High-Precision Lewis Weights via Fourth-Moment Control and Local Bregman Acceleration

Zhao Song*

August 24, 2026

Abstract

We study the high-precision computation of ℓ_p-Lewis weights for $p \geq 4$ in the black-box exact-real full-vector leverage-score oracle model, measuring complexity by the number of adaptive oracle rounds. In this model, Gribling, Sidford, and Zhang [GSZ26] obtained an $O(p^2 \log(m/\epsilon))$ bound for computing an ϵ-estimate. We improve this bound to $O(p \log(mp) + \sqrt{p} \log(1/\epsilon))$. To obtain this result, we isolate the normalized fourth-moment operator governing the nonlinear Hessian of their log-determinant matrix potential and prove that each relative-gradient step with denominator p resets the operator norm to a universal constant. This reset controls the entire update segment and yields an $O(p \log(mp))$ global entrance phase. After entering an $O(1/p)$ spectral neighborhood of the optimum, we switch to a restarted accelerated Bregman-gradient method for the vector potential.

1 Introduction

The notion of ℓ_p-Lewis weights originates in Lewis's study of finite-dimensional subspaces of L_p [Lew78]. Lewis weights assign an importance to each row of a matrix through a self-consistent diagonal rescaling, generalizing the role played by statistical leverage scores in Euclidean geometry. Cohen and Peng [CP15] brought this notion into randomized numerical linear algebra by showing that sampling rows according to their Lewis weights produces nearly optimal subspace embeddings for ℓ_p norms in input-sparsity time. Cohen and Peng also gave an iterative method for computing Lewis weights. Their fixed-point iteration is contractive for $p < 4$, yielding efficient high-accuracy algorithms in that regime. For $p \geq 4$, however, the same iteration is no longer contractive, and obtaining high-precision weights requires a different analysis. This change at $p = 4$ became the central algorithmic obstacle in extending the early Lewis-weight algorithms to all values of p.

Lee and Sidford [LS19] subsequently developed a convex formulation of Lewis weights and used Lewis-weight geometry to construct a nearly optimal self-concordant barrier for polytopes. Their accompanying computation procedures combine projected gradient descent, leverage-score estimation, and a homotopy initialization. This established polynomial-time computability of the barrier and high-accuracy Lewis weights, but the initialization has a dimension-dependent number of stages and therefore does not give the dimension-free round bound studied here for $p \geq 4$.

Fazel, Lee, Padmanabhan, and Sidford [FLPS22] obtained the first nearly dimension-free high-precision algorithm for $p \geq 4$, using $O(p^3 \log(m/\epsilon))$ adaptive full-vector leverage-score rounds. Gribling, Sidford, and Zhang [GSZ26] later improved the bound to $O(p^2 \log(m/\epsilon))$ by applying local relative smoothness to primal and dual formulations of the Lewis-weight optimization problem.

*magic.linuxkde@gmail.com. The author would like to thank Victor Reis and Omri Weinstein for very helpful discussions.

1.1 Our result

Table 1: Comparison of round bounds for high-precision ℓ_p-Lewis weights.

Year	Authors	Reference	Bound
2022	Fazel, Lee, Padmanabhan, and Sidford	[FLPS22]	$O(p^3 \log(m/\epsilon))$
2026	Gribling, Sidford, and Zhang	[GSZ26]	$O(p^2 \log(m/\epsilon))$
2026	This paper	Theorem 1.1	$O(p \log(mp) + \sqrt{p} \log(1/\epsilon))$

We state our main result as follows.

Theorem 1.1 (Main result, informal version of Theorem 4.4)**.** *In the black-box exact-real full-vector leverage-score oracle model, for every $A \in \mathbb{R}^{m \times n}$ with full column rank and no zero rows, $p \geq 4$, and $0 < \epsilon \leq 1$, there is an algorithm that computes an ϵ-estimate of the ℓ_p-Lewis weights using*

$$O(p \log(mp) + \sqrt{p} \log(1/\epsilon))$$

adaptive oracle rounds.

The previous best bound was $O(p^2 \log(m/\epsilon))$ due to [GSZ26]. Our Theorem 1.1 improves it to $O(p \log(mp) + \sqrt{p} \log(1/\epsilon))$.

The first term is the cost of entering an $O(1/p)$ spectral neighborhood of the optimum. Fourth-moment control improves this cost from $O(p^2 \log(mp))$ to $O(p \log(mp))$, independently of the dimension. Within this neighborhood, restarted Bregman acceleration reduces the error by a constant factor every $O(\sqrt{p})$ oracle rounds.

2 Technique Overview

Convex reformulation. We optimize a log-determinant potential over positive-definite matrices. Its unique minimizer exactly satisfies the fixed-point equations defining the ℓ_p-Lewis weights, so controlling the matrix objective controls the desired output coordinates. The relative-gradient update has a closed form and can be implemented by one full-vector leverage-score query. This is the matrix update used by Algorithm 2; the complete two-phase procedure is Algorithm 1, whose guarantee is stated in Theorem 4.4. The basic contraction mechanism is isolated in Lemma 4.5, while Lemma 4.6 bounds the distance from the canonical initial matrix to the optimum.

Fourth-moment global control. The main difficulty is that a coordinatewise smoothness estimate loses an additional factor of p. We instead measure the nonlinear part of the Hessian using an exact fourth-moment curvature operator. Lemma 4.7 gives both a variational description of its spectral radius and the resulting relative-Hessian bound. The key structural fact is Lemma 4.8: after every matrix update, a lifted-Gram representation and a row-sum estimate reset this spectral radius to a universal constant. Comparing the metric at an interior point of the update segment with the metric at its initial endpoint then shows that relative smoothness with denominator p is valid throughout the whole segment. Combining the reset, the segment comparison, and the initial-divergence estimate gives Lemma 4.9, which reaches an $O(1/p)$ spectral neighborhood of the optimum in $O(p \log(mp))$ oracle rounds.

Local vector geometry. Inside that neighborhood, we switch from the matrix variable to the vector potential whose coordinates encode the powered quadratic forms. The global output determines a multiplicative box that is known explicitly and contains the vector optimum. Lemma 4.10 proves that throughout this box the vector objective is strongly convex and smooth relative to

a separable power-type reference function, with condition number $O(p)$. The same lemma also compares the reference Hessian at any two points of the box. This uniform comparison yields the quadratic triangle-scaling estimate needed for acceleration and controls the two orientations of the Bregman divergence.

Boxed acceleration. Algorithm 3 performs accelerated Bregman steps while clipping every coordinate to the known box. The coordinatewise proximal update is explicit, and one oracle query supplies all quadratic forms needed in an iteration. Lemma 4.11 proves that an epoch of $O(\sqrt{p})$ iterations halves the Bregman divergence to the vector optimum. Restarting the epoch contributes $O(\sqrt{p}(\log n + \log p + \log(1/\epsilon)))$ oracle rounds. The $\sqrt{p}\log n$ and $\sqrt{p}\log p$ terms are absorbed by the global $O(p\log(mp))$ cost.

Output recovery. The local objective gap must ultimately imply coordinatewise accuracy after applying the final power. Lemma 4.12 identifies this gap exactly as the sum of a matrix Bregman divergence and a vector Bregman divergence. The matrix term yields a two-sided spectral comparison, which in turn controls every quadratic form; the chosen stopping threshold remains valid after raising these forms to the power $p/2$. Thus, the global and local costs add to $O(p\log(mp) + \sqrt{p}\log(1/\epsilon))$ adaptive oracle rounds.

3 Preliminaries

For a positive integer n, let $[n] := \{1, \ldots, n\}$. We write I for an identity matrix whose dimension is clear from context. For $v \in \mathbb{R}^m$, $\mathrm{Diag}(v)$ denotes the diagonal matrix with diagonal v. Vector inequalities, products, quotients, and powers are coordinatewise unless stated otherwise; in particular, $\mathbb{R}^m_{>0} := \{v \in \mathbb{R}^m : v_i > 0 \text{ for every } i \in [m]\}$.

For symmetric matrices, $X \succeq 0$ and $X \succ 0$ mean that X is positive semidefinite and positive definite, respectively, and $X \preceq Y$ denotes the Loewner order. We use $\lambda_{\min}(X)$ and $\lambda_{\max}(X)$ for the extreme eigenvalues, $\|X\|_{\mathrm{op}}$ and $\|X\|_{\mathrm{F}}$ for the operator and Frobenius norms, and $\det(X)$ and $\mathrm{tr}[X]$ for the determinant and trace. The inner products are $\langle x, y\rangle = x^\top y$ for vectors and $\langle X, Y\rangle = \mathrm{tr}[XY]$ for symmetric matrices. The symbol $\circ$ denotes the Hadamard product. For a twice differentiable function f on symmetric matrices, $\nabla^2 f(X)[H, H]$ denotes the Hessian quadratic form in direction H.

Throughout, we write $a_i^\top$ for the ith row of A. The matrix A is called *non-degenerate* if it has full column rank and no zero rows. For a full-column-rank matrix B with rows $b_i^\top$, its ith leverage score is

$$\tau_i(B) := b_i^\top (B^\top B)^{-1} b_i.$$

Definition 3.1 (ℓ_p-Lewis weights and ϵ-estimates)**.** *For $p > 0$, the ℓ_p-Lewis weights $\sigma_p(A)$ of a non-degenerate matrix $A \in \mathbb{R}^{m\times n}$ are the unique positive solution to*

$$\sigma_p(A)_i = \tau_i(\mathrm{Diag}(\sigma_p(A))^{1/2-1/p} A) \qquad \text{for every } i \in [m].$$

Existence and uniqueness follow from [LS19, Lemma 22]. For $0 < \epsilon \le 1$, a vector $\widehat{w} \in \mathbb{R}^m_{>0}$ is an ϵ-estimate of $\sigma_p(A)$ if

$$(1-\epsilon)\sigma_p(A)_i \le \widehat{w}_i \le (1+\epsilon)\sigma_p(A)_i \qquad \text{for every } i \in [m].$$

Definition 3.2 (Bregman divergence)**.** *For a differentiable convex function ψ, the Bregman divergence generated by ψ is*

$$D_\psi(x, y) := \psi(x) - \psi(y) - \langle \nabla\psi(y), x - y\rangle.$$

Thus, the gradient is evaluated at the second argument, and the order of the two arguments matters.

We use relative smoothness and strong convexity through a standard comparison between Hessians and Bregman divergences. The next lemma records this implication in the precise orientation needed later.

Lemma 3.3 (Hessian-to-Bregman comparison)**.** *Let f and h be twice continuously differentiable convex functions on an open convex set containing x and y, and let $0 \le \mu \le L$. Suppose that, for every z on the segment from x to y,*

$$\mu \nabla^2 h(z) \preceq \nabla^2 f(z) \preceq L \nabla^2 h(z).$$

Then

$$\mu D_h(y, x) \le D_f(y, x) \le L D_h(y, x).$$

Proof. Definition 3.2 and the fundamental theorem of calculus give, for $\psi \in \{f, h\}$,

$$D_\psi(y, x) = \int_0^1 (1-t) \nabla^2 \psi(x + t(y-x))[y-x, y-x] \mathrm{d}t.$$

For every $t \in [0, 1]$, apply the assumed Hessian comparison at $z := x + t(y - x)$ to the direction $y - x$. Multiplying the resulting scalar bounds by $1 - t \ge 0$ and integrating over $t \in [0, 1]$ proves the claim. □

Definition 3.4 (Black-box exact-real full-vector leverage-score oracle model)**.** *For a fixed non-degenerate matrix A, the algorithm is given m, n, p, ϵ and access to the primitive*

$$\mathcal{O}_A(v) := (v_i a_i^\top (A^\top \operatorname{Diag}(v) A)^{-1} a_i)_{i=1}^m, \qquad v \in \mathbb{R}^m_{>0}.$$

The algorithm may choose each query vector adaptively as a function of its input parameters and the responses to previous queries. One invocation of $\mathcal{O}_A$ returns the complete vector and costs one oracle round; partial-coordinate queries are not available, and A is accessed only through $\mathcal{O}_A$. Between oracle calls, arbitrary exact-real arithmetic on the parameters, query vectors, and previous answers is uncharged. More explicitly, p and ϵ are exact real inputs, and the permitted local primitives include field operations, x^c for $x > 0$ and exact-real c, natural logarithms, comparisons, minimum and maximum, and the floor and ceiling operations. Thus, our result bounds only the number of exact oracle rounds. It gives no running-time or bit-complexity guarantee and no robustness guarantee for finite-precision arithmetic or approximate or noisy oracle responses.

4 Fourth-moment global control and local Bregman acceleration

We prove Theorem 4.4 by combining a globally convergent matrix phase with a locally accelerated vector phase. Subsection 4.1 develops the matrix formulation and relative-gradient update, while Subsection 4.2 presents the complete algorithm and proves the main theorem. Subsection 4.3 establishes the fourth-moment global control, Subsection 4.4 develops the local vector geometry, Subsection 4.5 analyzes boxed Bregman acceleration, and Subsection 4.6 converts the final objective gap into coordinatewise Lewis-weight accuracy.

4.1 Matrix formulation and relative-gradient update

Let $A \in \mathbb{R}^{m \times n}$ be non-degenerate and let $p \ge 4$.

Definition 4.1 (Global parameters)**.** *Define $\beta := (p-2)/2 \ge 1$ and $\gamma := \beta + 1 = p/2 \ge 2$.*

These parameters govern the nonlinear power in the matrix potential and the conditioning of the local vector problem. We next introduce the matrix quantities used throughout the global phase.

Definition 4.2 (Matrix quantities)**.** *Let γ be as in Definition 4.1. For $M \succ 0$, define*

$$q_i(M) := a_i^\top M a_i, \qquad F(M) := -\log\det(M) + \frac{1}{\gamma}\sum_{i=1}^m q_i(M)^\gamma, \qquad h(M) := -\log\det(M).$$

We first show that F has a unique minimizer. Indeed, because A has full column rank, for every $M \succeq 0$,

$$\max_{i\in[m]} a_i^\top M a_i \geq \frac{1}{m}\sum_{i=1}^m a_i^\top M a_i = \frac{1}{m}\operatorname{tr}[A^\top A M] \geq \frac{\lambda_{\min}(A^\top A)}{m}\|M\|_{\mathrm{op}}.$$

Thus, we may take $c_A := \lambda_{\min}(A^\top A)/m > 0$. Moreover, $\det M \leq \|M\|_{\mathrm{op}}^n$, so

$$F(M) \geq -n\log\|M\|_{\mathrm{op}} + \frac{c_A^\gamma}{\gamma}\|M\|_{\mathrm{op}}^\gamma \longrightarrow +\infty$$

as $\|M\|_{\mathrm{op}} \to \infty$. If M approaches the boundary of the positive-definite cone while its operator norm stays bounded, then $-\log\det M \to +\infty$. Consequently, for every $c \in \mathbb{R}$, the sublevel set $\mathcal{S}_c := \{M \succ 0 : F(M) \leq c\}$ is bounded in operator norm and, by the preceding boundary blow-up, satisfies $\lambda_{\min}(M) \geq \eta_c$ for some $\eta_c > 0$. If $M_k \in \mathcal{S}_c$ converges in the ambient space of symmetric matrices, this lower bound implies that its limit remains positive definite, and continuity of F then places the limit in $\mathcal{S}_c$. Hence $\mathcal{S}_c$ is closed and bounded in a finite-dimensional space and therefore compact. Choose any $\overline{M} \succ 0$ and a minimizing sequence (M_k). After discarding finitely many terms, this sequence lies in the nonempty compact sublevel set $\mathcal{S}_{F(\overline{M})+1}$. Passing to a convergent subsequence and using the continuity of F therefore yields a minimizer. Finally, $-\log\det M$ is strictly convex and the remaining term is convex, so the minimizer is unique.

Definition 4.3 (Optimal matrix M_*)**.** *Let F be as in Definition 4.2. Define*

$$M_* := \operatorname*{argmin}_{M\succ 0} F(M).$$

By the preceding existence-and-uniqueness argument, M_* is well-defined.

The stationarity condition and its relation to the ℓ_p-Lewis weights are

$$M_*^{-1} = \sum_{i=1}^m q_i(M_*)^\beta a_i a_i^\top, \qquad \sigma_p(A)_i = q_i(M_*)^{\beta+1}. \tag{1}$$

To verify the second identity and the asserted uniqueness, put $w_i := q_i(M_*)^\gamma$ and $W := \mathrm{Diag}(w)$. Since $\gamma(1-2/p) = \beta$, Eq. (1) gives $M_* = (A^\top W^{1-2/p} A)^{-1}$, and

$$\tau_i(W^{1/2-1/p}A) = w_i^{1-2/p} q_i(M_*) = q_i(M_*)^{\beta+1} = w_i.$$

Conversely, any positive Lewis fixed point w defines $M := (A^\top W^{1-2/p}A)^{-1}$ and satisfies $w_i = w_i^{1-2/p} q_i(M)$. Hence $w_i = q_i(M)^\gamma$ and $M^{-1} = \sum_{i=1}^m q_i(M)^\beta a_i a_i^\top$. By uniqueness of the stationary point, $M = M_*$ and $w_i = q_i(M_*)^\gamma$.

For $L > 1$, consider the relative-gradient step

$$M^+ := \arg\min_{X \succ 0} \{F(M) + \langle \nabla F(M), X - M \rangle + L D_h(X, M)\}.$$

Its stationarity condition gives the explicit update

$$(M^+)^{-1} = (1 - \frac{1}{L}) M^{-1} + \frac{1}{L} \sum_{i=1}^{m} q_i(M)^\beta a_i a_i^\top. \tag{2}$$

To see that the minimum is attained, put

$$K_M := (1 - \frac{1}{L}) M^{-1} + \frac{1}{L} \sum_{i=1}^{m} q_i(M)^\beta a_i a_i^\top \succ 0.$$

Up to an additive constant, the minimized objective divided by L is $-\log\det X + \operatorname{tr}[K_M X]$. It is strictly convex and has the unique minimizer $X = K_M^{-1}$, which is exactly Eq. (2). If $M = (A^\top V A)^{-1}$ for $V := \operatorname{Diag}(v)$, then Eq. (2) is implemented by

$$v_i^+ := (1 + \frac{\Phi_i(v) - 1}{L}) v_i, \qquad \Phi_i(v) := \frac{q_i(M)^\beta}{v_i}. \tag{3}$$

The oracle returns $v_i q_i(M)$ as the ith leverage score of $V^{1/2} A$. Since $v_i > 0$ is known, one invocation recovers all $q_i(M)$ and implements the entire update.

4.2 Main algorithm and theorem

For clarity, we record each procedure next to the result that analyzes it. Algorithm 1 is the end-to-end procedure.

Algorithm 1 Two-phase high-precision computation of $\sigma_p(A)$

procedure HighPrecisionLewisWeights$(m, n, p, \epsilon, \mathcal{O}_A)$ ▷ Theorem 4.4
  $\beta \leftarrow (p-2)/2$
  $u \leftarrow$ GlobalEntrance$(m, p, \mathcal{O}_A)$ ▷ Algorithm 2
  $s \leftarrow$ BoxedAcceleratedRestarts$(m, n, p, \epsilon, u, \mathcal{O}_A)$ ▷ Algorithm 3
  $\ell^{\text{out}} \leftarrow \mathcal{O}_A(s)$
  $\widehat{w}_i \leftarrow (\ell_i^{\text{out}} / s_i)^{\beta+1}$ for every $i \in [m]$
  return $\widehat{w}$
end procedure

Theorem 4.4 (Main result, formal version of Theorem 1.1)**.** *In the black-box exact-real full-vector leverage-score oracle model of Definition 3.4, for every non-degenerate $A \in \mathbb{R}^{m \times n}$, $p \geq 4$, and $0 < \epsilon \leq 1$, the two-phase procedure* HighPrecisionLewisWeights *(Algorithm 1) outputs an ϵ-estimate of the ℓ_p-Lewis-weight vector $\sigma_p(A)$ from Definition 3.1 using*

$$O(p \log(mp) + \sqrt{p} \log(1/\epsilon))$$

adaptive oracle rounds.

Proof. The notation and auxiliary results invoked below are stated later in this section; in particular, set $\alpha := 1/\beta$ and use $M(v)$, $q_i(v)$, F_{vec}, h_{vec}, and v_* as defined there. Run Algorithm 1, using the subroutines in Algorithms 2 and 3. Let $v^{(T_{\mathrm{g}})}$ denote the terminal global iterate inside Algorithm 2, and let $s^{(R)}$ denote the vector returned by Algorithm 3. For the analysis of the global phase, let M_{basin} denote the matrix represented by $v^{(T_{\mathrm{g}})}$ in Lemma 4.9. Set $d_0 := 1/(65536\beta^2)$, $\eta := 1/(64\beta)$, and $\delta_0 := 1/(128\beta)$. Since $\beta \geq 1$, we have $0 < \delta_0 \leq 1/128 < 1/2$, so we may apply Eq. (6) with $(Y, X) := (M_*, M_{\mathrm{basin}})$. Because $D_h(M_*, M_{\mathrm{basin}}) \leq d_0 = \delta_0^2/4$, Eq. (6) gives

$$(1 - \delta_0)M_* \preceq M_{\mathrm{basin}} \preceq (1 + 2\delta_0)M_*.$$

Because $\eta = 2\delta_0$, weakening the lower bound gives

$$(1 - \eta)M_* \preceq M_{\mathrm{basin}} \preceq (1 + \eta)M_*.$$

Set $u_i := q_i(M_{\mathrm{basin}})^\beta$ and form the box $\mathcal{B}$ from Lemma 4.10-(a). Initialize the accelerated phase at $s^{(0)} := u$. The initial reverse divergence satisfies

$$D_{h_{\mathrm{vec}}}(v_*, u) \leq \frac{32\alpha}{63^3} \sum_{i=1}^{m} [v_*]_i^{1+\alpha} = \frac{32n}{63^3\beta} \leq n. \tag{4}$$

For each coordinate, let $h_i(t) := t^{1+\alpha}/(1+\alpha)$ and write $v_{*,i} := [v_*]_i$. Lemma 4.10-(a), together with $e^{1/64} \leq (1 - 1/64)^{-1} = 64/63$, gives

$$\frac{63}{64} v_{*,i} \leq u_i \leq \frac{64}{63} v_{*,i}.$$

Thus, every point between $v_{*,i}$ and u_i is at least $63v_{*,i}/64$, and $|v_{*,i} - u_i| \leq v_{*,i}/63$. Since

$$h_i''(x) = \alpha x^{\alpha-1} \leq \frac{64\alpha}{63} v_{*,i}^{\alpha-1},$$

the second-order integral formula yields

$$D_{h_i}(v_{*,i}, u_i) \leq \frac{32\alpha}{63^3} v_{*,i}^{1+\alpha}.$$

Summing this bound over i, using $\alpha = 1/\beta$, and using the normalization proved after Eq. (10) gives Eq. (4). We retain the simpler upper bound n in the restart schedule.

Starting from $s^{(r)}$, run one epoch as analyzed in Lemma 4.11. Set $s^{(r+1)} := x^{(K)}$. Then Lemma 4.11-(b) and Eq. (4) give

$$D_{h_{\mathrm{vec}}}(v_*, s^{(r)}) \leq 2^{-r} n.$$

Let

$$d_\epsilon := \frac{\epsilon^2}{256(\beta+1)^2}, \qquad R := 1 + \lceil \log_2(n/d_\epsilon) \rceil.$$

The final epoch has index $r := R - 1$, and its input satisfies

$$D_{h_{\mathrm{vec}}}(v_*, s^{(R-1)}) \leq 2^{-(R-1)} n \leq d_\epsilon.$$

Apply Lemma 4.11-(a) in this epoch. Its coefficient is at most $1/10$, so

$$F_{\mathrm{vec}}(s^{(R)}) - F_{\mathrm{vec}}(v_*) \leq d_\epsilon.$$

Lemma 4.12 therefore implies

$$D_h(M_*, M(s^{(R)})) \le d_\epsilon.$$

Set $\delta := \epsilon/(8(\beta+1))$ and $\tau := 2\delta$. Since $\beta + 1 \ge 2$ and $0 < \epsilon \le 1$, we have $0 < \delta \le 1/16 < 1/2$ and $0 < \tau \le 1/8 < 1$. Moreover, $d_\epsilon = \delta^2/4$, so Eq. (6) with $(Y, X) := (M_*, M(s^{(R)}))$ gives

$$(1-\delta)M_* \preceq M(s^{(R)}) \preceq (1+2\delta)M_*.$$

Since $\tau = 2\delta$, weakening the lower bound gives

$$(1-\tau)M_* \preceq M(s^{(R)}) \preceq (1+\tau)M_*.$$

Consequently, if $\widehat{w}_i := q_i(s^{(R)})^{\beta+1}$, then

$$(1-\tau)^{\beta+1} \le \frac{\widehat{w}_i}{\sigma_p(A)_i} \le (1+\tau)^{\beta+1}.$$

Bernoulli's inequality gives

$$(1-\tau)^{\beta+1} \ge 1 - (\beta+1)\tau = 1 - \frac{\epsilon}{4} \ge 1 - \epsilon.$$

For the upper bound,

$$(1+\tau)^{\beta+1} \le \exp((\beta+1)\tau) = e^{\epsilon/4} \le 1 + \epsilon.$$

The last inequality follows from

$$\log(1+\epsilon) = \int_0^\epsilon \frac{1}{1+t}\,\mathrm{d}t \ge \frac{\epsilon}{1+\epsilon} \ge \frac{\epsilon}{2} \ge \frac{\epsilon}{4}.$$

Thus, $\widehat{w}$ is an ϵ-estimate.

Let T_{g} be the number of relative-gradient iterations in Algorithm 2. Lemma 4.9 gives

$$T_{\mathrm{g}} = O(p \log(mp)).$$

According to Algorithm 3 and Lemma 4.11, there are

$$R = O(\log n + \log p + \log(1/\epsilon))$$

accelerated epochs, and each epoch uses $K = O(\sqrt{p})$ oracle rounds. One additional oracle invocation computes u, and one computes the output $\widehat{w}$. Hence, the total number of oracle rounds is

$$\underbrace{T_{\mathrm{g}} + 1}_{\text{Algorithm 2}} + \underbrace{KR}_{\text{Algorithm 3}} + 1 = \underbrace{O(p\log(mp))}_{T_g} + \underbrace{O(\sqrt{p}\log n) + O(\sqrt{p}\log p) + O(\sqrt{p}\log(1/\epsilon))}_{KR} + 2$$
$$= O(p\log(mp) + \sqrt{p}\log(1/\epsilon)).$$

The first step substitutes the bounds for T_{g}, R, and K. For the second step, $n \le m$ and $p \ge 4$ imply

$$\sqrt{p}\log n \le p\log(mp), \qquad \sqrt{p}\log p \le p\log(mp),$$

so the two middle terms are absorbed by $O(p\log(mp))$. This proves the claimed oracle-round bound. □

4.3 Fourth-moment control and global entrance

We use the following standard log-determinant estimate. If $D_h(Y, X) \le \delta^2/4$ and $\delta \le 1/2$, then

$$(1-\delta)X \preceq Y \preceq (1+\delta)X. \tag{5}$$

Indeed, let $\varphi(t) := t - \log t - 1$. This function is decreasing on $(0, 1]$ and increasing on $[1, \infty)$. For $0 < \delta \le 1/2$,

$$\varphi(1+\delta) = \int_0^\delta \frac{t}{1+t}\,\mathrm{d}t \ge \frac{\delta^2}{3}, \qquad \varphi(1-\delta) = \int_0^\delta \frac{t}{1-t}\,\mathrm{d}t \ge \frac{\delta^2}{2}.$$

The eigenvalues λ_j of $X^{-1/2}YX^{-1/2}$ satisfy $\sum_{j=1}^n \varphi(\lambda_j) \le \delta^2/4$. The monotonicity and the two boundary estimates therefore imply $|\lambda_j - 1| \le \delta$ for every j, proving Eq. (5). Rearranging the same inequalities gives

$$(1+\delta)^{-1}Y \preceq X \preceq (1-\delta)^{-1}Y.$$

Since $(1+\delta)^{-1} \ge 1-\delta$ and $(1-\delta)^{-1} \le 1+2\delta$ for $\delta \le 1/2$, we will also use the reversed form

$$(1-\delta)Y \preceq X \preceq (1+2\delta)Y. \tag{6}$$

The next one-step principle is a self-contained version of the local relative-smoothness argument of Gribling, Sidford, and Zhang [GSZ26, Proposition 4]. It shows that a relative-gradient update contracts the reverse Bregman divergence whenever the lower and upper relative models hold.

Lemma 4.5 (Local relative-gradient contraction)**.** *Let $\mathcal{U}$ be an open convex subset of a finite-dimensional inner-product space, let $\mathcal{C} \subseteq \mathcal{U}$ be nonempty and convex, and let $f, h : \mathcal{U} \to \mathbb{R}$ be differentiable convex functions. Suppose that*

$$f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + D_h(y, x) \qquad (x, y \in \mathcal{C}).$$

Fix $x \in \mathcal{C}$. For $L > 1$, assume that the following minimum is attained, and choose

$$x^+ \in \arg\min_{z\in\mathcal{C}} \{\langle \nabla f(x), z - x\rangle + LD_h(z, x)\}.$$

If the upper relative model

$$f(x^+) \le f(x) + \langle \nabla f(x), x^+ - x\rangle + LD_h(x^+, x)$$

holds and x_* minimizes f on $\mathcal{C}$, then

$$D_h(x_*, x^+) \le (1 - \frac{1}{L})D_h(x_*, x).$$

Proof. The prox objective is differentiable and convex. Since its minimum is attained at x^+ on the convex set $\mathcal{C}$, its one-sided directional derivative from x^+ toward every $z \in \mathcal{C}$ is nonnegative. Taking $z = x_*$ and using the Bregman three-point identity gives

$$\langle \nabla f(x), x^+ - x_*\rangle \le L(D_h(x_*, x) - D_h(x_*, x^+) - D_h(x^+, x)).$$

Combining this inequality with the upper relative model and then applying the lower relative model with $(x, y) := (x, x_*)$ yields

$$f(x^+) - f(x_*) \le (L-1)D_h(x_*, x) - LD_h(x_*, x^+).$$

The left-hand side is nonnegative, which proves the claim. □

The global phase begins at the unweighted Gram inverse $M^{(0)}$. The next lemma compares this initialization with M_* and supplies the finite divergence budget from which geometric contraction starts.

Lemma 4.6 (Initial matrix divergence)**.** *Let M_* be the optimal matrix from Definition 4.3. For $M^{(0)} := (A^\top A)^{-1}$,*

$$M_* \succeq M^{(0)} \qquad \text{and} \qquad D_h(M_*, M^{(0)}) \le m - n.$$

Proof. The quantities $q_i(M_*)^{\beta+1}$ are the leverage scores of $\mathrm{Diag}(q(M_*)^{\beta/2})A$, so they are at most one. Hence $q_i(M_*) \le 1$ and

$$M_*^{-1} = \sum_{i=1}^m q_i(M_*)^\beta a_i a_i^\top \preceq A^\top A = (M^{(0)})^{-1},$$

which proves $M_* \succeq M^{(0)}$. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of $(M^{(0)})^{-1/2} M_* (M^{(0)})^{-1/2}$; they are all at least one. Thus

$$\begin{aligned} D_h(M_*, M^{(0)}) &= \sum_{j=1}^n (\lambda_j - \log \lambda_j - 1) \le \sum_{j=1}^n (\lambda_j - 1) = \mathrm{tr}[(M^{(0)})^{-1} M_*] - n = \sum_{i=1}^m q_i(M_*) - n \\ &\le m - n. \end{aligned}$$

The first step follows from the spectral formula for the log-determinant Bregman divergence. The second step uses $\log \lambda_j \ge 0$. The third step uses that the trace is the sum of the eigenvalues, together with cyclicity of trace. The fourth step uses $(M^{(0)})^{-1} = A^\top A$ and the definition of q_i. The final step uses $q_i(M_*) \le 1$ for every $i \in [m]$. □

The nonlinear part of F produces fourth-order row interactions that are not visible in the log-determinant Hessian alone. The next lemma packages these interactions into an operator and expresses the relative curvature of F through its operator norm.

Lemma 4.7 (Exact fourth-moment curvature)**.** *Let β and γ be as in Definition 4.1. Fix $M \succ 0$ and set*

$$e_i := \frac{M^{1/2} a_i}{\sqrt{q_i(M)}}, \qquad w_i := q_i(M)^\gamma, \qquad P_i := e_i e_i^\top.$$

On the symmetric matrices with the Frobenius inner product, define

$$\mathcal{T}_M(X) := \sum_{i=1}^m w_i \langle P_i, X \rangle P_i, \qquad \Lambda(M) := \|\mathcal{T}_M\|_{\mathrm{op}}.$$

Then the following statements hold.

(a) The operator norm has the variational representation

$$\Lambda(M) = \sup_{\substack{H = H^\top \\ H \neq 0}} \frac{\sum_{i=1}^m q_i(M)^{\gamma-2} (a_i^\top H a_i)^2}{\mathrm{tr}[M^{-1} H M^{-1} H]}.$$

(b) For every symmetric H,

$$\nabla^2 F(M)[H, H] \le (1 + \beta \Lambda(M)) \nabla^2 h(M)[H, H].$$

Proof. **Proof of part (a).** Put $X := M^{-1/2} H M^{-1/2}$. Then

$$q_i(M)^{\gamma-2}(a_i^\top H a_i)^2 = w_i(e_i^\top X e_i)^2, \qquad \operatorname{tr}[M^{-1}HM^{-1}H] = \operatorname{tr}[X^2].$$

Therefore, the quotient in part (a) equals $\langle X, \mathcal{T}_M(X)\rangle / \operatorname{tr}[X^2]$. The operator $\mathcal{T}_M$ is self-adjoint and positive semidefinite, so its operator norm is the supremum of this quotient over nonzero symmetric X. The bijection $H = M^{1/2} X M^{1/2}$ proves part (a).

Proof of part (b). Direct differentiation gives

$$\begin{aligned}\nabla^2 F(M)[H,H] &= \operatorname{tr}[X^2] + \beta \sum_{i=1}^m q_i(M)^{\gamma-2}(a_i^\top H a_i)^2 \\ &= \operatorname{tr}[X^2] + \beta \sum_{i=1}^m w_i (e_i^\top X e_i)^2.\end{aligned}$$

By the definition of $\Lambda(M)$, the second sum is at most $\Lambda(M)\operatorname{tr}[X^2]$. Since $\nabla^2 h(M)[H,H] = \operatorname{tr}[X^2]$, this proves part (b). □

The curvature bound above is useful only if $\Lambda(M)$ remains controlled during the iteration. The next lemma shows that one relative-gradient step resets the endpoint curvature to an explicit bound depending only on γ and L, regardless of the incoming value of $\Lambda(M)$.

Lemma 4.8 (Universal fourth-moment reset)**.** *Let β and γ be as in Definition 4.1. Let $L > 1$, put $a := 1 - 1/L$, and let M^+ be the relative-gradient step in Eq. (2). Define $\kappa_{\gamma,L} := \frac{L}{\gamma}(\frac{\gamma-1}{a\gamma})^{\gamma-1}$. Then*

$$\Lambda(M^+) \le \kappa_{\gamma,L}.$$

The bound is universal: it assumes no bound on $\Lambda(M)$.

Proof. Write $M = SS^\top$ with $S := M^{1/2}$ and use the notation of Lemma 4.7. Set

$$C := \sum_{i=1}^m w_i P_i, \qquad R_M := aI + \frac{1}{L}C.$$

Eq. (2) gives $M^+ = S R_M^{-1} S^\top$. Put $B := S R_M^{-1/2}$, so $M^+ = BB^\top$. This factor may differ from the symmetric square root of M^+, but this does not change the spectrum of the curvature operator. Indeed, if the invertible matrices $B_1, B_2 \in \mathbb{R}^{n\times n}$ satisfy $B_1 B_1^\top = B_2 B_2^\top = M^+$, then $U := B_1^{-1} B_2$ satisfies

$$UU^\top = B_1^{-1} B_2 B_2^\top B_1^{-\top} = B_1^{-1} M^+ B_1^{-\top} = I.$$

Thus, U is orthogonal and $B_2 = B_1 U$. If $\mathcal{T}^{(j)}$ denotes the curvature operator formed using the normalized directions $e_i^{(j)} := B_j^\top a_i / \sqrt{q_i(M^+)}$, then

$$e_i^{(2)} = U^\top e_i^{(1)}, \qquad P_i^{(2)} := e_i^{(2)}[e_i^{(2)}]^\top = U^\top P_i^{(1)} U.$$

Define $\mathcal{Q}(X) := U^\top X U$ on the symmetric matrices. The map $\mathcal{Q}$ is a Frobenius isometry, with $\mathcal{Q}^{-1}(X) = UXU^\top$, and direct substitution gives

$$\mathcal{T}^{(2)} = \mathcal{Q} \circ \mathcal{T}^{(1)} \circ \mathcal{Q}^{-1}.$$

Taking $B_1 = (M^+)^{1/2}$ and $B_2 = B$ shows that the factor-B curvature operator is orthogonally conjugate to $\mathcal{T}_{M^+}$. We may consequently compute $\Lambda(M^+)$ using the factor B. Define

$$d_i := e_i^\top R_M^{-1} e_i, \qquad f_i := \frac{R_M^{-1/2} e_i}{\sqrt{d_i}}, \qquad w_i^+ := w_i d_i^\gamma.$$

Thus, $q_i(M^+) = q_i(M) d_i$, while f_i and w_i^+ are the normalized directions in the factor-B coordinates and the corresponding weights at M^+.

Define the linear map $\mathcal{Z}c := \sum_{i=1}^m c_i \sqrt{w_i^+} f_i f_i^\top$ from $\mathbb{R}^m$ to the symmetric matrices, let $\widetilde{\mathcal{T}}_{M^+} := \mathcal{Z}\mathcal{Z}^*$, and let $G^+ := \mathcal{Z}^*\mathcal{Z}$. By the preceding orthogonal-conjugacy observation, $\widetilde{\mathcal{T}}_{M^+}$ and $\mathcal{T}_{M^+}$ have the same spectrum. The nonzero eigenvalues of G^+ and $\widetilde{\mathcal{T}}_{M^+}$ also coincide. For completeness, if $G^+ c = \lambda c$ with $\lambda > 0$, then $\mathcal{Z}c \neq 0$ and $\widetilde{\mathcal{T}}_{M^+}(\mathcal{Z}c) = \lambda \mathcal{Z}c$. Conversely, if $\widetilde{\mathcal{T}}_{M^+}(X) = \lambda X$ with $\lambda > 0$, then $\mathcal{Z}^* X \neq 0$ and $G^+(\mathcal{Z}^* X) = \lambda \mathcal{Z}^* X$. Hence $\Lambda(M^+) = \lambda_{\max}(G^+)$. Moreover,

$$G_{ij}^+ = \sqrt{w_i w_j}\, d_i^{(\gamma-2)/2} d_j^{(\gamma-2)/2} (e_i^\top R_M^{-1} e_j)^2. \tag{7}$$

In particular, G^+ is entrywise nonnegative. For the positive vector $\xi_i := \sqrt{w_i} d_i^{-(\gamma-2)/2}$, Eq. (7) gives

$$\frac{(G^+\xi)_i}{\xi_i} = d_i^{\gamma-2} e_i^\top R_M^{-1} C R_M^{-1} e_i.$$

Since $C = L(R_M - aI)$, the right-hand side equals

$$L d_i^{\gamma-2}(d_i - a k_i), \qquad k_i := e_i^\top R_M^{-2} e_i.$$

Consequently,

$$\frac{(G^+\xi)_i}{\xi_i} \leq L d_i^{\gamma-1}(1 - a d_i) \leq \frac{L}{\gamma}\left(\frac{\gamma-1}{a\gamma}\right)^{\gamma-1} = \kappa_{\gamma,L},$$

The first step follows from Cauchy–Schwarz, which gives $k_i \geq d_i^2$. For the second step, $R_M \succeq aI$ gives $0 < d_i \leq 1/a$, and the function $L d^{\gamma-1}(1-ad)$ is maximized on this interval at $d = (\gamma-1)/(a\gamma)$. The final step follows from the definition of $\kappa_{\gamma,L}$. If $D := \mathrm{Diag}(\xi)$, then the nonnegative matrix $D^{-1} G^+ D$ has every row sum at most $\kappa_{\gamma,L}$. Its spectral radius, and hence $\lambda_{\max}(G^+)$, is at most $\kappa_{\gamma,L}$. This completes the proof. □

Algorithm 2 Global entrance phase

procedure GLOBALENTRANCE($m, p, \mathcal{O}_A$) ▷ Lemma 4.9
  $\beta \leftarrow (p-2)/2$ ▷ Definition 4.1
  $L_{\mathrm{g}} \leftarrow p$
  $T_{\mathrm{g}} \leftarrow \lceil L_{\mathrm{g}} \log(65536 m \beta^2) \rceil$
  $v^{(0)} \leftarrow \mathbf{1}$
  for $t = 0, \dots, T_{\mathrm{g}} - 1$ **do**
    $\ell^{(t)} \leftarrow \mathcal{O}_A(v^{(t)})$
    $q_i^{(t)} \leftarrow \ell_i^{(t)} / v_i^{(t)}$ for every $i \in [m]$
    $v_i^{(t+1)} \leftarrow (1 - L_{\mathrm{g}}^{-1}) v_i^{(t)} + L_{\mathrm{g}}^{-1} (q_i^{(t)})^{\beta}$ for every $i \in [m]$
  end for
  $\ell^{\mathrm{basin}} \leftarrow \mathcal{O}_A(v^{(T_{\mathrm{g}})})$
  $q_i^{\mathrm{basin}} \leftarrow \ell_i^{\mathrm{basin}} / v_i^{(T_{\mathrm{g}})}$ for every $i \in [m]$
  $u_i \leftarrow (q_i^{\mathrm{basin}})^{\beta}$ for every $i \in [m]$
  return u
end procedure

The procedure GLOBALENTRANCE repeatedly applies the relative-gradient update with denominator p. The next lemma combines the universal reset with segment control to prove geometric contraction of the matrix divergence and entrance into the local basin.

Lemma 4.9 (Global-entrance guarantee for Algorithm 2)**.** *Let β and γ be as in Definition 4.1. Let M_* be the optimal matrix from Definition 4.3. Consider the procedure* GLOBALENTRANCE *in Algorithm 2. Let $v^{(0)} := \mathbf{1}$, $M^{(0)} := (A^\top A)^{-1}$, and*

$$L_{\mathrm{g}} := 2\gamma = p, \qquad T_{\mathrm{g}} := \lceil L_{\mathrm{g}} \log(65536 m \beta^2) \rceil.$$

Run Eq. (3) with denominator L_{g} for T_{g} iterations and, for $0 \le t \le T_{\mathrm{g}}$, define

$$M^{(t)} := (A^\top \operatorname{Diag}(v^{(t)}) A)^{-1}.$$

Then

$$D_h(M_*, M^{(t)}) \le (m-n) \exp(-t/L_{\mathrm{g}}) \le m \exp(-t/L_{\mathrm{g}}),$$

so $D_h(M_*, M^{(T_{\mathrm{g}})}) \le 1/(65536\beta^2)$. Every iteration uses one oracle invocation.

Proof. We use the notation of Lemma 4.7 at $M^{(0)}$. The values $q_i := q_i(M^{(0)})$ are leverage scores, so $0 < q_i \le 1$ and $\sum_{i=1}^m q_i P_i = I$. For every symmetric X,

$$\sum_{i=1}^m q_i^{\gamma} (e_i^\top X e_i)^2 \le \sum_{i=1}^m q_i e_i^\top X^2 e_i = \operatorname{tr}[X^2].$$

Thus, $\Lambda(M^{(0)}) \le 1$.

Put $a := 1 - 1/(2\gamma) = (2\gamma - 1)/(2\gamma)$ and $\kappa := \kappa_{\gamma, 2\gamma}$. The scalar envelope used in the proof of Lemma 4.8 has value one at $d = 1$, while $\kappa = 2(1 - 1/(2\gamma - 1))^{\gamma - 1} < 2$. Thus, $1 \le \kappa < 2$. Consequently, every update segment starts at a matrix with curvature at most κ. For the segment starting at $M^{(0)}$, this follows from $\Lambda(M^{(0)}) \le 1 \le \kappa$; for every segment starting at $M^{(t)}$ with $t \ge 1$, it follows by applying the universal reset to the preceding update.

Consider one update from M to M^+ and the straight segment $M_\theta := (1-\theta)M + \theta M^+$, where $0 \le \theta \le 1$. In the notation of the reset proof, $M^+ = SR_M^{-1}S^\top$ and $R_M \succeq aI$. Hence

$$M^+ \preceq a^{-1}M, \qquad M_\theta \preceq a^{-1}M, \qquad M_\theta^{-1} \succeq aM^{-1}.$$

We claim that

$$\Lambda(M_\theta) \le a^{-\gamma}\Lambda(M). \tag{8}$$

Indeed, $q_i(M_\theta)^{\gamma-2} \le a^{-(\gamma-2)}q_i(M)^{\gamma-2}$. To compare the denominator in the definition of Λ, put $N := M^{-1}$, $N_\theta := M_\theta^{-1}$,

$$Q := N_\theta^{-1/2} N N_\theta^{-1/2} \preceq a^{-1}I, \qquad Z := N_\theta^{1/2} H N_\theta^{1/2}.$$

For every symmetric, possibly indefinite, H,

$$\operatorname{tr}[NHNH] = \operatorname{tr}[QZQZ] = \|Q^{1/2}ZQ^{1/2}\|_{\mathrm{F}}^2 \le a^{-2}\|Z\|_{\mathrm{F}}^2 = a^{-2}\operatorname{tr}[N_\theta H N_\theta H].$$

The first step follows from the definitions of Q and Z and cyclicity of trace. The second step uses $\|S\|_{\mathrm{F}}^2 = \operatorname{tr}[S^2]$ for the symmetric matrix $S := Q^{1/2}ZQ^{1/2}$. The third step uses $\|Q\|_{\mathrm{op}} \le a^{-1}$, which follows from $Q \preceq a^{-1}I$, together with submultiplicativity of the Frobenius norm. The final step follows from the definition of Z and cyclicity of trace. Combining the numerator and denominator comparisons proves Eq. (8).

The constants satisfy the exact identity

$$\frac{\kappa a^{-\gamma}}{(2\gamma-1)/(\gamma-1)} = (1 - \frac{1}{(2\gamma-1)^2})^\gamma < 1. \tag{9}$$

Hence every update segment satisfies $\Lambda(M_\theta) < (2\gamma-1)/(\gamma-1)$. Lemma 4.7-(b) therefore gives

$$\nabla^2 F(M_\theta) \preceq (1 + (\gamma-1)\Lambda(M_\theta))\nabla^2 h(M_\theta) \preceq 2\gamma\nabla^2 h(M_\theta).$$

Moreover, $\nabla^2 F \succeq \nabla^2 h$ globally. If $\Delta := M^+ - M$, then integrating the upper inequality on the same straight segment gives

$$D_F(M^+, M) = \int_0^1 (1-\theta)\nabla^2 F(M_\theta)[\Delta,\Delta]\,\mathrm{d}\theta \le 2\gamma D_h(M^+, M).$$

This is the upper relative model required by Lemma 4.5; the global lower Hessian inequality provides its lower relative model. Apply that lemma on the convex positive-definite cone with $x_* = M_*$. Since $L_{\mathrm{g}} = 2\gamma$, that lemma yields

$$D_h(M_*, M^{(t+1)}) \le (1 - \frac{1}{L_{\mathrm{g}}})D_h(M_*, M^{(t)}).$$

Lemma 4.6 now gives $D_h(M_*, M^{(t)}) \le (m-n)\exp(-t/L_{\mathrm{g}}) \le m\exp(-t/L_{\mathrm{g}})$. The stated value of T_{g} gives the required basin bound, and Eq. (3) shows that one oracle invocation implements each iteration. □

4.4 Local vector geometry

For the accelerated phase, set $\alpha := 1/\beta \in (0,1]$. For $v \in \mathbb{R}^m_{>0}$, let $V := \mathrm{Diag}(v)$ and define

$$M(v) := (A^\top V A)^{-1}, \qquad q_i(v) := a_i^\top M(v) a_i,$$

$$F_{\mathrm{vec}}(v) := -\log\det(A^\top V A) + h_{\mathrm{vec}}(v), \qquad h_{\mathrm{vec}}(v) := \frac{1}{1+\alpha}\sum_{i=1}^m v_i^{1+\alpha}.$$

Define v_* coordinatewise by the first identity in the following display:

$$[v_*]_i := q_i(M_*)^\beta, \qquad [v_*]_i^\alpha = q_i(M_*), \qquad M(v_*) = M_*. \tag{10}$$

Eq. (1) gives $M(v_*) = M_*$. Since $\alpha\beta = 1$, it also gives $[v_*]_i^\alpha = q_i(M_*)$. Direct differentiation therefore yields $\nabla F_{\mathrm{vec}}(v_*) = 0$. Moreover, $\nabla^2 h_{\mathrm{vec}}(v) = \alpha\,\mathrm{Diag}(v_i^{\alpha-1}) \succ 0$, while the log-determinant term is convex. Hence F_{vec} is strictly convex and v_* is its unique minimizer. Finally, multiplying Eq. (1) by M_* and taking the trace gives

$$\sum_{i=1}^m [v_*]_i^{1+\alpha} = \sum_{i=1}^m q_i(M_*)^{\beta+1} = \mathrm{tr}[M_* M_*^{-1}] = n.$$

For later use, define

$$\rho_i(v) := \frac{q_i(v)}{v_i^\alpha}, \qquad \rho_{\max}(v) := \max_{i\in[m]} \rho_i(v).$$

Let $P(v) := V^{1/2} A M(v) A^\top V^{1/2}$ and let $P(v)^{(2)} := P(v) \circ P(v)$. Direct differentiation gives

$$\nabla F_{\mathrm{vec}}(v) = v^\alpha - q(v), \qquad \nabla^2 F_{\mathrm{vec}}(v) = V^{-1} P(v)^{(2)} V^{-1} + \alpha V^{\alpha-1}. \tag{11}$$

Since $P(v)$ is an orthogonal projection,

$$\sum_{j=1}^m P(v)_{ij}^2 = P(v)_{ii}.$$

Thus, the difference in the next display is a weighted graph Laplacian. Indeed, for every $x \in \mathbb{R}^m$,

$$x^\top(\mathrm{Diag}(P(v)_{11},\dots,P(v)_{mm}) - P(v)^{(2)})x = \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m P(v)_{ij}^2 (x_i - x_j)^2 \geq 0.$$

It follows that $P(v)^{(2)} \preceq \mathrm{Diag}(P(v)_{11},\dots,P(v)_{mm})$. Consequently,

$$\nabla^2 h_{\mathrm{vec}}(v) \preceq \nabla^2 F_{\mathrm{vec}}(v) \preceq (1 + \beta\rho_{\max}(v))\nabla^2 h_{\mathrm{vec}}(v). \tag{12}$$

The global phase produces a matrix that is spectrally close to M_*, whereas the accelerated phase requires a coordinate box known to the algorithm. The next lemma converts spectral proximity into such a box and establishes the relative strong convexity, smoothness, triangle-scaling, and orientation bounds needed inside it.

Lemma 4.10 (A known local box)**.** *Let β and γ be as in Definition 4.1. Let M_* be the optimal matrix from Definition 4.3. Suppose*

$$(1-\eta)M_* \preceq M_{\mathrm{basin}} \preceq (1+\eta)M_*, \qquad \eta := \frac{1}{64\beta},$$

and set $u_i := q_i(M_{\mathrm{basin}})^\beta$. Then the following statements hold.

(a) For every $i \in [m]$,

$$\frac{63}{64} \le \frac{u_i}{[v_*]_i} \le e^{1/64}.$$

In particular, the known box

$$\mathcal{B} := \{v \in \mathbb{R}^m_{>0} : u/2 \le v \le 2u\}$$

contains v_*.

(b) On $\mathcal{B}$, F_{vec} is 1-strongly convex and L-smooth relative to h_{vec} for $L := 10\beta$.

(c) With $G := 4$, for all $x, z, \widetilde{z} \in \mathcal{B}$ and $\theta \in [0,1]$,

$$D_{h_{\mathrm{vec}}}((1-\theta)x + \theta\widetilde{z}, (1-\theta)x + \theta z) \le G\theta^2 D_{h_{\mathrm{vec}}}(\widetilde{z}, z).$$

(d) For the same G and all $a, b \in \mathcal{B}$,

$$D_{h_{\mathrm{vec}}}(a, b) \le G D_{h_{\mathrm{vec}}}(b, a).$$

Proof. **Proof of part (a).** The spectral assumption gives $1 - \eta \le q_i(M_{\mathrm{basin}})/q_i(M_*) \le 1 + \eta$. Raising this bound to the power β and applying Bernoulli's inequality and $1 + x \le e^x$ gives the claimed bounds on $u_i/[v_*]_i$. In particular, $u/2 \le v_* \le 2u$, so $v_* \in \mathcal{B}$. The same bounds imply

$$v_*/3 \le v \le 3v_* \tag{13}$$

for every $v \in \mathcal{B}$. This proves part (a).

Proof of part (b). Eq. (13) implies $M_*/3 \preceq M(v) \preceq 3M_*$, and hence $q_i(v)/q_i(M_*) \in [1/3, 3]$. Eq. (10) now gives $\rho_{\max}(v) \le 3^{1+\alpha} \le 9$. Eq. (12) and Lemma 3.3 prove 1-relative strong convexity and relative smoothness with $L = 1 + 9\beta \le 10\beta$. This proves part (b).

Proof of part (c). For any $r, s \in \mathcal{B}$, every coordinate ratio r_i/s_i belongs to $[1/4, 4]$. Since

$$\nabla^2 h_{\mathrm{vec}}(r) = \alpha \,\mathrm{Diag}(r_i^{\alpha-1}),$$

we have

$$\nabla^2 h_{\mathrm{vec}}(r) \preceq 4^{1-\alpha}\nabla^2 h_{\mathrm{vec}}(s) \preceq G\nabla^2 h_{\mathrm{vec}}(s).$$

The first inequality uses the coordinate-ratio bound, and the second uses $\alpha \le 1$ and $G = 4$. For any twice continuously differentiable convex ψ on an open set containing the segment $[a, b]$,

$$D_\psi(a, b) = \int_0^1 (1-t)(a-b)^\top \nabla^2\psi(b + t(a-b))(a-b)\,\mathrm{d}t.$$

For the claim in part (c), apply this formula to the segment from $(1-\theta)x + \theta z$ to $(1-\theta)x + \theta\widetilde{z}$. Its displacement is $\theta(\widetilde{z} - z)$. More explicitly, put

$$r_t := z + t(\widetilde{z} - z), \qquad \widehat{r}_t := (1-\theta)x + \theta r_t.$$

Both paths lie in $\mathcal{B}$, and the uniform Hessian comparison gives $\nabla^2 h_{\mathrm{vec}}(\widehat{r}_t) \preceq G\nabla^2 h_{\mathrm{vec}}(r_t)$ pointwise. Hence

$$D_{h_{\mathrm{vec}}}((1-\theta)x + \theta\widetilde{z}, (1-\theta)x + \theta z) = \theta^2 \int_0^1 (1-t)(\widetilde{z} - z)^\top \nabla^2 h_{\mathrm{vec}}(\widehat{r}_t)(\widetilde{z} - z)\mathrm{d}t \le G\theta^2 D_{h_{\mathrm{vec}}}(\widetilde{z}, z),$$

The first step applies the integral Bregman formula along $\widehat{r}_t$, and the second step uses the pointwise Hessian comparison followed by the same formula along r_t. This proves part (c).

Proof of part (d). Put $\Delta := a - b$ and $H_t := \nabla^2 h_{\mathrm{vec}}(b + t\Delta)$. The two orientations of the integral formula are

$$D_{h_{\mathrm{vec}}}(a, b) = \int_0^1 (1-t)\Delta^\top H_t \Delta \, \mathrm{d}t, \qquad D_{h_{\mathrm{vec}}}(b, a) = \int_0^1 t\Delta^\top H_t \Delta \, \mathrm{d}t.$$

The uniform Hessian comparison gives $H_t \preceq G H_{1-t}$. Therefore,

$$D_{h_{\mathrm{vec}}}(a, b) \le G \int_0^1 (1-t)\Delta^\top H_{1-t} \Delta \, \mathrm{d}t = G \int_0^1 u\Delta^\top H_u \Delta \, \mathrm{d}u = G D_{h_{\mathrm{vec}}}(b, a).$$

The first step follows from the integral representation of $D_{h_{\mathrm{vec}}}(a, b)$ and the pointwise comparison $H_t \preceq G H_{1-t}$. The second step uses the substitution $u := 1 - t$. The final step follows from the integral representation of $D_{h_{\mathrm{vec}}}(b, a)$ above. This proves part (d). □

4.5 Boxed Bregman acceleration

We next give the accelerated epoch used in the local phase.

Algorithm 3 Boxed accelerated restart phase

procedure BoxedAcceleratedRestarts($m, n, p, \epsilon, u, \mathcal{O}_A$) ▷ Lemma 4.11
  $\beta \leftarrow (p-2)/2$, $\alpha \leftarrow 1/\beta$
  $L \leftarrow 10\beta$, $G \leftarrow 4$, $K \leftarrow \lceil 40\sqrt{\beta} \rceil$
  $d_\epsilon \leftarrow \epsilon^2/(256(\beta+1)^2)$
  $R \leftarrow 1 + \lceil \log_2(n/d_\epsilon) \rceil$
  $\mathcal{B} \leftarrow \{v \in \mathbb{R}^m_{>0} : u/2 \le v \le 2u\}$, $s^{(0)} \leftarrow u$
  for $j = 0, \dots, R-1$ **do**
    $x^{(0)} \leftarrow s^{(j)}$, $z^{(0)} \leftarrow s^{(j)}$
    for $k = 0, \dots, K-1$ **do**
      $\theta_k \leftarrow 2/(k+2)$
      $y^{(k)} \leftarrow (1-\theta_k)x^{(k)} + \theta_k z^{(k)}$
      $\ell^{(j,k)} \leftarrow \mathcal{O}_A(y^{(k)})$
      $q_i^{(j,k)} \leftarrow \ell_i^{(j,k)}/y_i^{(k)}$ for every $i \in [m]$
      $r_i^{(j,k)} \leftarrow [z_i^{(k)}]^\alpha + (q_i^{(j,k)} - [y_i^{(k)}]^\alpha)/(\theta_k G L)$ for every $i \in [m]$
      for $i = 1, \dots, m$ **do**
        if $r_i^{(j,k)} \le 0$ **then**
          $z_i^{(k+1)} \leftarrow u_i/2$
        else
          $z_i^{(k+1)} \leftarrow \mathrm{clip}_{[u_i/2, 2u_i]}((r_i^{(j,k)})^{1/\alpha})$
        end if
      end for
      $x^{(k+1)} \leftarrow (1-\theta_k)x^{(k)} + \theta_k z^{(k+1)}$
    end for
    $s^{(j+1)} \leftarrow x^{(K)}$
  end for
  return $s^{(R)}$
end procedure

The procedure BoxedAcceleratedRestarts restarts an accelerated Bregman method inside the known box. The next lemma analyzes one epoch, proving feasibility, the accelerated objective-gap estimate, and a constant-factor contraction of the reverse divergence.

Lemma 4.11 (Boxed accelerated epoch for Algorithm 3)**.** *Let β and γ be as in Definition 4.1. Consider one epoch of the procedure* BoxedAcceleratedRestarts *in Algorithm 3. Assume the setting of Lemma 4.10, let $L := 10\beta$, $G := 4$, and set $K := \lceil 40\sqrt{\beta} \rceil$. Given $s \in \mathcal{B}$, initialize $x^{(0)} := s$ and $z^{(0)} := s$. For $k = 0, \dots, K-1$, set $\theta_k := 2/(k+2)$ and*

$$\begin{aligned} y^{(k)} &:= (1-\theta_k)x^{(k)} + \theta_k z^{(k)}, \\ z^{(k+1)} &:= \arg\min_{z \in \mathcal{B}} \{\langle \nabla F_{\mathrm{vec}}(y^{(k)}), z - y^{(k)} \rangle + \theta_k G L D_{h_{\mathrm{vec}}}(z, z^{(k)})\}, \\ x^{(k+1)} &:= (1-\theta_k)x^{(k)} + \theta_k z^{(k+1)}. \end{aligned} \tag{14}$$

Then the following statements hold.

(a) All iterates lie in $\mathcal{B}$, each iteration uses one oracle invocation, and

$$F_{\mathrm{vec}}(x^{(K)}) - F_{\mathrm{vec}}(v_*) \le \frac{4GL}{(K+1)^2} D_{h_{\mathrm{vec}}}(v_*, s).$$

(b)

$$D_{h_{\mathrm{vec}}}(v_*, x^{(K)}) \le \frac{1}{2} D_{h_{\mathrm{vec}}}(v_*, s).$$

Proof. **Proof of part (a).** The box $\mathcal{B}$ is convex. Hence the initialization, the convex-combination updates, and the constraint in Eq. (14) keep $x^{(k)}$, $y^{(k)}$, and $z^{(k)}$ in $\mathcal{B}$.

The prox step is explicit. Define

$$r_i^{(k)} := [z_i^{(k)}]^\alpha + \frac{q_i(y^{(k)}) - [y_i^{(k)}]^\alpha}{\theta_k G L}.$$

Then

$$z_i^{(k+1)} := \begin{cases} u_i/2, & r_i^{(k)} \le 0, \\ \mathrm{clip}_{[u_i/2, 2u_i]}((r_i^{(k)})^{1/\alpha}), & r_i^{(k)} > 0. \end{cases} \tag{15}$$

Here $\mathrm{clip}_{[c,d]}(t) := \min\{d, \max\{c, t\}\}$. Indeed, differentiating the scalar objective in Eq. (14) gives the unconstrained solution when $r_i^{(k)} > 0$; when $r_i^{(k)} \le 0$, that objective is increasing on the positive half-line, so its constrained minimizer is the lower endpoint. Finally, with $Y^{(k)} := \mathrm{Diag}(y^{(k)})$, the quantity $y_i^{(k)} q_i(y^{(k)})$ is the ith leverage score of $(Y^{(k)})^{1/2} A$. Since $y_i^{(k)} > 0$ is known, division by $y_i^{(k)}$ recovers $q_i(y^{(k)})$. Thus, all coordinates in Eq. (15) are obtained from one oracle invocation.

Write

$$\ell(w \mid y) := F_{\mathrm{vec}}(y) + \langle \nabla F_{\mathrm{vec}}(y), w - y \rangle.$$

Fix $w \in \mathcal{B}$ and put $g_k := \nabla F_{\mathrm{vec}}(y^{(k)})$. Lemma 4.10-(b), applied with $x := x^{(k+1)}$ and $y := y^{(k)}$, gives

$$\begin{aligned} F_{\mathrm{vec}}(x^{(k+1)}) &\le F_{\mathrm{vec}}(y^{(k)}) + \langle g_k, x^{(k+1)} - y^{(k)} \rangle + L D_{h_{\mathrm{vec}}}(x^{(k+1)}, y^{(k)}) \\ &= \ell(x^{(k+1)} \mid y^{(k)}) + L D_{h_{\mathrm{vec}}}(x^{(k+1)}, y^{(k)}). \end{aligned}$$

The definitions of the iterates give

$$y^{(k)} = (1-\theta_k)x^{(k)} + \theta_k z^{(k)}, \qquad x^{(k+1)} = (1-\theta_k)x^{(k)} + \theta_k z^{(k+1)}.$$

As proved above, $x^{(k)}, z^{(k)}, z^{(k+1)} \in \mathcal{B}$, and $\theta_k \in [0,1]$. Therefore, Lemma 4.10-(c), applied with $x := x^{(k)}$, $z := z^{(k)}$, and $\widetilde{z} := z^{(k+1)}$, gives

$$D_{h_{\mathrm{vec}}}(x^{(k+1)}, y^{(k)}) \le G\theta_k^2 D_{h_{\mathrm{vec}}}(z^{(k+1)}, z^{(k)}).$$

Substituting this estimate into the preceding relative-smoothness bound gives

$$F_{\mathrm{vec}}(x^{(k+1)}) \le \ell(x^{(k+1)} \mid y^{(k)}) + GL\theta_k^2 D_{h_{\mathrm{vec}}}(z^{(k+1)}, z^{(k)}).$$

Because $\ell(\cdot \mid y^{(k)})$ is affine and $x^{(k+1)} = (1-\theta_k)x^{(k)} + \theta_k z^{(k+1)}$,

$$\ell(x^{(k+1)} \mid y^{(k)}) = (1-\theta_k)\ell(x^{(k)} \mid y^{(k)}) + \theta_k \ell(z^{(k+1)} \mid y^{(k)}).$$

Convexity gives $\ell(x^{(k)} \mid y^{(k)}) \le F_{\mathrm{vec}}(x^{(k)})$ and $\ell(w \mid y^{(k)}) \le F_{\mathrm{vec}}(w)$, so

$$\ell(z^{(k+1)} \mid y^{(k)}) - F_{\mathrm{vec}}(w) \le \langle g_k, z^{(k+1)} - w \rangle.$$

The three-point optimality inequality for Eq. (14) gives

$$\langle g_k, z^{(k+1)} - w \rangle \le \theta_k GL(D_{h_{\mathrm{vec}}}(w, z^{(k)}) - D_{h_{\mathrm{vec}}}(w, z^{(k+1)}) - D_{h_{\mathrm{vec}}}(z^{(k+1)}, z^{(k)})).$$

Multiplying this inequality by θ_k and combining the preceding displays cancels the last divergence term and gives

$$F_{\mathrm{vec}}(x^{(k+1)}) - F_{\mathrm{vec}}(w) \le (1-\theta_k)(F_{\mathrm{vec}}(x^{(k)}) - F_{\mathrm{vec}}(w)) + GL\theta_k^2(D_{h_{\mathrm{vec}}}(w, z^{(k)}) - D_{h_{\mathrm{vec}}}(w, z^{(k+1)})). \tag{16}$$

The final divergence term cancels exactly with the triangle-scaling term.

Now set $w := v_*$. Since v_* minimizes F_{vec}, the gaps $A_k := F_{\mathrm{vec}}(x^{(k)}) - F_{\mathrm{vec}}(v_*)$ are nonnegative. Put $D_k := D_{h_{\mathrm{vec}}}(v_*, z^{(k)})$. Dividing Eq. (16) by θ_k^2 gives

$$\frac{A_{k+1}}{\theta_k^2} + GLD_{k+1} \le \frac{1-\theta_k}{\theta_k^2} A_k + GLD_k.$$

For $k = 0$, the right-hand side is GLD_0 because $\theta_0 = 1$. For $k \ge 1$, the nonnegativity of A_k and $(1-\theta_k)/\theta_k^2 \le 1/\theta_{k-1}^2$ imply

$$\frac{A_{k+1}}{\theta_k^2} + GLD_{k+1} \le \frac{A_k}{\theta_{k-1}^2} + GLD_k.$$

Induction therefore gives

$$\frac{A_K}{\theta_{K-1}^2} + GLD_K \le GLD_0.$$

Discarding the final nonnegative divergence and substituting $\theta_{K-1} = 2/(K+1)$ gives

$$F_{\mathrm{vec}}(x^{(K)}) - F_{\mathrm{vec}}(v_*) \le GL\theta_{K-1}^2 D_{h_{\mathrm{vec}}}(v_*, s) = \frac{4GL}{(K+1)^2} D_{h_{\mathrm{vec}}}(v_*, s).$$

This proves part (a).

Proof of part (b). Lemma 3.3, applied to Eq. (12), shows that F_{vec} is 1-strongly convex relative to h_{vec}. Together with Lemma 4.10-(d) and part (a), it gives

$$D_{h_{\mathrm{vec}}}(v_*, x^{(K)}) \le GD_{h_{\mathrm{vec}}}(x^{(K)}, v_*) \le G(F_{\mathrm{vec}}(x^{(K)}) - F_{\mathrm{vec}}(v_*)) \le \frac{4G^2L}{(K+1)^2} D_{h_{\mathrm{vec}}}(v_*, s).$$

The first step is Lemma 4.10-(d), the second is relative strong convexity, and the third is part (a). Since $G = 4$, $L = 10\beta$, and $K + 1 \ge 40\sqrt{\beta}$, the final coefficient is at most $2/5 < 1/2$. This proves part (b). □

4.6 Output recovery

The following identity transfers convergence of the vector potential directly to the matrix divergence used for the final coordinatewise estimate. It decomposes the objective gap into two nonnegative Bregman divergences, so each divergence is bounded by the same gap.

Lemma 4.12 (Exact dual-gap identity)**.** *Let M_* be the optimal matrix from Definition 4.3. For every $v \in \mathbb{R}^m_{>0}$,*

$$F_{\mathrm{vec}}(v) - F_{\mathrm{vec}}(v_*) = D_h(M_*, M(v)) + D_{h_{\mathrm{vec}}}(v, v_*).$$

Proof. Put $N(v) := A^\top V A = M(v)^{-1}$ and $N_* := M_*^{-1}$. Eq. (10) gives $M(v_*) = M_*$, and hence $N(v_*) = N_*$.

We first expand the matrix divergence. Since $h(M) = -\log\det M$ and $\nabla h(M) = -M^{-1}$,

$$\begin{aligned} D_h(M_*, M(v)) &= -\log\det M_* + \log\det M(v) + \mathrm{tr}[N(v)(M_* - M(v))] \\ &= \log\det N_* - \log\det N(v) + \mathrm{tr}[M_* N(v)] - n. \end{aligned}$$

The first step applies Definition 3.2 with $\psi := h$, $x := M_*$, and $y := M(v)$. The second step uses $M_* = N_*^{-1}$, $M(v) = N(v)^{-1}$, cyclicity of trace, and $\mathrm{tr}[N(v)M(v)] = n$.

We next expand the vector divergence. By the definition of h_{vec} and Eq. (10), for every $i \in [m]$,

$$[\nabla h_{\mathrm{vec}}(v_*)]_i = [v_*]_i^\alpha = q_i(M_*) = a_i^\top M_* a_i.$$

Therefore,

$$\begin{aligned} \langle \nabla h_{\mathrm{vec}}(v_*), v - v_* \rangle &= \sum_{i=1}^m q_i(M_*)(v_i - [v_*]_i) \\ &= \mathrm{tr}[M_* \sum_{i=1}^m (v_i - [v_*]_i) a_i a_i^\top] \\ &= \mathrm{tr}[M_*(N(v) - N_*)] \\ &= \mathrm{tr}[M_* N(v)] - n. \end{aligned}$$

Here the third step uses $N(v) = \sum_{i=1}^m v_i a_i a_i^\top$ and $N_* = N(v_*) = \sum_{i=1}^m [v_*]_i a_i a_i^\top$, while the last step uses $M_* N_* = I$. Applying Definition 3.2 with $\psi := h_{\mathrm{vec}}$, $x := v$, and $y := v_*$ gives

$$D_{h_{\mathrm{vec}}}(v, v_*) = h_{\mathrm{vec}}(v) - h_{\mathrm{vec}}(v_*) - \mathrm{tr}[M_* N(v)] + n.$$

Finally, adding the two divergence formulas cancels the trace and constant terms:

$$\begin{aligned} D_h(M_*, M(v)) + D_{h_{\mathrm{vec}}}(v, v_*) &= \log\det N_* - \log\det N(v) + h_{\mathrm{vec}}(v) - h_{\mathrm{vec}}(v_*) \\ &= F_{\mathrm{vec}}(v) - F_{\mathrm{vec}}(v_*). \end{aligned}$$

The last step follows from the definition $F_{\mathrm{vec}}(z) = -\log\det N(z) + h_{\mathrm{vec}}(z)$ and $N(v_*) = N_*$. This proves the claim. □

Acknowledgments

The AI tools being used is codex sol 5.6. The author use AI tools to polish the grammar issues of english.

References

[CP15] Michael B. Cohen and Richard Peng. ℓ_p row sampling by Lewis weights. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, pages 183–192. ACM, 2015.

[FLPS22] Maryam Fazel, Yin Tat Lee, Swati Padmanabhan, and Aaron Sidford. Computing Lewis weights to high precision. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2723–2742. SIAM, 2022.

[GSZ26] Sander Gribling, Aaron Sidford, and Chenyi Zhang. Computing Lewis weights to high precision using local relative smoothness. In Steve Hanneke and Tor Lattimore, editors, *Proceedings of Thirty Ninth Conference on Learning Theory*, volume 336 of *Proceedings of Machine Learning Research*, pages 2901–2939. PMLR, 29 Jun–03 Jul 2026.

[Lew78] D. R. Lewis. Finite dimensional subspaces of L_p. *Studia Mathematica*, 63(2):207–212, 1978.

[LS19] Yin Tat Lee and Aaron Sidford. Solving linear programs with $\sqrt{\text{rank}}$ linear system solves, 2019.