Matrix Probing and its Conditioning[*]

Jiawei Chiu [†] and Laurent Demanet [‡]

Abstract. When a matrix A with n columns is known to be well approximated by a linear combination of basis matrices $B_1, \dots, B_p$, we can apply A to a random vector and solve a linear system to recover this linear combination. The same technique can be used to recover an approximation to A^{-1}. A basic question is whether this linear system is invertible and well-conditioned. In this paper, we show that if the Gram matrix of the B_j's is sufficiently well-conditioned and each B_j has a high numerical rank, then $n \propto p \log^2 p$ will ensure that the linear system is well-conditioned with high probability. Our main application is probing linear operators with smooth pseudodifferential symbols such as the wave equation Hessian in seismic imaging [6]. We demonstrate numerically that matrix probing can also produce good preconditioners for inverting elliptic operators in variable media.

Acknowledgments. JC is supported by the A*STAR fellowship from Singapore. LD is supported in part by a grant from the National Science Foundation. We thank Holger Rauhut for the interesting discussions.

1. Introduction.

Randomized algorithms have their roots in numerical analysis, in the form of Monte Carlo integration, Monte Carlo Markov chains [1] and so on. These methods have widespread applications, from physics, econometrics to machine learning. However, they may often referred to as the method of last resort, because they are easy to implement but produce solutions of low and uncertain accuracy.

In the last few decades, a new breed of randomized algorithms has been developed by the computer science community. These algorithms remain easy to implement, sometimes parallelizable like the Monte Carlo methods, *and* in addition, have failure probabilities that are provably negligible and run no slower than some of the most sophisticated deterministic algorithms, e.g., Karger's algorithm for the min-cut problem [13]. The design of such algorithms revolves around a few principles such as "foiling an adversary", "abundance of witnessess" and "random sampling" [14, 17].

In recent years, these evolved randomized algorithms have found their way back to numerical analysis. One particularly attractive application is that we can learn the range of a matrix A by applying it to random vectors [11, 20]. This can then be used to obtain truncated singular value decompositions which are useful in data analysis and scientific computing.

Our work carries a similar flavor: often, the matrix A can be approximated as a linear combination of a small number of matrices and the idea is to obtain these coefficients by applying A to a random vector or just a few of them. We call this "forward matrix probing." What is even more interesting is that we can also probe for A^{-1} by applying A to a random vector. We call this "backward matrix probing" for a reason that will be clear soon.

Due to approximation errors, the output of "backward probing" C is only an approximate inverse. Nevertheless, as we will see in Section 4, C serves very well as a preconditioner for inverting A, and we believe that its performance could rival that of multigrid methods for elliptic operators in smooth media.

[*] Dated February 2011.

[†] Corresponding author. jiawei@mit.edu. Department of Mathematics, MIT, Cambridge, MA 02139, USA.

[‡] Department of Mathematics, MIT, Cambridge, MA 02139, USA.

1.1. Forward Matrix Probing. Let $\mathcal{B} = \{B_1, \dots, B_p\}$ where each $B_j \in \mathbb{C}^{m \times n}$ is called a basis matrix. Let u be a Gaussian or a Rademacher sequence, that is each component of u is independent and is either a standard normal variable or a symmetric Bernoulli variable taking ±1 with equal probability.

Define the matrix $L \in \mathbb{C}^{m \times p}$ such that its j-th column is $B_j u$. Let $A \in \mathbb{C}^{m \times n}$ be the matrix we want to probe and suppose A lies in the span of $\mathcal{B}$. Say

$$A = \sum_{i=1}^{p} c_i B_i \text{ for some } c_1, \dots, c_p \in \mathbb{C}.$$

Observe that $Au = \sum_{i=1}^{p} c_i (B_i u) = Lc$. Given the vector Au, we can obtain the coefficient vector $c^* = (c_1, \dots, c_p)$ simply by solving the linear system

$$Lc = Au. \tag{1.1}$$

In practice, A is not exactly in the span of a small $\mathcal{B}$ and Equation (1.1) has to be solved in a least squares sense, for example by applying the pseudoinverse of L to the vector Au, that is $c = L^+(Au) = (L^*L)^{-1}L^*(Au)$.

1.2. Conditioning of L. Whether Equation (1.1) can be solved accurately is an important issue and depends on $\mathrm{cond}(L)$, the condition number of L. This is the ratio between the largest and the smallest singular values of L and can be understood as how different L can stretch or shrink a vector. If we treat the rows of L as a frame, then $\mathrm{cond}(L)$ is also the square root of the ratio between the frame bounds.

Intuitively, whether $\mathrm{cond}(L)$ is small lies in the following two properties of $\mathcal{B}$.

1. The B_i's "act differently" in the sense that $\langle B_j, B_k \rangle \simeq \delta_{jk}$ for any $1 \le j, k \le p$.[1]
2. Each B_i has a high rank so that $B_1 u, \dots, B_p u \in \mathbb{C}^n$ exist in a high dimensional space.

When $\mathcal{B}$ possesses these two properties and p is sufficiently small compared to n, it makes sense that L's columns, $B_1 u, \dots, B_p u$, are likely to be independent, thus guaranteeing that L is invertible, at least.

We now make the above two properties more precise. Let

$$M = L^*L \in \mathbb{C}^{p \times p} \text{ and } N = \mathbb{E}M. \tag{1.2}$$

Clearly, $\mathrm{cond}(M) = \mathrm{cond}(L)^2$. If $\mathbb{E}M$ is ill-conditioned, there is little chance that M or L is well-conditioned. This can be related to Property 1 by observing that

$$N_{jk} = \mathbb{E}M_{jk} = \mathrm{Tr}({B_j}^* B_k) = \langle B_j, B_k \rangle\,. \tag{1.3}$$

If $\langle B_j, B_k \rangle \simeq \delta_{jk}$, then the Gram matrix N is approximately the identity matrix which is well-conditioned. Hence, a more quantitative way of putting Property 1 is that we have control over $\kappa(B)$ defined as follows.

Definition 1.1. *Let $\mathcal{B} = \{B_1, \dots, B_p\}$ be a set of matrices. Define its condition number $\kappa(\mathcal{B})$ as the condition number of the matrix $N \in \mathbb{C}^{p \times p}$ where $N_{jk} = \langle B_j, B_k \rangle$.*

[1] Note that $\langle \cdot, \cdot \rangle$ is the Frobenius inner product and δ_{jk} is the Kronecker delta.

On the other hand, Property 2 can be made precise by saying that we have control over $\lambda(\mathcal{B})$ as defined below.

Definition 1.2. *Let $A \in \mathbb{C}^{m\times n}$. Define its weak condition number*[2] *as*

$$\lambda(A) = \frac{\|A\| n^{1/2}}{\|A\|_F}.$$

Let $\mathcal{B}$ be a set of matrices. Define its (uniform) weak condition number as

$$\lambda(\mathcal{B}) = \max_{A\in\mathcal{B}} \lambda(A).$$

We justify the nomenclature as follows. Suppose $A \in \mathbb{C}^{n\times n}$ has condition number k, then $\|A\|_F^2 = \sum_{i=1}^n \sigma_i^2 \geq n\sigma_{\text{min}}^2 \geq n\|A\|^2/k^2$. Taking square root, we obtain $\lambda(A) \leq k$. In other words, any well-conditioned matrix is also weakly well-conditioned. And like the usual condition number, $\lambda(\mathcal{A}) \geq 1$ because we always have $\|A\|_F \leq n^{1/2}\|A\|$.

The numerical rank of a matrix A is $\|A\|_F^2/\|A\|^2 = n\lambda(A)^{-2}$, thus having a small $\lambda(A)$ is the same as having a high numerical rank. We also want to caution the reader that $\lambda(\mathcal{B})$ is defined very differently from $\kappa(\mathcal{B})$ and is not a weaker version of $\kappa(\mathcal{B})$.

Using classical concentration inequalties, it was shown [6] that when $\lambda(\mathcal{B})$ and $\kappa(\mathcal{B})$ are fixed, $p = \tilde{O}(n^{1/2})$[3] will ensure that L is well-conditioned with high probability.

In this paper, we establish a stronger result, namely that $p = \tilde{O}(n)$ suffices. The implication is that we can expect to recover $\tilde{O}(n)$ instead of $\tilde{O}(n^{1/2})$ coefficients. The exact statement is presented below.

Theorem 1.3 (Main Result). *Let $C_1, C_2 > 0$ be numbers given by Remark B.1 in the Appendix. Let $\mathcal{B} = \{B_1, \ldots, B_p\}$ where each $B_j \in \mathbb{C}^{m\times n}$. Define $L \in \mathbb{C}^{n\times p}$ such that its j-th column is $B_j u$ where u is either a Gaussian or Rademacher sequence. Let $M = L^*L$, $N = \mathbb{E}M$ $\kappa = \kappa(\mathcal{B})$ and $\lambda = \lambda(\mathcal{B})$. Suppose*

$$n = p\left(C\kappa\lambda\log p\right)^2 \text{ for some } C \geq 1.$$

Then

$$\mathbb{P}\left(\|M - N\| \geq \frac{t\|N\|}{\kappa}\right) \leq (2C_2 np)p^{-\alpha} \text{ where } \alpha = \frac{tC}{eC_1}.$$

The number C_1 is small. C_2 may be large but it poses no problem because $p^{-\alpha}$ decays very fast with larger p and C. With $t = 1/2$, we deduce that with high probability,

$$\text{cond}(M) \leq 2\kappa + 1.$$

In general, we let $0 < t < 1$ and for the probability bound to be useful, we need $\alpha > 2$, which implies $C > 2eC_1 > 1$. Therefore the assumption that $C \geq 1$ in the theorem can be considered redundant.

[2] Throughout the paper, $\|\cdot\|$ and $\|\cdot\|_F$ denote the spectral and Frobenius norms respectively.

[3] Note that $\tilde{O}(n)$ denotes $O(n\log^c n)$ for some $c > 0$. In other words, ignore log factors.

We remark that Rauhut and Tropp have a new result (a Bernstein-like tail bound) that may be used to refine the theorem. This will be briefly discussed in Section 4.1 where we conduct a numerical experiment.

Note that when u is a Gaussian sequence, M resembles a Wishart matrix for which the distribution of the smallest eigenvalue is well-studied [8]. However, each row of L is not independent, so results from random matrix theory cannot be used in this way.

An intermediate result in the proof of Theorem 1.3 is the following. It conveys the essence of Theorem 1.3 and may be easier to remember.

Theorem 1.4. *Assume the same set-up as in Theorem 1.3. Suppose $n = \tilde{O}(p)$. Then*

$$\mathbb{E}\|M - N\| \leq C(\log p)\|N\|(p/n)^{1/2}\lambda \textit{ for some } C > 0.$$

A numerical experiment in Section 4.1 suggests that the above is not tight with respect to p and n. We find that for $\mathbb{E}\|M - N\|/\|N\|$ to vanish as $p \to \infty$, n growing strictly faster than $p \log p$ may suffice, but Theorem 1.4 would require n to grow faster than $p \log^2 p$.

1.3. Multiple Probes.

Fix n and suppose $p > n$. L is not going to be well-conditioned or even invertible. One way around this is to probe A with multiple random vectors $u_1, \ldots, u_q \in \mathbb{C}^n$ at one go, that is to solve

$$L'c = A'u,$$

where the j-th column of L' and $A'u$ are respectively

$$\begin{pmatrix} B_j u_1 \\ \vdots \\ B_j u_q \end{pmatrix} \text{ and } \begin{pmatrix} A u_1 \\ \vdots \\ A u_q \end{pmatrix}.$$

For this to make sense, $A' = I_q \otimes A$ where I_q is the identity matrix of size q. Also define $B'_j = I_q \otimes B_j$ and treat the above as probing A' assuming that it lies in the span of $\mathcal{B}' = \{B'_1, \ldots, B'_p\}$.

Regarding the conditioning of L', we can apply Theorem 1.3 to A' and $\mathcal{B}'$. It is an easy exercise (cf. Proposition A.1) to see that the condition numbers are unchanged, that is $\kappa(\mathcal{B}) = \kappa(\mathcal{B}')$ and $\lambda(\mathcal{B}) = \lambda(\mathcal{B}')$. Applying Theorem 1.3 to A' and $\mathcal{B}'$, we deduce that $\operatorname{cond}(L) \leq 2\kappa + 1$ with high probability provided that

$$nq \propto p(\kappa\lambda \log p)^2.$$

Remember that A has only mn degrees of freedom; while we can increase q as much as we like to improve the conditioning of L, the problem set-up does not allow $p > mn$ coefficients. In general, when A has rank $\tilde{n}$, its degrees of freedom is $\tilde{n}(m + n - \tilde{n})$ by considering its SVD.

1.4. When to Probe.

Matrix probing is an especially useful technique when the following holds.

1. We know that the probed matrix A can be approximated by a small number of basis matrices that are specified in advance. This holds for operators with smooth pseudodifferential symbols, which will be studied in Section 3.

2. Each matrix B_i can be applied to a vector in $\tilde{O}(\max(m,n))$ time using only $\tilde{O}(\max(m,n))$ memory.

The second condition confers two benefits. First, the coefficients c can be recovered fast, assuming that u and Au are already provided. This is because L can be computed in $\tilde{O}(\max(m,n)p)$ time and Equation (1.1) can be solved in $O(mp^2+p^3)$ time by QR factorization or other methods. In the case where increasing m,n does not require a bigger $\mathcal{B}$ to approximate A, p can be treated as a constant and the recovery of c takes only $\tilde{O}(\max(m,n))$ time.

Second, given the coefficient vector c, A can be applied to any vector v by summing over B_iv's in $\tilde{O}(\max(m,n)p)$ time . This speeds up iterative methods such as GMRES and Arnoldi.

1.5. Backward Matrix Probing.

A compelling application of matrix probing is computing the pseudoinverse A^+ of a matrix $A \in \mathbb{C}^{m\times n}$ when A^+ is known to be well-approximated in the space of some $\mathcal{B} = \{B_1, \ldots, B_p\}$. This time, we probe A^+ by applying it to a random vector $v = Au$ where u is a Gaussian or Rademacher sequence that we generate.

Like in Section 1.1, define $L \in \mathbb{C}^{n\times p}$ such its j-th column is $B_jv = B_jAu$. Suppose $A^+ = \sum_{i=1}^p c_iB_i$ for some $c_1, \ldots, c_p \in \mathbb{C}$. Then the coefficient vector c can be obtained by solving

$$Lc = A^+v = A^+Au. \tag{1.4}$$

The right hand side is u projected onto $\mathrm{null}(A)^\perp$ where $\mathrm{null}(A)$ is the nullspace of A. When A is invertible, A^+Au is simply u. We call this "backward matrix probing" because the generated random vector u appears on the opposite side of the matrix being probed in Equation (1.4). The equation suggests the following framework for probing A^+.

Algorithm 1 (Backward Matrix Probing). *Suppose $A^+ = \sum_{i=1}^p c_iB_i$. The goal is to retrieve the coefficients $c_1, \ldots, c_p$.*

1. *Generate $u \sim N(0,1)^n$ iid.*
2. *Compute $v = Au$.*
3. *Filter away u's components in null(A). Call this $\tilde{u}$.*
4. *Compute L by setting its j-column to B_jv.*
5. *Solve for c the system $Lc = \tilde{u}$ in a least squares sense.*

For the conditioning of L, we may apply Theorem 1.3 with $\mathcal{B}$ replaced with $\mathcal{B}_A := \{B_1A, \ldots, B_pA\}$ since the j-th column of L is now B_jAu. Of course, $\kappa(\mathcal{B}_A)$ and $\lambda(\mathcal{B}_A)$ can be very different from $\kappa(\mathcal{B})$ and $\lambda(\mathcal{B})$; in fact, $\kappa(\mathcal{B}_A)$ and $\lambda(\mathcal{B}_A)$ seem much harder to control because it depends on A. Fortunately, as we shall see in Section 3.5, knowing the "order" of A^+ as a pseudodifferential operator helps in keeping these condition numbers small.

When A has a high dimensional nullspace but has comparable nonzero singular values, $\lambda(\mathcal{B}_A)$ may be much larger than is necessary. By a change of basis, we can obtain a tighter result. The proof is in Section 2.3

Corollary 1.5. *Let $C_1, C_2 > 0$ be numbers given by Remark B.1 in the Appendix. Let $A \in \mathbb{C}^{m\times n}$, $\tilde{n} = \mathrm{rank}(A)$ and $\mathcal{B}_A = \{B_1A, \ldots, B_pA\}$ where each $B_j \in \mathbb{C}^{n\times m}$. Define $L \in \mathbb{C}^{n\times p}$ such that its j-th column is B_jAu where $u \sim N(0,1)^n$ iid. Let $M = L^*L$, $N = \mathbb{E}M$, $\kappa = \kappa(\mathcal{B}_A)$ and $\lambda = (\tilde{n}/n)^{1/2}\lambda(\mathcal{B}_A)$. Suppose*

$$\tilde{n} = p\,(C\kappa\lambda\log p)^2 \text{ for some } C \geq 1.$$

Then

$$\mathbb{P}\left(\|M-N\| \geq \frac{t\|N\|}{\kappa}\right) \leq (2C_2\tilde{n}p)p^{-\alpha} \textit{ where } \alpha = \frac{tC}{eC_1}.$$

Notice that $\tilde{n} = \operatorname{rank}(A)$ has taken the role of n, and our new $\lambda = \max_{1\leq j\leq p} \tilde{n}^{1/2}\|B_jA\|/\|B_jA\|_F$ ignores the $n-\tilde{n}$ zero singular values of each B_jA. Clearly, λ can be much smaller than $\lambda(\mathcal{B}_A)$.

2. Conditioning of L and Proofs.

2.1. Proof of Theorem 1.3. Our proof is decoupled into two components: one linear algebraic and one probabilistic. The plan is to collect all the results that are linear algebraic, deterministic in nature, then appeal to a probabilistic result developed in the Appendix.

To facilitate the exposition, we use a different notation for this section. We use lower case letters as *superscripts* that run from 1 to p and Greek symbols as subscripts that run from 1 to n or m. For example, the set of basis matrices is now $\mathcal{B} = \{B^1, \ldots, B^p\}$.

Our linear algebraic results concern the following variables.

1. Let $T^{jk} = B^{j*}B^k \in \mathbb{C}^{n\times n}$ and $T_{\xi\eta} \in \mathbb{C}^{p\times p}$ such that the (j,k)-th entry of $T_{\xi\eta}$ is the (ξ,η)-th entry of T^{jk}.
2. Let $Q = \sum_{1\leq\xi,\eta\leq n} T^*_{\xi\eta}T_{\xi\eta}$.
3. Let $S = \sum_{j=1}^p B^jB^{j*} \in \mathbb{C}^{m\times m}$.
4. Let $F = (T_{\xi\eta})_{1\leq\xi,\eta\leq n} \in \mathbb{C}^{np\times np}$, a block matrix.

The reason for introducing T is that M can be written as a quadratic form in $T_{\xi\eta}$ with input u:

$$M = \sum_{1\leq\xi,\eta\leq n} u_\xi u_\eta T_{\xi\eta}.$$

Since u_ξ has unit variance and zero mean, $N = \mathbb{E}M = \sum_{\xi=1}^n T_{\xi\xi}$.

Probabilistic inequalties applied to M will involve $T_{\xi\eta}$, which must be related to $\mathcal{B}$. The connection between these n by n matrices and p by p matrices lies in the identity

$$T^{jk}_{\xi\eta} = \sum_{\zeta=1}^m \overline{B^j_{\zeta\xi}}B^k_{\zeta\eta}. \tag{2.1}$$

The linear algebraic results are contained in the following propositions.

Proposition 2.1. *For any* $1 \leq \xi, \eta \leq n$,

$$T_{\xi\eta} = T^*_{\eta\xi}.$$

Hence, $T_{\xi\xi}$ *and* N *are Hermitian. Moreover, they are positive semidefinite.*

Proof. Showing that $T_{\xi\eta} = T^*_{\eta\xi}$ is straightforward from Equation (2.1). We now check that $T_{\xi\xi}$ is positive semidefinite. Let $v \in \mathbb{C}^p$. By Equation (2.1), $v^*T_{\xi\xi}v = \sum_\zeta\sum_{jk}\overline{v^j}v^k\overline{B^j_{\zeta\xi}}B^k_{\zeta\xi} = \sum_\zeta\left|\sum_k v^kB^k_{\zeta\xi}\right|^2 \geq 0$. It follows that $N = \sum_\xi T_{\xi\xi}$ is also positive semidefinite. ■

Proposition 2.2.

$$Q^{jk} = \mathit{Tr}(B^{j*}SB^k) \textit{ and } Q = \sum_{1\leq\xi,\eta\leq n} T_{\xi\eta}T^*_{\xi\eta}.$$

Proof. By Equation (2.1), $Q^{jk} = \sum_l \langle T^{lj}, T^{lk} \rangle = \sum_l \mathrm{Tr}(B^{j*} B^l B^{l*} B^k)$. The summation and trace commute to give us the first identity. Similarly, the (j,k)-th entry of $\sum_{\xi\eta} T_{\xi\eta} T^*_{\xi\eta}$ is $\sum_l \langle T^{kl}, T^{jl} \rangle = \sum_l \mathrm{Tr}(B^{l*} B^k B^{j*} B^l)$. Cycle the terms in the trace to obtain Q^{jk}. ■

Proposition 2.3. *Let $u \in \mathbb{C}^p$ be a unit vector. Define $U = \sum_{k=1}^p u^k B^k \in \mathbb{C}^{m \times n}$. Then*

$$\|U\|_F^2 \leq \|N\|.$$

Proof. $\|U\|_F^2 = \mathrm{Tr}(U^* U) = \mathrm{Tr}(\sum_{jk} \overline{u^j} u^k B^{j*} B^k)$. The sum and trace commute and due to Equation (1.3), $\|U\|_F^2 = \sum_{jk} \overline{u^j} u^k N^{jk} \leq \|N\|$. ■

Proposition 2.4.

$$\|Q\| \leq \|S\| \, \|N\| \, .$$

Proof. Q is Hermitian, so $\|Q\| = \max_u u^* Q u$ where $u \in \mathbb{C}^p$ has unit norm. Now let u be an arbitrary unit vector and define $U = \sum_{k=1}^p u^k B^k$. By Proposition 2.2, $u^* Q u = \sum_{jk} \overline{u^j} u^k Q^{jk} = \mathrm{Tr}(\sum_{jk} \overline{u^j} u^k B^{j*} S B^k) = \mathrm{Tr}(U^* S U)$. Since S is positive definite, it follows from "$\|AB\|_F \leq \|A\| \|B\|_F$" that $u^* Q u = \|S^{1/2} U\|_F^2 \leq \|S\| \|U\|_F^2$. By Proposition 2.3, $u^* Q u \leq \|S\| \|N\|$. ■

Proposition 2.5. *For any $1 \leq j \leq p$,*

$$\|B^j\| \leq \lambda n^{-1/2} \|N\|^{1/2}.$$

It follows that

$$\|Q\| = \| \sum_{\xi\eta} T_{\xi\eta} T^*_{\xi\eta} \| \leq p \lambda^2 \|N\|^2 / n.$$

Proof. We begin by noting that $\|N\| \geq \max_j |N^{jj}| = \max_j \langle B^j, B^j \rangle = \max_j \|B^j\|_F^2$. From Definition 1.2, $\|B^j\| \leq \lambda n^{-1/2} \|B^j\|_F \leq \lambda n^{-1/2} \|N\|^{1/2}$ for any $1 \leq j \leq p$, which is our first inequality. It follows that $\|S\| \leq \sum_{j=1}^p \|B^j\|^2 \leq p \lambda^2 \|N\| / n$. Apply Propositions 2.4 and 2.2 to obtain the second inequality. ■

Proposition 2.6. *F is Hermitian and*

$$\|F\| \leq \lambda^2 \|N\| (p/n).$$

Proof. That F is Hermitian follows from Proposition 2.1. Define $F' = (T^{jk})$ another block matrix. Since reindexing the rows and columns of F does not change its norm, $\|F\| = \|F'\|$. By Proposition 2.5, $\|F'\|^2 \leq \sum_{j,k=1}^p \|T^{jk}\|^2 \leq \sum_{j,k=1}^p \|B^j\|^2 \|B^k\|^2 \leq \lambda^4 \|N\|^2 (p/n)^2$. ■

We now combine the above linear algebraic results with a probabilistic result in Appendix B. Prepare to apply Proposition B.6 with A_{ij} replaced with $T_{\xi\eta}$. Note that $R = \sum_{\xi\eta} T_{\xi\eta} T^*_{\xi\eta} = Q$ by Proposition 2.2. Bound σ using Propositions 2.5 and 2.6:

$$\begin{aligned} \sigma &= C_1 \max(\|Q\|^{1/2}, \|R\|^{1/2}, \|F\|) \\ &\leq C_1 \|N\| \max((p/n)^{1/2} \lambda, (p/n) \lambda^2) \\ &\leq C_1 \|N\| (p/n)^{1/2} \lambda. \end{aligned}$$

The last step goes through because our assumption on n guarantees that $(p/n)^{1/2}\lambda \leq 1$. Finally, apply Proposition B.6 with $t\|N\|/\kappa = e\sigma u$.

2.2. Sketch of the Proof for Theorem 1.4.

Follow the proof of Proposition B.6. Letting $s = \log p$, we obtain

$$\begin{aligned}\mathbb{E}\|M-N\| &\leq (\mathbb{E}\|M-N\|^s)^{1/s}\\ &\leq C_1(2C_2np)^{1/s}s\max(\|Q\|^{1/2},\|R\|^{1/2},\|F\|)\\ &\leq C(\log p)\|N\|(p/n)^{1/2}\lambda.\end{aligned}$$

2.3. Proof of Corollary 1.5.

Let $u \sim N(0,1)^n$ iid. Say A has a singular value decomposition $E\Lambda F^*$ where Λ is diagonal. Do a change of basis by letting $u' = F^*u \sim N(0,1)^n$ iid, $B'_j = F^*B_jE$ and $\mathcal{B}'_\Lambda = \{B'_1\Lambda, \ldots, B'_p\Lambda\}$. Equation (1.1) is reduced to $L'c = \Lambda u'$ where the j-th column of L' is $B'_j\Lambda u'$.

Since Frobenius inner products, $\|\cdot\|$ and $\|\cdot\|_F$ are all preserved under unitary transformations, it is clear that $\kappa(\mathcal{B}'_\Lambda) = \kappa(\mathcal{B}_A)$ and $\lambda(\mathcal{B}'_\Lambda) = \lambda(\mathcal{B}_A)$. Essentially, for our purpose here, we may pretend that $A = \Lambda$.

Let $\tilde{n} = \operatorname{rank}(A)$. If A has a large nullspace, i.e., $\tilde{n} \ll \min(m,n)$, then $B'_j\Lambda$ has $n-\tilde{n}$ columns of zeros and many components of u' are never transmitted to the B'_j's anyway. We may therefore truncate the length of u' to $\tilde{n}$, let $\tilde{B}_j \in \mathbb{C}^{n\times\tilde{n}}$ be $B'_j\Lambda$ with its columns of zeros chopped away and apply Theorem 1.3 with $\mathcal{B}$ replaced with $\tilde{\mathcal{B}} := \{\tilde{B}_1, \ldots, \tilde{B}_p\}$. Observe that $\kappa(\tilde{\mathcal{B}}) = \kappa(\mathcal{B}'_\Lambda)$, whereas $\lambda(\tilde{\mathcal{B}}) = (\tilde{n}/n)^{1/2}\lambda(\mathcal{B}'_\Lambda)$ because $\|\tilde{B}_j\|_F = \|B'_j\Lambda\|_F$ and $\|\tilde{B}_j\| = \|B'_j\Lambda\|$ but $\tilde{B}_j$ has $\tilde{n}$ instead of n columns. The proof is complete.

3. Probing Operators with Smooth Symbols.

3.1. Basics and Assumptions.

We begin by defining what a pseudodifferential symbol is.

Definition 3.1.*Every linear operator A is associated with a pseudodifferential symbol $a(x,\xi)$ such that for any $u : \mathbb{R}^d \to \mathbb{R}$,*

$$Au(x) = \int_{\xi\in\mathbb{R}^d} e^{2\pi i\xi\cdot x}a(x,\xi)\hat{u}(\xi)d\xi \tag{3.1}$$

where $\hat{u}$ is the Fourier transform of u, that is $\hat{u}(\xi) = \int_{x\in\mathbb{R}^d} u(x)e^{-2\pi i\xi\cdot x}dx$.

We refrain from calling A a "pseudodifferential operator" at this point because its symbol has to satisfy some additional constraints that will be covered in Section 3.5. What is worth noting here is the Schwartz kernel theorem which shows that every linear operator $A : \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ has a symbol representation as in Equation (3.1) and in that integral, $a(x,\xi) \in \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ acts as a distribution. Recall that $\mathcal{S}$ is the Schwartz space and $\mathcal{S}'$ is its dual or the space of tempered distributions. The interested reader may refer to [9] or [19] for a deeper discourse.

The term "pseudodifferential" arises from the fact that differential operators have very simple symbols. For example, the Laplacian has the symbol $a(x,\xi) = -4\pi^2\|\xi\|^2$. A more elaborate example is

$$Au(x) = u(x) - \nabla\cdot\alpha(x)\operatorname{grad}u(x) \text{ for some } \alpha(x) \in C^1(\mathbb{R}^d).$$

Its symbol is

$$a(x,\xi) = 1 + \alpha(x)(4\pi^2\|\xi\|^2) - \sum_{k=1}^{d}(2\pi i\xi_k)\partial_{x_k}\alpha(x). \tag{3.2}$$

Clearly, if the media $\alpha(x)$ is smooth, so is the symbol $a(x,\xi)$ smooth in both x and ξ, an important property which will be used in Section 3.3.

For practical reasons, we make the following assumptions about $u : \mathbb{R}^d \to \mathbb{R}$ on which symbols are applied.

1. u is periodic with period 1, so only $\xi \in \mathbb{Z}^d$ will be considered in the Fourier domain.
2. u is bandlimited, say $\hat{u}$ is supported on $\Xi := [-\xi_0, \xi_0]^d \subseteq \mathbb{Z}^d$. Any summation over the Fourier domain is by default over Ξ.[4]
3. $a(x,\xi)$ and $u(x)$ are only evaluated at $x \in X \subset [0,1]^d$ which are points uniformly spaced apart. Any summation over x is by default over X.

Subsequently, Equation (3.1) reduces to a discrete and finite form:

$$Au(x) = \sum_{\xi\in\Xi} e^{2\pi i\xi\cdot x} a(x,\xi)\hat{u}(\xi). \tag{3.3}$$

We like to call $a(x,\xi)$ a "discrete symbol." Some tools are already available for manipulating such symbols [7].

3.2. User Friendly Representations of Symbols.

Given a linear operator A, it is useful to relate its symbol $a(x,\xi)$ to its matrix representation in the Fourier basis. This helps us understand the symbol as a matrix and also exposes easy ways of computing the symbols of A^{-1}, A^* and AB using standard linear algebra software.

By a matrix representation $(A_{\eta\xi})$ in Fourier basis, we mean of course that $\widehat{Au}(\eta) = \sum_\xi A_{\eta\xi}\hat{u}(\xi)$ for any η. We also introduce a more compact form of the symbol: $\hat{a}(j,\xi) = \int_x a(x,\xi)e^{-2\pi i j\cdot x}dx$. The next few results are pedagogical and listed for future reference.

Proposition 3.2. *Let A be a linear operator with symbol $a(x,\xi)$. Let $(A_{\eta\xi})$ and $\hat{a}(j,\xi)$ be as defined above. Then*

$$A_{\eta\xi} = \int_x a(x,\xi)e^{-2\pi i(\eta-\xi)x}dx; \quad a(x,\xi) = e^{-2\pi i\xi x}\sum_\eta e^{2\pi i\eta x}A_{\eta\xi};$$

$$A_{\eta\xi} = \hat{a}(\eta-\xi,\xi); \quad \hat{a}(j,\xi) = A_{j+\xi,\xi}.$$

Proof. Let $\eta = \xi + j$ and apply the definitions. ■

Proposition 3.3 (Trace). *Let A be a linear operator with symbol $a(x,\xi)$. Then*

$$Tr(A) = \sum_\xi \hat{a}(0,\xi) = \sum_\xi \int_x a(x,\xi)dx.$$

[4] To have an even number of points per dimension, one can use $\Xi = [-\xi_0, \xi_0 - 1]^d$ for example. We leave this generalization to the reader and continue to assume $\xi \in [-\xi_0, \xi_0]^d$.

Proposition 3.4 (Adjoint). *Let A and $C = A^*$ be linear operators with symbols $a(x,\xi), c(x,\xi)$. Then*

$$\hat{c}(j,\xi) = \overline{\hat{a}(-j, j+\xi)}; \quad c(x,\xi) = \sum_{\eta} \int_{y} \overline{a(y,\eta)} e^{2\pi i(\eta-\xi)(x-y)} dy.$$

Proposition 3.5 (Composition). *Let A, B and $C = AB$ be linear operators with symbols $a(x,\xi), b(x,\xi), c(x,\xi)$. Then*

$$\hat{c}(j,\xi) = \sum_{\zeta} \hat{a}(j+\xi-\zeta,\zeta)\hat{b}(\zeta-\xi,\xi);$$

$$c(x,\xi) = \sum_{\zeta} \int_{y} e^{2\pi i(\zeta-\xi)(x-y)} a(x,\zeta) b(y,\xi) dy.$$

We leave it to the reader to verify the above results.

3.3. Symbol Expansions. The idea is that when a linear operator A has a smooth symbol $a(x,\xi)$, only a few basis functions are needed to approximate a, and correspondingly only a small $\mathcal{B}$ is needed to represent A. This is not new, see for example [7]. In this paper, we consider the separable expansion

$$a(x,\xi) = \sum_{jk} c_{jk} e_j(x) g_k(\xi).$$

This is the same as expanding A as $\sum_{jk} c_{jk} B_{jk}$ where the symbol for B_{jk} is $e_j(x)g_k(\xi)$. With an abuse of notation, let B_{jk} also denote its matrix representation *in Fourier basis*. Given our assumption that $\xi \in [-\xi_0, \xi_0]^d$, we have $B_{jk} \in \mathbb{C}^{n\times n}$ where $n = (2\xi_0+1)^d$. As its symbol is separable, B_{jk} can be factorized as

$$B_{jk} = \mathcal{F} \operatorname{diag}(e_j(x)) \mathcal{F}^{-1} \operatorname{diag}(g_k(\xi)) \tag{3.4}$$

where $\mathcal{F}$ is the unitary Fourier matrix. An alternative way of viewing B_{jk} is that it takes its input $\hat{u}(\xi)$, multiply by $g_k(\xi)$ and convolve it with $\hat{e}_j(\eta)$, the Fourier transform of $e_j(x)$. There is also an obvious algorithm to apply B_{jk} to $u(x)$ in $\tilde{O}(n)$ time as outlined below. As mentioned in Section 1.4, this speeds up the recovery of the coefficients c and makes matrix probing a cheap operation.

Algorithm 2.*Given vector $u(x)$, apply the symbol $e_j(x)g_k(\xi)$.*

1. *Perform FFT on u to obtain $\hat{u}(\xi)$.*
2. *Multiply $\hat{u}(\xi)$ by $g_k(\xi)$ elementwise.*
3. *Perform IFFT on the previous result, obtaining $\sum_{\xi} e^{2\pi i\xi\cdot x} g_k(\xi)\hat{u}(\xi)$.*
4. *Multiply the previous result by $e_j(x)$ elementwise.*

Recall that for L to be well-conditioned with high probability, we need to check whether N, as defined in Equation (1.3), is well-conditioned, or in a rough sense whether $\langle B_j, B_k\rangle \simeq \delta_{jk}$. For separable symbols, this inner product is easy to compute.

Proposition 3.6. *Let $B_{jk}, B_{j'k'} \in \mathbb{C}^{n\times n}$ be matrix representations (in Fourier basis) of linear operators with symbols $e_j(x)g_k(\xi)$ and $e_{j'}(x)g_{k'}(\xi)$. Then*

$$\langle B_{jk}, B_{j'k'}\rangle = \langle e_j, e_{j'}\rangle \langle g_k, g_{k'}\rangle$$

where $\langle e_j, e_{j'} \rangle = \frac{1}{n}\sum_{i=1}^{n} \overline{e_j(x_i)} e_{j'}(x_i)$ *and* $x_1, \ldots, x_n$ *are points in* $[0,1]^d$ *uniformly spaced, and* $\langle g_k, g_{k'} \rangle = \sum_\xi \overline{g_k(\xi)} g_k(\xi)$.

Proof. Apply Propositions 3.3, 3.4 and 3.5 with the symbols in the $\hat{a}(\eta,\xi)$ form. ■

To compute $\lambda(\mathcal{B})$ as in Definition 1.2, we examine the spectrum of B_{jk} for every j,k. A simple and relevant result is as follows.

Proposition 3.7. *Assume the same set-up as in Proposition 3.6. Then*

$$\sigma_{\min}(B_{jk}) \geq \min_x |e_j(x)| \min_\xi |g_k(\xi)|; \quad \sigma_{\max}(B_{jk}) \leq \max_x |e_j(x)| \max_\xi |g_k(\xi)|.$$

Proof. In Equation (3.4), $\mathcal{F} \operatorname{diag}(e^j(x))\mathcal{F}^{-1}$ has singular values $|e_j(x)|$ as x varies over X, defined at the end of Section 3.1. The result follows from the min-max theorem. ■

As an example, suppose $a(x,\xi)$ is smooth and periodic in both x and ξ. It is well-known that a Fourier series is good expansion scheme because the smoother $a(x,\xi)$ is as a periodic function in x, the faster its Fourier coefficients decay, and less is lost when we truncate the Fourier series. Hence, we pick[5]

$$e_j(x) = e^{2\pi i j\cdot x}; \quad g_k(\xi) = e^{2\pi i k\cdot \varphi(\xi)}, \tag{3.5}$$

where $\varphi(\xi) = (\xi+\xi_0)/(2\xi_0+1)$ maps ξ into $[0,1]^d$.

Due to Proposition 3.6, $N = \mathbb{E}M$ is a multiple of the identity matrix and $\kappa(\mathcal{B}) = 1$ where $\mathcal{B} = \{B_{jk}\}$. It is also immediate from Proposition 3.7 that $\lambda(B_{jk}) = 1$ for every j,k, and $\lambda(\mathcal{B}) = 1$. The optimal condition numbers of this $\mathcal{B}$ make it suitable for matrix probing.

3.4. Chebyshev Expansion of Symbols. The symbols of differential operators are polynomials in ξ and nonperiodic. When probing these operators, a Chebyshev expansion in ξ is in principle favored over a Fourier expansion, which may suffer from Gibbs phenomenon. However, as we shall see, $\kappa(\mathcal{B})$ grows with p and can lead to ill-conditioning.

For simplicity, assume that the symbol is periodic in x and that $e_j(x) = e^{2\pi i j\cdot x}$. Applying Proposition 3.2, we see that B_{jk} is a matrix with a displaced diagonal and its singular values are $(g_k(\xi))_{\xi\in\Xi}$. (Recall that we denote the matrix representation (in Fourier basis) of B_{jk} as B_{jk} as well.)

Let T_k be the k-th Chebyshev polynomial. In 1D, we can pick

$$g_k(\xi) = T_k(\xi/\xi_0) \text{ for } k = 1, \ldots, K. \tag{3.6}$$

Define $\|T_k\|_2 = (\int_{z=-1}^{1} T_k(z)^2 dz)^{1/2}$. Notice that there is no $(1-z^2)^{-1/2}$ weight factor. By approximating sums with integrals, $\lambda(B_{jk}) \simeq \sqrt{2}\|T_k\|_2^{-1} = \left(\frac{4k^2-1}{2k^2-1}\right)^{1/2}$. In practice, this approximation becomes very accurate with larger n and we see no need to be rigorous here. As k increases, the above approaches $\sqrt{2}$. More importantly, $\lambda(B_{jk}) \leq \lambda(B_{j1})$ for any j,k, so

$$\lambda(\mathcal{B}) = \sqrt{3}.$$

Applying the same technique to approximate the sum $\langle g_k, g_{k'} \rangle$, we find that $\langle g_k, g_{k'} \rangle \propto (1-(k+k')^2)^{-1} + (1-(k-k')^2)^{-1}$ when $k+k'$ is even, and zero otherwise. We then compute

[5] Actually, $\exp(2\pi i k\xi_0/(2\xi_0+1))$ does not vary with ξ, and we can use $\varphi(\xi) = \xi/(2\xi_0+1)$.

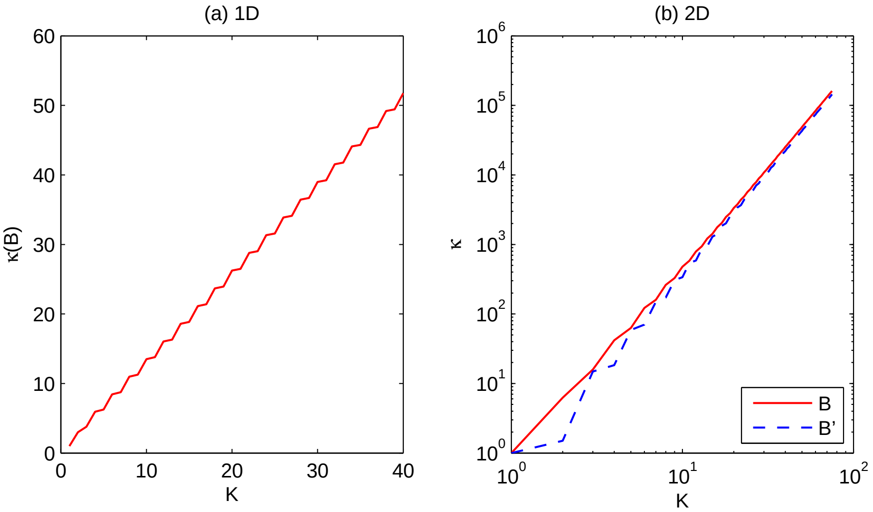

Figure 3.1. *Let K be the number of Chebyshev polynomials used in the expansion of the symbol, see Equation (3.6) and (3.7). Observe that in 1D, $\kappa(\mathcal{B}) = O(K)$ while in 2D, $\kappa(\mathcal{B}) = O(K^3)$. These condition numbers mean that we cannot expect to retrieve $p = \tilde{O}(n)$ parameters unless K is fixed and independent of p, n.*

$N = \mathbb{E}M$ for various K and plot $\kappa(\mathcal{B})$ versus K, the number of Chebyshev polynomials. As shown in Figure 3.1(a),

$$\kappa(\mathcal{B}) \simeq 1.3K.$$

This means that if we expect to recover $p = \tilde{O}(n)$ coefficients, we must keep K fixed. Otherwise, if $p = K^2$, only $p = \tilde{O}(n^{1/2})$ are guaranteed to be recovered by Theorem 1.3.

In 2D, a plausible expansion is

$$g_k(\xi) = e^{ik_1 \arg \xi} T_{k_2}(\varphi(\|\xi\|)) \text{ for } 1 \le k_2 \le K \tag{3.7}$$

where $k = (k_1, k_2)$ and $\varphi(r) = (\sqrt{2}r/\xi_0) - 1$ maps $\|\xi\|$ into $[-1, 1]$. We call this the "Chebyshev on a disk" expansion.

The quantity $\lambda(B_{jk})$ is approximately $2\left(\int_{x=-1}^{1}\int_{y=-1}^{1} T_k(\psi(x,y))^2 dx\, dy\right)^{-1/2}$ where $\psi(x,y) = (2x^2 + 2y^2)^{1/2} - 1$. The integral is evaluated numerically and appears to converge[6] to $\sqrt{2}$ for large k_2. Also, $k_2 = 1$ again produces the worst $\lambda(B_{jk})$ and

$$\lambda(\mathcal{B}) \le 2.43.^{7}$$

As for $\kappa(\mathcal{B})$, observe that when $k_1 \ne k_1'$, $\left\langle g_{k_1 k_2}, g_{k_1' k_2'} \right\rangle = \pm 1$ due to symmetry[8], whereas when $k_1 = k_1'$, the inner product is proportional to n and is much larger. As a result, the g_k's

[6] This is because when we truncate the disk of radius $\xi_0\sqrt{2}$ to a square of length $2\xi_0$, most is lost along the vertical axis and away from the diagonals. However, for large k, T_k oscillates very much and the truncation does not matter. If we pretend that the square is a disk, then we are back in the 1D case where the answer approaches $\sqrt{2}$ for large k.

[7] The exact value is $2(4 - \frac{8}{3}\sqrt{2}\sinh^{-1}(1))^{-1/2}$.

[8] The ξ and $-\xi$ terms cancel each other. Only $\xi = 0$ contributes to the sum.

with different k_1's hardly interact and in studying $\kappa(\mathcal{B})$, one may assume that $k_1 = k_1' = 0$. To improve $\kappa(\mathcal{B})$, we can normalize g_k such that the diagonal entries of N are all ones, that is $g_k'(\xi) = g_k(\xi)/\|g_k(\xi)\|$.

This yields another set of basis matrices $\mathcal{B}'$. Figure 3.1(b) reveals that

$$\kappa(\mathcal{B}) = O(K^3) \text{ and } \kappa(\mathcal{B}') \simeq \kappa(\mathcal{B}).$$

The latter can be explained as follows: we saw earlier that $\langle B_{jk}, B_{jk} \rangle$ converges as k_2 increases, so the diagonal entries of N are about the same and the normalization is only a minor correction.

If $a(x,\xi)$ is expanded using the same number of basis functions in each direction of x and ξ, i.e., $K = p^{1/4}$, then Theorem 1.3 suggests that only $p = \tilde{O}(n^{2/5})$ coefficients can be recovered.

To recap, for both 1D and 2D, $\lambda(\mathcal{B})$ is a small number but $\kappa(\mathcal{B})$ increases with K. Fortunately, if we know that the operator being probed is a second order derivative for example, we can fix $K = 2$.

Numerically, we have observed that the Chebyshev expansion can produce dramatically better results than the Fourier expansion of the symbol. More details can be found in Section 4.3.

3.5. Order of an Operator.

In standard texts, A is said to be a pseudodifferential operator of order w if its symbol $a(x,\xi)$ is in $C^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ and for any multi-indices α, β, there exists a constant $C_{\alpha\beta}$ such that

$$|\partial_\xi^\alpha \partial_x^\beta a(x,\xi)| \le C_{\alpha\beta} \langle \xi \rangle^{w-|\alpha|} \text{ for all } \xi, \text{ where } \langle \xi \rangle = 1 + \|\xi\|.$$

Letting $\alpha = \beta = 0$, we see that such operators have symbols that grow or decay as $(1+\|\xi\|)^w$. As an example, the Laplacian is of order 2. The factor 1 prevents $\langle \xi \rangle$ from blowing up when $\xi = 0$. There is nothing special about it and if we take extra care when evaluating the symbol at $\xi = 0$, we can use

$$\langle \xi \rangle = \|\xi\|.$$

For forward matrix probing, if it is known a priori that $a(x,\xi)$ behaves like $\langle \xi \rangle^w$, it makes sense to expand $a(x,\xi) \langle \xi \rangle^{-w}$ instead. Another way of viewing this is that the symbol of the operator B_{jk} is modified from $e_j(x) g_k(\xi)$ to $e_j(x) g_k(\xi) \langle \xi \rangle^w$ to suit A better.

For backward matrix probing, if A is of order z, then A^{-1} is of order $-z$ and we should replace the symbol of B_{jk} with $e_j(x) g_k(\xi) \langle \xi \rangle^{-w}$. We believe that this small correction has an impact on the accuracy of matrix probing, as well as the condition numbers $\kappa(\mathcal{B}_A)$ and $\lambda(\mathcal{B}_A)$.

Recall that an element of $\mathcal{B}_A$ is $B_{jk}A$. If A is of order w and B_{jk} is of order 0, then $B_{jk}A$ is of order w and $\lambda(B_{jk}A)$ will grow with n^w, which will adversely affect the conditioning of matrix probing. However, by multiplying the symbol of B_{jk} by $\langle \xi \rangle^{-w}$, we can expect $B_{jk}A$ to be order 0 and that $\lambda(B_{jk}A)$ is independent of the size of the problem n. The argument is heuristical but we will support it with some numerical evidence in Section 4.3.

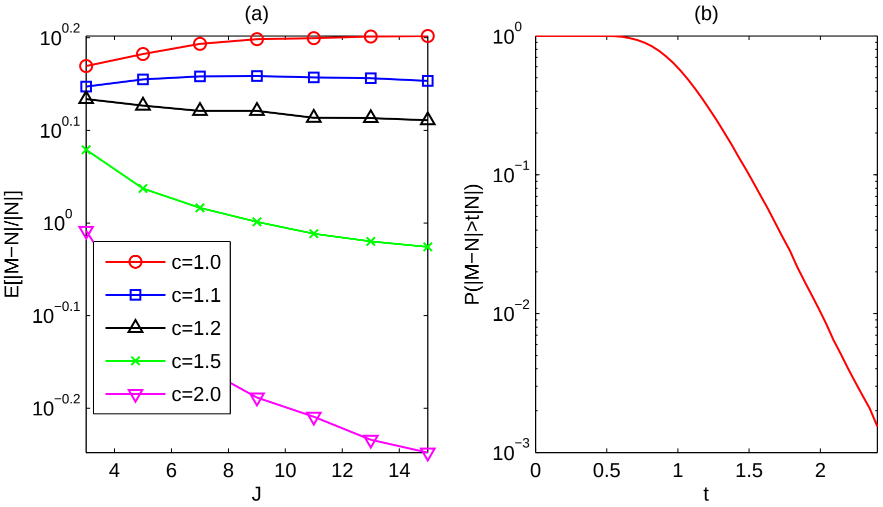

Figure 4.1. *Consider the Fourier expansion of the symbol. J is the number of basis functions in x and ξ, so $p = J^2$. Let $n = \log^c p$. Figure (a) shows that the estimated $\mathbb{E}\|M - N\|/\|N\|$ decays for $c \geq 1.1$ which suggests that Theorem 1.4 is not tight. In Figure (b), we estimate $\mathbb{P}(\|M - N\|/\|N\| > t)$ by sampling $\|M - N\|/\|N\|$ 10^5 times. The tail probability appears to be subgaussian for small t but exponential for larger t.*

4. Numerical Examples.

We carry out four different experiments. The first experiment suggests that Theorem 1.4 is not tight. The second experiment presents the output of backward probing in a visual way. In the third experiment, we explore the limitations of backward probing and also tests the Chebyshev expansion of symbols. The last experiment involves the forward probing of the foveation operator, which is related to human vision.

4.1. 1D Statistical Study.

We are interested in whether the probability bound in Theorem 1.3 is tight with respect to p and n, but as the tail probabilities are small and hard to estimate, we opt to study its first moment instead. In particular, if Theorem 1.4 captures exactly the dependence of $\mathbb{E}\|M - N\|/\|N\|$ on p and n, then we would need n to grow faster than $p \log^2 p$ for $\mathbb{E}\|M - N\|/\|N\|$ to vanish, assuming $\lambda(\mathcal{B})$ is fixed.

For simplicity, we use the Fourier expansion of the symbol in 1D so that $\lambda(\mathcal{B}) = \kappa(\mathcal{B}) = 1$. Let J be the number of basis functions in both x and ξ and $p = J^2$. Figure 4.1(a) suggests that $\mathbb{E}\|M - N\|/\|N\|$ decays to zero when $n = p \log^c p$ and $c > 1$. It follows from the previous paragraph that Theorem 1.4 cannot be tight.

Nevertheless, Theorem 1.4 is optimal in the following sense. Imagine a more general bound

$$\mathbb{E}\frac{\|M - N\|}{\|N\|} \leq \log^{\alpha} p \left(\frac{p}{n}\right)^{\beta} \text{ for some } \alpha, \beta > 0. \tag{4.1}$$

In Figure 4.2(a), we see that for various values of p/n, $\alpha = 1$ since the graphs are linear. On the other hand, if we fix p and vary n, the log-log graph of Figure 4.2(b) shows that $\beta = 1/2$. Therefore, any bound in the form of Equation (4.1) is no better than Theorem 1.4.

Next, we fix $p = 25, n = 51$ and sample $\|M - N\|/\|N\|$ many times to estimate the tail probabilities. In Figure 4.1(b), we see that the tail probability of $\mathbb{P}(\|M - N\|/\|N\| > t)$

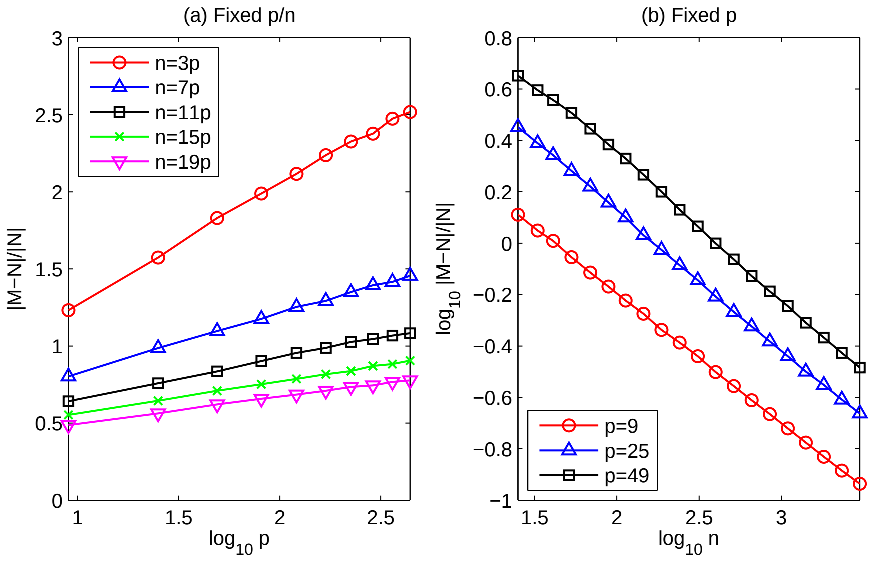

Figure 4.2. *Consider bounding* $\mathbb{E}\|M-N\|/\|N\|$ *by* $(\log^\alpha p)(p/n)^\beta$*. In Figure (a), the estimated* $\mathbb{E}\|M-N\|/\|N\|$ *depends linearly on* $\log p$*, so* $\alpha \geq 1$*. In Figure (b), we fix* p *and find that for large* n*,* $\beta = 1/2$*. The conclusion is that the bound in Theorem 1.4 has the best* α, β*.*

decays as $\exp(-c_1 t)$ when t is big, and as $\exp(-c_2 t^2)$ when t is small, for some positive numbers c_1, c_2. This behavior may be explained by Rauhut and Tropp's yet published result.

4.2. Elliptic Equation in 1D. We find it instructive to consider a 1D example of matrix probing because it is easy to visualize the symbol $a(x,\xi)$. Consider the operator

$$Au(x) = -\frac{d}{dx}\alpha(x)\frac{du(x)}{dx} \text{ where } \alpha(x) = 1 + 0.4\cos(4\pi x) + 0.2\cos(6\pi x). \tag{4.2}$$

Note that we use periodic boundaries and A is positive semidefinite with a one dimensional nullspace consisting of constant functions.

We probe for A^+ using Algorithm 1 and the Fourier expansion of its symbol or Equation (3.5). Since A is of order 2, we premultiply $g_k(\xi)$ by $\langle\xi\rangle^{-2}$ as explained in Section 3.5.

In the experiment, $n = 201$ and there are two other parameters J, K which are the number of e_j's and g_k's used in Equation (3.5). To be clear, $-\frac{J-1}{2} \leq j \leq \frac{J-1}{2}$ and $-\frac{K-1}{2} \leq k \leq \frac{K-1}{2}$.

Let C be the output of matrix probing. In Figure 4.3(b), we see that $J = K = 5$ is not enough to represent A^+ properly. This is expected because our media $\alpha(x)$ has a bandwidth of 7. We expect $J = K = 13$ to do better, but the much larger p leads to overfitting and a poor result, as is evident from the wobbles in the symbol of C in Figure 4.3(c). Probing with four random vectors, we obtain a much better result as shown in Figure 4.3(d).

4.3. Elliptic Equation in 2D. In this section, we extend the previous set-up to 2D and address a different set of questions. Consider the operator A defined as

$$Au(x) = -\nabla\cdot\alpha(x)\nabla u(x) \text{ where } \alpha(x) = \frac{1}{T} + \cos^2(\pi\gamma x_1)\sin^2(\pi\gamma x_2). \tag{4.3}$$

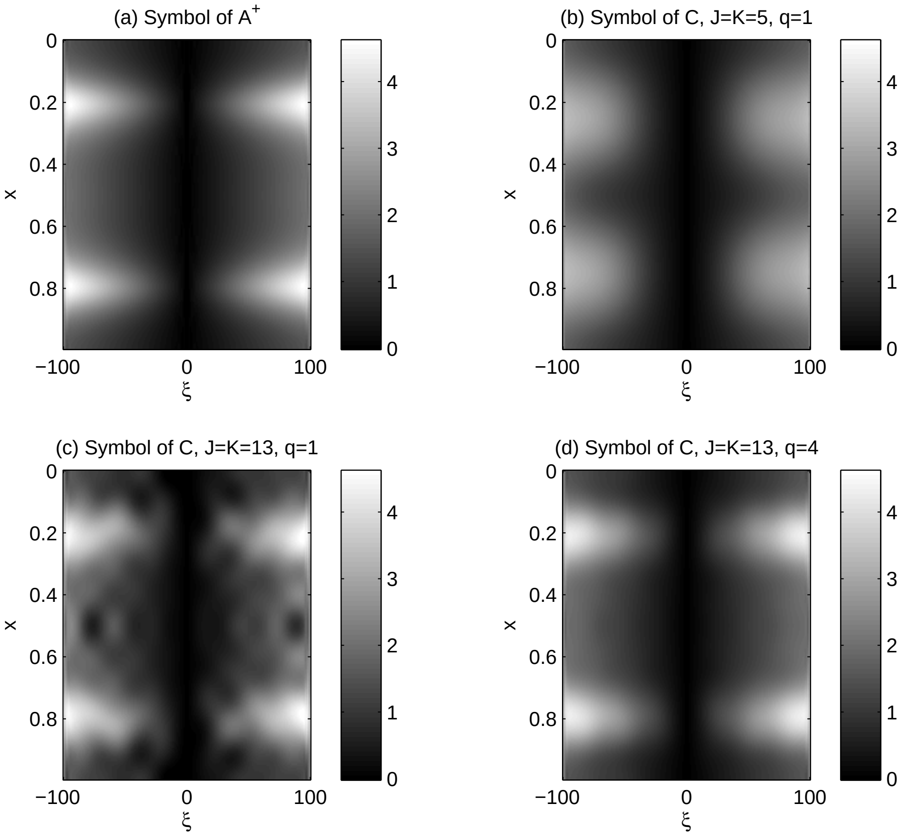

Figure 4.3. *Let A be the 1D elliptic operator in Equation (4.2) and A^+ be its pseudoinverse. Let C be the output of backward matrix probing with the following parameters: q is the number of random vectors applied to A^+; J, K are the number of e_j's and g_k's used to expand the symbol of A^+ in Equation (3.5). Figure (a) is the symbol of A^+. Figure (b) is the symbol of C with $J = K = 5$. It lacks the sharp features of Figure (a) because $\mathcal{B}$ is too small to represent A^+ well. With $J = K = 13$, probing with only one random vector leads to ill-conditioning and an inaccurate result in Figure (b). In Figure (c), four random vectors are used and a much better result is obtained. Note that the symbols are multipled by $\langle \xi \rangle^3$ for better visual contrast.*

The positive value T is called the contrast while the positive integer γ is the roughness of the media, since the bandwidth of $\alpha(x)$ is $2\gamma + 1$. Again, we assume periodic boundary conditions such that A's nullspace is exactly the set of constant functions.

Let C be the output of the backward probing of A. As we shall see, the quality of C drops as we increase the contrast T or the roughness γ.

Fix $n = 101^2$ and expand the symbol using Equation (3.5). Let $J = K$ be the number of basis functions used to expand the symbol in each of its four dimensions, that is $p = J^4$.

In Figure 4.4(b), we see that between $J = 2\gamma - 1$ and $J = 2\gamma + 1$, the bandwidth of the media, there is a marked improvement in the preconditioner, as measured by the ratio $\mathrm{cond}(CA)/\mathrm{cond}(A)$.[9]

[9] Since A has one zero singular value, $\mathrm{cond}(A)$ actually refers to the ratio between its largest singular value and its second smallest singular value. The same applies to CA.

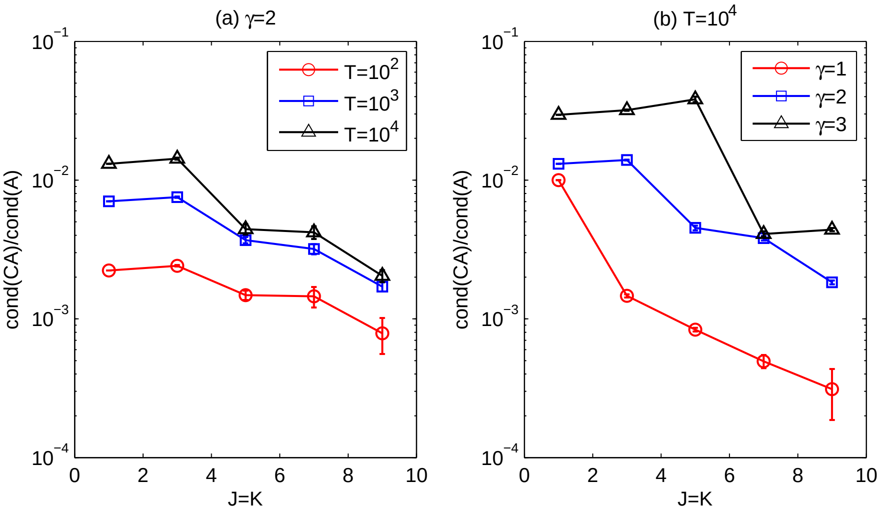

Figure 4.4. *Let A be the operator defined in Equation (4.3) and C be the output of backward probing. In Figure (b), we fix $T = 10^4$ and find that as J goes from $2\gamma-1$ to $2\gamma+1$, the bandwidth of the media, the quality of the preconditioner C improves by a factor between $10^{0.5}$ and 10. In Figure (a), we fix $\gamma = 2$ and find that increasing the contrast worsens cond$(CA)/$cond(A). Nevertheless, the improvement between $J = 3$ and $J = 5$ becomes more distinct. The error bars correspond to $\hat{\sigma}$ where $\hat{\sigma}^2$ is the estimated variance. They indicate that C is not just good on average, but good with high probability.*

On the other hand, Figure 4.4(a) shows that as the contrast increases, the preconditioner C degrades in performance, but the improvement between $J = 2\gamma - 1$ and $2\gamma + 1$ becomes more pronounced.

The error bars in Figure 4.4 are not error margins but $\hat{\sigma}$ where $\hat{\sigma}^2$ is the unbiased estimator of the variance. They indicate that cond$(CA)/$cond(A) is tightly concentrated around its mean, provided J is not too much larger than is necessary. For instance, for $\gamma = 1$, $J = 3$ already works well but pushing to $J = 9$ leads to greater uncertainty.

Next, we consider *forward probing* of A using the "Chebyshev on a disk" expansion or Equation (3.7). Let m be the order correction, that is we multiply $g_k(\xi)$ by $\langle\xi\rangle^m = \|\xi\|^m$. Let C be the output of the probing and K be the number of Chebyshev polynomials used.

Fix $n = 55^2$, $T = 10$, $\gamma = 2$ and $J = 5$. For $m = 0$ and $K = 3$, i.e., no order correction and using up to quadratic polynomials in ξ, we obtain a relative error $\|C - A\|/\|A\|$ that is less than 10^{-14}. On the other hand, using Fourier expansion, with $K = 5$ in the sense that $-\frac{K-1}{2} \leq k_1, k_2 \leq \frac{K-1}{2}$, the relative error is on the order of 10^{-1}. The point is that in this case, A has an exact "Chebyshev on a disk" representation and probing using the correct $\mathcal{B}$ enables us to retrieve the coefficients with negligible errors.

Finally, we consider backward probing with the Chebyshev expansion. We use $J = 5$, $\gamma = 2$ and $T = 10$. Figure 4.5 shows that when $m = -2$, the condition numbers $\lambda(\mathcal{B}_A)$ and $\kappa(\mathcal{B}_A)$ are minimized and hardly increases with n. This emphasizes the importance of knowing the order of the operator being probed.

4.4. Foveation. In this section, we forward-probe for the foveation operator, a space-variant imaging operator [4], which is particularly interesting as a model for human vision.

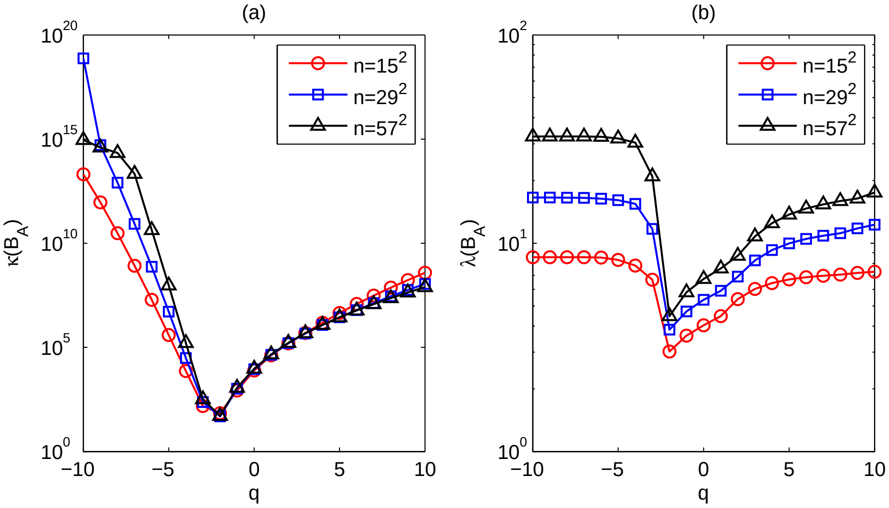

Figure 4.5. *Consider the backward probing of A in Equation (4.3), a pseudodifferential oeprator of order 2. Perform order correction by multiplying $g_k(\xi)$ by $\langle\xi\rangle^q$ in the expansion of the symbol. See Section 3.5. Observe that at $q = -2$, the condition numbers $\lambda(\mathcal{B}_A)$ and $\kappa(\mathcal{B}_A)$ are minimized and hardly grow with n.*

Formally, we may treat the foveation operator A as a Gaussian blur with a width or standard deviation that varies over space, that is

$$Au(x) = \int_{\mathbb{R}^2} K(x,y)u(y)dy \text{ where } K(x,y) = \frac{1}{w(x)\sqrt{2\pi}} \exp\left(\frac{-\|x-y\|^2}{2w^2(x)}\right), \tag{4.4}$$

where $w(x)$ is the width function which returns only positive real numbers.

The resolution of the output image is highest at the point where $w(x)$ is minimal. Call this point x_0. It is the point of fixation, corresponding to the center of the fovea. For our experiment, the width function takes the form of $w(x) = (\alpha\|x - x_0\|^2 + \beta)^{1/2}$. Our images are 201×201 and treated as functions on the unit square. We choose $x_0 = (0.5, 0.5)$ and $\alpha, \beta > 0$ such that $w(x_0) = 0.003$ and $w(1,1) = 0.012$.

The symbol of A is $a(x,\xi) = \exp(-2\pi^2 w(x)^2\|\xi\|^2)$, and we choose to use a Fourier series or Equation (3.5) for expanding it. Let C be the output of matrix probing and z be a standard test image. Figure 4.6(c) shows that the relative ℓ^2 error $\|Cz - Az\|_{\ell^2}/\|Az\|_{\ell^2}$ decreases exponentially as p increases. In general, forward probing yields great results like this because we know its symbol well and can choose an appropriate $\mathcal{B}$.

4.5. Inverting the wave equation Hessian. In seismology, it is common to recover the model parameters m, which describe the subsurface, by minimizing the least squares misfit between the observed data and $F(m)$ where F, the forward model, predicts data from m.

Methods to solve this problem can be broadly categorized into two classes: steepest descent or Newton's method. The former takes more iterations to converge but each iteration is computationally cheaper. The latter requires the inversion of the Hessian of the objective function, but achieves quadratic convergence near the optimal point.

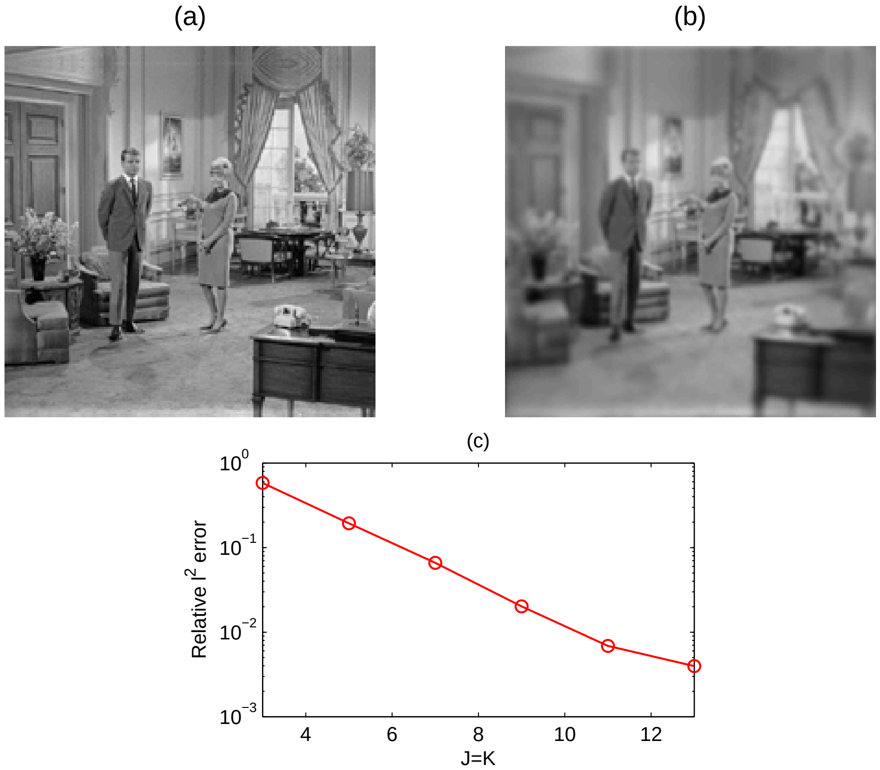

Figure 4.6. *Let A be the foveation operator in Equation (4.4) and C be the output of the forward probing of A. Figure (a) is the test image z. Figure (b) is Cz and it shows that C behaves like the foveation operator as expected. Figure (c) shows that the relative ℓ^2 error (see text) decreases exponentially with the number of parameters $p = J^4$.*

In another paper, we use matrix probing to precondition the inversion of the Hessian. Removing the nullspace component from the noise vector is more tricky (see Algorithm 1) and involves checking whether "a curvelet is visible to any receiver" via raytracing. For details on this more elaborate application, please refer to [6].

5. Conclusion.

When a matrix A with n columns belongs to a specified p-dimensional subspace, say $A = \sum_{i=1}^{p} c_i B_i$, we can probe it with a few random vectors to recover the coefficient vector c.

Let q be the number of random vectors used, κ be the condition number of the Gram matrix of $B_1, \dots, B_p$ and λ be the "weak condition number" of each B_i (cf. Definition 1.2) which is related to the numerical rank. From Theorem 1.3 and Section 1.3, we learn that when $nq \propto p(\kappa\lambda \log p)^2$, then the linear system that has to be solved to recover c (cf. Equation (1.1)) will be well-conditioned with high probability.

The same technique can be used to compute an approximate A^{-1}, or a preconditioner for inverting A. In [6], we used it to invert the wave equation Hessian — here we demonstrate that it can also be used to invert elliptic operators in smooth media (cf. Sections 4.2 and 4.3).

Some possible future work include the following.

1. Consider operators that have a sparse representation in a much larger $\mathcal{B}$ and try to recover the coefficients by compressed sensing. This is similar in spirit to "exact matrix completion" by Candès and Recht [3]. Hopefully, fewer random vectors will be needed.
2. Build a framework for probing $f(A)$ interpreted as a Cauchy integral

$$f(A) = \frac{1}{2\pi i} \oint_{\Gamma} f(z)(zI - A)^{-1} dz,$$

where Γ is a closed curve enclosing the eigenvalues of A. For more on approximating matrix functions, see [10, 12].
3. Consider expansion schemes for symbols that highly oscillate or have singularities that are well-understood.

Appendix A. Linear Algebra.

Recall the definitions of $\kappa(\mathcal{B})$ and $\lambda(\mathcal{B})$ at the beginning of the paper. The following concerns probing with multiple vectors (cf. Section 1.3).

Proposition A.1. *Let $I_q \in \mathbb{C}^{q\times q}$ be the identity. Let $\mathcal{B} = \{B_1, \dots, B_p\}$. Let $B'_j = I_q \otimes B_j$ and $\mathcal{B}' = \{B'_1, \dots, B'_p\}$. Then $\kappa(\mathcal{B}) = \kappa(\mathcal{B}')$ and $\lambda(\mathcal{B}) = \lambda(\mathcal{B}')$.*

Proof. Define $N \in \mathbb{C}^{p\times p}$ such that $N_{jk} = \langle B_j, B_k\rangle$. Define $N' \in \mathbb{C}^{p\times p}$ such that $N'_{jk} = \left\langle B'_j, B'_k \right\rangle$. Clearly, $N' = qN$, so their condition numbers are the same and $\kappa(\mathcal{B}) = \kappa(\mathcal{B}')$.

For any $A = B_j \in \mathbb{C}^{m\times n}$ and $A' = B'_j$, we have $\frac{\|A'\|(nq)^{1/2}}{\|A'\|_F} = \frac{\|A\|(nq)^{1/2}}{\|A\|_F q^{1/2}} = \frac{\|A\| n^{1/2}}{\|A\|_F}$. Hence, $\lambda(\mathcal{B}) = \lambda(\mathcal{B}')$. ■

Appendix B. Probabilistic Tools. In this section, we present some probabilistic results used in our proofs. The first theorem is used to decouple homogeneous Rademacher chaos of order 2 and can be found in [5, 18] for example.

Theorem B.1. *Let (u_i) and $(\tilde{u}_i)$ be two independent sequences of real-valued random variables and A_{ij} be in a Banach space where $1 \le i, j \le n$. There exists universal constants $C_1, C_2 > 0$ such that for any $s \ge 1$,*

$$\left(\mathbb{E}\left\| \sum_{1\le i\ne j\le n} u_i u_j A_{ij} \right\|^s\right)^{1/s} \le C_1 C_2^{1/s} \left(\mathbb{E}\left\| \sum_{1\le i\ne j\le n} u_i \tilde{u}_j A_{ij} \right\|^s\right)^{1/s}. \tag{B.1}$$

A homogeneous *Gaussian* chaos is one that involves only products of Hermite polynomials with the same total degree. For instance, a homogeneous Gaussian chaos of order 2 takes the form $\sum_{1\le i\ne j\le n} g_i g_j A_{ij} + \sum_{i=1}^n (g_i^2 - 1) A_{ii}$. It can be decoupled according to Arcones and Giné [2].

Theorem B.2. *Let (u_i) and $(\tilde{u}_i)$ be two independent Gaussian sequences and A_{ij} be in a Banach space where $1 \le i, j \le n$. There exists universal constants $C_1, C_2 > 0$ such that for any $s \ge 1$,*

$$\left(\mathbb{E}\left\| \sum_{1\le i\ne j\le n} u_i u_j A_{ij} + \sum_{i=1}^{n} (u_i^2 - 1) A_{ii} \right\|^s\right)^{1/s} \le C_1 C_2^{1/s} \left(\mathbb{E}\left\| \sum_{1\le i, j\le s} u_i \tilde{u}_j A_{ij} \right\|^s\right)^{1/s}.$$

Remark B.1. *For Rademacher chaos, $C_1 = 4$ and $C_2 = 1$. For Gaussian chaos, it is clear from [2] that we can pick $C_1 = 2^{1/2}$ and $C_2 = 2^{14}$. Better constants may be available.*

We now proceed to the Khintchine inequalties. Let $\|\cdot\|$ and $\|\cdot\|_{C_s}$ denote the s-Schatten norm. Recall that $\|A\|_{C_s} = (\sum_i |\sigma_i|^s)^{1/s}$ where σ_i is a singular value of A. The following is due to Lust-Piquard and Pisier [15, 16].

Theorem B.3. *Let $s \geq 2$ and (u_i) be a Rademacher or Gaussian sequence. Then for any set of matrices $\{A_i\}_{1\leq i\leq n}$,*

$$\left(\mathbb{E}\left\|\sum_{i=1}^n u_i A_i\right\|_{C_s}^s\right)^{1/s} \leq s^{1/2} \max\left(\left\|(\sum_{i=1}^n A_i^* A_i)^{1/2}\right\|_{C_s}, \left\|(\sum_{i=1}^n A_i A_i^*)^{1/2}\right\|_{C_s}\right).$$

In [18], Theorem B.3 is applied twice in a clever way to obtain a Khintchine inequality for a decoupled chaos of order 2.

Theorem B.4. *Let $s \geq 2$ and (u_i) and $(\tilde{u}_i)$ be two independent Rademacher or Gaussian sequences. For any set of matrices $\{A_{ij}\}_{1\leq i,j\leq n}$,*

$$\left(\mathbb{E}\left\|\sum_{1\leq i,j\leq n} u_i \tilde{u}_j A_{ij}\right\|_{C_s}^s\right)^{1/s} \leq 2^{1/s} s \max(\|Q^{1/2}\|_{C_s}, \|R^{1/2}\|_{C_s}, \|F\|_{C_s})$$

where $Q = \sum_{1\leq i,j\leq n} A_{ij}^* A_{ij}$ and $R = \sum_{1\leq i,j\leq n} A_{ij} A_{ij}^*$ and F is the block matrix $(A_{ij})_{1\leq i,j\leq n}$.

For Rademacher and Gaussian chaos, higher moments are controlled by lower moments, a property known as "hypercontractivity" [2, 5]. This leads to exponential tail bounds by Markov's inequality as we illustrate below.

Proposition B.5. *Let X be a nonnegative random variable. Let $\sigma, c, \alpha > 0$. Suppose $(\mathbb{E}X^s)^{1/s} \leq \sigma c^{1/s} s^{1/\alpha}$ for all $s_0 \leq s < \infty$. Then for any $k > 0$ and $u \geq s_0^{1/\alpha}$,*

$$\mathbb{P}\left(X \geq e^k \sigma u\right) \leq c \exp(-k u^\alpha).$$

Proof. By Markov's inequality, for any $s > 0$,

$$\mathbb{P}\left(X \geq e^k \sigma u\right) \leq \frac{\mathbb{E}X^s}{(e^k \sigma u)^s} \leq c\left(\frac{\sigma s^{1/\alpha}}{e^k \sigma u}\right)^s.$$

Pick $s = u^\alpha \geq s_0$ to complete the proof. ■

Proposition B.6. *Let (u_i) be a Rademacher or Gaussian sequence and C_1, C_2 be constants obtained from Theorem B.1 or B.2. Let $\{A_{ij}\}_{1\leq i,j\leq n}$ be a set of p by p matrices, and assume that the diagonal entries A_{ii} are positive semidefinite. Define $M = \sum_i u_i u_j A_{ij}$ and $\sigma = C_1 \max(\|Q\|^{1/2}, \|R\|^{1/2}, \|F\|)$ where Q, R, F are as defined in Theorem B.4. Then*

$$\mathbb{P}\left(\|M - \mathbb{E}M\| \geq e\sigma u\right) \leq (2C_2 np) \exp(-u).$$

Proof. We will prove the Gaussian case first. Recall that the s-Schatten and spectral norms are equivalent: for any $A \in \mathbb{C}^{r\times r}$, $\|A\| \leq \|A\|_{C_s} \leq r^{1/s}\|A\|$. Apply the decoupling inequality, that is Theorem B.2, and deduce that for any $s \geq 2$,

$$(\mathbb{E}\,\|M-N\|^s)^{1/s} \leq C_1 C_2^{1/s}\left(\mathbb{E}\left\|\sum_{1\leq i,j\leq n} u_i\tilde{u}_j A_{ij}\right\|_{C_s}^s\right)^{1/s}.$$

Invoke Khintchine's inequality, that is Theorem B.4, and obtain

$$\begin{aligned}(\mathbb{E}\,\|M-N\|^s)^{1/s} &\leq C_1(2C_2)^{1/s} s\max(\|Q^{1/2}\|_{C_s}, \|R^{1/2}\|_{C_s}, \|F\|_{C_s})\\ &\leq C_1(2C_2np)^{1/s} s\max(\|Q\|^{1/2}, \|R\|^{1/2}, \|F\|)\\ &\leq \sigma(2C_2np)^{1/s}s.\end{aligned}$$

Apply Proposition B.5 with $c = 2C_2np$ and $k = \alpha = 1$ to complete the proof for the Gaussian case. For the Rademacher case, we take similar steps. First, decouple $(\mathbb{E}\|M-N\|^s)^{1/s}$ using Theorem B.1. This leaves us a sum that excludes the A_{ii}'s. Apply Khintchine's inequality with the A_{ii}'s zeroed. Of course, Q, R, F in Proposition B.4 will not contain any A_{ii}'s, but this does not matter because $A_{ii}^* A_{ii}$ and $A_{ii}A_{ii}^*$ and A_{ii} are all positive semidefinite for any $1 \leq i \leq n$ and we can add them back. For example, $\|(A_{ij})_{1\leq i\neq j\leq n}\| \leq \|(A_{ij})_{1\leq i,j\leq n}\|$ as block matrices. ■

REFERENCES

[1] C. Andrieu, N. De Freitas, A. Doucet, and M.I. Jordan, *An introduction to MCMC for machine learning*, Machine learning, 50 (2003), pp. 5–43.

[2] M.A. Arcones and E. Giné, *On decoupling, series expansions, and tail behavior of chaos processes*, Journal of Theoretical Probability, 6 (1993), pp. 101–122.

[3] E.J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9 (2009), pp. 717–772.

[4] Ee-Chien Chang, Stphane Mallat, and Chee Yap, *Wavelet foveation*, Applied and Computational Harmonic Analysis, 9 (2000), pp. 312 – 335.

[5] V. De la Peña and E. Giné, *Decoupling: from dependence to independence*, Springer Verlag, 1999.

[6] L. Demanet, P.D. Létourneau, N. Boumal, H. Calandra, J. Chiu, and S. Snelson, *Matrix probing: a randomized preconditioner for the wave-equation Hessian*, (preprint).

[7] L. Demanet and L. Ying, *Discrete symbol calculus*, SIAM Review, (preprint).

[8] A. Edelman, *Eigenvalues and condition numbers of random matrices*, PhD thesis, Massachusetts Institute of Technology, 1989.

[9] G.B. Folland, *Introduction to partial differential equations*, Princeton Univ Pr, 1995.

[10] Nicholas Hale, Nicholas J. Higham, and Lloyd N. Trefethen, *Computing $a^\alpha, \log(a)$, and related matrix functions by contour integrals*, SIAM Journal on Numerical Analysis, 46 (2008), pp. 2505–2523.

[11] N. Halko, P.G. Martinsson, and J.A. Tropp, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, (preprint).

[12] N.J. Higham, *Functions of matrices: theory and computation*, Society for Industrial Mathematics, 2008.

[13] D.R. Karger and C. Stein, *A new approach to the minimum cut problem*, Journal of the ACM (JACM), 43 (1996), pp. 601–640.

[14] R.M. Karp, *An introduction to randomized algorithms*, Discrete Applied Mathematics, 34 (1991), pp. 165–201.

[15] F. Lust-Piquard, *Inégalités de Khintchine dans Cp*, CR Acad. Sci. Paris, 303 (1986), pp. 289–292.

[16] F. Lust-Piquard and G. Pisier, *Non commutative Khintchine and Paley inequalities*, Arkiv för Matematik, 29 (1991), pp. 241–260.

[17] R. Motwani and P. Raghavan, *Randomized algorithms*, ACM Computing Surveys (CSUR), 28 (1996), pp. 33–37.

[18] H. Rauhut, *Circulant and Toeplitz matrices in compressed sensing*, Proc. SPARS, 9 (2009).

[19] M.A. Shubin, *Pseudodifferential operators and spectral theory*, Springer Verlag, 2001.

[20] F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert, *A fast randomized algorithm for the approximation of matrices*, Applied and Computational Harmonic Analysis, 25 (2008), pp. 335–366.