Proceedings of Machine Learning Research vol 336:1–23, 2026 39th Annual Conference on Learning Theory

Spectral Valleys and Sharp Failures in Greedy Determinant Maximization

Rajiv Khanna RAJIVAK@PURDUE.EDU
Purdue University

Editors: Steve Hanneke and Tor Lattimore

Abstract

The classic greedy algorithm is a widely used method for determinant maximization. Although worst-case theory predicts exponentially poor performance, greedy methods are often observed to perform substantially better in practice. This work explains this discrepancy through a finer spectrum-dependent analysis of the greedy algorithm.

Specifically, we develop a sharp spectrum-dependent characterization of the greedy vs optimal determinant gap by analyzing greedy selection over structured spectral landscapes. Our main result is an upper bound that decomposes this gap in terms of stable-rank windows. When the target cardinality lies within a wide spectral valley, greedy admits guarantees exponentially stronger than the classical bound; when such valleys vanish due to sharp spectral drops, greedy necessarily encounters failure cliffs, matching known worst-case constructions. This yields a spectral-landscape-dependent characterization that explains sharp regime changes in greedy performance as the target cardinality varies for the input matrix.

Finally, we show that several practical statistical models—including isotropic random features, near-identity kernels, and spiked-plus-noise spectra—provably induce these spectral success valleys, yielding strictly stronger guarantees than the worst-case theory. Together, our results provide a tight, beyond-worst-case understanding of greedy determinant maximization.

1. Introduction

Cardinality-constrained determinant maximization is a canonical discrete optimization problem at the interface of numerical linear algebra, optimal experimental design, and modern subset selection in machine learning. Given a data matrix $A = [v_1, \dots, v_n] \in \mathbb{R}^{d\times n}$ with Gram matrix $M := A^\top A \succeq 0$, the objective $\det(M_{SS})$ underlies D-optimal design (Pukelsheim, 2006; Atkinson et al., 2007), maximum-volume submatrix selection (Çivril and Magdon-Ismail, 2009), and MAP inference for determinantal point processes (Kulesza and Taskar, 2012). The problem is NP-hard and admits strong worst-case inapproximability phenomena (Çivril and Magdon-Ismail, 2009), yet it remains a workhorse primitive in scientific computing and learning pipelines.

A prominently used algorithm is the classic *greedy pivoting*: at each step, add the column that maximizes the Schur-complement pivot, or equivalently, the residual norm under orthogonalization. The greedy algorithm has clear advantages: the method is easy to implement, deterministic, streaming-friendly, and reasonably fast. Further, in practice it often produces high-quality subsets. Theoretical analysis of the algorithm, however, paints a more pessimistic picture. The greedy selection rule admits only factorial-type guarantees in the worst case: the determinant-gap factor is on the order of $(k!)^2$ (Çivril and

Magdon-Ismail, 2009), which was recently improved to $4^k k!$ (Gollapudi et al., 2023), where k is the number of selection rounds. Explicit counterexamples show that the method can indeed be exponentially suboptimal (Çivril and Magdon-Ismail, 2009; Engler, 1997). From the vantage point of worst-case approximation, it is therefore unclear why classical greedy pivoting should be trusted as a general-purpose method for determinant maximization.

To understand this gap between theory and practice, this paper undertakes a *finer-grained* analysis of greedy determinant maximization over *structured spectral landscapes*. Our key technical observation is that the quality of the greedy choices is governed by a collection of *stable-rank windows* associated with the spectrum of M. For $s \geq 0$, define the order-s stable rank

$$\mathrm{sr}_s(M) \;:=\; \lambda_{s+1}^{-1} \sum_{i>s} \lambda_i, \qquad t_s \;:=\; s + \mathrm{sr}_s(M),$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of M. The window $[s, t_s]$ should be interpreted as the range over which the spectral tail mass is comparable to $(t_s - s)\lambda_{s+1}$. Instead of providing a general guarantee applicable for any k, we provide adaptive guarantees depending on where k falls within this spectral window. Within such a window, the tail energy cannot drop too rapidly, and the greedy pivots are forced to stay large (yielding better guarantees); near the right edge, when the *right-edge slack* $\delta = t_s - k$ vanishes, the spectral bound on the pivot product deteriorates sharply which can make greedy make exponentially bad choices. This yields a concrete valley–cliff picture: *success valleys* correspond to sizes k lying deep inside at least one stable-rank window, while *failure cliffs* correspond to sharp spectral drops that collapse all relevant windows near k. This mechanism we discover is deterministic and purely spectrum-driven and independent of any constructive probabilistic assumptions.

Our main result is a windowed master theorem that upper bounds the greedy–optimal log-determinant gap by an explicit *right-edge penalty* in each spectral window. More specifically, the theorem decomposes the gap into (i) a term measuring how sharply the spectrum drops in the head relative to λ_{s+1} and (ii) a right-edge term $\Psi_{\mathrm{right}}(m, \delta) = m - \delta \log(1 + m/\delta)$ capturing how close k is to the window endpoint. Finally, for each k we take the minimum over all s which makes the bound intrinsically landscape-dependent: different windows can dominate at different cardinalities, potentially leading to sharp regime changes as k varies. This discovers a deterministic phenomenon: the same greedy algorithm can behave near-optimally for some k and catastrophically for others on the *same* input matrix. We also complement the windowed main theorem with a tightness result showing that failure cliffs are not mere artifacts of analysis. Finally, to connect the landscape theory to familiar worst-case guarantees and to enable direct quantitative comparisons, we derive a concise stable-rank refinement by specializing the main result to the global spectral window with $s = 0$. We summarize our contributions below.

1.1. Contributions

- **Master theorem over stable-rank windows.** We prove an assumption-free, windowed upper bound on the greedy–optimal determinant gap. The bound is expressed as a minimum over stable-rank windows and includes an explicit right-edge penalty $\Psi_{\mathrm{right}}(m, \delta)$, yielding a sharp valley–cliff characterization of greedy performance.

- **Tightness via spectral cliffs.** We show that the failure mechanism predicted by the master theorem is unavoidable: there exist classical counterexamples (Engler, 1997) that correspond exactly to sharp spectral cliffs and induce an exponential lower bound on greedy–optimal gap, matching the scale of existing lower bounds for this problem.
- **Unified stable-rank guarantee and improvement over factorial theory.** We derive a concise stable-rank specialization (order-0 window) that yields an explicit bound in terms of the global stable rank $\mathrm{sr}_0(M)$, which we show to be provably exponentially sharper in broad-spectrum valleys.
- **Representative statistical regimes.** We provide representative propositions showing that broad success valleys arise in standard model classes (isotropic subgaussian/random features, near-identity kernels, spiked-plus-noise spectra, polynomial decay), thereby connecting the deterministic landscape theory to widely used constructions in machine learning.

2. Related Work

Determinant maximization and maximum-volume selection. Selecting a subset of columns to maximize a determinant appears under several equivalent guises, including maximum-volume submatrix selection, maximum-volume simplex selection, and the largest simplex problem. The problem is NP-hard and admits strong inapproximability phenomena in the dimension parameter. Çivril and Magdon-Ismail (2009) give a systematic treatment of the computational landscape, relating several maximum-volume formulations and establishing worst-case guarantees for the classical greedy algorithm. Beyond worst-case hardness, determinant maximization is also central in optimal experimental design (D-optimality), where one maximizes the determinant of an information matrix; see, e.g., Pukelsheim (2006). In machine learning, the same determinant objective arises as MAP inference for determinantal point processes (DPPs), a probabilistic model favoring diverse subsets (Kulesza and Taskar, 2012), and this view has also helped motivate the use of determinant regularizers in subset selection pipelines.

Greedy algorithms and factorial-type worst-case bounds. The classic greedy algorithm considered here is a pivoting procedure—equivalently Gram–Schmidt or QR factorization with column pivoting—that selects at each step the column with maximum residual norm. As shown by Çivril and Magdon-Ismail (2009), such methods admit only factorial-type worst-case guarantees for maximum-volume objectives. Explicit adversarial examples in Engler (1997) further demonstrate how QR with column pivoting can lead to exponential failures. Motivated by the empirical rarity of these failures, we provide a spectrum-dependent characterization of the greedy–optimal determinant gap that interpolates between two extremes of broad valleys and sharp cliffs. When the spectrum has broad valleys, our bounds yield exponentially stronger guarantees than prior theory and when it exhibits sharp right-edge drops, the analysis certifies unavoidable failures by the greedy algorithm consistent with known lower bounds. Related work in distributed settings (Mahabadi et al., 2019; Gollapudi et al., 2023) studies greedy variants under communication constraints. In contrast, we focus on the single-machine setting and explain, in a spectrum-sensitive way, when and why greedy pivoting can substantially outperform worst-case predictions.

Statistical models with structured spectra. Finally, our "spectral landscape" conditions are motivated by standard statistical constructions in which eigenvalues exhibit predictable structure. Random feature models (Rahimi and Recht, 2007) and their kernelized limits lead to Gram matrices whose spectra can be analyzed via random matrix theory; see, e.g., El Karoui (2010) for foundational results on kernel random matrices in high dimensions. Low-rank-plus-noise and approximate factor models (Bai and Ng, 2002) likewise induce spiked spectra with stable-rank plateaus separated by sharp drops. In these regimes, our results certify the existence of broad success valleys for a range of subset sizes, yielding provable improvements over worst-case factorial-type bounds and tighten the gap between pessimistic theory and the widely observed practical empirical effectiveness of greedy determinant maximization.

Spectral analyses for CSSP. Several recent results in modern randomized numerical linear algebra relate subset selection quality to spectral decay and effective dimension. For column subset selection (CSSP) and Nyström approximation, Derezinski et al. (2020) give improved guarantees and identify a "multiple descent" phenomenon as the target rank varies, sharpening classical worst-case bounds by accounting for spectral structure. In kernel learning, Bach (2013) provides a beyond-worst-case statistical analysis of low-rank Nyström approximations in terms of degrees of freedom, and Kumar et al. (2012) study practical sampling strategies for Nyström with theoretical and empirical evaluation. On the deterministic side, pivoted Cholesky decompositions are widely used as adaptive, data-dependent low-rank approximations for PSD kernels; Harbrecht et al. (2012) give trace-norm error control and convergence guarantees under eigenvalue decay assumptions. Finally, *volume sampling* provides a randomized mechanism that draws subsets with probability proportional to a determinant e.g. Deshpande et al. (2006) develop guarantees for matrix approximation via volume-based sampling schemes. Our proof techniques are informed by this body of work: we bound greedy progress through the spectrum of a residual (Nyström-like) matrix and explicitly quantify how spectral tail mass governs the product of greedy pivots. The conceptual difference is the objective: rather than bounding additive approximation error (e.g., in spectral/Frobenius/trace norms), we obtain a *multiplicative* characterization of the greedy–optimal *determinant* gap, including tightness via worst-case cliff constructions.

(Weak) submodularity perspectives on greedy selection. Greedy approximation guarantees are often derived via submodularity or its relaxations. Das and Kempe (2011) introduced the submodularity ratio as a spectrum-aware condition under which greedy yields provable approximations for feature selection and sparse approximation objectives. Bian et al. (2017) further develop guarantees for greedy maximization of non-submodular functions via (sub)modularity ratio and curvature parameters, and discuss applications that include determinant-style objectives. A complementary line of work shows that restricted strong convexity (RSC) conditions can imply weak submodularity for objectives arising in statistical estimation; see Elenberg et al. (2018); Khanna et al. (2017a,b). While these frameworks provide powerful general-purpose tools, the resulting guarantees typically depend on global parameters (e.g., a uniform submodularity ratio) that may be overly pessimistic for determinant maximization, where curvature can vary sharply along the greedy trajectory. Our analysis is orthogonal: we avoid global weak-submodularity constants and instead derive a spectrum-structured bound in terms of stable-rank windows. This yields a tight beyond-worst-case theory that precisely identifies when greedy behaves near-optimally (success valleys) and when it provably cannot (failure cliffs).

3. Setup and Notation

Let $A = [v_1, v_2, \ldots, v_n] \in \mathbb{R}^{d\times n}$ be a data matrix with column vectors v_i, and let $M := A^\top A \succeq 0$ be the $n \times n$ Gram matrix of A. For any subset $S \subseteq [n]$, we denote by $A_S \in \mathbb{R}^{d\times|S|}$ the submatrix of A consisting of the columns indexed by S. Similarly, M_{SS} denotes the $|S| \times |S|$ principal submatrix of M with rows and columns indexed by S.

Determinant maximization. For a subset S of size k, the squared k-dimensional volume of the parallelotope spanned by the vectors $\{v_i : i \in S\}$ is given by the determinant of the corresponding Gram submatrix:

$$\mathrm{Vol}(S)^2 \;=\; \det(A_S^\top A_S) \;=\; \det(M_{SS})\,.$$

The *determinant maximization problem* seeks a subset S of k columns maximizing this volume. We denote by S^* an optimal choice of size k, i.e.

$$S^* \;\in\; \arg\max_{|S|=k}\; \det(M_{SS})\,.$$

Greedy Selection via Schur Pivots. We consider the standard greedy procedure for approximately solving the above selection problem. Starting with $S_0 = \emptyset$, the algorithm iteratively builds up a selected set. At iteration r (with current set S_r of size r), for each index $j \notin S_r$ we compute the *Schur complement pivot* value

$$p_r(j) \;=\; M_{jj} \;-\; M_{j\,S_r}\, M_{S_rS_r}^{-1}\, M_{S_rj}\,,$$

which represents the additional contribution to the determinant if column j were added. Equivalently, this is also the residual of v_j orthogonal to $\mathrm{span}\{v_i : i \in S_r\}$. The next column selected is $i_{r+1} := \arg\max_{j\notin S_r}\; p_r(j)$, and we update $S_{r+1} = S_r \cup \{\, i_{r+1}\}$. Repeating this for $r = 0, 1, \ldots, k-1$ yields a greedy chosen set $S_{\mathrm{gr}} := S_k$. With slight overload of notation, we write $p_{r+1} := p_r(i_{r+1})$ for the pivot value actually picked at the $(r+1)$-th step. A classical fact is that greedy determinant maximization is a *pivot-product* procedure:

$$\det(M_{S_{\mathrm{gr}}S_{\mathrm{gr}}}) \;=\; \prod_{r=0}^{k-1} p_{r+1}. \tag{1}$$

Eigenvalues and stable-rank windows. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ denote the eigenvalues of M in descending order. For an index s with $0 \leq s < \mathrm{rank}(M)$, we define the *stable rank of order s* as

$$\mathrm{sr}_s(M) \;:=\; \frac{\sum_{i>s}\lambda_i}{\lambda_{s+1}}.$$

In words, $\mathrm{sr}_s(M)$ measures the effective number of eigenvalues from λ_{s+1} onward, relative to the magnitude of λ_{s+1}. For $s = 0$, this reduces to the usual *stable rank* of M: $\mathrm{sr}_0(M) = \sum_{i\geq 1}\lambda_i \,/\, \lambda_1$. We then define the *stable-rank window endpoint* as $t_s := \; s + \mathrm{sr}_s(M)$.

Falling Factorial. For any integer $m \geq 1$ and real $x > m-1$, we define the *falling factorial*

$$(x)_m := x\,(x-1)\cdots(x-m+1).$$

4. Main Results

Goal. We study the approximation quality of greedy determinant maximization through the ratio

$$\mathsf{gap}(M,k) \;:=\; \frac{\det(M_{S^\star S^\star})}{\det(M_{S_{\mathrm{gr}}S_{\mathrm{gr}}})}.$$

where S_{gr} is the size-k subset returned by the classic greedy algorithm while $S^\star \in \arg\max_{|S|=k} \det(M_{SS})$. Equivalently, we bound $\log \mathsf{gap}(M,k)$.

Our bounds are spectrum-dependent and are expressed in terms of *stable-rank windows* of M. Intuitively, for any fixed $s \in \{0,1,\ldots,k-1\}$ and with $s < k \leq t_s$, the interval $[s,t_s]$ describes an *admissible spectral window* that characterizes the upper bound on gap based on where k falls within $[s,t_s]$.

Right-edge penalty. A key quantity in our analysis is the *right-edge slack* $\delta := t_s - k$, measuring how close k lies to the window endpoint. The loss incurred as k approaches t_s is captured by the explicit penalty (with $m \geq 1,\ \delta > 0$)

$$\Psi_{\mathrm{right}}(m,\delta) \;:=\; m - \delta \log\!\Big(1 + \frac{m}{\delta}\Big), \tag{2}$$

which will appear in our main upper bounds. Our analysis is centered around understanding when this product remains large and when it collapses. We show that these regimes are governed by a *structured spectral landscape* determined by stable-rank windows of the spectrum of M.

4.1. Success Spectral Valleys: A windowed upper bound

Our first result is a general *windowed* upper bound: it holds for all matrices and decomposes the greedy–optimal gap into a *head spikiness* term and a *right-edge* term.

Theorem 1 (Upper Bound) *For any fixed $s \in \{0,1,\ldots,k-1\}$ such that $k < t_s$, define $t_s := s + \mathrm{sr}_s(M)$ as the edge of the stable rank window. Set $m := k - s$ and $\delta := t_s - k$. Then,*

$$\log \mathsf{gap}(M,k) \leq s \log\!\Big(\frac{H_s}{\lambda_{s+1}}\Big) \;+\; \log(n)_k - \log(s!) -\; m \log m +\; \Psi_{\mathrm{right}}(m,\delta), \tag{3}$$

where $H_s := \frac{1}{s}\sum_{i=1}^{s} \lambda_i$ (with the convention $H_0 := 1$), $(n)_k := n(n-1)\cdots(n-k+1)$ and Ψ_{right} as defined in (2). Taking a minimum over all s gives us the tightest bound.

Discussion. Theorem 1 is the central result of the paper and can be viewed as a master theorem: it provides an assumption-free, fine-grained and spectrum-dependent characterization of greedy performance from which all subsequent bounds are derived. Unlike classical worst-case bounds that depend only on k, the bound depends on *where* k falls relative to stable-rank windows, and it explicitly identifies a right-edge mechanism (via Ψ_{right}) that forces pivot-product collapse near sharp spectral drops. For a fixed s and its corresponding spectral window $[s, t_s]$, the only dependence on k is through $m := k - s$ and the right-edge slack $\delta := t_s - k$. Theorem 1 yields a spectral "landscape" for greedy: deep inside the window (large δ), the right-edge penalty $\Psi_{\text{right}}(m, \delta)$ is small, and the bound decreases as the window provides more slack. As $k \uparrow t_s$ (so $\delta \downarrow 0$), the penalty term increases sharply creating a *cliff* at the window boundary predicting a sharp degradation. Since the theorem holds for arbitrary s, the minimization over s selects the most favorable window, yielding a landscape-dependent bound. When the spectrum contains multiple sharp drops (hence multiple competing window endpoints), the minimizing window index can change with k, producing multiple success valleys separated by failure cliffs. This "multi-window" behavior is the fine-grained mechanism behind regime changes in greedy performance.

For a fixed window index s, the bound (3) is small when two conditions hold: (i) *head-flatness* (the head average H_s is comparable to λ_{s+1}), and (ii) *right-edge slack* (the target size k lies well inside the window, so $t_s - k$ is not small). While Theorem 1 is assumption-free, in the sequel we will show how some practical models can satisfy these favorable conditions leading to bounds that are exponentially better than the worst case. Towards this goal, in Section 4.3, we will provide a simplification of this theorem in the global window (i.e. $s = 0$), which will lead to an explicit characterization of better-than-worst case guarantees. Further, in Section 5, we will explore statistical constructions often used in machine learning to again show provable upper bounds that are better than the known failure condition lower bounds.

4.2. Failure cliffs: Tightness via a worst-case construction

The upper bound in Section 4.1 predicts that if no window has right-edge slack (i.e., all spectral windows collapse near k), then the guarantee can deteriorate rapidly. Our next result instantiates this "cliff" phenomenon: there exist matrices with a sharp spectral drop for which greedy pivoting is exponentially suboptimal. The construction we use is a classical Hadamard-based QRCP worst-case instance borrowed from Engler (1997) to certify tightness of our landscape characterization.

Theorem 2 (Tightness: Spectral Cliffs of Exponential Failure) *The spectral-cliff regime identified by Theorem 1 is unavoidable: for infinitely many $m \geq 2$, there exists an infinite family of matrices $C^{(r)} \in \mathbb{R}^{m \times N_r}$ (with $r \to \infty$) such that:*

1. ***Spectral Cliff.*** *Let $K^{(r)} := (C^{(r)})^\top C^{(r)}$ be the Gram matrix. Then $K^{(r)}$ exhibits a* sharp spectral cliff *at m, so that for the target cardinality $k = m$ there is* no stable-rank slack*: every admissible stable-rank window endpoint satisfies $t_s = m = k$.*

2. ***Exponential gap for greedy.*** *Greedy maximum-volume selection outputs $J_{\text{gr}} = [m]$, while there exists a size-m subset $J^\star$ such that*

$$\lim_{r \to \infty} \frac{\det\big((C^{(r)}_{J^\star})^\top C^{(r)}_{J^\star}\big)}{\det\big((C^{(r)}_{J_{\text{gr}}})^\top C^{(r)}_{J_{\text{gr}}}\big)} \;=\; \exp(\Theta(m)).$$

Discussion. The construction underlying Theorem 2 is classical and has appeared in earlier analyses (Engler, 1997). Our contribution is not the construction itself, but the observation that this example corresponds precisely to a *sharp spectral drop*, i.e., the collapse of all stable-rank windows at the target cardinality. This is used by Theorem 2 to establish the tightness of Theorem 1, specifically to show that the failure mode predicted by our right-edge analysis is unavoidable in general. Further, our lower bounds in Theorem 2 match the exponential scale of previous worst-case analysis of the greedy algorithm (Çivril and Magdon-Ismail, 2009). While, the lower-bound construction of Çivril and Magdon-Ismail (2009) is a recursive geometric argument to ensure that the greedy makes progressively suboptimal choices, our construction is different – we use the spectral valley/cliff phenomenon to derive the same $\exp(\Theta(k))$ scale of the approximation gap.

4.3. Order-0 Stable Rank Refinement: A Beyond Worst-case Guarantee

Theorem 1 provides a fine-grained, windowed control of the greedy–optimal determinant gap and pinpoints a single obstruction—*right-edge collapse of stable-rank windows*—that induces failure cliffs, with tightness ratified by Theorem 2. While this structural characterization is the appropriate lens for understanding spectral regime changes across k, its statement is deliberately general and involves a minimization over different values of window indices s for the tightest possible bound. For practical use and for direct comparison to the existing factorial-type bounds, it is valuable to extract a simplified and more concise spectrum-only refinement that depends on a single scalar summary of the spectrum.

In this subsection, we derive precisely such a specialization: a *stable-rank guarantee* obtained by restricting the windowed landscape analysis to the global window with $s = 0$. The resulting bound becomes exponentially sharper than the classical worst case bound of $4^k k!$ (Gollapudi et al., 2023) in broad spectral valleys where the stable rank is large.

Order-0 stable rank. Recall $\mathrm{sr}_0(M) = \mathrm{tr}(M)/\lambda_1(M)$. When $\mathrm{sr}_0(M) \gg k$, the target size lies deep inside the global spectral window, corresponding to a broad success valley. When $\mathrm{sr}_0(M) \approx k$, stable-rank slack vanishes and the refined bound necessarily reverts to worst-case theory, consistent with the cliff construction in Theorem 2.

Theorem 3 (Unified spectral bound (offline setting)) *Assume* $\mathrm{rank}(M) \geq k$ *and* $\mathrm{sr}_0(M) \geq k$. *Then*

$$\mathsf{gap}(M, k) \;\leq\; \min\Big\{4^k (k!),\; \frac{(n)_k}{(\mathrm{sr}_0(M))_k}\Big\}.$$

Discussion. The unified bound in Theorem 3 uses $\min\{4^k(k!), \cdot\}$ to combine two *independent* bounds on the greedy pivot sequence: (i) the classical OPT-relative argument (Gollapudi et al., 2023) yielding a factorial-type guarantee[1], and (ii) our spectrum-driven window argument yielding valley-sensitive improvements. Since our bounds can be loose in certain adversarial settings, we present the unified bound. The two proof techniques leading to these two bounds are incomparable in general and it is unclear to us whether it is possible to combine them into a single bound without taking an explicit minimum. Understanding whether such a combination is possible, perhaps under additional structural assumptions on M, is an interesting direction for future work. Importantly, our bounds yield explicit improvements

1. The bound derived by Gollapudi et al. (2023) is slightly sharper than $4^k k!$. We use the latter throughout because it leads to simpler comparisons and cleaner exposition.

over the factorial type guarantee in broad valley settings. e.g. (a) if $\mathrm{sr}_0(M) \geq n/\kappa$ for some $\kappa \geq 1$ and $\frac{e\kappa}{2} < k \leq \mathrm{sr}_0(M)/2$, then from Theorem 3 the $\mathsf{gap}(M,k) \leq (2\kappa)^k$, which is exponentially smaller than $4^k k!$ for sufficiently large k. Another example arises when k is bounded away from the edge of the failure cliff: if $\mathrm{sr}_0(M) - k + 1 > \frac{en}{4k}$, then using the fact that $\frac{(n)_k}{(\mathrm{sr}_0(M))_k} \leq \frac{n^k}{(\mathrm{sr}_0(M)-k+1)^k}$ (see Lemma 14) directly gives $\mathsf{gap}(M,k) < \left(\frac{4k}{e}\right)^k \leq 4^k k!$, where the last inequality follows from Stirling's inequality.

5. Representative statistical regimes

In this section, we present concrete regimes in which the spectrum-dependent guarantee in Theorems 1,3 yields an approximation factor that is *provably* smaller than the factorial-based worst-case bound of $4^k(k!)$ for forward greedy. We will repeatedly use the elementary lower bound $k! \geq (k/e)^k$. Thus any bound of the form $\mathsf{gap}(M,k) \leq C^k$ is strictly stronger than $4^k(k!)$ for sufficiently large k.

5.1. Sub-gaussian Design

Construction. Let $X \in \mathbb{R}^{D\times n}$ have independent mean-zero subgaussian entries with variance $1/D$ (e.g. $X_{ij} \sim \mathcal{N}(0, 1/D)$). Define the Gram matrix $M := X^\top X \succeq 0$. This regime subsumes random-feature embeddings for kernel methods when the feature dimension D is at least a constant multiple of n (e.g. random Fourier features).

Proposition 4 *Assume $D \geq cn$ for a sufficiently large absolute constant c. Then with probability at least $1 - 2e^{-\Omega(n)}$ there exists an absolute constant $\kappa \geq 1$ such that*

$$\mathrm{sr}_0(M) = \frac{\mathrm{tr}(M)}{\lambda_1(M)} \;\geq\; \frac{n}{\kappa}.$$

Consequently, for every $k \leq \mathrm{sr}_0(M)/2$,

$$\mathsf{gap}(M,k) \;\leq\; \min\left\{(4^k k!),\; (2\kappa)^k\right\}. \tag{4}$$

In particular, whenever $k > e\kappa/2$, the bound $(2\kappa)^k$ is strictly smaller than $(4^k k!)$.

Proof Note that $\mathrm{tr}(M) = \|X\|_F^2 = \sum_{j=1}^n \|x_j\|_2^2$ concentrates around its mean n; in particular $\mathrm{tr}(M) \geq n/2$ with probability at least $1 - e^{-\Omega(n)}$ by a standard Bernstein bound for sub-exponential random variables (Vershynin, 2018, Ch. 2). Second, $\lambda_1(M) = \|X\|^2$ and the spectral norm of a $D \times n$ subgaussian matrix with variance $1/D$ satisfies $\|X\| \leq C$ with probability at least $1 - e^{-\Omega(n)}$ when $D \gtrsim n$ (Vershynin, 2018, Ch. 4); hence $\lambda_1(M) \leq C^2$. Combining, at the intersection of both these events $\mathrm{sr}_0(M) \geq (n/2)/C^2 = n/\kappa$ with $\kappa := 2C^2$ and failure probability $2e^{-\Omega(n)}$. Applying Theorem 3 leads to the result. ■

5.2. Near-Identity Kernels

Construction. Let $M \succeq 0$ be a kernel gram matrix with $M_{ii} = 1$ and $|M_{ij}| \leq \rho$, $\forall i \neq j$ for small ρ. This regime corresponds to kernel matrices that are close to diagonal, which arises for RBF kernels under strong separation or small bandwidth; in such settings we show below that the greedy selection is provably in a broad success valley for a wide range of k.

Proposition 5 *Assume $M_{ii} = 1$ and $|M_{ij}| \leq \rho$ for all $i \neq j$, with $\rho \leq c/(n-1)$ for some $c \in (0,1)$. Then $\lambda_1(M) \leq 1 + (n-1)\rho \leq 1 + c$ and hence*

$$\mathrm{sr}_0(M) = \frac{\mathrm{tr}(M)}{\lambda_1(M)} = \frac{n}{\lambda_1(M)} \ \geq \ \frac{n}{1+c}.$$

Consequently, for every $k \leq \mathrm{sr}_0(M)/2$,

$$\mathsf{gap}(M,k) \ \leq \ \min\left\{(4^k k!),\ (2(1+c))^k\right\},$$

and for $k > e(1+c)/2$ the spectral bound is strictly smaller than $(4^k k!)$.

Proof Gershgorin's theorem gives $\lambda_1(M) \leq 1 + max_i \sum_{j \neq i} |M_{ij}| \leq 1 + (n-1)\rho \leq 1 + c$. Apply Theorem 3 and the same simplification as in Proposition 4. ∎

5.3. Spiked-plus-Noise Spectra

Construction. Consider a spiked-plus-noise PSD matrix

$$M \ = \ \sigma^2 I_n + \sum_{j=1}^{r} \theta_j u_j u_j^\top, \qquad \text{with } \sigma^2 > 0, \quad \theta_j \geq 0, \quad u_j \text{ orthonormal.}$$

Let $\theta_{\max} := \max_j \theta_j$. This regime captures covariance/information matrices in factor-type models where a few components are present but do not dominate the noise: $\theta_{\max} \leq C\sigma^2$ for constant C. In such settings, greedy determinant maximization provably improves over factorial worst-case bounds for a wide range of subset sizes.

Proposition 6 *If $\theta_{\max} \leq C\,\sigma^2$ for some constant $C \geq 0$, then*

$$\mathrm{sr}_0(M) = \frac{\mathrm{tr}(M)}{\lambda_1(M)} \ \geq \ \frac{n\sigma^2}{\sigma^2 + \theta_{\max}} \ \geq \ \frac{n}{1+C}.$$

Consequently, for every $k \leq \mathrm{sr}_0(M)/2$,

$$\mathsf{gap}(M,k) \ \leq \ \min\left\{(4^k k!),\ (2(1+C))^k\right\},$$

and for $k > e(1+C)/2$ the spectral bound is strictly smaller than $(4^k k!)$.

Proof We have $\mathrm{tr}(M) = n\sigma^2 + \sum_{j=1}^{r} \theta_j \geq n\sigma^2$ and $\lambda_1(M) \leq \sigma^2 + \theta_{\max}$. Thus $\mathrm{sr}_0(M) = \mathrm{tr}(M)/\lambda_1(M) \geq n\sigma^2/(\sigma^2 + \theta_{\max}) \geq n/(1+C)$. The result then follows from Theorem 3. ∎

5.4. Polynomial Eigenvalue Decay

Polynomial eigenvalue decay is an often-used modeling assumption for covariance and kernel matrices with limited intrinsic dimension. In this regime, the stable-rank landscape predicts intermediate-k success valleys in which greedy determinant maximization provably improves over worst-case factorial bounds. The key point is that, under sufficiently slow polynomial decay, the order-0 stable rank $\mathrm{sr}_0(M)$ grows polynomially in n, and a "slack" condition ensuring wide spectral valleys holds for an interval of k values.

Construction. Assume the eigenvalues satisfy $\lambda_i(M) = c\, i^{-p}$ for some $c > 0$ and $p \in (0, 1/2)$.

Proposition 7 *Let $p \in (0, 1/2)$ and assume $\lambda_i(M) = c\, i^{-p}$ for $i = 1, \ldots, n$. Let*

$$T \;:=\; \mathrm{sr}_0(M) \;=\; \frac{\mathrm{tr}(M)}{\lambda_1(M)} \;=\; \sum_{i=1}^{n} i^{-p}.$$

Then there exists $n_0(p)$ such that for all $n \geq n_0(p)$ and all integers k satisfying

$$e(1-p)n^p < k \leq \frac{n^{1-p}}{4(1-p)}, \tag{5}$$

we have $T \geq k$ and

$$\mathsf{gap}(M, k) \;<\; (4^k k!).$$

Proof

(a) Lower bound Since $p \in (0, 1)$,

$$T = \sum_{i=1}^{n} i^{-p} \;\geq\; \int_1^{n+1} x^{-p}\, dx = \frac{(n+1)^{1-p} - 1}{1-p}.$$

For all n sufficiently large, $(n+1)^{1-p} - 1 \geq \frac{1}{2} n^{1-p}$, hence

$$T \;\geq\; \frac{n^{1-p}}{2(1-p)}. \tag{6}$$

(b)Existence of Slack The upper bound $k \leq \frac{n^{1-p}}{4(1-p)}$ (from (5)) together with (6) imply $k \leq T/2$, hence $T - k + 1 \geq T/2$. Using (6) again,

$$T - k + 1 \;\geq\; \frac{T}{2} \;\geq\; \frac{n^{1-p}}{4(1-p)}. \tag{7}$$

Also from the lower bound on k from (5), we get

$$\frac{en}{k} \;\leq\; \frac{en}{e(1-p)\, n^p} = \frac{n^{1-p}}{(1-p)}.$$

Combining with (7) yields

$$T - k + 1 \;>\; \frac{en}{4k},$$

which means we can now apply Lemma 14 to get the final result. ■

Remark 8 *The interval (5) is nonempty for large n precisely because $p < 1 - p$ when $p < 1/2$, so $n^p \ll n^{1-p}$. For faster decay ($p \geq 1/2$) the order-0 stable rank grows too slowly to produce a comparable intermediate-k interval from the $s = 0$ specialization; in such cases, the windowed master theorem (Theorem 1) and its landscape envelope may still yield informative valleys.*

5.5. Exponential Eigenvalue Decay

In contrast to polynomial decay, *geometric/exponential* eigenvalue decay typically yields *narrow* stable-rank windows. This regime is therefore best viewed as a sanity check: our landscape theory predicts that broad success valleys are *not* expected when the spectrum drops exponentially fast, except possibly for very small k (or when the decay rate itself scales with n).

Construction. Assume the eigenvalues satisfy a geometric decay law

$$\lambda_i \;=\; \lambda_1\, \rho^{\,i-1} \qquad \text{for some } \rho \in (0,1), \tag{8}$$

for $i = 1, 2, \ldots, n$ (up to the normalization constant λ_1).

Proposition 9 (Stable-rank windows have constant width under geometric decay) *Assume (8). Then for every $s \in \{0, 1, \ldots, n-1\}$,*

$$\mathrm{sr}_s(M) \;\leq\; \frac{1}{1-\rho}.$$

Equivalently, every stable-rank window endpoint satisfies

$$t_s \;=\; s + \mathrm{sr}_s(M) \;\leq\; s + \frac{1}{1-\rho},$$

so the window width $t_s - s$ is $O(1)$ for fixed $\rho < 1$. In particular, the sufficient condition $k < t_s$ in Theorem 1 can hold only when $k - s \leq \frac{1}{1-\rho}$, i.e., only within $O(1)$ distance of the window start.

Proof By (8),

$$\sum_{i>s} \lambda_i = \sum_{i=s+1}^{n} \lambda_1 \rho^{i-1} = \lambda_1 \rho^s \sum_{j=0}^{n-s-1} \rho^j, \qquad \lambda_{s+1} = \lambda_1 \rho^s.$$

Dividing yields $\mathrm{sr}_s(M) = \sum_{j=0}^{n-s-1} \rho^j = (1-\rho^{n-s})/(1-\rho) \leq 1/(1-\rho)$. ■

Implication for our bounds. Proposition 9 shows that exponential decay with fixed ratio $\rho < 1$ produces *rapid cliffs*: once k exceeds $s + O(1)$, right-edge slack $t_s - k$ becomes negligible and the right-edge penalty in Theorem 1 necessarily grows. Moreover, Theorem 3 requires $\mathrm{sr}_0(M) \geq k$, but here $\mathrm{sr}_0(M) \leq 1/(1-\rho)$, so the stable-rank refinement is informative only for $k = O(1)$ when $\rho < 1$ is a fixed constant.

Remark 10 (Slow exponential decay) *If the decay rate depends on n, e.g. $\rho = \rho_n \uparrow 1$ with $1 - \rho_n = \Theta(1/n)$, then $\mathrm{sr}_0(M) \asymp 1/(1-\rho_n) = \Theta(n)$ and the stable-rank guarantee yields a broad success valley. This "slow exponential" regime is effectively broad-valley spectrum at the relevant scale.*

6. Conclusion and Future Work

Conclusion. We developed a fine-grained, spectrum-dependent theory of greedy determinant maximization over structured spectral landscapes. Our main result expresses the greedy–optimal determinant gap through stable-rank windows and an explicit right-edge penalty, yielding a concrete valley–cliff picture: greedy succeeds in broad spectral valleys with nontrivial window slack and necessarily fails near spectral cliffs where all windows collapse. A classical QRCP counterexample serves as a tightness certificate, showing that the cliff mechanism identified by the theory is unavoidable. We further derived a concise stable-rank refinement that yields provably sharper guarantees in broad-spectrum regimes, including isotropic subgaussian/random-feature models, near-identity kernels, spiked-plus-noise spectra, and polynomial decay. Together, these results provide a deterministic, structurally tight explanation for the widely observed empirical effectiveness of greedy pivoting in determinant-based subset selection.

Future work. Several directions appear particularly promising.

- **Local search and exchange stability.** A natural refinement of forward greedy is swap-based local improvement (e.g., 1-swap or p-exchange). Understanding how local optimality conditions interact with spectral windows may yield sharper rates, potentially reducing global n-dependence in benign regimes while retaining cliff tightness.
- **Eigenvector regularity and leverage-based refinements.** Our core results are expressed in terms of eigenvalues only. In regimes where the relevant spectral subspaces are delocalized, leverage-score/coherence bounds may sharpen the trace-to-pivot step by effectively replacing n with an intrinsic support size. Formalizing this refinement is a compelling next step.
- **Beyond cardinality constraints.** Determinant maximization arises under richer constraints (e.g., matroid and partition constraints, budgeted selection, or streaming/composable models). Extending the spectral-landscape methodology to such constraint families, and identifying the correct notion of "right-edge slack" in these settings, could broaden the reach of the valley–cliff characterization.
- **Landscape-aware algorithms and adaptive k.** The landscape envelope suggests that different windows dominate at different cardinalities. Designing algorithms that adaptively identify the "active" window and exploit it (e.g., by choosing k or stopping rules based on estimated slack) may lead to robust, parameter-free variants with improved performance profiles.

References

Anthony C. Atkinson, Alexander N. Donev, and Randall D. Tobias. *Optimum Experimental Designs, with SAS*. Oxford University Press, 2007.

Francis Bach. Sharp analysis of low-rank kernel matrix approximations. In Shai Shalev-Shwartz and Ingo Steinwart, editors, *Proceedings of the 26th Annual Conference on Learning Theory*, volume 30 of *Proceedings of Machine Learning Research*, pages 185–209. PMLR, 2013.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002. doi: 10.1111/1468-0262.00273.

Andrew Bian, Joachim M. Buhmann, Andreas Krause, and Sebastian Tschiatschek. Guarantees for greedy maximization of non-submodular functions with applications. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

Ali Çivril and Malik Magdon-Ismail. On selecting a maximum volume sub-matrix of a matrix and related problems. *Theoretical Computer Science*, 410(47–49):4801–4811, 2009. doi: 10.1016/j.tcs.2009.06.018.

Abhimanyu Das and David Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 2011.

Michal Derezinski, Rajiv Khanna, and Michael W. Mahoney. Improved guarantees and a multiple-descent curve for column subset selection and the Nyström method. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Amit Deshpande, Luis Rademacher, Santosh S. Vempala, and Grant Wang. Matrix approximation and projective clustering via volume sampling. *Theory of Computing*, 2 (12):225–247, 2006. doi: 10.4086/toc.2006.v002a012. URL https://theoryofcomputing.org/articles/v002a012.

Noureddine El Karoui. The spectrum of kernel random matrices. *The Annals of Statistics*, 38(1):1–50, 2010. doi: 10.1214/08-AOS648.

Ethan Elenberg, Rajiv Khanna, Alexandros G. Dimakis, and Sahand Negahban. Restricted strong convexity implies weak submodularity. *Annals of Statistics*, 2018.

Hans Engler. The behavior of the QR-factorization algorithm with column pivoting. *Applied Mathematics Letters*, 10(6):7–11, 1997. doi: 10.1016/S0893-9659(97)00098-0.

Siddharth Gollapudi, Sepideh Mahabadi, and Varun Sivashankar. Composable coresets for determinant maximization: Greedy is almost optimal. *Neurips*, 2023.

Helmut Harbrecht, Michael Peters, and Reinhold Schneider. On the low-rank approximation by the pivoted Cholesky decomposition. *Applied Numerical Mathematics*, 62(4):428–440, 2012. doi: 10.1016/j.apnum.2011.10.001.

Rajiv Khanna, Ethan Elenberg, Alex Dimakis, Sahand Negahban, and Joydeep Ghosh. Scalable Greedy Feature Selection via Weak Submodularity. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1560–1568. PMLR, 20–22 Apr 2017a.

Rajiv Khanna, Ethan R. Elenberg, Alexandros G. Dimakis, Joydeep Ghosh, and Sahand Negahban. On approximation guarantees for greedy low rank optimization. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1837–1846. PMLR, 06–11 Aug 2017b.

Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning*, 5(2–3):123–286, 2012. doi: 10.1561/2200000044.

Sanjiv Kumar, Mehryar Mohri, and Ameet Talwalkar. Sampling methods for the Nyström method. *Journal of Machine Learning Research*, 13(34):981–1006, 2012.

Sepideh Mahabadi, Piotr Indyk, Shayan Oveis Gharan, and Alireza Rezaei. Composable core-sets for determinant maximization: A simple near-optimal algorithm. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4254–4263. PMLR, 2019.

Friedrich Pukelsheim. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics (SIAM), 2006. doi: 10.1137/1.9780898719109.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, pages 1177–1184, 2007.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018. doi: 10.1017/9781108231596.

Appendix A. Proofs

A.1. Theorem 1

Proof [Proof of Theorem 1] Fix $s \in \{0, 1, \ldots, k-1\}$ such that $k < t_s$ and set $m := k - s$ and $T := t_s - s = \mathrm{sr}_s(M)$, so that $T = m + \delta$ with $\delta := t_s - k > 0$. Let S_r denote the greedy set after r steps ($S_0 = \emptyset$), and define the step-r pivot for $j \notin S_r$ by

$$p_r(j) := M_{jj} - M_{jS_r} M_{S_r S_r}^{-1} M_{S_r j}, \qquad p_{r+1} := \max_{j \notin S_r} p_r(j).$$

Recall that by the pivot-product identity (Eq (1) in Section 3), we have

$$\det(M_{S_{\mathrm{gr}} S_{\mathrm{gr}}}) = \prod_{r=0}^{k-1} p_{r+1}.$$

Step (i): Lower bounding the greedy choice Define the Nyström/Schur residual at step r by

$$R_r := M - M_{:,S_r} M_{S_r S_r}^{-1} M_{S_r,:}.$$

Then $R_r \succeq 0$ and $p_r(j) = (R_r)_{jj}$ for $j \notin S_r$, while $(R_r)_{jj} = 0$ for $j \in S_r$. Hence

$$p_{r+1} = \max_{j \notin S_r} (R_r)_{jj} \;\geq\; \frac{1}{n-r} \sum_{j \notin S_r} (R_r)_{jj} = \frac{1}{n-r} \operatorname{tr}(R_r).$$

Moreover, writing $\widetilde{M}_r := M_{:,S_r} M_{S_r S_r}^{-1} M_{S_r,:}$, we have $R_r = M - \widetilde{M}_r$, $\widetilde{M}_r \succeq 0$, and $\operatorname{rank}(\widetilde{M}_r) \leq r$. Since $R_r \succeq 0$ we also have $\widetilde{M}_r \preceq M$, hence $\lambda_i(\widetilde{M}_r) \leq \lambda_i(M)$ for all i and $\lambda_i(\widetilde{M}_r) = 0$ for $i > r$. Therefore

$$\operatorname{tr}(R_r) = \operatorname{tr}(M) - \operatorname{tr}(\widetilde{M}_r) \geq \sum_{i>r} \lambda_i(M).$$

Combining the two displays yields the universal per-step bound

$$p_{r+1} \;\geq\; \frac{1}{n-r} \sum_{i>r} \lambda_i(M) \qquad (0 \leq r \leq k-1). \tag{9}$$

Step (ii): Stable-rank window control of the tail. We next lower bound $\sum_{i>r} \lambda_i(M)$ in terms of λ_{s+1} and the window endpoint t_s.

(i) Early steps $0 \leq r \leq s-1$. Since $\lambda_{r+1} \geq \cdots \geq \lambda_s \geq \lambda_{s+1}$,

$$\sum_{i>r} \lambda_i(M) \;\geq\; \sum_{i=r+1}^{s} \lambda_i(M) \;\geq\; (s-r)\lambda_{s+1}.$$

Plugging into (9) gives

$$p_{r+1} \;\geq\; \frac{s-r}{n-r} \lambda_{s+1} \qquad (0 \leq r \leq s-1),$$

and hence

$$\prod_{r=0}^{s-1} p_{r+1} \;\geq\; \lambda_{s+1}^{s} \cdot \frac{s!}{(n)_s}. \tag{10}$$

(ii) Window steps $s \le r \le k-1$. By definition of $T = \mathrm{sr}_s(M)$, we have $\sum_{i>s} \lambda_i(M) = T\lambda_{s+1}$.

For $r \ge s$,

$$\begin{aligned}\sum_{i>r} \lambda_i(M) &= \sum_{i>s} \lambda_i(M) - \sum_{i=s+1}^{r} \lambda_i(M) \\ &\ge T\lambda_{s+1} - (r-s)\lambda_{s+1} \\ &= (t_s - r)\lambda_{s+1},\end{aligned}$$

since $\lambda_i(M) \le \lambda_{s+1}$ for all $i \ge s+1$. Plugging into (9) yields

$$p_{r+1} \;\ge\; \frac{t_s - r}{n-r}\,\lambda_{s+1} \qquad (s \le r \le k-1).$$

Therefore

$$\prod_{r=s}^{k-1} p_{r+1} \;\ge\; \lambda_{s+1}^m \cdot \frac{\prod_{r=s}^{k-1}(t_s - r)}{\prod_{r=s}^{k-1}(n-r)} = \lambda_{s+1}^m \cdot \frac{(T)_m}{(n-s)_m}. \tag{11}$$

Multiplying (10) and (11), using $(n)_s(n-s)_m = (n)_k$, we obtain the greedy determinant lower bound

$$\det(M_{S_{\mathrm{gr}}S_{\mathrm{gr}}}) = \prod_{r=0}^{k-1} p_{r+1} \;\ge\; \lambda_{s+1}^k \cdot \frac{s!\,(T)_m}{(n)_k}. \tag{12}$$

Step (iii): An AM–GM upper bound on $\det(M_{S^\star S^\star})$. For any k-subset S, interlacing implies $\det(M_{SS}) \le \prod_{i=1}^k \lambda_i(M)$, hence $\det(M_{S^\star S^\star}) \le \prod_{i=1}^k \lambda_i(M)$. Fix $s < k$. By AM–GM,

$$\prod_{i=1}^{s} \lambda_i(M) \;\le\; \Big(\frac{1}{s}\sum_{i=1}^{s} \lambda_i(M)\Big)^s = H_s^s,$$

(with the convention $H_0 := 1$), and

$$\begin{aligned}\prod_{i=s+1}^{k} \lambda_i(M) &\le \Big(\frac{1}{m}\sum_{i=s+1}^{k} \lambda_i(M)\Big)^m \\ &\le \Big(\frac{1}{m}\sum_{i>s} \lambda_i(M)\Big)^m \\ &= \Big(\frac{T\lambda_{s+1}}{m}\Big)^m.\end{aligned}$$

Therefore

$$\det(M_{S^\star S^\star}) \;\le\; H_s^s\Big(\frac{T\lambda_{s+1}}{m}\Big)^m = \lambda_{s+1}^k \cdot \Big(\frac{H_s}{\lambda_{s+1}}\Big)^s \cdot \Big(\frac{T}{m}\Big)^m. \tag{13}$$

Step (iv): Combine and invoke the right-edge penalty. Dividing (13) by (12) gives

$$\frac{\det(M_{S^\star S^\star})}{\det(M_{S_{\mathrm{gr}}S_{\mathrm{gr}}})} \;\le\; \Big(\frac{H_s}{\lambda_{s+1}}\Big)^s \cdot \frac{(n)_k}{s!\,(T)_m} \cdot \Big(\frac{T}{m}\Big)^m.$$

Taking logarithms yields

$$\log \mathsf{gap}(M,k) \;\le\; s\log\Big(\frac{H_s}{\lambda_{s+1}}\Big) + \log(n)_k - \log(s!) - m\log m + \big(m\log T - \log((T)_m)\big).$$

Since $T = m + \delta$ with $\delta > 0$, Lemma 11 implies $m\log T - \log((T)_m) \le \Psi_{\mathrm{right}}(m,\delta)$. Substituting this into the equation above proves the claimed bound for the fixed s.

Finally, minimizing over all admissible s yields the stated minimum form. ■

A.1.1. Helpful Lemmas

Lemma 11 (Right-edge slack) *For any $m \ge 1$ and $\delta > 0$, set $T := m + \delta$. Then*

$$m\log T - \log\big((T)_m\big) \;\le\; \Psi_{\mathrm{right}}(m,\delta) \;:=\; m - \delta\log\Big(1 + \frac{m}{\delta}\Big).$$

Equivalently,

$$(T)_m \;\ge\; T^m \exp\big(-\Psi_{\mathrm{right}}(m,\delta)\big).$$

Proof Write $(T)_m = \prod_{j=0}^{m-1}(T-j) = \prod_{u=1}^{m}(\delta+u)$. Since $x \mapsto \log(\delta+x)$ is increasing, $\log(\delta+u) \ge \int_{u-1}^{u} \log(\delta+x)\,dx$. Summing over $u = 1,\dots,m$ gives

$$\log((T)_m) \ge \int_0^m \log(\delta+x)\,dx.$$

Evaluating the integral yields

$$\int_0^m \log(\delta+x)\,dx = \;\big[(\delta+x)\log(\delta+x) - (\delta+x)\big]_0^m \;=\; T\log T - \delta\log\delta - m.$$

Rearrange to obtain

$$m\log T - \log((T)_m) \le\; m - \delta\log\Big(\frac{T}{\delta}\Big) \;=\; m - \delta\log\Big(1+\frac{m}{\delta}\Big) \;=\; \Psi_{\mathrm{right}}(m,\delta).$$

■

A.2. Proof of Theorem 2

This section presents a complete proof of Theorem 2. We follow a classical Hadamard-based worst-case instance for QR factorization with column pivoting (QRCP) due to Engler (1997), and we emphasize how it realizes a *spectral cliff* i.e., the collapse of all stable-rank windows at the target cardinality.

A.2.1. Hadamard-based construction

Our focus is Hadamard matrices of size $m \times m$. Such matrices can be constructed, for example, by Sylvester's construction, for all $m = 2^\ell$.

Let $H_m \in \{\pm 1\}^{m \times m}$ be a Hadamard matrix, so that

$$H_m H_m^\top \;=\; mI_m \qquad \text{and hence} \qquad |\det(H_m)| \;=\; m^{m/2}. \tag{14}$$

Fix an integer $r \geq 1$ and define

$$N \;=\; N_r \;:=\; m - 1 + rm.$$

We now define a matrix $C^{(r)} \in \mathbb{R}^{m \times N}$ in two blocks: the first $m-1$ columns form a diagonal block with carefully chosen magnitudes, and the remaining rm columns are r repeated Hadamard blocks.

Define scalars $d_i, f_i > 0$ for $i \in \{1, \dots, m\}$ as follows. Set $f_1^2 = \frac{1}{N-m+1}$, and

$$d_i^2 := \frac{(N-m+2)^{\,i-1} - (N-m)(N-m+1)^{\,i-2}}{(N-m+2)^{\,i-1}}, \forall i \in [m] \tag{15}$$

$$f_i^2 := \frac{1}{N-m+1}\big(1 - d_i^2\big) = \frac{(N-m)(N-m+1)^{\,i-3}}{(N-m+2)^{\,i-1}} \forall i \in \{2, 3, \dots, m\}. \tag{16}$$

We now define $C^{(r)} = (c_{ij}) \in \mathbb{R}^{m \times N}$ columnwise. For $1 \leq j \leq m-1$, set

$$c_{ij} \;:=\; d_{m-i+1}\,\delta_{ij}, \qquad 1 \leq i \leq m, \tag{17}$$

where $\delta_{ij} = 1$ for $i = j$ and 0 otherwise; so the first $m-1$ columns form a diagonal matrix (in the first $m-1$ rows) with entries $d_m, d_{m-1}, \dots, d_2$. For the remaining columns $m \leq j \leq N$, write $j = m - 1 + (b-1)m + \ell$ with block index $b \in \{1, \dots, r\}$ and within-block index $\ell \in \{1, \dots, m\}$, and set

$$c_{ij} \;:=\; f_{m-i+1}\,(H_m)_{i\ell}, \qquad 1 \leq i \leq m. \tag{18}$$

Equivalently, the last rm columns consist of r repeated copies of $\operatorname{diag}(f_m, f_{m-1}, \dots, f_1)\,H_m$.

The following result follows from Theorem 4 of Engler (1997). We quote it in the form needed for our determinant-gap proof.

Proposition 12 ((Engler, 1997), Theorem 4) *For the matrix $C^{(r)}$ defined above, we have*

$$C^{(r)}(C^{(r)})^\top \;=\; I_m. \tag{19}$$

Moreover, as $r \to \infty$ (equivalently $N \to \infty$),

$$\sqrt{N}\,d_i \to \sqrt{i}, \qquad \sqrt{N}\,f_i \to 1 \qquad (1 \leq i \leq m), \tag{20}$$

and

$$\binom{N}{m}\,\det\big(C^{(r)}_{[m]}(C^{(r)}_{[m]})^\top\big) \;\to\; 1, \tag{21}$$

where $[m] = \{1, 2, \dots, m\}$ and $C^{(r)}_{[m]}$ denotes the $m \times m$ submatrix formed by the first m columns.

A.2.2. Greedy choice and determinant ratio

Lemma 13 (Asymptotic determinant ratio) *Let* $J_{\mathrm{gr}} := [m]$ *and let* $J^\star := \{N-m+1,\dots,N\}$ *denote the last* m *columns of* $C^{(r)}$*. Then*

$$\lim_{r\to\infty} \frac{\det\big((C^{(r)}_{J^\star})^\top C^{(r)}_{J^\star}\big)}{\det\big((C^{(r)}_{J_{\mathrm{gr}}})^\top C^{(r)}_{J_{\mathrm{gr}}}\big)} \;=\; \frac{m^m}{m!}. \tag{22}$$

Proof
(a) Determinant of the greedy set $J_{\mathrm{gr}} = [m]$.
Since $C^{(r)}$ satisfies standard assumptions (Theorem 4 of Engler (1997)), the greedy algorithm chooses the indices $J_{gr} = [m]$ (Proposition 1 of Engler (1997)). By (21) and $\det(C_{[m]}C_{[m]}^\top) = \det(C_{[m]})^2$,

$$|\det(C^{(r)}_{[m]})| \;\sim\; \binom{N}{m}^{-1/2}.$$

Using $\binom{N}{m} \to N^m/m!$ as $N\to\infty$ gives

$$N^{m/2}\,|\det(C^{(r)}_{[m]})| \;\to\; \sqrt{m!}. \tag{23}$$

(b) Determinant of the set $J^\star$. By construction (18), the last m columns form one Hadamard block:

$$C^{(r)}_{J^\star} \;=\; \mathrm{diag}(f_m, f_{m-1},\dots,f_1)\,H_m.$$

Therefore,

$$|\det(C^{(r)}_{J^\star})| \;=\; \Big(\prod_{i=1}^m f_i\Big)\,|\det(H_m)|.$$

By (20), $\sqrt{N}\,f_i \to 1$ for each i, so $\prod_{i=1}^m f_i \sim N^{-m/2}$. Together with $|\det(H_m)| = m^{m/2}$ from (14), we obtain

$$N^{m/2}\,|\det(C^{(r)}_{J^\star})| \;\to\; m^{m/2}. \tag{24}$$

(c) Determinant ratio. Combining (23) and (24) yields

$$\frac{|\det(C^{(r)}_{J^\star})|}{|\det(C^{(r)}_{[m]})|} \;\to\; \frac{m^{m/2}}{\sqrt{m!}}.$$

Thus, the corresponding Gram determinants satisfy

$$\frac{\det\big((C^{(r)}_{J^\star})^\top C^{(r)}_{J^\star}\big)}{\det\big((C^{(r)}_{[m]})^\top C^{(r)}_{[m]}\big)} = \Big(\frac{\det(C^{(r)}_{J^\star})}{\det(C^{(r)}_{[m]})}\Big)^2 \;\to\; \frac{m^m}{m!},$$

which proves (22).

∎

A.2.3. Spectral cliff and exponential failure

Proof [Proof of the spectral cliff claim in Theorem 2] By (19), $C^{(r)}(C^{(r)})^\top = I_m$. Hence the Gram matrix $K^{(r)} := (C^{(r)})^\top C^{(r)}$ has rank m and its m nonzero eigenvalues coincide with those of $C^{(r)}(C^{(r)})^\top$, i.e., are all equal to 1. Therefore $K^{(r)}$ exhibits a sharp spectral drop at m. In particular, for target cardinality $k = m$ every stable-rank window endpoint satisfies $t_s(K^{(r)}) = m$ and thus $t_s(K^{(r)}) - k = 0$ (no stable-rank slack). ■

Proof [Proof of Non-asymptotic exponential bound in Theorem 2] By Lemma 13, the ratio converges to $m^m/m!$. Using Stirling's bound $m! \le \sqrt{2\pi m}(m/e)^m e^{1/(12m)}$ yields

$$\frac{m^m}{m!} \;\ge\; \frac{e^m}{\sqrt{2\pi m}}\, e^{-1/(12m)} \;\gtrsim\; \frac{e^m}{\sqrt{m}},$$

and therefore for all sufficiently large r the same lower bound (up to constants) holds for the finite-r ratio in Theorem 2. ■

A.3. Proof of Theorem 3

Proof We prove the two bounds appearing in the minimum separately.

(i) Classical factorial-type bound. The inequality $\mathrm{gap}(M, k) \le 4^k k!$ is the worst-case guarantee for forward greedy maximum-volume selection (equivalently QRCP/pivoted Cholesky); see Gollapudi et al. (2023) for a proof.

(ii) Stable-rank bound. Assume $\mathrm{rank}(M) \ge k$ and $\mathrm{sr}_0(M) \ge k$, where $\mathrm{sr}_0(M) := \mathrm{tr}(M)/\lambda_1(M)$ and $\lambda_1(M)$ is the largest eigenvalue of M. We show that

$$\mathrm{gap}(M, k) \;\le\; \frac{(n)_k}{(\mathrm{sr}_0(M))_k}.$$

Step 1: upper bound the optimum by λ_1^k. For any k-subset S, Cauchy interlacing implies $\det(M_{SS}) \le \prod_{i=1}^k \lambda_i(M) \le \lambda_1(M)^k$. Hence

$$\det(M_{S^\star S^\star}) \;\le\; \lambda_1(M)^k. \tag{25}$$

Step 2: lower bound the greedy determinant by a pivot product. Let S_r denote the greedy set after r steps ($S_0 = \emptyset$), so $|S_r| = r$ and $S_{\mathrm{gr}} = S_k$. At step r, define the Schur pivot

$$p_r(j) := M_{jj} - M_{jS_r} M_{S_r S_r}^{-1} M_{S_r j}, \qquad j \notin S_r,$$

and let $p_{r+1} := \max_{j \notin S_r} p_r(j)$ be the selected pivot. By the pivot-product identity (Eq. (1) in the setup),

$$\det(M_{S_{\mathrm{gr}} S_{\mathrm{gr}}}) \;=\; \prod_{r=0}^{k-1} p_{r+1}. \tag{26}$$

Step 3: per-step spectral lower bound on p_{r+1}. Define the Nyström/Schur residual

$$R_r \;:=\; M - M_{:,S_r}\, M_{S_r S_r}^{-1}\, M_{S_r,:} \;\succeq\; 0.$$

Then $p_r(j) = (R_r)_{jj}$ for $j \notin S_r$ and $(R_r)_{jj} = 0$ for $j \in S_r$, hence

$$p_{r+1} = \max_{j \notin S_r} (R_r)_{jj} \;\geq\; \frac{1}{n-r}\,\mathrm{tr}(R_r).$$

Write $\widetilde{M}_r := M_{:,S_r} M_{S_r S_r}^{-1} M_{S_r,:}$, so $R_r = M - \widetilde{M}_r$ and $\mathrm{rank}(\widetilde{M}_r) \leq r$. Since $R_r \succeq 0$ we have $\widetilde{M}_r \preceq M$, and therefore

$$\mathrm{tr}(R_r) = \mathrm{tr}(M) - \mathrm{tr}(\widetilde{M}_r) \;\geq\; \sum_{i>r} \lambda_i(M).$$

Combining yields

$$p_{r+1} \;\geq\; \frac{1}{n-r} \sum_{i>r} \lambda_i(M). \tag{27}$$

Now use $\sum_{i>0} \lambda_i(M) = \mathrm{tr}(M) = \lambda_1(M)\mathrm{sr}_0(M)$ and $\lambda_i(M) \leq \lambda_1(M)$ for all i:

$$\sum_{i>r} \lambda_i(M) = \sum_{i>0} \lambda_i(M) - \sum_{i=1}^{r} \lambda_i(M) \;\geq\; \lambda_1(M)\mathrm{sr}_0(M) - r\lambda_1(M) = \lambda_1(M)\big(\mathrm{sr}_0(M) - r\big).$$

Substituting into (27) gives

$$p_{r+1} \;\geq\; \lambda_1(M)\,\frac{\mathrm{sr}_0(M) - r}{n-r}. \tag{28}$$

Step 4: multiply over r. Since $\mathrm{sr}_0(M) \geq k$, the factors $\mathrm{sr}_0(M) - r$ are positive for $r = 0, \ldots, k-1$. Multiplying (28) over $r = 0, \ldots, k-1$ and using (26) yields

$$\det(M_{S_{\mathrm{gr}} S_{\mathrm{gr}}}) \;\geq\; \lambda_1(M)^k \cdot \prod_{r=0}^{k-1} \frac{\mathrm{sr}_0(M) - r}{n-r} = \lambda_1(M)^k \cdot \frac{(\mathrm{sr}_0(M))_k}{(n)_k}.$$

Combining with (25) gives

$$\mathrm{gap}(M,k) = \frac{\det(M_{S^\star S^\star})}{\det(M_{S_{\mathrm{gr}} S_{\mathrm{gr}}})} \;\leq\; \frac{\lambda_1(M)^k}{\lambda_1(M)^k (\mathrm{sr}_0(M))_k/(n)_k} = \frac{(n)_k}{(\mathrm{sr}_0(M))_k}.$$

Combining parts(i) and (ii) above gives us the final result. ■

Appendix B. Additional Lemmas

Lemma 14 (Sufficiency of Slack) *Let $T := \mathrm{sr}_0(M)$ and assume $T \geq k$ (so $(T)_k > 0$). If*

$$T - k + 1 \;>\; \frac{e\,n}{4k}, \tag{29}$$

then the stable-rank term from Theorem 3 satisfies

$$\frac{(n)_k}{(T)_k} \;<\; (4^k k!),$$

and hence

$$\mathsf{gap}(M,k) \;\leq\; \frac{(n)_k}{(T)_k} \;<\; (4^k k!).$$

Proof Since $T \geq k$, we have $(T)_k = \prod_{j=0}^{k-1}(T-j) > 0$. Using the simple and crude bounds

$$(n)_k = \prod_{j=0}^{k-1}(n-j) \;\leq\; n^k, \qquad (T)_k = \prod_{j=0}^{k-1}(T-j) \;\geq\; (T-k+1)^k,$$

we obtain

$$\frac{(n)_k}{(T)_k} \;\leq\; \Big(\frac{n}{T-k+1}\Big)^k. \tag{30}$$

From (29), we get $n/(T-k+1) < 4k/e$, which gives

$$\Big(\frac{n}{T-k+1}\Big)^k \;<\; \Big(\frac{4k}{e}\Big)^k.$$

Finally, from the Stirling inequality $k! \geq (k/e)^k$ we get

$$\frac{(n)_k}{(T)_k} \;<\; (4^k k!),$$

which proves the result. The bound on $\mathsf{gap}(M,k)$ follows from Theorem 3, which gives $\mathsf{gap}(M,k) \leq (n)_k/(T)_k$ when $T = \mathrm{sr}_0(M) \geq k$. ■