

Filtered Stable Rank and Tail-Local Limits on Oversampling in Randomized SVD

Effective Rank RSVD Study Group

June 25, 2026

Abstract

Randomized SVD is often run with a fixed oversampling parameter, even though approximation quality is driven by the spectrum and by the power-iteration count. We test whether the stable rank of a matrix can govern the oversampling needed to reach a target relative error, and whether this yields an a-priori rule that is independent of ambient size and target rank. The answer is conditional. In finite Gaussian range finding, global stable rank alone does not remove target-local tail dependence: for power-law spectra with exponent greater than one half, stable rank can remain bounded while a normalized tail quantity grows with the target rank. We define a power-filtered rank that contracts under subspace iteration, derive a conditional oversampling threshold under an explicit intrinsic-tail premise, and validate the algebra with seven symbolic checks. Executed experiments on 1,728 synthetic randomized SVD runs and 648 real-matrix runs show mixed empirical support. Normalizing by filtered rank improves collapse only slightly on average, with mean relative spread improvement 0.019 over global stable rank. Conditional tail fits are mostly decreasing in oversampling per filtered rank, but a finite two-parameter rule has only 80% training and held-out coverage and no feasible cell in the tested calibration grid. On SuiteSparse SNAP/ca-GrQc, MovieLens 100K, and a Fashion-MNIST train slice, matched-accuracy transfer rows show zero positive sketch-column savings. The supported conclusion is therefore not a universal stable-rank-only law. Effective rank is useful when paired with power filtering and tail-local diagnostics, and the resulting rule should be treated as a scoped calibration rather than a general replacement for existing oversampling heuristics.

1 Introduction

Low-rank approximation is one of the main ways large matrices become tractable in machine learning, scientific computing, recommender systems, compression, and exploratory data analysis. Principal components, latent-factor models, kernel approximations, and model-compression routines all depend on the same operational question: how many matrix-vector products or sketch columns are enough to retain the part of the spectrum that matters? Randomized SVD addresses this question by replacing a full deterministic factorization with a smaller randomized range-finding problem. The Halko-Martinsson-Tropp algorithm chooses a target rank k , an oversampling parameter p , a sketch size $\ell = k + p$, and often a small number q of power iterations before computing a basis for the sampled range [12, 24, 41]. Its appeal is practical as much as theoretical: it can turn expensive dense factorizations into a sequence of matrix products and a small SVD, and software packages expose the method through a few parameters that users often leave at default values [8, 26, 39, 45].

The oversampling parameter is the least interpretable of these user choices. Fixed recommendations such as $p = 5$ or $p = 10$ are robust enough to be useful, but they do not explain why two matrices of the same size and target rank can have very different returns from extra samples. Classical analysis already shows that the answer cannot be ambient dimension alone: the singular-value tail, the gap near k , the sketch distribution, and power iteration all enter existing bounds [12, 24, 36]. At the same time, randomized numerical linear algebra has repeatedly found that intrinsic dimension can replace ambient dimension in matrix concentration, approximate matrix products, and stable-rank-sensitive algorithms [4, 14, 22, 40, 46]. This motivates a sharper question: can effective rank, rather than size or target rank, determine how much oversampling randomized SVD needs?

We study that question for finite matrices with controlled spectra and Gaussian sketches. The target is not an adaptive residual estimator, but an a-priori rule that is available before the randomized range finder is run. Such a rule would be useful when factorization plans must be set before data movement, when repeated decompositions are scheduled across many matrices, or when a production interface should choose a conservative oversampling value without exposing all numerical details to the user. The original optimistic form of the question asks whether p is governed primarily by the stable rank $\|A\|_F^2/\|A\|_2^2$ and a tolerance ϵ , essentially independent of k and of the ambient dimensions once the matrix is much larger than its intrinsic dimension. Our results narrow this framing. Stable rank is a relevant spectral summary, but it is not enough by itself because it is global, while randomized SVD error is judged against the tail beyond a target rank.

This paper makes four contributions.

- We define a power-filtered stable rank $r_q(A)$ and a target-local normalized tail $T_k(A)$ for finite nonzero matrices, separating global spectrum size from the tail that sets rank- k error.
- We prove that bounded global stable rank does not preclude target-local tail growth for finite power-law spectra, so a stable-rank-only oversampling law cannot hold without additional regime restrictions.
- We derive a conditional chance-threshold bound for the minimal oversampling p^* under an explicit intrinsic-tail premise and a monotone two-parameter rule $\hat{p}_\theta$ for finite-grid calibration.
- We validate the formal and empirical claims with symbolic checks, synthetic controlled spectra, calibrated holdout rows, a boundary atlas, and real-transfer matrices from graph, recommender, and image domains.

The empirical findings are intentionally reported with their negative edges. Synthetic controlled spectra support the usefulness of q -aware normalization, but only weakly as a universal collapse coordinate. The calibrated rule is conservative and sometimes insufficient, with no feasible cell in the tested two-parameter grid. Real-transfer comparisons satisfy a matched-accuracy gate, yet do not save sketch columns against fixed- p comparators. These outcomes matter because they show where effective rank helps and where it cannot substitute for target-local information. The central conclusion is a claims-limiting one: the relevant intrinsic dimension for randomized SVD oversampling is filtered and tail-aware, not merely the stable rank of the original matrix.

2 Related Work

Randomized low-rank approximation has a long line of algorithms that differ in sketch type, pass count, reconstruction, and error target. Early Monte Carlo and projection methods established that random sampling and random projections can produce useful low-rank approximations at much lower cost than deterministic factorizations [9, 20, 33, 38, 47]. The HMT framework unified many of these ideas into a two-stage range finder followed by deterministic compression, with explicit roles for k , p , ℓ , and optional power iteration [12]. Later surveys and implementation papers refined this picture by emphasizing pass efficiency, sparse embeddings, block methods, and software structure [7, 23, 24, 26]. The strength of this lineage is that it gives robust algorithms and interpretable error comparisons against A_k . Its limitation for our question is that the standard bounds and defaults do not deliver a simple a-priori formula for p from an intrinsic rank statistic alone.

A second body of work explains why sketch size can depend on intrinsic structure rather than only on ambient dimension. Stable rank appears in approximate matrix multiplication, matrix sampling, geometric functional analysis, and non-Gaussian low-rank approximation [4, 14, 35, 37]. Matrix concentration inequalities provide the technical template for replacing dimension with variance or intrinsic-dimension parameters in tail bounds [40]. Effective rank also appears in covariance estimation, where trace-to-operator-norm ratios quantify the complexity of a population covariance operator [2, 17, 18]. These results make stable rank a natural candidate for randomized SVD oversampling. They also mark the boundary of what can be borrowed. Approximate matrix products, covariance operators, and non-Gaussian sketch analyses do not directly prove a stable-rank-only HMT residual bound. In this paper, they motivate a conditional concentration premise, while the manuscript-defined quantities $r_q(A)$ and $T_k(A)$ handle the specific rank- k residual structure.

Power iteration and Krylov methods form a third comparison axis. Randomized subspace iteration replaces $A\Omega$ with $(AA^*)^q A\Omega$, amplifying singular values as σ_i^{2q+1} and improving separation when the spectrum decays slowly [11, 12, 36]. Block Krylov methods can be stronger when spectral-norm accuracy is demanding, and recent work continues to study block size, pass budgets, and model-compression settings [3, 27, 32, 41]. These methods show why a rule based on $r_{\text{eff}}(A)$ alone is suspect: the relevant spectrum after q power iterations is not the original spectrum but the filtered profile. Our filtered rank $r_q(A)$ is not a new algorithm; it is a diagnostic and rule input aligned with the spectrum actually sampled by the power method.

A fourth line of work uses adaptivity rather than a one-shot oversampling choice. Fixed-precision randomized low-rank approximation, rank-revealing randomized SVD, numerical-rank estimation, and fixed-threshold methods choose rank or sketch size after observing residual information [15, 21, 25, 49]. These methods attack the same practical pain point as an a-priori rule but with a different contract. They can be more reliable because they observe the matrix through incremental sketches, yet they may require additional passes or stopping logic. Our study does not try to dominate adaptive methods. It asks how far a precomputed rule can go and where its evidence fails. The comparison to fixed- p and k -proportional baselines is therefore a conservative first benchmark, not a claim that adaptivity is unnecessary.

Recent specialized variants also warn against overgeneralization. Noise sensitivity can alter randomized SVD behavior even when the nominal singular spectrum looks favorable [34]. Row-aware and structured variants account for geometry or matrix class beyond singular values [28–31]. Explanations of approximate low rank in data matrices emphasize latent structure, smoothness, and physical or statistical mechanisms that stable rank alone may not capture [6, 44]. These works motivate the failure atlas in this paper: coherence, preprocessing, noise, and exact-rank denominators are not nuisances to hide after the fact, but validity conditions that determine whether an oversampling rule is meaningful.

The real-matrix validation sources are deliberately varied. SuiteSparse supplies sparse graph matrices and a well-documented collection interface [5, 16, 19]. MovieLens 100K supplies a sparse rectangular user-item matrix whose missing entries are not observed zeros and whose historical collection process matters for interpretation [10, 13]. Fashion-MNIST supplies a dense image matrix with fixed dimensions, preprocessing conventions, and repository-level loader details [48, 50]. These datasets are not evidence that a universal oversampling law holds. They test whether a rule calibrated on synthetic spectra transfers across graph, recommender, and image matrix regimes with declared preprocessing.

3 Problem Setting and Scope

The problem setting fixes the mathematical objects before introducing any rule. This is necessary because the same word “rank” can refer to a target rank, an intrinsic-dimension proxy, a numerical rank, or the rank of a postprocessed approximation. Randomized SVD also mixes deterministic matrix approximation with probability over the sketch. The definitions below separate these layers: the HMT range finder and optimal rank- k comparator are inherited from randomized low-rank approximation theory [12, 24], while the filtered rank, target-local tail, threshold, and calibrated rule are introduced here for this oversampling question.

3.1 Randomized SVD Setting

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $A \in \mathbb{F}^{m \times n}$ be a finite nonzero matrix with $m, n \in \mathbb{N}_{\geq 1}$ and $d = \min(m, n) \geq 2$. Write its singular values as $\sigma_1(A) \geq \sigma_2(A) \geq \dots \geq \sigma_d(A) \geq 0$, and assume $\sigma_1(A) > 0$. For a target rank $k \in \{1, \dots, d-1\}$ and an oversampling value $p \in \{0, \dots, d-k\}$, the sketch size is $\ell = k + p$. Unless stated otherwise, $\Omega \in \mathbb{F}^{n \times \ell}$ has independent standard Gaussian entries, interpreted as real or complex according to $\mathbb{F}$. For a power-iteration count $q \in \mathbb{N}_0$, the sampled matrix is

$$Y_q = (AA^*)^q A\Omega, \quad Q_{p,q} = \text{orth}(Y_q), \quad (1)$$

where $Q_{p,q}$ has orthonormal columns spanning the numerical or exact range of Y_q . The approximation used in the analysis is

$$\tilde{A}_{k,p,q} = Q_{p,q} [Q_{p,q}^* A]_k, \quad (2)$$

where $[B]_k$ denotes a best rank- k truncation of B in the spectral or Frobenius norm. The construction in (1) and (2) follows the HMT randomized range-finder setting and its common postprocessing variant

[12, 24, 41]. The notation A_k denotes a best rank- k SVD truncation of A , so $\|A - A_k\|_2 = \sigma_{k+1}(A)$ and $\|A - A_k\|_F = (\sum_{i=k+1}^d \sigma_i(A)^2)^{1/2}$.

The performance metric is relative excess error above the optimal rank- k comparator. For $\nu \in \{2, F\}$, define

$$\Delta_\nu(A, k, p, q; \Omega) = \frac{\|A - \tilde{A}_{k,p,q}\|_\nu - \|A - A_k\|_\nu}{\|A - A_k\|_\nu}, \quad (3)$$

whenever $\|A - A_k\|_\nu > 0$. If $\text{rank}(A) \leq k$, the denominator is zero and the relative-excess claim is outside scope. This exclusion is not cosmetic: exact-rank matrices can have stable rank and randomized sketches, but the relative accuracy question changes into an absolute-error or exact-recovery question. The method and experiments therefore record such cases as boundary conditions rather than as successes.

3.2 Intrinsic Dimension and Tail Quantities

The standard stable rank of a nonzero matrix is

$$r_{\text{eff}}(A) = \frac{\|A\|_F^2}{\|A\|_2^2} = \frac{\sum_{i=1}^d \sigma_i(A)^2}{\sigma_1(A)^2}. \quad (4)$$

This definition is inherited from stable-rank and intrinsic-dimension literature [4, 14, 40]. In this work we define a filtered stable rank aligned with the power sample in (1):

$$r_q(A) = \frac{\sum_{i=1}^d \sigma_i(A)^{4q+2}}{\sigma_1(A)^{4q+2}}, \quad q \in \mathbb{N}_0. \quad (5)$$

For $q = 0$, (5) equals (4). For $q > 0$, it is the stable rank of the singular-value profile after the power filter $\sigma_i \mapsto \sigma_i^{2q+1}$. The exponent $4q + 2$ comes from squaring the filtered singular values. We also define a target-local normalized tail

$$T_k(A) = \frac{\sum_{i=k+1}^d \sigma_i(A)^2}{\sigma_{k+1}(A)^2}, \quad (6)$$

whenever $\sigma_{k+1}(A) > 0$. This object is not a new error metric; it is a dimensionless way to compare the Frobenius tail to the spectral denominator at the target boundary. The distinction between $r_q(A)$ and $T_k(A)$ is central. Stable rank measures the global energy relative to the top singular value, while randomized rank- k error is controlled by the part of the spectrum beyond k .

For controlled synthetic spectra we use two finite families. A power-law spectrum $P(\alpha, d)$ has $\sigma_i = Ci^{-\alpha}$ for $C > 0$, $\alpha > 0$, and $i = 1, \dots, d$. An exponential spectrum $E(\beta, d)$ has $\sigma_i = C \exp[-\beta(i-1)]$ for $C > 0$, $\beta > 0$, and $i = 1, \dots, d$. These definitions are manuscript-specific test families, not claims about all matrices. They isolate how $r_q(A)$, $T_k(A)$, and oversampling interact when singular vectors do not introduce additional leverage effects. Real matrices are interpreted only after preprocessing has defined the actual numerical matrix A .

4 Formal Analysis

The formal analysis has two roles. First, it identifies a deterministic obstruction to a stable-rank-only interpretation, so the negative result does not depend on Monte Carlo noise or implementation choices. Second, it states a conditional positive result: if randomized SVD residual exceedance decays with oversampling per filtered rank in a declared regime, then an a-priori threshold follows algebraically. The conditional premise is motivated by concentration and intrinsic-dimension theory [2, 17, 18, 40], but it is not presented as an established theorem for all HMT residuals.

4.1 Finite Spectral Formulas

The first formal point is that filtered rank and target-local tail behave differently even in simple spectra. For $P(\alpha, d)$, substituting $\sigma_i = Ci^{-\alpha}$ into (5) and (6) cancels C and gives

$$r_q(P) = \sum_{i=1}^d i^{-(4q+2)\alpha}, \quad T_k(P) = (k+1)^{2\alpha} \sum_{i=k+1}^d i^{-2\alpha}. \quad (7)$$

For $E(\beta, d)$, the corresponding geometric sums are

$$r_q(E) = \frac{1 - \exp[-(4q+2)\beta d]}{1 - \exp[-(4q+2)\beta]}, \quad T_k(E) = \frac{1 - \exp[-2\beta(d-k)]}{1 - \exp[-2\beta]}. \quad (8)$$

Equation (7) and (8) are used in the methodology to generate deterministic tail diagnostics before any randomized SVD run. They also explain why power iteration can change the apparent intrinsic dimension: increasing q raises the exponent in r_q , often pushing the filtered profile toward rank one when the leading singular value is separated.

Lemma 1 (Tail-local obstruction for finite power laws). *Let A have singular values $\sigma_i = Ci^{-\alpha}$ for $C > 0$, $i = 1, \dots, d$, and $\alpha > 1/2$. If $k \in \{1, \dots, d-1\}$ and $d+1 \geq 2(k+1)$, then*

$$r_{\text{eff}}(A) \leq 1 + \frac{1}{2\alpha - 1}, \quad T_k(A) \geq \frac{k+1}{2\alpha - 1} (1 - 2^{1-2\alpha}). \quad (9)$$

Proof. For $\alpha > 1/2$, the function $x \mapsto x^{-2\alpha}$ is positive and decreasing on $(0, \infty)$. The stable rank is $\sum_{i=1}^d i^{-2\alpha}$, which is at most $1 + \int_1^d x^{-2\alpha} dx$ and therefore at most $1 + 1/(2\alpha - 1)$. For the lower bound, (6) gives $T_k(A) = (k+1)^{2\alpha} \sum_{i=k+1}^d i^{-2\alpha}$. Since the summand is decreasing and $d+1 \geq 2(k+1)$, the sum from $i = k+1$ to d is at least $\int_{k+1}^{2(k+1)} x^{-2\alpha} dx$. Evaluating the integral gives $[(2(k+1))^{1-2\alpha} - (k+1)^{1-2\alpha}]/(1-2\alpha)$, which equals $(k+1)^{1-2\alpha}(1 - 2^{1-2\alpha})/(2\alpha - 1)$. Multiplication by $(k+1)^{2\alpha}$ yields the stated bound. $\square$

Lemma 1 is the main negative formal result. It shows that $r_{\text{eff}}(A)$ can be bounded by a constant depending on α while $T_k(A)$ grows at least linearly in k over a valid finite range. This does not disprove every stable-rank-sensitive rule, but it rules out a broad interpretation in which the original global stable rank alone controls all target ranks under rapid decay. The role of $T_k(A)$ is therefore not optional in this analysis; it identifies a target-local obstruction that survives global intrinsic-dimension control. The two inequalities in (9) are the concrete obstruction used in the later failure atlas: the first bounds the global intrinsic dimension, while the second leaves a target-dependent tail term that an oversampling rule must still confront.

4.2 Chance Thresholds and Conditional Rule

Given $\epsilon \in (0, 1)$ and $\delta \in (0, 1)$, define the minimal feasible oversampling for metric $\nu \in \{2, F\}$ as

$$p_\nu^*(\epsilon, \delta; A, k, q) = \min \left\{ p \in \{0, \dots, d-k\} : \Pr_\Omega[\Delta_\nu(A, k, p, q; \Omega) \leq \epsilon] \geq 1 - \delta \right\}, \quad (10)$$

with $p_\nu^* = +\infty$ if no feasible p satisfies the probability condition. We derive an upper bound under a named premise rather than importing it from the literature.

Theorem 1 (Conditional filtered-rank oversampling bound). *Fix a valid regime $\mathcal{R}$ of finite nonzero matrices, target ranks, power counts, Gaussian sketches, and a metric $\nu \in \{2, F\}$ such that $\|A - A_k\|_\nu > 0$. Suppose there exist constants $C_\nu \geq 1$, $c_\nu > 0$, and $\lambda_\nu \geq 0$ such that every $(A, k, q) \in \mathcal{R}$ satisfies*

$$\Pr_\Omega[\Delta_\nu(A, k, p, q; \Omega) > \epsilon] \leq C_\nu \epsilon^{-\lambda_\nu} \exp\left(-c_\nu \frac{p}{r_q(A)}\right) \quad (11)$$

for all $\epsilon \in (0, 1)$ and feasible p . If

$$p \geq \frac{r_q(A)}{c_\nu} [\log C_\nu + \lambda_\nu \log(1/\epsilon) + \log(1/\delta)] \quad (12)$$

and $p \leq d-k$, then $\Pr_\Omega[\Delta_\nu(A, k, p, q; \Omega) \leq \epsilon] \geq 1 - \delta$. Consequently, the ceiling of the right side of (12), when feasible, is an upper bound on (10).

Proof. Because $C_\nu \geq 1$, $\lambda_\nu \geq 0$, and $\epsilon, \delta \in (0, 1)$, the bracketed term in (12) is nonnegative. Substituting the lower bound on p into the right side of (11) gives

$$C_\nu \epsilon^{-\lambda_\nu} \exp(-\log C_\nu - \lambda_\nu \log(1/\epsilon) - \log(1/\delta)) = \delta.$$

Thus $\Pr_\Omega[\Delta_\nu > \epsilon] \leq \delta$, which is equivalent to $\Pr_\Omega[\Delta_\nu \leq \epsilon] \geq 1 - \delta$. If the ceiling of (12) is feasible, it belongs to the set minimized in (10), so the minimal feasible value is no larger. $\square$

Theorem 1 is deliberately conditional. The premise in (11) resembles intrinsic-dimension concentration but is not an established HMT theorem. It is a testable bridge: if residual exceedance behaves like an exponential function of $p/r_q(A)$ in a declared regime, then (12) yields a rule. The experiments below estimate whether such behavior is plausible on finite grids and record where it fails.

Definition 1 (Two-parameter calibrated rule). *For $\theta = (a, b)$ with $a \geq 0$ and $b \geq 1$, define*

$$\hat{p}_\theta(\epsilon, \delta; A, k, q) = \text{clip}_{\{0, \dots, d-k\}} \left[a r_q(A) \log \frac{b}{\epsilon \delta} \right]. \quad (13)$$

The expression in (13) is monotone in $r_q(A)$ and becomes no larger when ϵ is relaxed. The ceiling and clipping are order-preserving maps, so the monotonicity survives integer feasibility. A finite calibration grid can therefore search over a and b while preserving a simple a-priori form. If the feasible subset is empty, the finite-grid optimization has no constrained solution; a slack-penalized fallback must be reported as such.

5 Validation Protocol

The validation protocol follows the formal quantities above and separates four evidence sources. The SymPy theorem audit checks the finite-sum identities, the power-law antiderivative, the conditional implication in Theorem 1, monotonicity of the continuous proxy in (13), and finite-grid existence logic. Synthetic controlled spectra execute Gaussian HMT randomized SVD on diagonal power-law and exponential matrices with six fixed seeds. The Calibrated rule holdout grid fits the two-parameter rule on finite threshold rows and evaluates held-out groups without refitting. The Boundary and failure atlas records regimes where a claim should be weakened, including slow power laws, high T_k/r_q , unbracketed thresholds, exact-rank denominators, and preprocessing sensitivity.

Real-transfer validation uses three public matrix sources. SuiteSparse SNAP/ca-GrQc is represented as a binary symmetric adjacency matrix with the diagonal set to zero [5, 16, 19]. MovieLens 100K is represented as a sparse rectangular rating matrix with observed ratings centered by the observed global mean and missing entries left as structural zeros [10, 13]. The Fashion-MNIST train slice uses the first 2,000 training images, scales pixels to $[0, 1]$, and column-centers the resulting dense 2000×784 matrix [48, 50]. These preprocessing choices define A , and therefore define $r_{\text{eff}}(A)$, $r_q(A)$, A_k , and all relative errors.

Algorithm 1 summarizes the empirical contract. The synthetic executed grid contains 1,728 randomized SVD rows over two dimensions, two target ranks, three power counts, six seeds, and two spectrum families. The real executed grid contains 648 rows over three matrices, two target ranks, three power counts, six seeds, and six oversampling values. The threshold table contains 336 chance-threshold rows. Uncertainty is summarized by fixed-seed means, standard deviations, empirical hit rates, binned residual spread, and train/held-out coverage. The experiments do not claim asymptotic constants; they test whether the finite behavior supports the rule form suggested by (12).

Table 1 is the compact evidence ledger for the results. It defines the scale of each executed component before the figures are interpreted. The presence of unbracketed threshold rows and an infeasible calibration grid is part of the result, because both directly limit the strength of any a-priori oversampling rule.

6 Results

The results are organized from deterministic evidence to randomized evidence. The first subsection reports the tail-local obstruction, which is the strongest supported claim. The second subsection tests whether filtered rank improves empirical collapse on synthetic spectra. The third subsection evaluates whether a

Algorithm 1 Finite validation of filtered-rank oversampling rules. The procedure first computes deterministic spectrum diagnostics, then runs randomized SVD sweeps, then fits and tests an a-priori rule. Each stage records boundary cases before interpreting a success rate.

- 1: Choose matrix family or real matrix, preprocessing operator, target ranks $k \in \{10, 50\}$, power counts $q \in \{0, 1, 2\}$, metrics $\nu \in \{2, F\}$, and tolerances $\epsilon \in \{10^{-2}, 10^{-3}\}$.
- 2: Compute $r_{\text{eff}}(A)$, $r_q(A)$, $T_k(A)$ when defined, and the positive denominator condition for (3).
- 3: For $p \in \{0, 5, 10, 20, 40, 64\}$ on real matrices and the finer synthetic sweep, run Gaussian randomized SVD over fixed seeds and estimate Δ_ν .
- 4: Estimate p_ν^* by the empirical chance condition in (10); mark rows unbracketed when the finite p grid never reaches the requested hit rate.
- 5: Fit $\hat{p}_\theta$ on training rows over a finite grid of a and b ; if no feasible cell exists, report the slack-penalized minimizer.
- 6: Compare transfer rows to fixed- p and k -proportional baselines only after the calibrated rule matches or improves the comparator error.

Evidence component	Executed object	Count or outcome
Symbolic theorem audit	Algebraic and finite-grid checks	7 of 7 passed
Synthetic controlled spectra	Gaussian randomized SVD rows	1,728
Real-transfer matrices	Gaussian randomized SVD rows	648
Chance thresholds	Empirical p^* rows	204 bracketed, 132 unbracketed
Calibration grid	Finite $\theta = (a, b)$ cells	0 feasible, slack fallback
Matched transfer	Accuracy-gated savings rows	54 matched, 0 positive savings

Table 1: Executed validation components used in the manuscript results. The table reports what was actually computed rather than desired follow-up experiments, and it distinguishes bracketed thresholds from cases where the finite oversampling grid did not reach the requested probability.

calibrated a-priori rule transfers to held-out and real matrices. This ordering matters because a positive real-transfer result would not override the formal obstruction, and a negative transfer result does not invalidate the conditional theorem under its stated premise. It also keeps supported, mixed, and negative evidence visually separate.

6.1 Stable Rank Does Not Remove Target-Local Dependence

The formal obstruction appears before any Monte Carlo variability enters the problem. Lemma 1 proves that, for finite power-law spectra with $\alpha > 1/2$, global stable rank can remain bounded while $T_k(A)$ grows with k . The symbolic theorem audit verified the finite power-law cancellation, the exponential geometric sums, the power-law antiderivative, and the conditional threshold algebra. This supports the finite-dimensional claim that stable rank alone cannot remove target-rank and tail-local dependence.

Figure 1 visualizes this separation. The deterministic grid is not a randomized SVD benchmark; it is a spectral diagnostic that explains why an error rule based only on $r_{\text{eff}}(A)$ would be under-specified. The figure also motivates $r_q(A)$: for spectra where power iteration sharply contracts the filtered profile, oversampling per original stable rank is the wrong horizontal coordinate. The formal and deterministic evidence therefore supports a corrected hypothesis, not the original stable-rank-only version.

6.2 Collapse Improves Only Conditionally

The Synthetic controlled spectra evidence asks whether randomized SVD errors collapse when plotted against $p/r_q(A)$ instead of $p/r_{\text{eff}}(A)$. The executed grid uses diagonal spectra, Gaussian sketches, six seeds, and the relative excess metric in (3). This setting intentionally removes singular-vector coherence so that spectral effects can be isolated. If a simple filtered-rank law were dominant, residual spread after normalization would drop consistently across dimensions, target ranks, power counts, and spectrum families.

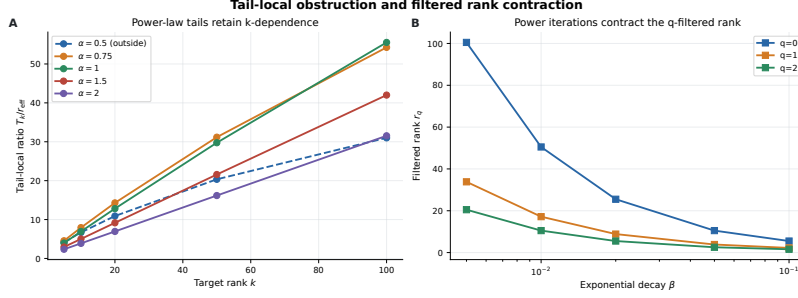


Figure 1: Deterministic tail diagnostics for the formal obstruction. The left panel shows that target-local tail quantities can grow even when the global stable rank is bounded, while the right panel shows how power filtering contracts the effective rank for exponential spectra; the dashed boundary marks the slow-tail regime outside the stated power-law theorem.

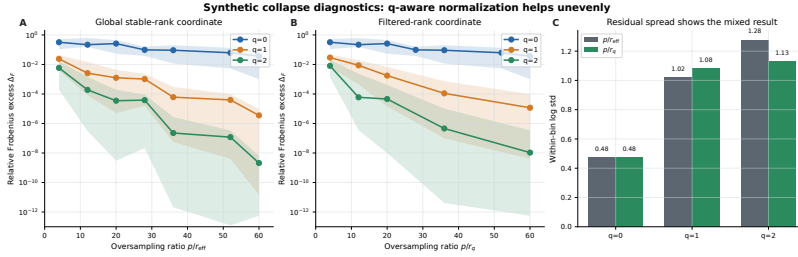


Figure 2: Collapse diagnostics for synthetic controlled spectra. The panels compare oversampling normalized by global stable rank and by filtered rank, showing that the filtered coordinate helps most clearly at larger power counts but does not eliminate family, target-rank, or tolerance effects across the grid.

The observed collapse is mixed. Across the synthetic summaries, the mean relative spread improvement of $p/r_q(A)$ over $p/r_{\text{eff}}(A)$ is 0.019, which is positive but small. Conditional tail fits are more encouraging: five of six fitted exceedance curves have strengthened support, and all six slopes of log exceedance against $p/r_q(A)$ are negative. The slopes range from about -0.0011 to -0.1661 , with R^2 values from 0.43 to 0.99. The weak $q = 2$, $\epsilon = 10^{-2}$ fit is a warning that low exceedance rates and finite-grid binning can make a clean exponential tail hard to identify. Figure 2 therefore supports the conditional premise behind Theorem 1 only as a scoped empirical approximation.

6.3 Calibration and Transfer Are Conservative

The Calibrated rule holdout grid evaluates the two-parameter rule in (13). The finite grid over $\theta = (a, b)$ had no feasible cell satisfying all groupwise empirical coverage constraints. The reported rule therefore uses the predeclared slack-penalized fallback with $a = 1.2$ and $b = 2$. Its training coverage is 80%, and held-out coverage is also 80%. These values are useful evidence because they show that the rule shape is not vacuous, but they are not strong enough to validate a universal a-priori guarantee.

Figure 3 shows the key calibration facts. Negative fitted slopes are consistent with decreasing exceedance as $p/r_q(A)$ grows, but the finite rule does not fully cover strict tolerance rows. The p^* grid reinforces the point: out of 336 chance-threshold rows, 204 are bracketed and 132 remain unbracketed within the finite oversampling sweep. The unbracketed cases are concentrated at $q = 0$ and at strict tolerance, where 106 of 112 $q = 0$ threshold rows are unbracketed and 76 of 168 $\epsilon = 10^{-3}$ rows are unbracketed. These are not implementation failures; they are evidence that a bounded sweep and a two-parameter rule cannot close every hard finite case.

Real-transfer comparisons use the matched-accuracy gate before measuring savings. This means a rule row is counted only if its mean Frobenius relative excess is no worse than the comparator row. All 54 real-transfer rows satisfy this matched-accuracy gate, but none has positive sketch-column savings. The average

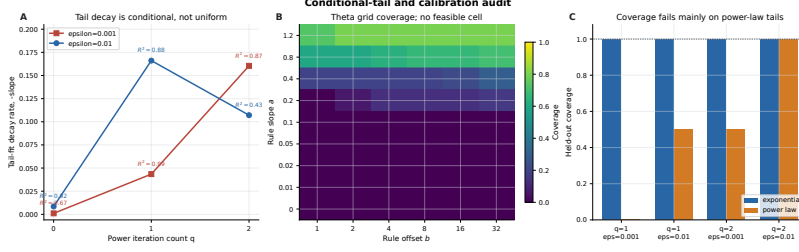


Figure 3: Conditional rule audit for the calibrated oversampling formula. The panels report fitted exceedance slopes, finite-grid calibration coverage, and held-out failures, making clear that the rule is conditional and that the tested two-parameter grid did not contain a fully feasible solution.

Matrix	Shape	Preprocessing	r_{eff} estimate	Mean savings
SuiteSparse SNAP/ca-GrQc	5242×5242	Binary symmetric adjacency	13.92	-26.3
MovieLens 100K	943×1682	Centered observed ratings	15.83	-26.3
Fashion-MNIST train slice	2000×784	Scaled and column-centered pixels	3.40	-15.0

Table 2: Real-transfer matrices and matched-accuracy savings summary. The mean savings column is measured after the calibrated rule has matched comparator accuracy, so negative values mean the rule used more sketch columns rather than fewer.

savings are negative: -15.0 columns for the Fashion-MNIST train slice and about -26.3 columns for each of SuiteSparse SNAP/ca-GrQc and MovieLens 100K. Table 2 reports the source matrices and preprocessing choices, and the appendix gives the savings distribution. The evidence supports a conservative statement: the calibrated rule can match accuracy in these rows, but it does so by spending as many or more sketch columns than the fixed- p comparators.

7 Discussion

The main lesson is that effective rank is most useful when the word “effective” is tied to the sampled spectrum and to the target residual. The original stable rank $r_{\text{eff}}(A)$ is invariant under singular-vector rotations and easy to compute from norms, which makes it attractive for a one-line rule. But randomized SVD with target rank k does not ask how much total energy lies under $\sigma_1(A)$; it asks how accurately a random subspace captures the part of A that matters after the first k singular directions are removed. The target-local tail $T_k(A)$ and the filtered rank $r_q(A)$ expose this mismatch. They do not produce a universal theorem by themselves, but they make the validity regime explicit.

The empirical evidence fits this interpretation. Power iteration can make $r_q(A)$ much smaller than $r_{\text{eff}}(A)$, especially for exponential spectra and for real matrices with dominant leading components. That contraction explains why $p/r_q(A)$ can be a better coordinate than $p/r_{\text{eff}}(A)$. However, the small average collapse improvement and the unbracketed threshold rows show that filtered rank is still a summary, not a full sufficient statistic. Residual behavior also depends on the target rank, the denominator norm, the spectrum family, and the strictness of the tolerance. This is consistent with modern randomized numerical linear algebra, where sketch size interacts with subspace geometry, tail energy, pass count, and the reconstruction metric [1, 42, 43, 46].

The calibrated rule should be read as an interpretable diagnostic baseline. Its monotonicity is valuable because it behaves sensibly as filtered rank increases or tolerance tightens. Its failure to find a feasible two-parameter grid is also valuable because it prevents an overconfident deployment claim. In the tested real-transfer rows, the rule matches accuracy but spends extra columns. That outcome may still be acceptable in settings where conservative accuracy is preferred over compute savings, but it does not meet the stronger goal of reducing sketch columns at matched accuracy. A practical implementation should therefore expose the rule as a starting value or safety-biased planner, not as a replacement for residual estimation or adaptive

rank discovery.

The results also clarify how future theory should be shaped. A useful theorem would not simply replace ambient dimension by $r_{\text{eff}}(A)$ in an HMT-style bound. It would state a regime in which an exceedance probability has filtered-rank scaling, and it would include target-local tail or gap terms that Lemma 1 shows cannot be ignored. Matrix concentration and covariance effective-rank theory provide a template for such a result [2, 17, 18, 40], but the randomized SVD residual contains a projection, a postprocessed rank- k truncation, and a denominator tied to A_k . Those ingredients are precisely where a direct transfer of stable-rank concentration can fail.

8 Limitations and Future Work

The evidence narrows the original conjecture rather than closing it in the affirmative. This section states the limits that affect interpretation of the formal result, the finite experiments, and the proposed rule. The limitations are not all equal: some are mathematical exclusions required for the definitions to make sense, while others are empirical coverage gaps that can be reduced by larger validation. The future-work directions are therefore written as concrete extensions of the present evidence rather than as speculative applications. Each item affects the strength of the oversampling recommendation.

8.1 Limitations

The study is finite and scoped. Synthetic controlled spectra use diagonal matrices, so they isolate singular-value effects but do not test leverage, coherence, or adversarial singular-vector geometry. Gaussian sketches are used throughout the formal and empirical analysis; sparse embeddings, structured transforms, and sampling sketches require separate validation. The power-law obstruction is stated for $\alpha > 1/2$ and $d + 1 \geq 2(k + 1)$, and it explicitly excludes the slow-tail boundary $\alpha \leq 1/2$. Relative excess error is undefined when $\text{rank}(A) \leq k$, so exact-rank and zero-denominator cases are not part of the main guarantee.

The empirical grid is also bounded. The largest synthetic dimensions in the executed randomized SVD grid are 500 and 1000, and the real matrices cover one graph, one recommender matrix, and one dense image slice. The finite oversampling grid leaves 132 threshold rows unbracketed, which means the experiments cannot estimate the true p^* for those rows. The calibrated two-parameter grid has no fully feasible cell, so reported calibration uses a slack-penalized fallback. These limitations reduce the strength of the rule claim but do not weaken the formal obstruction: the negative result about stable rank and target-local tail is deterministic under its stated assumptions.

8.2 Future Work

The next step is a larger validation that separates rule shape from finite-grid truncation. Larger dimensions, wider p sweeps, and more spectrum families would reveal whether unbracketed rows are mainly a strict-tolerance artifact or a structural failure of the two-parameter rule. A second direction is to include coherence-controlled matrices with identical singular values, because stable rank and filtered rank are blind to singular-vector geometry while sampling and sparse sketches are not. A third direction is to test adaptive comparators such as fixed-precision randomized algorithms, rank-revealing randomized SVD, and numerical-rank estimators under matched pass and sketch-column budgets [15, 21, 25, 49]. Finally, theory should target a conditional statement that includes $r_q(A)$, $T_k(A)$, spectral gaps, and the metric denominator, rather than searching for a stable-rank-only law that the finite obstruction already rules out.

9 Conclusion

This paper tested whether stable rank governs the oversampling returns of randomized SVD strongly enough to yield an a-priori rule. The supported answer is narrower than the motivating conjecture. Stable rank is an important intrinsic-dimension signal, but global stable rank alone cannot remove target-local dependence. A filtered rank $r_q(A)$ better matches the spectrum sampled by power iteration, and a conditional tail premise yields a clean threshold form. In executed validation, this form provides useful diagnostics but not a universal

rule: collapse improves only weakly on average, calibration is slack-penalized, strict threshold rows are often unbracketed, and real-transfer savings are nonpositive after matching accuracy. The main practical recommendation is therefore conservative. Use filtered rank and tail diagnostics to inform oversampling, but keep target rank, tolerance, power iteration, preprocessing, and residual validation in the decision loop.

References

- [1] Elvar K. Bjarkason. Pass-efficient randomized algorithms for low-rank matrix approximation using any number of views. <https://doi.org/10.1137/18M118966X>, 2019. SIAM Journal on Scientific Computing.
- [2] Florentina Bunea and Luo Xiao. The sample covariance matrix and reduced effective rank population matrices, with applications to fpca. <https://doi.org/10.3150/14-BEJ602>, 2015. Bernoulli.
- [3] Tyler Chen, Ethan N. Epperly, Raphael A. Meyer, Christopher Musco, and Akash Rao. Does block size matter in randomized block krylov low-rank approximation? <https://doi.org/10.48550/arXiv.2508.06486>, 2025. SODA 2026 preprint.
- [4] Michael B. Cohen, Jelani Nelson, and David P. Woodruff. Optimal approximate matrix product in terms of stable rank. <https://doi.org/10.4230/LIPIcs.ICALP.2016.11>, 2016. ICALP 2016.
- [5] Timothy A. Davis and Yifan Hu. Suitesparse matrix collection. <https://sparse.tamu.edu/about>, 2011. ACM Transactions on Mathematical Software and SuiteSparse Matrix Collection website.
- [6] Michal Dereziński and Michael W. Mahoney. Recent and upcoming developments in randomized numerical linear algebra for machine learning. <https://doi.org/10.48550/arXiv.2406.11151>, 2024. KDD 2024 / arXiv version.
- [7] Petros Drineas and Michael W. Mahoney. Randnla: Randomized numerical linear algebra. <https://doi.org/10.1145/2842602>, 2016. Communications of the ACM.
- [8] N. Benjamin Erichson, Sergey Voronin, Steven L. Brunton, and J. Nathan Kutz. Randomized matrix decompositions using r. <https://doi.org/10.18637/jss.v089.i11>, 2019. Journal of Statistical Software.
- [9] Alan Frieze, Ravi Kannan, and Santosh Vempala. Fast monte-carlo algorithms for finding low-rank approximations. <https://doi.org/10.1145/1039488.1039494>, 2004. Journal of the ACM.
- [10] GroupLens Research. Movielens 100k dataset. <https://grouplens.org/datasets/movielens/100k/>, 1998. GroupLens Research, University of Minnesota.
- [11] Ming Gu. Subspace iteration randomization and singular value problems. <https://doi.org/10.48550/arXiv.1408.2208>, 2015. SIAM Journal on Scientific Computing / arXiv version.
- [12] Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. <https://doi.org/10.1137/090771806>, 2011. SIAM Review.
- [13] F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. <https://doi.org/10.1145/2827872>, 2015. ACM Transactions on Interactive Intelligent Systems.
- [14] Ilse C. F. Ipsen and Arvind K. Saibaba. Stable rank and intrinsic dimension of real and complex matrices. <https://doi.org/10.1137/24M1681537>, 2025. SIAM Journal on Matrix Analysis and Applications.
- [15] Hao Ji, Wenjian Yu, and Yaohang Li. A rank revealing randomized singular value decomposition (r3svd) algorithm for low-rank matrix approximations. <https://doi.org/10.48550/arXiv.1605.08134>, 2016. arXiv preprint.
- [16] Scott P. Kolodziej, Mohsen Azaveh, Matthew Bullock, Jarrett David, Timothy A. Davis, Matthew Henderson, Yifan Hu, and Read Sandstrom. The suitesparse matrix collection website interface. <https://doi.org/10.21105/joss.01244>, 2019. Journal of Open Source Software.
- [17] Vladimir Koltchinskii. Estimation of trace functionals and spectral measures of covariance operators. <https://doi.org/10.48550/arXiv.2402.11321>, 2024. arXiv preprint.

- [18] Vladimir Koltchinskii and Karim Lounici. Concentration inequalities and moment bounds for sample covariance operators. <https://doi.org/10.3150/15-BEJ730>, 2017. Bernoulli.
- [19] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Snap/ca-grqc collaboration network of arxiv general relativity. <https://sparse.tamu.edu/SNAP/ca-GrQc>, 2007. SuiteSparse Matrix Collection and ACM Transactions on Knowledge Discovery from Data.
- [20] Edo Liberty, Franco Woolfe, Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. Randomized algorithms for the low-rank approximation of matrices. <https://doi.org/10.1073/pnas.0709640104>, 2007. Proceedings of the National Academy of Sciences.
- [21] Qiaohua Liu and Yuejuan Yu. Efficient adaptive randomized algorithms for fixed-threshold low-rank matrix approximation. <https://doi.org/10.48550/arXiv.2508.07553>, 2025. arXiv preprint.
- [22] Michael W. Mahoney. Randomized algorithms for matrices and data. <https://doi.org/10.1561/22000000035>, 2011. Foundations and Trends in Machine Learning.
- [23] Per-Gunnar Martinsson. Randomized methods for matrix computations. <https://doi.org/10.48550/arXiv.1607.01649>, 2019. The Mathematics of Data, IAS/Park City Mathematics Series.
- [24] Per-Gunnar Martinsson and Joel A. Tropp. Randomized numerical linear algebra: Foundations and algorithms. <https://doi.org/10.1017/S0962492920000021>, 2020. Acta Numerica.
- [25] Maike Meier and Yuji Nakatsukasa. Fast randomized numerical rank estimation for numerically low-rank matrices. <https://doi.org/10.1016/j.laa.2024.01.001>, 2024. Linear Algebra and its Applications.
- [26] Riley Murray, James Demmel, Michael W. Mahoney, N. Benjamin Erichson, Maksim Melnichenko, Osman Asif Malik, Laura Grigori, Piotr Luszczek, Michal Dereziński, Miles E. Lopes, Tianyu Liang, Hengrui Luo, and Jack Dongarra. Randomized numerical linear algebra: A perspective on the field with an eye to software. <https://doi.org/10.48550/arXiv.2302.11474>, 2023. UCB/EECS Technical Report.
- [27] Cameron Musco and Christopher Musco. Randomized block krylov methods for stronger and faster approximate singular value decomposition. <https://arxiv.org/abs/1504.05477>, 2015. NeurIPS 2015.
- [28] Yuji Nakatsukasa and Taejun Park. Randomized low-rank approximation for symmetric indefinite matrices. <https://doi.org/10.1137/22M1538648>, 2023. SIAM Journal on Matrix Analysis and Applications.
- [29] Davide Palitta and Sascha Portaro. Row-aware randomized svd with applications. <https://doi.org/10.48550/arXiv.2408.04503>, 2024. arXiv preprint.
- [30] David Persson and Daniel Kressner. Randomized low-rank approximation of monotone matrix functions. <https://doi.org/10.1137/22M1523923>, 2023. SIAM Journal on Matrix Analysis and Applications.
- [31] David Persson, Nicolas Boullé, and Daniel Kressner. Randomized nyström approximation of non-negative self-adjoint operators. <https://doi.org/10.48550/arXiv.2404.00960>, 2024. arXiv preprint / SIAM version.
- [32] Farhad Pourkamali-Anaraki. Low-rank compression of pretrained models via randomized subspace iteration. <https://doi.org/10.48550/arXiv.2604.02659>, 2026. arXiv preprint.
- [33] Vladimir Rokhlin, Arthur Szlam, and Mark Tygert. A randomized algorithm for principal component analysis. <https://doi.org/10.1137/080736417>, 2009. SIAM Journal on Matrix Analysis and Applications.
- [34] Elad Romanov. On the noise sensitivity of the randomized svd. <https://doi.org/10.1109/TIT.2024.3450412>, 2024. IEEE Transactions on Information Theory / arXiv version.

- [35] Mark Rudelson and Roman Vershynin. Sampling from large matrices: An approach through geometric functional analysis. <https://doi.org/10.1145/1255443.1255449>, 2007. Journal of the ACM.
- [36] Arvind K. Saibaba. Analysis of randomized subspace iteration: Canonical angles and unitarily invariant norms. <https://doi.org/10.1137/18M1179432>, 2019. SIAM Journal on Matrix Analysis and Applications.
- [37] Arvind K. Saibaba and Agnieszka Miedlar. Randomized low-rank approximations beyond gaussian random matrices. <https://doi.org/10.1137/23M1593255>, 2025. SIAM Journal on Mathematics of Data Science.
- [38] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. <https://doi.org/10.1109/FOCS.2006.37>, 2006. FOCS 2006.
- [39] scikit-learn developers. sklearn.utils.extmath.randomized_svd. https://scikit-learn.org/stable/modules/generated/sklearn.utils.extmath.randomized_svd.html, 2026. scikit-learn API documentation.
- [40] Joel A. Tropp. An introduction to matrix concentration inequalities. <https://doi.org/10.1561/22000000048>, 2015. Foundations and Trends in Machine Learning.
- [41] Joel A. Tropp and Robert J. Webber. Randomized algorithms for low-rank matrix approximation: Design, analysis, and applications. <https://doi.org/10.48550/arXiv.2306.12418>, 2023. arXiv preprint.
- [42] Joel A. Tropp, Alp Yurtsever, Madeleine Udell, and Volkan Cevher. Practical sketching algorithms for low-rank matrix approximation. <https://doi.org/10.1137/17M111590>, 2017. SIAM Journal on Matrix Analysis and Applications.
- [43] Joel A. Tropp, Alp Yurtsever, Madeleine Udell, and Volkan Cevher. Streaming low-rank matrix approximation with an application to scientific simulation. <https://doi.org/10.1137/18M1201068>, 2019. SIAM Journal on Scientific Computing.
- [44] Madeleine Udell and Alex Townsend. Why are big data matrices approximately low rank? <https://doi.org/10.1137/18M1183480>, 2019. SIAM Journal on Mathematics of Data Science.
- [45] Sergey Voronin and Per-Gunnar Martinsson. Rsvdpack: An implementation of randomized algorithms for computing the singular value, interpolative, and cur decompositions of matrices on multi-core and gpu architectures. <https://doi.org/10.48550/arXiv.1502.05366>, 2015. arXiv preprint.
- [46] David P. Woodruff. Sketching as a tool for numerical linear algebra. <https://doi.org/10.1561/04000000060>, 2014. Foundations and Trends in Theoretical Computer Science.
- [47] Franco Woolfe, Edo Liberty, Vladimir Rokhlin, and Mark Tygert. A fast randomized algorithm for the approximation of matrices. <https://doi.org/10.1016/j.acha.2007.12.002>, 2008. Applied and Computational Harmonic Analysis.
- [48] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. <https://arxiv.org/abs/1708.07747>, 2017. arXiv:1708.07747.
- [49] Wenjian Yu, Yu Gu, and Yaohang Li. Efficient randomized algorithms for the fixed-precision low-rank matrix approximation. <https://doi.org/10.1137/17M1141977>, 2018. SIAM Journal on Matrix Analysis and Applications.
- [50] Zalando Research. Fashion-mnist github repository and dataset files. <https://github.com/zalando-research/fashion-mnist>, 2017. GitHub repository.

A Extended Derivation Details

The derivations in the main text use finite-dimensional objects throughout. No statement is made for compact operators, streaming limits, or asymptotic random matrix models unless those objects can be reduced to the finite singular-value sequence already defined. The field $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, and all norms are the spectral norm or Frobenius norm. The stable rank in (4) is inherited from prior intrinsic-dimension work, while $r_q(A)$, $T_k(A)$, Δ_ν , p_ν^* , and $\hat{p}_\theta$ are defined in this work for the randomized SVD question.

For power-law spectra, the cancellation of C in (7) is important because it shows that oversampling diagnostics are scale-free under relative error. If every singular value is multiplied by the same positive constant, $r_q(A)$, $T_k(A)$, and Δ_ν are unchanged. The obstruction in Lemma 1 uses the standard integral comparison for a decreasing positive function. The condition $\alpha > 1/2$ is not a convenience; it is the regime in which the infinite power-law energy is summable and stable rank can be bounded independently of d . At $\alpha = 1/2$, the antiderivative becomes logarithmic, and below that boundary the stable rank grows with the dimension. The condition $d + 1 \geq 2(k + 1)$ guarantees that the integration interval used in the lower bound lies within the available finite tail.

For exponential spectra, (8) shows why power iteration can quickly collapse the filtered rank. The ratio in the geometric sum is $\exp[-(4q + 2)\beta]$, which decreases with q . This explains the empirical behavior in which $q = 1$ or $q = 2$ can make a matrix look intrinsically much lower rank to the sampled range finder. The same formula also explains a boundary case: as β approaches zero, the ratio approaches one and the spectrum approaches a flat profile, so filtered-rank contraction disappears. This boundary is visible in the failure atlas because flat or slow spectra require substantially more oversampling to reach strict relative excess targets.

The conditional theorem uses a premise rather than an established bound. The premise in (11) says that exceedance decays exponentially in $p/r_q(A)$ after tolerance and metric constants are separated. If this premise holds, the algebraic threshold follows exactly. If the premise fails, the theorem says nothing. This distinction matters for interpretation: symbolic validation can verify the implication from the premise to the threshold, but it cannot prove that Gaussian randomized SVD residuals satisfy the premise for every matrix. The experiments therefore estimate the premise through fitted exceedance slopes and through held-out rule coverage, while the manuscript keeps the probabilistic claim conditional.

B Symbolic Audit and Equation Use

The symbolic theorem audit verified seven checks. Four checks belong to the deterministic formulas: power-law cancellation in r_q , the exponential geometric sum for r_q , the exponential geometric sum for T_k , and the antiderivative used in the power-law obstruction. One check belongs to the conditional theorem: substituting (12) into (11) reduces the exceedance upper bound to δ . Two checks belong to calibration: the continuous proxy in (13) is monotone in filtered rank under $a \geq 0$, $b \geq 1$, and $\epsilon, \delta \in (0, 1)$, and the finite-grid existence statement is valid whenever the feasible set is nonempty.

The equation roles are as follows. Equation (1) defines the sampled subspace used by the Gaussian HMT range finder. Equation (2) defines the rank- k approximation whose residual is measured. Equation (3) defines the relative excess error and its denominator restrictions. Equation (4) is the borrowed stable-rank definition. Equation (5) and (6) are manuscript definitions used to separate power-filtered global energy from target-local tail mass. Equation (7) and (8) support deterministic diagnostics for the synthetic spectra. Equation (9) states the paired bounded-stable-rank and growing-tail inequalities that motivate the atlas of failure modes. Equation (10) defines the chance-constrained oversampling threshold, (11) states the conditional exceedance premise, (12) gives the derived upper bound, and (13) gives the calibrated rule evaluated in experiments.

The audit also identifies cases outside the formal context. The zero matrix is excluded because $\sigma_1(A) = 0$ makes stable rank undefined. If $\sigma_{k+1}(A) = 0$, then $T_k(A)$ and relative excess in the spectral norm are not defined in the same way as the main text. Oversampling values greater than $d - k$ are infeasible because ℓ would exceed the smaller matrix dimension. Non-Gaussian sketches require their own assumptions. These exclusions are repeated here because they are easy to overlook when a simple oversampling formula is written as a single line.

Check family	Mathematical object	Status
Finite sums	$r_q(P)$ and $r_q(E)$	Passed
Tail formulas	$T_k(E)$ and power-law antiderivative	Passed
Conditional theorem	Exceedance threshold implication	Passed
Rule monotonicity	Continuous proxy before ceiling and clipping	Passed
Finite calibration	Existence or slack fallback on a finite grid	Passed

Table 3: Summary of symbolic checks supporting the formal statements. The table distinguishes algebraic verification from the empirical question of whether the conditional exceedance premise holds for randomized SVD residuals in a broad matrix class.

C Extended Empirical Protocol

The synthetic validation uses diagonal matrices with finite power-law and exponential spectra. Diagonal matrices are not meant to approximate every application matrix. They remove singular-vector geometry so the experiment can focus on spectral decay, target rank, power iteration, and oversampling. For each valid row, the code computes $r_{\text{eff}}(A)$, $r_q(A)$, $T_k(A)$, the gap ratio σ_{k+1}/σ_k when defined, and the optimal rank- k denominators. Randomized SVD is then run with fixed seeds 11, 17, 23, 31, 43, 59. The synthetic randomized grid contains 1,728 rows, and the deterministic symbolic grid used for boundary diagnostics contains a broader set of dimensions, target ranks, and decay parameters.

The chance-threshold analysis estimates p_ν^* from empirical hit rates. A row is bracketed when the finite oversampling sweep contains at least one p whose hit rate reaches $1 - \delta$ for the requested tolerance. A row is unbracketed when the maximum tested p does not reach that hit rate. This convention is conservative: an unbracketed row might become bracketed under a wider sweep, but the current evidence cannot assign it a finite threshold. In the executed table, 204 of 336 threshold rows are bracketed. Bracketing improves with power iteration: $q = 0$ has only 6 bracketed rows, $q = 1$ has 90, and $q = 2$ has 108. This pattern supports the idea that power filtering matters, while also showing that strict targets remain hard.

The finite calibration grid evaluates $\hat{p}_\theta$ over admissible a and b values. Because no grid cell satisfies all groupwise coverage constraints, the reported rule is the slack-penalized fallback with $a = 1.2$ and $b = 2$. The fallback is not described as a feasible constrained optimum. Its value is diagnostic: it reveals how much of the desired rule behavior can be obtained with a simple monotone formula. Held-out rows are not used to refit θ , and their 0.8 coverage is reported directly.

Real-transfer experiments are designed to avoid a common ambiguity in compute-savings claims. A rule can always save columns by using fewer samples, but that is meaningless if accuracy degrades. The protocol first checks whether the rule’s mean Frobenius relative excess is no worse than the comparator’s mean error. Only then does it compute sketch-column savings. Under this matched-accuracy gate, all 54 real-transfer rows match accuracy, but none saves columns. The result is therefore conservative accuracy, not efficient planning.

D Additional Diagnostics and Reproducibility

Figure 4 gives the real-transfer savings evidence that is summarized in the main text. The negative bars should not be read as accuracy failures. They show that the calibrated rule often buys lower or matched error by using a larger sketch than fixed $p = 5$, fixed $p = 10$, or the simple k -proportional comparator. This distinction is important for deployment: a conservative planner may be useful when accuracy matters more than cost, but it does not satisfy the stronger objective of reducing compute at matched accuracy.

Figure 5 records claim-limiting cases. Slow power laws outside the $\alpha > 1/2$ theorem regime weaken size-independence. High T_k/r_q shows that tail-local obstruction can persist after filtered-rank control. Unbracketed threshold rows indicate that $p \leq 64$ does not close all strict tolerance cases. Exact-rank or zero-denominator cases require a different metric. Preprocessing sensitivity matters because centering, scaling, zero filling, and graph symmetrization change the matrix and therefore change all rank and error quantities.

Reproducibility follows from fixed random seeds, explicit sweep grids, and recorded preprocessing. The

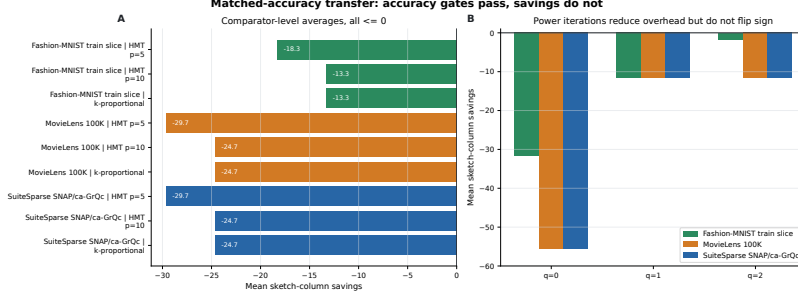


Figure 4: Matched-accuracy savings on real matrices. Bars below zero indicate that the calibrated rule used more sketch columns than the comparator after matching accuracy, so the plot supports a conservative-transfer interpretation rather than a compute-savings claim.

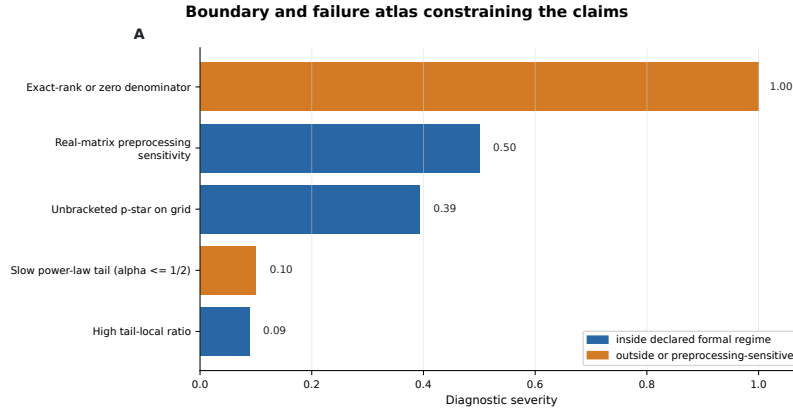


Figure 5: Boundary and failure atlas for the oversampling rule. Each diagnostic summarizes a condition that limits the main claims, including slow power-law tails, high target-local tail ratios, unbracketed thresholds, exact-rank denominators, and preprocessing sensitivity.

synthetic runs use Gaussian sketches and deterministic diagonal spectra. Real graph validation uses a binary symmetric adjacency with the diagonal removed. Real recommender validation centers observed ratings by the observed global mean and treats missing pairs as structural zeros. Real image validation scales pixels to $[0, 1]$ and centers columns. Spectral denominators for real matrices use the available numerical estimates recorded with the experiment, so real-matrix claims are tied to those numerical approximations. The symbolic checks were executed separately from the randomized simulations, which makes the deterministic formula validation independent of Monte Carlo noise.

Uncertainty is interpreted at the level supported by the executed sweep. Seed-to-seed variation is summarized for each randomized row, while threshold and calibration claims depend on empirical hit rates across the same fixed seed set. The finite sweep creates right-censoring for rows that never reach the target tolerance; those rows are not converted into guessed thresholds. Instead, they enter the failure atlas and the calibration summary as unbracketed cases. This convention is stricter than fitting a smooth curve through missing high- p outcomes, and it is why the rule is described as conservative and mixed rather than as an extrapolated guarantee.

Reproducibility materials include the manuscript source, bibliography, vector figures with previews, source tables used to populate manuscript summaries, experiment code, and symbolic verification report. The manuscript cites dataset and method sources directly rather than citing local files. Code and tabular outputs support reproducibility, while the paper itself presents the evidence through equations, theorem statements, tables, figures, and prose.