$, $\kappa = 2a/(1+a)$, $f_1(x) = 0$, and $f_2(x) = \kappa x^2/2$. Their mean is $F(x) = \kappa x^2/4$. Unit-step SGD with fresh independent with-replacement samples obeys $x^+ = (1 - \overline{H}_m)x$, with single-sample curvature $H \in \{0, \kappa\}$ equiprobably. Hence

$$\mathbb{E}[(x^+)^2 \mid x] = \{(1-\kappa/2)^2 + \kappa^2/(4m)\}x^2 = \frac{1+\gamma/m}{(1+a)^2}x^2.$$

Iteration yields $\mathbb{E}F(x_N) = C_0(1+a)^{-2N}\prod_i(1+\gamma/m_i)$. Any $A > 0$ is obtained by $C_0 = A(1+a)^{2N}$; the batch-one multiplier $(1+a^2)/(1+a)^2$ is strictly below one. □

Corollary A.7 (Density, tolerance, and absolute scale)**.** *For fixed $\gamma > 0$, uniformly in N and $0 < \theta < 1$ as $b \to \infty$,*

$$G = \frac{N\gamma\theta(1-\theta)}{2b^3}[1 + O_\gamma(b^{-1})]. \tag{A.15}$$

Along integer-budget sequences with $\theta \to \theta_0 \in (0,1)$, the regimes $N/b^3 \to 0, c, \infty$ give family ratios tending to 1, $\exp\{c\gamma\theta_0(1-\theta_0)/2\}$, *and ∞, respectively. If the lower bound in* (A.13) *exceeds* $\log(1+\varepsilon)$*, the family cannot certify any integer plan at relative tolerance ε. If its upper bound is at most that threshold, the optimal support certifies a balanced plan.*

Proof. Uniformly on $[b, b+1]$, $\gamma v/(v+\gamma)^2 = (\gamma/b)[1 + O_\gamma(b^{-1})]$ and $h^2 = b^{-2}[1 + O(b^{-1})]$. Divide the two-sided bound by its positive leading term. The limits follow by exponentiation; the tolerance statements use the attained family value and the integer optimality of balanced plans. □

For fixed initial risk C_0 in this realization, $A = A_N = C_0(1+\sqrt{\gamma})^{-2N}$. Thus an exponentially large relative gap can coexist with exponentially small absolute risk. A nonnegative constant floor τ adds to both optimal values by Corollary A.2, and

$$\frac{\tau + J^*_{\mathrm{int}}}{\tau + L_{\mathrm{fam}}} - 1 = \frac{J^*_{\mathrm{int}} - L_{\mathrm{fam}}}{\tau + L_{\mathrm{fam}}}.$$

The critical density $b \asymp N^{1/3}$ assumes fixed γ and θ away from its endpoints. Integer density has exactly zero gap and is handled separately.

B SUPPORT LADDER AND EXECUTION DETAILS

For h_z the tangent at z, convexity gives $h_z \le \phi$ globally. Hence the maximum of finitely many tangents is a valid lower certificate and $L(Z)$ is monotone under adding supports. Let $x \in \log D$ and choose $z \in Z$ with projected distance at most $\eta_V(Z)$. Smoothness gives

$$\phi(x) - h_z(x) \le \frac{\mathsf{L}_V}{2}\|\Pi_V(x-z)\|_2^2. \tag{B.1}$$

Taking the worst x proves the first bound in (4.4). The metric farthest-point traversal has the standard 2-approximation for discrete k-center, yielding (4.5). It reads only the projected geometry of $X = \log D$; the risk values are used after the support set is fixed.

The executable path keeps support, candidate, and reference evaluations separate. A first access to a K-support certificate costs K support evaluations plus one candidate evaluation. A feasible integer schedule can be recovered from a hull anchor by filling fractional slack and applying sum-preserving reciprocal exchanges. Its full interval proof is intentionally not part of the mainline; boundary comparisons are reported as unresolved when the stated tolerance cannot separate the endpoints.

B.1 Protected execution from rational cells

The recovery bound becomes executable once cell membership, the budget, and endpoint comparisons all have explicit representations. This section gives such a representation, including a selection allowance when finite-precision comparisons cannot separate the two endpoints. It applies to

$$J(m) = \sum_{S \subseteq [N]} c_S \prod_{i \in S} m_i^{-1}, \qquad c_S \geq 0, \quad J > 0,$$

on $\mathcal{D} = \{m \in \mathbb{Z}^N : \ell_i \leq m_i \leq u_i,\ \sum_i m_i \leq B\}$, with integer B, ℓ_i, u_i and $\ell_i \geq 1$.

An exact input representation. Choose integer k_i with $\ell_i \leq k_i \leq u_i$ and rational $a_i \in [0,1]$ satisfying $\sum_i (k_i + a_i) \leq B$. If $k_i = u_i$, require $a_i = 0$ and keep this coordinate fixed. On the other coordinates define

$$z_i = (1-a_i)\log k_i + a_i \log(k_i+1), \qquad t_i(a_i) = \frac{1-a_i}{k_i} + \frac{a_i}{k_i+1}, \qquad Q(a) = \sum_S c_S \prod_{i \in S} t_i(a_i). \tag{B.2}$$

At a fixed upper node use $z_i = \log k_i$ and $t_i = 1/k_i$.

Proposition (exact cell recovery). The input represents a point $z \in P = \operatorname{conv}\{\log m : m \in \mathcal{D}\}$. There is an algorithm using exact endpoint selection, coordinate increases, and at most $N-1$ pair exchanges that returns $\widehat{m} \in \mathcal{D}$ with $J(\widehat{m}) \leq Q(a)$. Its feasibility arithmetic is rational; its potential comparisons are also rational when the coefficients c_S are rational. For any independent lower bound $L > 0$ on $\min_{\mathcal{D}} J$, $J(\widehat{m})/L \leq Q(a)/L$.

Proof. Put $R = B - \sum_i k_i$, an integer. The polytope

$$\{b : 0 \leq b_i \leq 1,\ \sum_i b_i \leq R,\ b_i = 0 \text{ when } k_i = u_i\}$$

has only binary vertices: two fractional coordinates admit a small sum-preserving perturbation in both directions; a single fractional coordinate leaves slack in the integer right-hand-side inequality and also admits a perturbation. Every binary vertex gives a feasible plan $k + b$. The affine image $b_i \mapsto (1-b_i)\log k_i + b_i \log(k_i+1)$ therefore proves $z \in P$.

Compute $r = \lceil \sum_i a_i \rceil$ exactly. As $r \leq R$, increasing fractional coordinates until their total reaches r is feasible. There is enough capacity within the currently fractional coordinates; if their number is s, their total capacity is $s - \sum_{i:\, 0<a_i<1} a_i$, which is at least the required increase. Every such increase decreases each affected t_i, so nonnegative coefficients imply that Q cannot increase.

If fractional coordinates remain, their integer total guarantees that at least two, say i, j, remain. Replace (a_i, a_j) by $(a_i + x, a_j - x)$ for

$$x_- = \max\{-a_i, a_j - 1\} \leq x \leq x_+ = \min\{1 - a_i, a_j\}.$$

All points on this segment obey the same budget. Writing $\beta_i = 1/[k_i(k_i+1)]$, the terms containing both i, j contribute the quadratic coefficient $-\beta_i\beta_j \sum_{S \supseteq \{i,j\}} c_S \prod_{h \in S \setminus \{i,j\}} t_h \leq 0$; all remaining terms are affine. Thus Q is concave in x, and at least one endpoint has potential no greater than the current point. Choose an exact minimum endpoint. Each exchange fixes at least one additional coordinate to zero or one. After at most $N-1$ exchanges, a is binary, $\widehat{m} = k + a$ is feasible, and $t_i = 1/\widehat{m}_i$ yields $J(\widehat{m}) = Q(a) \leq Q(a^{\text{initial}})$. The ratio statement follows by division. $\square$

Proposition (protected endpoint comparisons). Suppose an exchange's endpoint values V_-, V_+ have outward enclosures $[\underline{V}_-, \overline{V}_-]$ and $[\underline{V}_+, \overline{V}_+]$. Choose the endpoint s with smaller upper endpoint, denote the other one by o, and record

$$\delta = \max\{0, \overline{V}_s - \underline{V}_o\}. \tag{B.3}$$

If the initial potential has an upper bound U_Q and the returned integer plan is independently evaluated with upper bound U_e, then

$$J(\widehat{m}) \le \min\left\{U_e,\ U_Q + \sum_{\text{exchanges } j} \delta_j\right\}. \tag{B.4}$$

In particular, divided by an independently valid $L > 0$, the right-hand side is a protected quality-ratio bound. A zero allowance on every exchange preserves the initial analytic potential bound up to its evaluation enclosure.

Proof. If the selected endpoint is a true minimum, then $V_s \le \min\{V_-, V_+\} + \delta$ immediately. Otherwise $V_s - V_o \le \overline{V}_s - \underline{V}_o \le \delta$, proving the same inequality. Exchange concavity bounds the true minimum by the preceding potential. Summing these inequalities, together with nonincrease during the initial fill, gives $J(\widehat{m}) \le U_Q + \sum_j \delta_j$. The independent final evaluation supplies $J(\widehat{m}) \le U_e$. □

Moving an anchor requires a new evaluation. Converting a rational mixture of integer log-plans to rational cell fractions generally changes the log-anchor: ratios of logarithms need not be rational. Consequently, the old anchor's family-search endpoints cannot be reused as if they had been evaluated at the new point. A simple perturbation control is

$$|\phi(z) - \phi(\widetilde{z})| \le \|z - \widetilde{z}\|_1, \qquad \phi(z) = \log J(e^z).$$

Indeed, $-\partial_i \phi$ is the total normalized weight of terms containing i, hence belongs to $[0, 1]$; integrate the gradient along the segment. An implementation must either reevaluate the new feasible point or propagate a rigorously enclosed displacement. The execution audit below starts with declared rational cells and does not transfer any old family-search interval.

Executed stress audit. The standard-library audit recorded every fill, both endpoints of every exchange, the rational budget, all allowances, and an independent final integer-risk evaluation; the corresponding interval checker is included in the source tree. Exact reciprocal-polynomial arithmetic supplies the reference; a separate positive interval evaluator rounds every operation outward to 80, 120, 180, or 240 binary significand bits. For two source-positive examples it evaluates the recurrence $M_{i+1} = A_i M_i + (G_i M_i + W_i) t_i$ implicitly, with rational $A_i, G_i, W_i > 0$, and checks against an independently expanded polynomial.

The 11 cases cover $N = 1$, integer anchors, a fixed upper node, slack budget, fraction sums 1 ± 2^{-180}, tied endpoints, zero coefficients, an unresolved near-tie, and three- and six-step recurrences with nonzero source terms. There are 11 exact and 44 interval chains, 48 protected exchanges, 184 initial/final/endpoint enclosures, and 90 feasible cell vertices checked. All final feasibility, enclosure, and allowance bounds pass. Three injected errors—a fractional upper node, an overspent budget, and an interval missing its exact reference—are rejected.

The near-tie is intentionally consequential. Set $k = (1, 1)$, $a = (1/2, 1/2)$, $B = 3$, and $J(m) = (1 + 2^{-300})/m_1 + 1/m_2$. The initial potential is $3/2 + 3 \cdot 2^{-302}$. At all four tested precisions, a tied upper-endpoint comparison selects $(1, 2)$, whose risk is $3/2 + 2^{-300}$ and therefore exceeds the exact initial potential by 2^{-302}. The saved nonzero allowance covers this difference in every run. Exact endpoint selection returns $(2, 1)$ and preserves the analytic bound. Thus the audit checks the finite-precision allowance rather than incorrectly asserting that every floating comparison preserves the unadjusted recovery factor. It establishes execution correctness on these cases, not a gain in scheduling quality or an automatic conversion of arbitrary search anchors.

C STABLE-SGD DERIVATION

For the two-point curvature law, a minibatch multiplier has conditional second moment $t^2(1 + \gamma/m)$. Iterating independent batches gives (5.1). For $B = Nb + r$, an exchange of one sample from a larger

Artifact	Role
`frontier data`	exact product/noncommuting integer optima and endpoint intervals used by the frontier.
`frontier generator`	finite-support values and the two-panel capacity figure.
`Digits check`	corrected public-Digits implementation sanity check; it is not a transfer claim.
`interval check`	interval and defect check for the upper-model limitation.
`reproduction script`	reruns the deterministic numerical examples from the source tree.

batch to a smaller batch decreases the product, so optimal batches differ by at most one. Writing $s = \log m$ and $f(s) = \log(1 + \gamma e^{-s})$, the integer and hull values differ by the Jensen gap on one cell of width $h = \log(1 + 1/b)$:

$$\frac{N\theta(1-\theta)h^2\kappa_-}{2} \le \Delta \le \frac{N\theta(1-\theta)h^2\kappa_+}{2}, \qquad \kappa_- \le f''(s) \le \kappa_+. \tag{C.1}$$

Since $h = b^{-1}(1 + O(b^{-1}))$ and $\kappa_\pm = \gamma b^{-1}(1 + O_\gamma(b^{-1}))$, this yields the stated $N\gamma\theta(1-\theta)/(2b^3)$ law. Adding bounded residual noise changes the log gap by a bounded additive term on a fixed box; it does not alter the exact positive reciprocal theorem.

Interval interpretation. When lower endpoints L_f, L_e are outward rounded and $U_e \le (1 + \varepsilon)L_f$, the candidate is certified. If an independent integer lower bound satisfies $L_I > (1 + \varepsilon)U_f$, the family is empty at that tolerance; otherwise the output remains unresolved.

Two-dimensional regression for the support quantifier. For $J(m) = m_1^{-1} + m_2^{-1}$ and $D = \{(1,1),(1,2),(2,1)\}$, the reusable capacity is $L_{\mathrm{fam}} = \sqrt{2}$. The valid integer anchor $z = \log(1,2)$ gives $L(\{z\}) = (3/2)2^{-1/3} = 1.190550789 < \sqrt{2}$. Thus an arbitrary support set cannot be claimed to dominate the optimal single reusable monomial; only the finite-domain bound and nested-support monotonicity stated in Theorem 4.1 are used.

D AUDIT MAP AND SUPPLEMENTAL CHECKS

The following table identifies what is evidence for the mainline and what is only an implementation or auxiliary check.

The Digits check evaluates log/raw consistency, direct recurrence, empirical covariance normalization, tangent validity, and log-space gate arithmetic. Its derived record has zero false-safe rows and zero unresolved rows; these facts support code credibility but are not needed for the capacity theorem. The nonlinear audit similarly checks interval dominance and exact zero variance records without being used to claim true-risk transfer.

Scope boundaries. For a weighted domain, the correct object remains $P_w = \mathrm{conv}(\log D_w)$, but the explicit unit-cost hull oracle does not transfer automatically. For a nonlinear upper model $\mathcal{C}_N \ge R_N$, certificate capacity alone does not imply a true-risk guarantee; a separate slack and coverage theorem is required. These are boundaries of the claim, not failed experiments.

The submitted supplement contains only the four sections above so that it functions as proof and audit material for the one capacity thesis rather than as a second paper.