SELECTIVE CERTIFICATES FOR REDUCED-SPACE ROBUST SCHATTEN-p OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reduced-space optimization can be stationary in its working coordinates while a finite-amplitude direction outside that space still decreases the objective. We turn this failure mode into a selective certificate interface for robust Schatten-p self-representation. The exact split $r_Z^2 = a^2 + e^2$ separates in-space from omitted original-space error, while a bounded top-k normal scan tests finite-amplitude escapes. The controller returns a certificate for the declared $(\Phi_\beta, \mathcal{D}, [0, T], \eta)$ object or a typed abstention when coverage, rank, amplitude, or resources are insufficient. We prove the residual split, a finite nonnegative-Huber tail closure with optional Lipschitz tightening, a strict grid-error bound, and separate descent and compute ledgers. An incomplete tail scan remains an explicitly scoped prefix. The image panel reports both $\Delta\Phi$ and ΔF, their envelope gap, majorant acceptance diagnostics, and full/prefix certificate counts. This separation is the point: compressed computation should expose exactly what it certifies and abstain when the requested original-space condition has not been covered.

1 INTRODUCTION

Reduced spectral coordinates are attractive because they replace an $N \times N$ self-representation with a small working factor. They also create a failure mode that ordinary stopping rules cannot see: the iterate can be stationary in the coordinates being optimized while an unrepresented, finite-amplitude direction still lowers the original objective. For a nonconvex robust spectral model, this is a correctness question rather than a choice of implementation width. A solver must say whether it has checked the original variable, how much of it was checked, and what it does when the check is incomplete.

The failure is already visible in the smallest example. Let $X = I_2$, $L = 0$, and $Z_0 = 0$. For $0 < p < 1$, $0 < \lambda < \beta$, and $0 < \mu < \lambda - \lambda^2/(2\beta)$, the normal path $Z(t) = te_1e_1^\top$ has $\Delta\Phi_\beta(1) = -.0895$ for the parameters used in our witness, although the limiting subdifferential at Z_0 contains zero. Thus a zero stationarity residual does not certify finite-amplitude no-escape. The condition $0 < \lambda < \beta$ is the sufficient branch condition for this saturated-Huber witness; the complete subdifferential calculation is in Proposition B.1.

Our answer is a selective certificate interface. Write the computed matrix as $Z = QH$, with Q orthonormal, and let a denote the residual inside the working space. The omitted original-space component is $e = \|(I - QQ^\top)X^\top\Lambda_\beta V\|_F$. The exact identity

$$r_Z^2 = a^2 + e^2$$

turns an invisible blind spot into a decision: Continue when the represented coordinates dominate, and Expand when omitted error dominates. After the first-order test is small, a finite top-k normal scan checks whether a finite amplitude still lowers the robust envelope. The scan is finite and auditable: its amplitude interval, direction family, grid error, and arithmetic error are all part of the output. If the declared family or interval is not enough for the requested claim, the correct output is abstention.

The decisive advance is a certificate contract that makes this blind spot observable, bounds the finite scan, and abstains when the requested original-space claim has not been covered.

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.

Contributions. **(1) Original-space omission certificate.** We prove the exact residual split and the corresponding Continue/Expand rule for a nonsmooth Schatten-p spectral objective. Expansion is a measurable allocation of computation, not an untracked increase in width.

(2) Finite-gain selective certificate. We derive the exact robust normal scan, an always-finite tail closure from the nonnegative Huber envelope, and a strict grid bound based on the Hölder continuity of t^p and the Lipschitz robust envelope. A top-k screened direction set yields a finite-family-prefix or finite-family-full certificate; an ambient statement is available only with an explicit coverage premise. Descent charges and compute costs are reported as separate ledgers.

(3) Certificate-or-abstention contract. We make coverage, rank, scan-domain, and resource failures typed outputs. The contract separates optimization evidence from downstream task quality: an accepted activation must lower the declared envelope, but need not improve every readout.

The interface applies whenever a solver optimizes in a compressed space but is judged in the original variable. The omitted residual determines whether to expand; the finite-gain probe determines whether small stationarity is enough; the budget records the cost of making that decision. This is the useful unit of generality for ICLR readers, while all identities and guarantees below are proved for the robust self-representation model.

Relation to prior work. Reduced active-subspace methods and FaNCL establish that spectral computation can be carried out without materializing the full matrix (Hsieh & Olsen, 2014; Yao et al., 2015; 2019). GSVT and Schatten-p proximal rules provide the fixed-space spectral update (Lu et al., 2015; Giampouras et al., 2020). Finite-amplitude escape is a known limitation of residual-only reasoning; our contribution is to make its omitted component observable, its robust gain computable on a declared domain, and its uncertainty an explicit abstention outcome. The detailed comparison is deferred to Appendix A so that the main text stays on this single evidence chain.

The experiments follow that chain. On six frozen image inputs and three shared seeds, Act-ON and Act-OFF use the same Continue/Expand worker and both keep correction disabled. The revised panel reports full/prefix candidate certificates, typed abstentions, ΔF, $\Delta\Phi$, majorant slacks, capture ratios, and compute cost. This is the expected selective behavior: the method certifies a declared optimization object rather than claiming a universal task improvement. A small finite-grid panel separately calibrates how screened direction families compare with a dense direction reference.

2 Model, outputs, and requested accuracy

Let $X = [x_1, \dots, x_N] \in \mathbb{R}^{d\times N}$ contain fixed observations or features. We study

$$\min_{Z,E} \ \mu \sum_i \sigma_i(Z)^p + \lambda\|E\|_{2,1} + \frac{\gamma}{2}\operatorname{tr}(ZLZ^\top) \quad \text{subject to } XZ + E = X. \tag{1}$$

Here $0 < p < 1$, $\mu, \lambda > 0$, $\gamma \geq 0$, $L = L^\top \succeq 0$ is fixed, and $\|E\|_{2,1} = \sum_j \|E_{:j}\|_2$ penalizes column corruption. The spectral term favors compact shared structure. For a graph Laplacian, $\operatorname{tr}(ZLZ^\top) = \frac{1}{2}\sum_{ij} w_{ij}\|z_i - z_j\|_2^2$, so neighboring samples are encouraged to have similar coefficients. The model combines established robust spectral and graph penalties (Zhang et al., 2019; Yin et al., 2016). Coefficients are unconstrained, including their diagonal. Eliminating E defines the original objective $F(Z)$. We intentionally keep the diagonal free because the paper studies computation allocation in the unconstrained coefficient model. A zero- diagonal variant would need projected updates and a separate GSVT/row-space derivation; it is outside the present claims and has no reported ablation.

Reader map. The symbols used below have one fixed role: $X \in \mathbb{R}^{d\times N}$ is the data matrix, $Z \in \mathbb{R}^{N\times N}$ is the self-representation, and $Q \in \mathbb{R}^{N\times h}$ is the current orthonormal working space. Its projector is $P = QQ^\top$; a measures error inside Q, e measures the omitted original-variable error, and r_Z is their ambient stationarity residual. The continuation stage is β, the inverse step scale is s, and t is an activation amplitude. *Continue* updates Z inside Q; *Expand* changes Q; *Activate* tests a finite top-k normal direction family. The first main theorem turns (a, e, r_Z) into a work/expansion decision, and the second turns a finite scan into a candidate certificate or abstention. The lift produces (E_β, Λ_β); r_E, r_F measure the lifted blocks, F is the nonsmoothed objective, V is the right singular factor, and $V_{\min(8,r)}$ is the task readout. The pipeline

is lift → residual split → Continue/Expand → finite-gain scan → certificate or abstention. A returned state reports $r_Z \le \varepsilon_Z$, $r_F \le \varepsilon_F$ (checked at the final continuation stage), the candidate gain and its scan error, the observed domain $[0, T_{\mathrm{obs}}]$, and a certificate scope. An ambient no-escape statement additionally requires a declared direction coverage factor. A rank-zero state with no accepted finite-gain move returns the typed status ZERO-RANK/NO-GAIN rather than a successful task readout.

Three dimensions with different roles. We store $Z = QH$, where $Q^\top Q = I_h$, and write its thin SVD as $Z = U\Sigma V^\top$ with rank $r \le h$. The *working width* h determines computation; the *rank* r is selected by optimization; the *readout dimension* $k \le r$ is fixed by the task interface. Thus expanding h, changing r, and choosing k are separate computation, optimization, and task decisions. Our primary clustering readout row-normalizes $V_{\min(8,r)}$ and uses K_{cls}-means; a zero row maps to the zero vector and is excluded from undefined angle comparisons. The outlier score is $q_j = \|x_j - Xz_j\|_2$. Both are available without storing Z: $Z^\top Z = H^\top H$ and $XZ = (XQ)H$. Evaluation labels do not enter optimization, readout selection, or stopping. Class-based task construction and development-set choices are specified in the experimental protocol.

Smooth the residual, retain the spectral penalty. For $C = X - XZ$, define

$$\Phi_\beta(Z) = \mu \sum_i \sigma_i(Z)^p + \psi_\beta(Z),\ \psi_\beta(Z) = \frac{\gamma}{2}\operatorname{tr}(ZLZ^\top) + \min_E \left\{ \lambda\|E\|_{2,1} + \frac{\beta}{2}\|E - C\|_F^2 \right\}. \quad (2)$$

Column shrinkage gives $E_{\beta,:j} = (1 - \lambda/(\beta\|C_{:j}\|_2))_+ C_{:j}$ (with zero at a zero column). Set $\Lambda_\beta = \beta(E_\beta - C)$ and $G_\beta = \nabla\psi_\beta = \gamma ZL + X^\top\Lambda_\beta$. The convex smooth part has Lipschitz constant at most $\rho_\beta = \beta\|X\|_2^2 + \gamma\lambda_{\max}(L)$.

Using the limiting subdifferential, the ambient envelope stationarity residual is $r_Z = \operatorname{dist}(0, \partial\Phi_\beta(Z))$. It belongs to the smoothed envelope Φ_β, not to nonsmoothed F; the latter is controlled by the displayed feasibility and envelope-gap bounds. The lifted variables also satisfy

$$\begin{gathered} r_E = \operatorname{dist}(0, \lambda\partial\|E_\beta\|_{2,1} + \Lambda_\beta) = 0,\ r_F = \|XZ + E_\beta - X\|_F \le \lambda\sqrt{N}/\beta, \\ 0 \le F(Z) - \Phi_\beta(Z) \le C_\lambda/\beta,\ C_\lambda = \lambda^2 N/2. \end{gathered} \quad (3)$$

Thus ε_Z specifies stationarity and ε_F specifies constraint accuracy; choosing $\beta_T \ge \lambda\sqrt{N}/\varepsilon_F$ handles the latter. These are approximate KKT criteria for (1); task quality is measured separately. The non-Lipschitz penalty makes $Z = 0$ stationary. The sufficient initial gap $\Gamma = F(0) - [\Phi_{\beta_0}(Z^0) + C_\lambda/\beta_0] > 0$ excludes zero along descent and continuation (Appendix 28). We record this condition and actual nonzero rank separately, counting a zero-rank task output as a failure.

3 Selective certificates for compressed spectral computation

A measurable original-space error. Let P_X project onto $\operatorname{range}(X^\top)$. Projection preserves XZ and decreases the objective. For $M > 0$ and $\|Z\|_2 \le M$,

$$\Phi_\beta(Z) - \Phi_\beta(P_X Z) \ge \frac{\mu p}{2} M^{p-2}\|(I - P_X)Z\|_F^2. \quad (4)$$

The proof, including the $M = 0$ trivial case, is given in Appendix B.2. Thus a complete row-space basis is exact, but a small working basis needs an original-space audit. Write $Z = QH = U\Sigma V^\top$, with $Q^\top Q = I$, and define

$$a = \operatorname{dist}(0, \partial_H \Phi_\beta(QH)),\ W = (I - QQ^\top)X^\top\Lambda_\beta V,\ e = \|W\|_F.$$

Theorem 3.1 (Omission certificate and expansion)**.** *The original residual obeys the exact identity*

$$r_Z^2 = a^2 + e^2. \quad (5)$$

If an orthonormal block $B \perp Q$ is appended and zero rows are added to H, Z is unchanged while $e_+^2 = e^2 - \|B^\top W\|_F^2$ and $a_+^2 = a^2 + \|B^\top W\|_F^2$.

Key identity. The minimum-norm spectral subgradient splits into orthogonal Q and $(I - QQ^\top)$ blocks; Pythagoras gives the first identity and projecting the external block onto B gives the second. The full subdifferential, including rank zero and repeated singular values, is recorded in Appendix B.1. □

When $a \le \theta r_Z$ and $r_Z > \varepsilon_Z$, the omitted component is the measurable reason to Expand; otherwise Continue with the current coordinates. A leading block of W is accepted only when $\|B^\top W\|_F^2 \ge \zeta e^2$. The implementation checks this ratio after projection and re-orthogonalization; if the available width cannot meet ζ, it records an expansion-capture failure and does not expand. This makes expansion a charged allocation of computation rather than an unreported width change.

Finite-amplitude gain and its declared domain. For a normal pair $u \perp U$, $v \perp V$, let $C = X - XZ$, $w = Xu$ and $\ell = v^\top Lv$. Along $Z(t) = Z + tuv^\top$,

$$\Delta\Phi_\beta(t; u, v) = \mu t^p + \frac{\gamma\ell}{2}t^2 + \sum_j \{h_\beta(\|C_{:j} - tv_j w\|_2) - h_\beta(\|C_{:j}\|_2)\}. \tag{6}$$

Here h_β abbreviates the scalar Huber envelope

$$h_{\lambda,\beta}(q) = \begin{cases} \beta q^2/2, & q \le \lambda/\beta, \\ \lambda q - \lambda^2/(2\beta), & q > \lambda/\beta, \end{cases} \qquad q \ge 0. \tag{7}$$

The same definition is used in every scan and audit record. Define $H_0 = \sum_j h_{\lambda,\beta}(\|C_{:j}\|_2)$ and $a_{uv}(t) = \mu t^p + c_{uv}t^2/2$, where

$$b_{uv} = \lambda\|Xu\|_2\|v\|_1, \qquad c_{uv} = \gamma v^\top Lv, \tag{8}$$

Nonnegativity of the Huber envelope gives the baseline bound

$$\Delta\Phi_\beta(t; u, v) \ge a_{uv}(t) - H_0. \tag{9}$$

For a declared tail loss $\tau_{uv} \ge 0$, define

$$T_{uv}(\tau_{uv}) = a_{uv}^{-1}([H_0 - \tau_{uv}]_+). \tag{10}$$

Because $\mu > 0$ and $p > 0$, a_{uv} is continuous, strictly increasing on $[0, \infty)$, and diverges. Hence this unique root is finite, including when $\gamma = 0$, $Xu = 0$, or $H_0 = 0$, and obeys $T_{uv} \le ([H_0 - \tau_{uv}]_+/\mu)^{1/p}$. For every $t \ge T_{uv}$, Eq. (9) gives $\Delta\Phi_\beta(t; u, v) \ge -\tau_{uv}$. Huber Lipschitzness optionally tightens the lower bound to $a_{uv}(t) - \min\{H_0, b_{uv}t\}$; a failed tightening returns to the baseline bound and cannot create an unbounded status. A finite-family-full result requires scanning each screened direction through its own bound and checking both the family scan error and $\max_d \tau_d \le \varepsilon_A$.

The certificate object is explicitly $(\Phi_\beta, \mathcal{D}_t, [0, T_{\text{obs}}], \eta_{\text{scan}})$; it is not a certificate for F unless the logged envelope gap $\Delta_\beta = F - \Phi_\beta$ is added to the reported tolerance.

For a screened direction family $\mathcal{D}_t$, define

$$\begin{aligned} G_t(u, v) &= \max_{0 \le \tau \le T_{\text{obs}}} [-\Delta\Phi_\beta(\tau; u, v)]_+, \\ A_t^{\mathcal{D}} &= \max_{(u,v)\in\mathcal{D}_t} G_t(u, v), \; A_t^\star(T_{\text{obs}}) = \sup_{u\perp U, v\perp V} G_t(u, v). \end{aligned} \tag{11}$$

The code uses the first k projected singular pairs of the normal gradient as a finite family, evaluates every pair on the same grid, and records the winning `direction_rank` and family size. It records `candidate_gain` $A_t^{\mathcal{D}}$ and leaves `ambient_reference` unset unless an ambient search is actually run. In particular, $\eta_{\text{dir}} = 0$ means exact evaluation of the screened pair, not zero error over all directions. The runtime aggregates the whole declared family before assigning a scope:

$$\eta_{\text{family}} = \max_{d\in\mathcal{D}_t} \eta_{\text{scan,d}}, \qquad U_{\text{family}} = \max_{d\in\mathcal{D}_t} \{\widehat{G}_d + \eta_{\text{scan,d}}, \tau_d\},$$

and `finite-family-full` requires $\text{all}_d(\texttt{tail_scanned}_d \wedge \texttt{tail_bound_exists}_d)$. The stopping test uses $U_{\text{family}} \le \varepsilon_A$ together with this closure flag; a winning direction alone can never promote the family.

Algorithm 1 Selective certificate controller.

Require: $Z^0 = Q_0 H^0$, thresholds $(\varepsilon_Z, \varepsilon_A)$, declared candidate directions, T_{obs}, a descent ledger, and a compute ledger.
Compute (a, e, r_Z, W) and the strict scan error ledger.
if $r_Z > \varepsilon_Z$ and $a \le \theta r_Z$ **then**
    EXPAND with a block capturing ζe^2.
end if
if $r_Z > \varepsilon_Z$ and $a > \theta r_Z$ **then**
    CONTINUE with an accepted GSVT step.
end if
if the candidate scan has certified positive gain **then**
    ACTIVATE the best accepted normal move and charge its decrease.
end if
if coverage, amplitude, rank, readout, or resource conditions fail **then**
    ABSTAIN with the typed failure and the spent ledger.
end if
Return factors, residual split, candidate gain, scope, error terms, and task readout.

Strict grid error. For grid spacing h and $w = Xu$, the Huber envelope is λ-Lipschitz, $|t^p - s^p| \le |t - s|^p$, and $|q_j(t) - q_j(s)| \le |t - s|\,|v_j|\|w\|_2$. Therefore, on $[0, T_{\mathrm{obs}}]$,

$$\eta_{\mathrm{grid}}^{\mathrm{cert}} = \mu(h/2)^p + \frac{h}{2}\left(\gamma \ell T_{\mathrm{obs}} + \lambda \|Xu\|_2 \|v\|_1\right). \tag{12}$$

The finite-difference slope is logged only as $\eta_{\mathrm{grid}}^{\mathrm{empirical}}$; it is not used as a certificate. With $\eta_{\mathrm{scan}} = \eta_{\mathrm{dir}} + \eta_{\mathrm{grid}}^{\mathrm{cert}} + \eta_{\mathrm{arith}}$,

Theorem 3.2 (Candidate or ambient certificate)**.** *If $\widehat{A}_t^{\mathcal{D}}$ is within η_{scan} of $A_t^{\mathcal{D}}$ and $A_t^{\mathcal{D}} \ge \kappa_{\mathrm{cov}} A_t^{\star}(T_{\mathrm{obs}})$, then*

$$A_t^{\star}(T_{\mathrm{obs}}) \le \frac{\widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}}}{\kappa_{\mathrm{cov}}}.$$

Thus a candidate no-gain decision requires $\widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}} \le \varepsilon_A$; a global no-escape decision at tolerance ε_A additionally requires $\widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}} \le \kappa_{\mathrm{cov}} \varepsilon_A$. Without the coverage premise the output remains scoped to the finite family. If any screened direction has $T_{\mathrm{obs}} < T_{uv}(\tau_{uv})$, the output carries the truncation flag; this check is direction dependent. Extending a finite-family result to the whole amplitude half-line additionally requires $\max_d \tau_d \le \varepsilon_A$ and the family scan-error test.

Proof. The scan error gives $A_t^{\mathcal{D}} \le \widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}}$; the coverage inequality then gives the displayed ambient bound. □

Operations kept outside the main certificate. A rank-preserving correction is an optional appendix operation, not a second main contribution. Its local quadratic bound is

$$\Phi_\beta(Z(T)) - \Phi_\beta(Z) \le -\langle R, T\rangle + \tfrac{1}{2}\langle T, \mathcal{A}_c[T]\rangle. \tag{13}$$

It preserves the full right row space, whereas a normal activation adds $vv^\top$ to the right projector. Truncated readout conditions and the correction proof are in Appendix C; they are not used to upgrade a candidate scan to an ambient certificate.

Ordinary spectral updates. At scale s, form the exact scalar GSVT trial $H^+ = \mathrm{GSVT}_{\mu/s}(H - Q^\top G_\beta / s)$ and accept it when

$$\psi_\beta(Z + D) \le \psi_\beta(Z) + \langle G_\beta, D\rangle + \frac{\nu s}{2}\|D\|_F^2,\ 0 < \nu < 1. \tag{14}$$

A rejection multiplies s by $q_s > 1$ and is recorded with its majorant slack, accepted scale, and trial count; each trial has its own compute cost. The one-step rank and exact full-objective scan checks used by the appendix audit are stated in Section 4.

4 WORK ACCOUNTING AND TYPED ABSTENTION

The controller distinguishes model assumptions, oracle operations, and runtime checks. Assume X is fixed, $L \succeq 0$, $0 < p < 1$, and $0 < \beta_0 \le \beta_1 \le \cdots \le \beta_T$ with $\beta_T \ge \beta_0$; the continuation schedule is finite, ordinary trials pass the displayed majorant test, expansions capture at least ζe^2, and every accepted activation is rechecked in both Φ_β and F. All terms in Φ_β are nonnegative, so $\Phi_\beta \ge 0$. These are exact-oracle conditions; replay checks test the saved instances but do not turn a nonconvex run into a global optimum.

Let

$$B_T = \Phi_{\beta_0}(Z^0) + C_\lambda(\beta_0^{-1} - \beta_T^{-1}),\ \bar{s}_T = \max\{s_0, q_s \rho_{\beta_T}/\nu\}.$$

Theorem 4.1 (Base Continue/Expand work)**.** *Under the stated oracle conditions, each fixed-space stage either reaches $r_Z \le \varepsilon_t$ or returns a typed expansion, rank, or resource failure. If m, K, J count accepted expansions, accepted positive-displacement ordinary steps, and rejected trials, then*

$$m \le \operatorname{rank}(X) - h_0,\ K \le T + 1 + m + \left\lceil \frac{2B_T \bar{s}_T}{(1-\nu)\theta^2 \varepsilon_Z^2} \right\rceil, J \le \left\lceil \frac{\log \max\{1, \rho_{\beta_T}/(\nu s_0)\}}{\log q_s} \right\rceil. \quad (15)$$

The bound charges accepted decreases and continuation once; zero trials and failed checks are recorded rather than silently counted as convergence. This is the descent ledger B_Φ. The implementation reports an independent compute ledger C_{spent} that counts scan evaluations, rejected majorant trials, expansion SVD work, and accepted moves; it is not subtracted from B_Φ.

Proof. Each accepted ordinary step decreases the envelope by at least $(1-\nu)\theta^2\varepsilon_Z^2/(2\bar{s}_T)$ unless it finishes a stage. There are at most $T + 1 + m$ exceptional outputs, at most m expansions, and the scale can grow only geometrically. Telescoping the continuation changes gives B_T. □

Conditional budget versus replay floor. The displayed work bound is a conditional symbolic statement: it requires a global $\bar{s}_T$ and the exact-oracle conditions above. It is an accepted-operation accounting theorem under those conditions, not an end-to-end complexity guarantee for the nonconvex controller. The frozen replay does not instantiate such a global constant. It therefore enforces and audits the numerical floor $\delta_{\mathrm{ord,min}} = 10^{-10}$, while retaining δ_{ord} as a diagnostic rather than presenting it as an empirical iteration certificate.

Theorem 4.2 (Certificate-or-abstention budget)**.** *Let $\delta_{\mathrm{act}} > 0$ be the minimum accepted candidate gain. Let $N_{\mathrm{ord}}^{\mathrm{cont}}$ count accepted ordinary Continue steps, and let N_{act} count accepted activations whose recheck satisfies $\Delta\Phi_i \ge \delta_{\mathrm{act}} > 0$. With $\eta_{\mathrm{scan}} = \eta_{\mathrm{dir}} + \eta_{\mathrm{grid}}^{\mathrm{cert}} + \eta_{\mathrm{arith}}$,*

$$N_{\mathrm{ord}}^{\mathrm{cont}} \delta_{\mathrm{ord}} + N_{\mathrm{act}} \delta_{\mathrm{act}} \le B_T,\ \delta_{\mathrm{ord}} = \frac{(1-\nu)\theta^2\varepsilon_Z^2}{2\bar{s}_T}. \quad (16)$$

A successful return reports r_Z, r_F, candidate gain, $[0, T_{\mathrm{obs}}]$, all error terms, and its scope. A screened direction set gives a finite-family result. It is a prefix result when any direction has $T_{\mathrm{obs}} < T_{uv}(\tau_{uv})$; a full-family pass additionally requires $\max_d\{\widehat{G}_d + \eta_{\mathrm{scan,d}}, \tau_d\} \le \varepsilon_A$. It becomes an ambient certificate only with a stated coverage factor κ_{cov}. The object certified here is Φ_β. Since $0 \le F - \Phi_\beta \le \Delta_\beta = \lambda^2 N/(2\beta)$, an F-level no-gain claim must use the conservative tolerance $\varepsilon_A + \eta_{\mathrm{scan}} + \Delta_\beta$ and report Δ_β. If the scan domain is truncated, the direction coverage is unavailable, the rank is zero, a readout margin fails, or the budget is exhausted, the output is respectively 'TRUNCATED-AMPLITUDE', 'COVERAGE', 'ZERO-RANK', 'READOUT-MARGIN', or 'RESOURCE' abstention. None of these outputs is labelled convergence. In the frozen replay, the symbolic Lipschitz quantity $\bar{s}_T$ is not instantiated globally. The worker therefore enforces the declared numerical floor $\delta_{\mathrm{ord,min}} = 10^{-10}$ on every accepted ordinary step and records an 'ORDINARY–DELTA–FAILURE' otherwise; the theorem value δ_{ord} remains a diagnostic until a global $\bar{s}_T$ is supplied.

Proof. Every accepted move pays its displayed charge and expansions leave the matrix unchanged while consuming at most $\operatorname{rank}(X) - h_0$ directions. Summation of charges gives the budget. The certificate scope follows directly from Theorem 3.2. □

Theorem 4.3 (Residual-only insufficiency; finite-scan sufficiency)**.** *For a current state, let $G_t(u,v)$ be the finite-amplitude gain in* (11)*, let $A_t^{\mathcal{D}} = \max_{(u,v)\in\mathcal{D}_t} G_t(u,v)$, and let $A_t^\star = \sup_{u\perp U, v\perp V} G_t(u,v)$ on a declared amplitude interval containing $t=1$. Then:*

1. *A residual-only stopping rule cannot imply $A_t^\star = 0$: the parameterized $X = I_2$, $L = 0$, $Z = 0$ witness in Proposition B.1 has $r_Z = 0$ and a strictly positive finite gain.*
2. *If a finite family satisfies $|\widehat{A}_t^{\mathcal{D}} - A_t^{\mathcal{D}}| \le \eta_{\mathrm{scan}}$, then the candidate no-gain condition is $\widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}} \le \varepsilon_A$, which implies $A_t^{\mathcal{D}} \le \varepsilon_A$. If, in addition, a declared coverage premise $A_t^{\mathcal{D}} \ge \kappa_{\mathrm{cov}} A_t^\star$ holds for $\kappa_{\mathrm{cov}} > 0$, then*

$$A_t^\star \le \frac{\widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}}}{\kappa_{\mathrm{cov}}}. \tag{17}$$

In particular, a global no-escape claim at tolerance ε_A requires the stronger inequality $\widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}} \le \kappa_{\mathrm{cov}}\varepsilon_A$.

Thus residual-only reasoning is insufficient, while finite scanning plus an explicit coverage premise is sufficient for the stated ambient bound.

Proof. The first claim is exactly the strict decrease in Proposition B.1. For the second, the scan error gives $A_t^{\mathcal{D}} \le \widehat{A}_t^{\mathcal{D}} + \eta_{\mathrm{scan}}$; substitute the stopping inequality and divide the coverage premise by κ_{cov}. No coverage factor is inferred from a finite calibration. □

Certificate interface. The four public states are therefore CONTINUE, EXPAND, ACTIVATE, and CERTIFY-or-ABSTAIN. This is the main output contract. Each record carries both the descent ledger B_Φ and the independent compute total C_{spent}, so a low objective cost is never presented as a low runtime cost; correction and detailed readout geometry are optional appendix operations rather than independent claims.

Conditional readout and rank checks. The deployed readout uses V_k, $k = \min(8, r)$. If a correction induces $A^+ = A + E_A$ and $\delta_k = \lambda_k(A) - \lambda_{k+1}(A) > 0$, then

$$\|\sin\Theta(V_k^+, V_k)\|_2 \le \frac{2\|E_A\|_2}{\delta_k} \quad (\|E_A\|_2 < \delta_k/2). \tag{18}$$

If $m_k = \min_i \|V_{i,1:k}\|_2 > 0$, row normalization and a declared decision margin give the sufficient bound

$$\max_i \left\| \frac{V^+_{i,1:k}}{\|V^+_{i,1:k}\|_2} - \frac{V_{i,1:k}R}{\|V_{i,1:k}\|_2} \right\|_2 \le \frac{4\sqrt{2k}}{m_k}\frac{\|E_A\|_2}{\delta_k}. \tag{19}$$

When either margin is missing, the interface records READOUT-MARGIN abstention. For the one-step rank audit, use

$$\min_{i\le r}\{s\sigma_i + \mu p\sigma_i^{p-1}\} > b(s) + r_Z,\ \|G_N\|_2 < b(s) - r_Z. \tag{20}$$

For a normal pair, write $C = X - XZ$, $w = Xu$, $n_j = \|C_{:j}\|_2$, and $q_j(t) = \|C_{:j} - tv_j w\|_2$; h_β denotes the scalar Huber envelope defined in Eq. (7). The exact scan used by the audit is

$$\Delta F(t) = \mu t^p + \frac{\gamma\ell}{2}t^2 + \lambda\sum_j (q_j(t) - n_j),\ \Delta\Phi_\beta(t) = \mu t^p + \frac{\gamma\ell}{2}t^2 + \sum_j (h_\beta(q_j(t)) - h_\beta(n_j)). \tag{21}$$

These conditions are checks on the declared operation and do not certify task quality by themselves.

5 Experiments: selective certificates and task readouts

We evaluate the certificate interface with a frozen panel and two mechanism audits. All image runs use six public-data feature pools, three shared initialization seeds, $p = .5$, $\mu = .02$, $\gamma = .03$, $\lambda = N^{-1/2}$, $(\varepsilon_Z, \varepsilon_F) = (.003, .05)$, and $\beta = 20$. Act-OFF and Act-ON share the Continue/Expand worker and disable correction; Act-ON alone scans a top-4 normal family. Ordinary trials use the smooth-majorant test with $\nu = .9$ and $q_s = 1.5$.

Table 1: Scope ledger. Endpoint evaluations and smoke checks are reported separately. No ambient coverage factor is inferred. – indicates that the row is not split by arm.

quantity	Act-OFF	Act-ON	all rows
saved endpoint evaluations	18	18	36
one-iteration smoke runs	6	6	12
accepted activations	0	55	55
finite-family-full	0	7	7
tail-closed/resource abstention	–		21
prefix/truncated	–		8
unbounded-tail rows	–		0

Table 2: Descent and compute ledgers for the saved endpoint panel.

quantity	Act-OFF	Act-ON
trajectories	18	18
accepted activations	0	55
expansion events	396	
median C_{spent}	2.97×10^6	3.47×10^6

Why residual-only stopping is insufficient. For the I_2 witness of Section 1 with $(p, \mu, \lambda, \beta) = (.5, .01, .1, 10)$, the limiting-subdifferential residual is $r_Z = 0$, yet $\Delta\Phi_\beta(1) = -.0895$; this fixes the distinction between stationarity and no finite-gain escape.

Scope of claims. The public interface has four typed outputs. Continue/Expand reports the exact $r_Z^2 = a^2 + e^2$ split and a post-projection capture ratio; a finite-family-prefix state bounds only the observed interval and declared family; a finite-family-full state additionally closes every screened direction and the family error budget; an ambient no-escape state would require a coverage premise that is not assumed here. The audit counts are summarized in Table 1; aggregate rows are reported over all 36 endpoint records.

Endpoint panel and independent evaluation. The saved endpoint panel contains 36 trajectories and 12 one-iteration smoke checks. The family tail evaluation uses the nonnegative-Huber baseline $a_{uv}(T_{uv}) = [H_0 - \tau_{uv}]_+$, with $a_{uv}(t) = \mu t^p + c_{uv} t^2/2$, and aggregates all screened directions before assigning a scope. Every saved direction has a finite baseline closure; the resulting scope counts are the finite-family-full, tail-closed, and prefix rows of Table 1, with no unbounded classification. These are scoped finite-family outcomes for Φ_β on the declared family only; they are not ambient claims.

Table 2 gives the corresponding descent and compute ledgers. Every accepted activation decreases both logged objectives: the mean accepted drop is $\Delta\Phi = .02163$ and $\Delta F = .02163$ (median .01742), with no negative ΔF. At $\beta = 20$ the endpoint gap $F - \Phi_\beta = 0.025$ is reported explicitly. Across accepted ordinary steps, all majorant gaps are at most -8.6×10^{-11} after backtracking; 76 rejected trials are recorded. The 396 expansion events have capture ratio $\|B^\top W\|_F^2 / \|W\|_F^2$ from 0.402 to 0.978 (median 0.656), above the declared target $\zeta = .4$. Ordinary updates enforce $\delta_{\mathrm{ord,min}} = 10^{-10}$ and no row violated this floor; the symbolic theorem value remains diagnostic because a global $\bar{s}_T$ is not instantiated for the frozen panel.

The paired Act-ON minus Act-OFF envelope change is -0.00482 (18 input–seed units; bootstrap 95% interval $[-0.00682, -0.00308]$). The corresponding task changes are ARI $+0.00493$, AP $+0.03130$, and AUROC -0.00498; none has a 6-bundle interval excluding zero. These task readouts are not optimization witnesses.

Published baseline and matched statistical unit. We ran the published robust low-rank representation (LRR) program $\min_{Z,E} \|Z\|_* + \lambda \|E\|_{2,1}$ subject to $X = XZ + E$ (Liu et al., 2013) on the same six frozen bundles. LRR-u uses the standard unit-column normalization; LRR-r reruns the identical solver on raw columns. Each has three deterministic bookkeeping repeats per bundle and 18/18 converged solves. Table 3 first averages repeats within a bundle, then uses 10,000 paired

Table 3: Absolute downstream readouts and 6-bundle paired contrasts. LRR-u/r denote unit-column/raw input. Brackets are 95% percentile intervals for the indicated controller minus the LRR variant; $n = 6$ bundles. The ON–OFF contrast is reported separately in Appendix Table 5 and is the only activation contrast.

	Act-OFF	Act-ON	LRR-u	LRR-r	OFF–LRR-u [95% CI]	OFF–LRR-r [95% CI]	ON–LRR-u [95% CI]	ON–LRR-r [95% CI]
ARI	.6793	.6842	.2878	.3770	+.3914 [.2533,.5434]	+.3023 [.1346,.4559]	+.3964 [.2471,.5522]	+.3072 [.1461,.4613]
AP	.4806	.5119	.2081	.3030	+.2725 [-.0112,.5527]	+.1776 [-.0364,.4189]	+.3038 [-.0018,.5801]	+.2089 [-.0364,.4819]
AUROC	.7335	.7285	.3803	.4326	+.3531 [.0505,.6640]	+.3009 [.0149,.5868]	+.3481 [.0294,.6632]	+.2959 [-.0013,.5982]

Table 4: Executed finite-domain grid replay and declared sensitivity axes. The grid row is the saved endpoint diagnostic; the other axes remain protocol parameters without additional measurements.

axis	values	recorded output
grid count	129, 257, 513, 1025	η_{cert} medians .02252/.01226/.00684/.00392; empirical .01731/.00866/.00433/.00217
tail tolerance	$10^{-8}, 10^{-10}, 10^{-12}$	protocol axis
continuation	baseline β plus one sensitivity value	protocol axis

bootstrap resamples of the six bundles (seed 1729); it never treats the 18 LRR rows as independent inputs.

The table is an external-method contrast, not an activation ablation: the Act-ON–Act-OFF interval is reported independently and crosses zero for all three task readouts. The published LRR objective differs from our raw-X Schatten-p, graph-regularized Huber-envelope objective; LRR-u additionally changes input scale. The comparison therefore concerns downstream readout, not common- objective optimality. Raw-panel F, Φ_{20}, r_Z, r_F evaluated at LRR's returned Z are descriptive audit values (Appendix D.4). Certificate availability is a separate interface capability: `no-finite-family-scan` means no declared scan was run, not a failed LRR solve. LRR wall-clock and our C_{spent} count different resources; we make no cross-method timing claim.

Tail, margin, and grid audit. The baseline tail closure uses $H_0 = \sum_j h_{\lambda,\beta}(q_j(0))$, $a_{uv}(t) = \mu t^p + c_{uv} t^2/2$, and a declared loss τ_{uv}. The family upper bound is aggregated over every screened direction as $U_{\mathrm{family}} = \max_d\{\widehat{G}_d + \eta_{\mathrm{scan,d}}, \tau_d\}$. The numerical decision uses the recomputed U_{family}; rows failing the full scan or error budget abstain. The strict grid error is $\eta^{\mathrm{cert}}_{\mathrm{grid}} = \mu(h/2)^p + \frac{h}{2}(\gamma \ell T_{\mathrm{obs}} + \lambda \|Xu\|_2 \|v\|_1)$. The sensitivity axes are finite-domain resolution diagnostics; they do not establish ambient coverage. In the executed replay, the four grid counts use the same 36 saved endpoint states and $T_{\mathrm{obs}} = 15$; the reported medians are numerical-resolution summaries, not certificate counts. Scope labels continue to use the family-aggregate error and tail-loss checks.

Same-objective finite oracle. On 12 rank-one 3×3 states with the same robust objective, the two projected leading pairs attain mean/minimum gain recall $0.9369/0.7689$ against a dense finite angular oracle while using $512\times$ fewer objective evaluations. The full protocol and coverage/feasibility calibrations remain in Appendices D.2 and D.3; this finite reference is not an ambient theorem.

Interpretation. The evidence supports a scope-aware certificate interface: the implementation enforces and logs the theorem's majorant, expansion capture, F-level recheck, tail domain, and compute cost. A finite-family-full certificate is a positive, finite-domain result for Φ_β only after the corrected amplitude closure; missing coverage or closure is reported as typed abstention, not a global no-escape statement.

Conclusion. Reduced-space convergence becomes an original-space claim only after the omitted residual and finite-amplitude escape are checked. Our interface makes that distinction executable: $r_Z^2 = a^2 + e^2$ chooses Continue or Expand, the bounded scan reports gain with a strict error, and separate ledgers report descent and compute cost. It returns a certificate for the declared $(\Phi_\beta, \mathcal{D}, [0, T], \eta)$ object or a typed ABSTAIN when coverage, rank, amplitude, readout, or resources are missing. No ambient no-escape statement is made without a declared coverage premise and a closed amplitude domain, and downstream task metrics stay separate from optimization evidence.

REPRODUCIBILITY AND DATA AVAILABILITY

Code, frozen inputs, and run records for all reported experiments are included in the supplementary materials, together with machine-readable checks that tie the saved records to the stated claims. Appendix D specifies fixed features, initialization pairing, timing scopes, estimator formulas, and failure handling. Complete proofs are given in Appendices B and C. Public dataset and encoder sources are recorded with the feature procedure; no unpublished patient data are used.

AI USE STATEMENT

Generative AI tools were used to aid manuscript restructuring, language editing, and documentation. We have reviewed all AI-assisted work. The mathematical results, proofs, computational checks, and reported numbers were produced and verified by the authors, who take full responsibility for the final content of the paper.

ETHICS STATEMENT

The experiments use controlled synthetic inputs and established public benchmarks to study numerical optimization, clustering, and outlier ranking. Outlier labels in these tasks are defined by the experimental class split; they are not judgments about people. The reported scores are benchmark measurements and are not calibrated decision rules for clinical or other consequential deployment.

REFERENCES

Paris Giampouras, René Vidal, Athanasios Rontogiannis, and Benjamin D. Haeffele. A novel variational form of the Schatten-p quasi-norm. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21453–21463, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f53eb4122d5e2ce81a12093f8f9ce922-Abstract.html.

Cho-Jui Hsieh and Peder A. Olsen. Nuclear norm minimization via active subspace selection. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 575–583. PMLR, 2014. URL https://proceedings.mlr.press/v32/hsiehb14.html. Part 1.

Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013. doi: 10.1109/TPAMI.2012.88. URL https://doi.org/10.1109/TPAMI.2012.88.

Canyi Lu, Changbo Zhu, Chunyan Xu, Shuicheng Yan, and Zhouchen Lin. Generalized singular value thresholding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1):1805–1811, 2015. doi: 10.1609/aaai.v29i1.9464. URL https://doi.org/10.1609/aaai.v29i1.9464.

Weiping Shen, Linglingzhi Zhu, Yaohua Hu, Chong Li, and Xiaoqi Yang. Dynamic proximal gradient algorithms for Schatten-p quasi-norm regularized problems, 2026. URL https://arxiv.org/abs/2603.00333v1. arXiv preprint arXiv:2603.00333v1.

Quanming Yao, James T. Kwok, and Wenliang Zhong. Fast low-rank matrix learning with nonconvex regularization. In *2015 IEEE International Conference on Data Mining*, pp. 539–548. IEEE, 2015. doi: 10.1109/ICDM.2015.9. URL https://cse.hkust.edu.hk/~jamesk/papers/icdm15.pdf.

Quanming Yao, James T. Kwok, Taifeng Wang, and Tie-Yan Liu. Large-scale low-rank matrix learning with nonconvex regularizers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(11):2628–2643, 2019. doi: 10.1109/TPAMI.2018.2858249. URL https://arxiv.org/abs/1708.00146v3. Author manuscript arXiv:1708.00146v3, revised July 2018.

Ming Yin, Junbin Gao, and Zhouchen Lin. Laplacian regularized low-rank representation and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):504–517, 2016. doi: 10.1109/TPAMI.2015.2462360. URL https://doi.org/10.1109/TPAMI.2015.2462360.

Hengmin Zhang, Jian Yang, Fanhua Shang, Chen Gong, and Zhenyu Zhang. LRR for subspace segmentation via tractable Schatten-p norm minimization and factorization. *IEEE Transactions on Cybernetics*, 49(5):1722–1734, 2019. doi: 10.1109/TCYB.2018.2811764. URL https://doi.org/10.1109/TCYB.2018.2811764.

A POSITIONING AND SCOPE

Reduced-space spectral methods avoid materializing a full self-representation matrix and are therefore natural for large sample sets (Hsieh & Olsen, 2014; Yao et al., 2015; 2019). GSVT and Schatten-p proximal rules provide exact fixed-space spectral updates (Lu et al., 2015; Giampouras et al., 2020); dynamic proximal methods refine the same fixed-space optimization family (Shen et al., 2026); and FaNCL supplies a related active-space computational interface. These lines of work address how to compute inside a chosen space. Our question is different: when is stationarity in that space evidence about the original matrix problem?

The distinction matters because a reduced residual can vanish while an omitted normal direction has a finite-amplitude decrease. The paper makes that missing information observable through the exact split $r_Z^2 = a^2 + e^2$, then couples it to a declared finite scan and a typed certificate-or-abstention output. The result is deliberately narrower than a global optimization claim: a screened direction set and a truncated amplitude interval produce a finite-family-prefix certificate, while an ambient statement requires an explicit coverage factor.

The experiments consequently separate optimization evidence from task readouts. The Act-OFF/Act-ON panel tests whether the finite-amplitude event is visible and charged; the finite-grid panel calibrates the coverage risk; the continuation and FaNCL panels check feasibility and stop-rule alignment. Additional tables and complete run ledgers are available with the reproducibility materials.

B PROOFS FOR THE SELECTIVE CERTIFICATE INTERFACE

All matrices are real; unspecified matrix norms below are Frobenius norms. Let $g(Z) = \mu \sum_i \sigma_i(Z)^p$, $f(E) = \lambda \|E\|_{2,1}$, and

$$F(Z) = g(Z) + \frac{\gamma}{2} \operatorname{tr}(ZLZ^\top) + f(X - XZ).$$

For the quadratic envelope $m_\beta(C) = \min_E \{f(E) + \beta \|E - C\|_F^2/2\}$, define $\Phi_\beta(Z) = g(Z) + \gamma \operatorname{tr}(ZLZ^\top)/2 + m_\beta(X - XZ)$. Columnwise shrinkage gives

$$E_{\beta,:j} = \left(1 - \frac{\lambda}{\beta \|C_{:j}\|_2}\right)_+ C_{:j}, \quad \Lambda_\beta = \beta(E_\beta - C), \quad G_\beta = \gamma ZL + X^\top \Lambda_\beta. \tag{22}$$

The envelope is smooth with Lipschitz bound $\rho_\beta = \beta \|X\|_2^2 + \gamma \lambda_{\max}(L)$, and

$$r_F = \|XZ + E_\beta - X\|_F \le \lambda \sqrt{N}/\beta, \; 0 \le F(Z) - \Phi_\beta(Z) \le \frac{\lambda^2 N}{2\beta}. \tag{23}$$

These are feasibility and envelope-gap statements; r_Z is the stationarity residual of Φ_β, not of the nonsmoothed objective F.

Proposition B.1 (Residual-only stopping can miss finite gain)**.** *Let $X = I_2$, $L = 0$, $Z_0 = 0$, $0 < p < 1$, $0 < \lambda < \beta$, and $0 < \mu < \lambda - \lambda^2/(2\beta)$. Then $r_Z(Z_0) = 0$, but for $u = v = e_1$,*

$$\Phi_\beta(e_1 e_1^\top) - \Phi_\beta(0) = \mu - \lambda + \frac{\lambda^2}{2\beta} < 0.$$

Proof. At the origin the limiting subdifferential of g is the whole matrix space: $g(Y)/\|Y\|_F \ge \mu \|Y\|_F^{p-1} \to \infty$. Hence the smooth envelope gradient can be canceled and $r_Z = 0$. At amplitude one the first residual column is removed and the Huber value is $\lambda - \lambda^2/(2\beta)$, which gives the displayed decrease. □

B.1 SPECTRAL CALCULUS USED BY THE CONTROLLER

For a positive-rank thin SVD $Z = U\Sigma V^\top$, put $D_p = \mu p U \Sigma^{p-1} V^\top$ and $G_N = (I - UU^\top) G (I - VV^\top)$. The limiting spectral subdifferential is

$$\partial g(Z) = D_p + \{A : U^\top A = 0, \; AV = 0\}, \operatorname{dist}(0, \partial g(Z) + G) = \|D_p + G - G_N\|_F.$$

At $Z = 0$, $\partial g(0)$ is the whole space. The formula follows by splitting the rank-r tangent and normal blocks; the singular-value term dominates every fixed linear functional in the zero-rank limit. Repeated positive singular values do not change the block formula.

For $s > 0$, define

$$\tau_p(s) = \left[\frac{2\mu(1-p)}{s}\right]^{1/(2-p)}, \ \vartheta_p(s) = \tau_p(s) + \frac{\mu p}{s}\tau_p(s)^{p-1}.$$

The global scalar minimizer of $\mu x^p + s(x-y)^2/2$ is zero for $y \le \vartheta_p(s)$ and otherwise is the larger positive root of $x + (\mu p/s)x^{p-1} = y$. Applying this rule to singular values gives the global GSVT trial used by the ordinary-update branch. The result follows from the scalar comparison and von Neumann's trace inequality.

B.2 Original-space omission and expansion

Let $Z = QH = U\Sigma V^\top$, $Q^\top Q = I$, and define

$$a = \operatorname{dist}(0, \partial_H \Phi_\beta(QH)), \ W = (I - QQ^\top)X^\top \Lambda_\beta V, \ e = \|W\|_F.$$

The internal and external normal blocks are orthogonal, so

$$r_Z^2 = a^2 + e^2. \tag{24}$$

If $B \perp Q$ is appended and zero rows are added to H, the represented matrix and objective do not change, while

$$e_+^2 = e^2 - \|B^\top W\|_F^2, \ a_+^2 = a^2 + \|B^\top W\|_F^2. \tag{25}$$

Thus $a \le \theta r_Z$ implies a nonzero omitted component and justifies EXPAND; otherwise the internal residual is large enough for CONTINUE. The identity is Pythagoras applied to the minimum-norm spectral subgradient.

Proposition B.2 (Quantitative row-space omission bound)**.** *Assume $M > 0$ and $\|Z\|_2 \le M$. With P_X the orthogonal projector onto* $\operatorname{range}(X^\top)$*,*

$$\Phi_\beta(Z) - \Phi_\beta(P_X Z) \ge \frac{\mu p}{2} M^{p-2} \|(I - P_X)Z\|_F^2. \tag{26}$$

When $M = 0$, the conclusion is trivial because $Z = 0$.

Proof. Put $q = p/2$, $S = Z^\top P_X Z$, and $T = Z^\top (I - P_X)Z$. Then $S, T \succeq 0$, $S + T = Z^\top Z \preceq M^2 I$, and the spectral term is $\mu \operatorname{tr}(\cdot)^q$. For $\varphi(t) = \operatorname{tr}(S + tT)^q$, differentiate first at $S + \epsilon I$ and let $\epsilon \downarrow 0$ to obtain

$$\varphi'(t) = q \operatorname{tr}\left[(S + tT)^{q-1} T\right].$$

Because $q - 1 < 0$ and $S + tT \preceq M^2 I$, operator monotonicity gives $(S + tT)^{q-1} \succeq M^{2q-2} I$. Hence

$$\operatorname{tr}(S+T)^q - \operatorname{tr} S^q = \int_0^1 \varphi'(t)\, dt \ge q M^{2q-2} \operatorname{tr} T \quad = \frac{p}{2} M^{p-2} \|(I - P_X)Z\|_F^2.$$

The residual term is unchanged because $XP_X = X$, and the graph term does not increase under the left orthogonal projection because $L \succeq 0$. Multiplying the spectral inequality by μ proves the claim. □

B.3 Finite certificates and separate ledgers

At stage t, use the exact GSVT trial and accept it only after the displayed smooth majorant test in the main text. Let B_T be the initial envelope plus the telescoping continuation increment, and let $\bar{s}_T$ bound all carried inverse step sizes. Every accepted ordinary step that neither ends a stage nor triggers expansion decreases the envelope by at least

$$\delta_{\mathrm{ord}} = \frac{(1-\nu)\theta^2 \varepsilon_Z^2}{2\bar{s}_T}.$$

There are at most $T + 1 + m$ exceptional outputs for m expansions, and rejected trials are bounded by the geometric scale schedule. Summing charges gives

$$K \le T + 1 + m + \left\lceil \frac{2B_T \bar{s}_T}{(1-\nu)\theta^2 \varepsilon_Z^2} \right\rceil, \; m \le \operatorname{rank}(X) - h_0. \tag{27}$$

The same ledger adds $N_{\mathrm{act}}\delta_{\mathrm{act}}$ for accepted finite-gain activations. If the budget, rank, amplitude, coverage, or readout condition fails, the controller returns the achieved quantities and a typed abstention; it does not relabel the run as convergence.

Nonzero output check. The sufficient condition

$$F(0) - \left\{ \Phi_{\beta_0}(Z^0) + \frac{\lambda^2 N}{2\beta_0} \right\} > 0 \tag{28}$$

excludes the zero output along an accepted finite run by the same telescoping argument. It is a check on the original objective, not a rank or task-quality claim.

C Exact normal scan and certificate scope

Let $Z = U\Sigma V^\top$, $u \perp U$, $v \perp V$, and $Z_t = Z + tuv^\top$. Write $C = X - XZ$, $w = Xu$, $\ell = v^\top Lv$, and

$$n_j^2 = \|C_{:j}\|_2^2, \; \alpha_j = v_j C_{:j}^\top w, \; \omega_j = v_j^2 \|w\|_2^2, \; q_j(t) = \sqrt{n_j^2 - 2\alpha_j t + \omega_j t^2}.$$

The orthogonal rank-one factors add one singular value t and eliminate the graph cross term. Therefore the original and envelope changes are exactly

$$\begin{aligned} F(Z_t) - F(Z) &= \mu t^p + \frac{\gamma \ell}{2} t^2 + \lambda \sum_j (q_j(t) - n_j), \\ \Phi_\beta(Z_t) - \Phi_\beta(Z) &= \mu t^p + \frac{\gamma \ell}{2} t^2 + \sum_j \{ h_{\lambda,\beta}(q_j(t)) - h_{\lambda,\beta}(n_j) \}. \end{aligned} \tag{29}$$

After preparing Xu, $C^\top Xu$, and Lv, each amplitude costs $O(N)$; the scan is therefore auditable without constructing the full matrix.

The implementation declares $H_0 = \sum_j h_{\lambda,\beta}(n_j)$, an allowed tail loss $\tau_{uv} \ge 0$, and the observed prefix $[0, T_{\mathrm{obs}}]$. With $c_{uv} = \gamma v^\top Lv$ and $a_{uv}(t) = \mu t^p + c_{uv} t^2/2$, nonnegativity of the Huber envelope gives $\Delta\Phi_\beta(t) \ge a_{uv}(t) - H_0$. The baseline bound is the unique nonnegative solution of

$$a_{uv}(T_{uv}) = [H_0 - \tau_{uv}]_+, \qquad T_{uv} \le ([H_0 - \tau_{uv}]_+ / \mu)^{1/p}.$$

It is finite even when $c_{uv} = 0$ or $Xu = 0$ and yields $\Delta\Phi_\beta(t) \ge -\tau_{uv}$ for all $t \ge T_{uv}$. The optional bound $a_{uv}(t) - \min\{H_0, b_{uv} t\}$, with $b_{uv} = \lambda \|Xu\|_2 \|v\|_1$, can shorten the scan; a failed tightening falls back to the baseline. The record exposes the bound type, amplitude, assumptions, tail loss, and scope. A full-family pass requires every direction to be scanned through its bound and $\max_d \{\widehat{G}_d + \eta_{\mathrm{scan,d}}, \tau_d\} \le \varepsilon_A$. A screened family yields a candidate gain; it becomes an ambient bound only if an explicit coverage factor κ_{cov} is supplied and $\widehat{A} + \eta_{\mathrm{cert}} \le \kappa_{\mathrm{cov}} \varepsilon_A$. A finite-grid diagnostic estimates coverage risk but never substitutes for the premise. The image panel reports typed prefix or resource abstention when its observed scan or error budget cannot satisfy the full-family test.

Correction and detailed readout perturbation bounds are outside this appendix: both arms disable correction, and the main certificate does not use a task metric as an optimization witness.

D Selective-certificate protocol and diagnostics

This appendix records the evidence needed to evaluate the paper's central claim: reduced-space convergence must be paired with a finite-amplitude scan, a declared coverage domain, and typed abstention when either is missing. The endpoint records and generated tables are retained with the reproducibility materials; they support the claims stated in this appendix.

Table 5: Paired Act-ON minus Act-OFF changes over 18 shared input–seed units. The envelope interval uses the original paired bootstrap; task intervals resample the six bundle means to avoid treating deterministic repeats as independent.

quantity	mean change	95% interval
envelope $\Delta\Phi$	-0.00482	$[-0.00682, -0.00308]$
ARI	$+0.00493$	$[-0.00781, 0.01744]$
AP	$+0.03130$	$[-0.02958, 0.10759]$
AUROC	-0.00498	$[-0.03362, 0.02558]$

D.1 Paired image protocol

The primary panel uses six frozen image pools (three Fashion-MNIST and three CIFAR-10 mixtures), each with $N = 300$ columns and $d = 128$ retained coordinates. Features are frozen DINOv2 ViT-S/14 outputs followed by a development-pool PCA and column normalization. The same unlabeled pool, ten-neighbor graph, initialization, and objective worker are supplied to both arms; labels are used only for ARI, AP, and AUROC readouts.

Each input has three shared initialization seeds, giving 18 paired input–seed units and 36 trajectories. Both arms use Continue/Expand with $p = 1/2$, $\mu = .02$, $\gamma = .03$, $\lambda = N^{-1/2}$, $(\varepsilon_Z, \varepsilon_F) = (.003, .05)$, and $\beta = 20$. Correction is disabled in both arms. Act-OFF never scans a normal direction; Act-ON scans a top-4 finite normal direction family and may accept a finite-amplitude activation. The worker records both the descent ledger B_Φ and a separate compute total C_{spent} for grid evaluations, expansion SVDs, ordinary trials, and rejected majorant trials. This isolates the activation interface rather than comparing two mixed controllers.

The new run reports activation events, full/prefix certificate states, and typed abstentions for every trajectory. Every accepted event is re-evaluated in both the original objective and its envelope; the event log includes ΔF, $\Delta\Phi$, and the gap $F - \Phi_\beta$. The envelope contrast is therefore an optimization result, while the task intervals do not support a task-level improvement claim.

D.2 Exact scan, strict grid error, and certificate scope

For a state $Z = U\Sigma V^\top$ and a screened normal pair $u \perp U$, $v \perp V$, put $C = X - XZ$, $w = Xu$, $\ell = v^\top Lv$, and $q_j(t) = \|C_{:j} - tv_j w\|_2$. The evaluator computes the original smoothed objective without a finite-difference gradient:

$$\Delta\Phi_\beta(t) = \mu t^p + \frac{\gamma\ell}{2}t^2 + \sum_j \{h_{\lambda,\beta}(q_j(t)) - h_{\lambda,\beta}(q_j(0))\}, \tag{30}$$

where

$$h_{\lambda,\beta}(q) = \begin{cases} \beta q^2/2, & q \le \lambda/\beta, \\ \lambda q - \lambda^2/(2\beta), & q > \lambda/\beta. \end{cases}$$

For grid spacing h on the observed prefix $[0, T_{\text{obs}}]$, the logged certificate error is the strict bound

$$\eta^{\text{cert}}_{\text{grid}} = \mu(h/2)^p + \frac{h}{2}\left(\gamma\ell T_{\text{obs}} + \lambda\|Xu\|_2\|v\|_1\right). \tag{31}$$

The finite-difference slope is diagnostic only and never substitutes for this bound. The tail audit sets $H_0 = \sum_j h_{\lambda,\beta}(q_j(0))$ and uses

$$a_{uv}(T_{uv}) = [H_0 - \tau_{uv}]_+, \quad a_{uv}(t) = \mu t^p + \frac{c_{uv}}{2}t^2, \quad c_{uv} = \gamma v^\top Lv. \tag{32}$$

Since the Huber envelope is nonnegative, $\Delta\Phi_\beta(t) \ge a_{uv}(t) - H_0$, so the unique root is finite and $\Delta\Phi_\beta(t) \ge -\tau_{uv}$ for every $t \ge T_{uv}$. This remains valid for $c_{uv} = 0$, $Xu = 0$, or $H_0 = 0$. The Lipschitz bound $a_{uv}(t) - \min\{H_0, b_{uv}t\}$, with $b_{uv} = \lambda\|Xu\|_2\|v\|_1$, is an optional tightening; failure of its root solver falls back to the baseline and does not imply an unbounded tail. Candidate rows also expose `amplitude_bound_type`, `amplitude_bound`, `bound_assumptions`, and `scope`.

Table 6: Finite-amplitude closure and scope audit. A full-family result requires a finite bound and a completed scan for all four screened directions. A dash denotes rows not split by arm.

scope	Act-OFF	Act-ON	all rows
saved endpoint evaluation	18	18	36
finite-family-full	0	7	7
tail-closed/resource abstention	–		21
prefix/truncated	–		8
unbounded-tail rows	–		0

Table 6 reports the independent evaluation of the 36 saved endpoints: The runtime computes $U_{\mathrm{family}} = \max_d\{\widehat{G}_d + \eta_{\mathrm{scan,d}}, \tau_d\}$ and requires every direction to be scanned through its finite tail bound before assigning `finite-family-full`. The numerical margin is $\varepsilon_A - U_{\mathrm{family}}$, computed after family aggregation. The saved endpoint panel is evaluated deterministically from the stored states; 12 one-iteration smoke checks are reported separately. No κ_{cov} is supplied, so no endpoint is promoted to an ambient no-escape certificate. Table 4 reports the saved finite-domain grid diagnostics; tail-tolerance and continuation sensitivity remain declared protocol axes without additional measurements.

At every attempted expansion the replay constructs the block after projection, checks the realized capture ratio, and rejects the operation when it is below $\zeta = .4$. Ordinary updates enforce the numerical floor $\delta_{\mathrm{ord,min}} = 10^{-10}$; the symbolic δ_{ord} from the work bound is retained as a diagnostic because no global $\bar{s}_T$ is supplied.

D.3 Coverage, penalty feasibility, and external stop diagnostic

The coverage diagnostic uses 24 synthetic states, a dense finite angular grid of 2,304 screened-direction pairs and 41 amplitudes per state, and reports $\kappa_{\mathrm{cov}} = A_t^{\mathcal{D}}/A_t^{\star}$ against that finite reference. Its minimum, fifth percentile, and median are 0.5796, 0.6146, and 0.8973. This calibrates the risk of promoting a screened candidate; it is not a global coverage theorem. The controller therefore requires an explicitly declared coverage factor before returning an ambient certificate. The top-4 synthetic normal-gradient calibration (50 states) has median coverage 1.00 and minimum 0.900, versus $0.991/0.755$ for the leading pair; we use this only to justify a finite-family probe, not to infer global coverage.

The lifted feasibility relation is checked independently on 90 Act-OFF runs (six inputs, three seeds, and $\beta \in \{8, 12, 20, 40, 80\}$). The mean r_F values are $(.125, .0833, .0500, .0250, .0125)$ in that order, and every run satisfies $\beta r_F/(\lambda\sqrt{N}) \le 1.0000000000000002$. This supports the finite-penalty feasibility bound, not task-quality monotonicity.

The separately scoped FaNCL-core transfer panel contains 36 rectangular paired units. Its Python port gives mean native relative-decrease-stop residual $r_Z = .018781$ (maximum .019665), whereas the common target $r_Z \le 10^{-3}$ takes 70.83 iterations on average. The panel diagnoses stop-rule alignment only: the supplied comparator is a Python port of the published control flow; no native MATLAB/MEX timing or cross-language speed claim is made.

D.4 Published robust-LRR baseline

As an external method check, we reimplemented the robust low-rank representation solver of Liu et al. (Liu et al., 2013) from its published program

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_{2,1} \quad \text{s.t.} \quad X = XZ + E.$$

The protocol uses the same six frozen public-image feature bundles and the same three deterministic bookkeeping rows per bundle as the main panel. The published preprocessing normalizes each column to unit ℓ_2 norm; the raw-input ablation reruns the identical solver on raw columns. Both use $\lambda = 1/\sqrt{\max(d, N)}$, zero ALM initialization, $\mu_0 = .01$, $\rho = 1.5$, tolerance 10^{-6}, and at most 300 iterations. The readout is the fixed four-cluster KMeans on row-normalized top-eight right singular vectors; the anomaly score is the column ℓ_2 norm of E.

All 18 normalized-input solves and all 18 raw-input solves converged. Mean rank/readout ARI/AP/AUROC are $45.17/(.28785, .20809, .38035)$ for normalized input and $5.67/(.37698, .30304, .43260)$ for raw input. This sizeable preprocessing sensitivity is why Table 3 reports both variants. The source records wall-clock within each LRR solve, but the manuscript makes no cross-method timing claim: C_{spent} is a different compute ledger. Re-evaluation at the published normalized LRR Z on the manuscript's raw panel gives descriptive means $F = 1.35115$, $\Phi_{20} = 1.32615$, $r_Z = 1.883 \times 10^3$, and $r_F = .05$; these do not compare common-objective optima. Both LRR variants retain `unavailable-external-baseline` scope and `no-finite-family-scan` status on all rows. Numerical solver convergence is distinct from availability of the paper's certificate.

D.5 Reproduction boundary

The supplementary materials retain the frozen inputs, endpoint records, executable scan and baseline drivers, and machine-readable summaries. This appendix specifies the protocol and evidence needed to interpret individual endpoints.