Linear dimension reduction approximately preserving a function of the 1-norm

Michael P. Casey

U. S. Air Force Research Laboratory, e-mail: mpcasey@alumni.duke.edu

Abstract: For any finite point set in D-dimensional space equipped with the 1-norm, we present random linear embeddings to k-dimensional space, with a new metric, having the following properties. For any pair of points from the point set that are not too close, the distance between their images is a strictly concave increasing function of their original distance, up to multiplicative error. The target dimension k need only be quadratic in the logarithm of the size of the point set to ensure the result holds with high probability. The linear embeddings are random matrices composed of standard Cauchy random variables, and the proofs rely on Chernoff bounds for sums of iid random variables. The new metric is translation invariant, but is not induced by a norm.

1. Introduction

The Johnson-Lindenstrauss lemma [8] states that for a finite set of points $P \subset \mathbb{R}^D$ and $0 < \epsilon < 1$, there are random linear maps $F : \mathbb{R}^D \to \mathbb{R}^k$ satisfying, for any $x, y \in P$,

$$(1 - \epsilon) \|x - y\|_2 \leq \|F(x) - F(y)\|_2 \leq (1 + \epsilon) \|x - y\|_2$$

with high probability, provided $k = \Theta(\epsilon^{-2} \ln |P|)$. It is sufficient to draw the entries of F i.i.d. sub-Gaussian [13]. These random linear projections have provided improved worst case performance bounds for many problems in theoretical computer science, machine learning, and numerical linear algebra. Ailon and Chazelle [1] show how F may be computed quickly and apply it to the approximate nearest-neighbor problem, working on the projected points $F(P)$. Vempala [19] gives a review of problems that may be reduced to analyzing a set of points $P \subset \mathbb{R}^D$, so that after the random projection $F : \mathbb{R}^D \to \mathbb{R}^k$ is applied, the recovery of approximate solutions is possible with time and space bounds depending on k, the target dimension, instead of D, the ambient dimension.

In numerical linear algebra, Drineas et al. [5] use the lemma to approximate the leverage scores of a given matrix A; such scores are used to inform subsampling schemes for A, resulting in sketches $\tilde{A}$ of smaller dimensions that preserve desired properties of A. Drineas and Mahoney [6] give a further review of using randomness in numerical linear algebra.

The Johnson-Lindenstrauss lemma is a metric embedding result; the map F sends the finite metric space $P \subset \mathbb{R}^D$ induced by the 2-norm to a corresponding metric space $F(P) \subset \mathbb{R}^k$, also induced by the 2-norm, such that distances are preserved well. Ailon and Chazelle [1] also show that equipping the target space $\mathbb{R}^k$ with the 1-norm is also possible; the target dimension is still proportional to $\ln|P|$, but the dependence on ϵ may be a bit worse. However, analogous results using the 1-norm on the domain do not hold. For example, in [2] and [10], specific N-point subsets of $\mathbb{R}^D$ equipped with the 1-norm are shown to embed only in $\mathbb{R}^k$ with $k = N^{1/c^2}$ if one requires

$$\|x - y\|_1 \leq \|F(x) - F(y)\|_1 \leq c\,\|x - y\|_1\,.$$

In particular, Brinkman and Charikar [2] show the target dimension k must be at least $N^{1/2 - O(\epsilon \ln(1/\epsilon))}$ if one wants $c = 1 + \epsilon$.

In light of these negative results, people have tried estimating $\|x - y\|_1$ from the coordinates of $F(x) - F(y)$. When the entries of F are i.i.d. standard Cauchy random variables, the coordinates are distributed i.i.d. like $\|x - y\|_1 X$ with $X \sim \text{Cauchy}(1)$. The median of $\|x - y\|_1 |X|$ is $\|x - y\|_1$, so estimating the median from the coordinates of $F(x) - F(y)$ would estimate the distance this way. Indyk [7] considers the sample median as an estimator, while Li, Hastie, and Church [12] consider 1-homogeneous functions of these coordinates for estimators. None of the estimators considered are metrics on $\mathbb{R}^k$. For nearest neighbor methods, we should like to have a metric on the target space $\mathbb{R}^k$ and prefer a low number of coordinates for each point.

Relaxing the problem as follows, we wish to find linear maps $F : \mathbb{R}^D \to \mathbb{R}^k$ satisfying, for any $x, y \in P$,

$$(1 - \epsilon)\mu(\|x - y\|_1) \leq \rho(F(x), F(y)) \leq (1 + \epsilon)\mu(\|x - y\|_1)$$

with high probability. We have changed the metric on $\mathbb{R}^k$ to ρ instead of the one induced by the 1-norm, and we have introduced a nonlinear function μ in place of the identity function. We want $k = \Theta(\epsilon^{-2} \ln^c |P|)$, with $c < 4$ or better.

Here, $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ is a concave increasing function with $\mu(0) = 0$. Such μ are called "metric preserving" by Corazza [4], for the following reason:

$$\mu(\|x - y\|_1) \leq \mu(\|x - z\|_1) + \mu(\|z - y\|_1) \quad \text{for any} \quad x, y, z \in \mathbb{R}^D,$$

that is, they admit a new metric on the space that is "compatible" with the old one. In particular, spheres for the new metric about a particular point $y \in \mathbb{R}^D$, that is, the level sets $\{x \in \mathbb{R}^D \mid \mu \circ \|x - y\|_1 = t\}$, look like scaled versions of spheres for the 1-norm (crosspolytopes) about that point; the scaling however is nonlinear. The 1-norm is used here as an example, but any other input metric

will still satisfy the triangle inequality under such μ. Not all metric preserving functions are concave increasing, but such a choice ensures the new metric generates the same topology as the old one.

For us, the linear map $F : \mathbb{R}^D \to \mathbb{R}^k$ will have entries $F_{ij} \overset{\text{i.i.d.}}{\sim}$ Cauchy (1), and we introduce the metric ρ on $\mathbb{R}^k$ using an auxiliary function ξ:

$$\rho(x, y) := \frac{1}{k} \sum_{i=1}^{k} \xi(|x_i - y_i|)$$

with

$$\xi(\lambda) := \ln(1 + \sqrt{\lambda}) + \frac{1}{2} \ln(1 + \lambda) \quad \text{and} \quad \mu(\lambda) := \mathbb{E}\xi(\lambda F_{11})$$

for $\lambda > 0$. Our main theorem has several regimes depending on how big $\|x - y\|_1$ can be. (See theorems 3.0.1, 3.0.3, and 3.0.9.) However, the primary result is as follows.

Theorem 1.0.1. *Let F, ρ, and μ be as above. Given N points $P \subset \mathbb{R}^D$ and $\epsilon \in (0, 1)$,*

$$\mu\left(\frac{\|x - y\|_1}{1 + \epsilon}\right) \leq \rho(F(x), F(y)) \leq \mu((1 + \epsilon) \|x - y\|_1)$$

for all $x, y \in P$ with $\|x - y\|_1 \geq \sqrt{1 + \epsilon}$, provided

$$k = \frac{C}{\epsilon^2(1 - \epsilon)^2} \ln N.$$

Independent of its interest as an analog of the Johnson-Lindenstrauss lemma, theorem 1.0.1 also contributes to the study of p-stable projections. In fact, we make the following conjecture for $1 < p < 2$ upon replacing the entries F_{ij} of F by i.i.d. standard p-stable random variables and setting $\mu(\lambda) = \mathbb{E}\xi(\lambda F_{11})$. Just like the 1.0.1, the conjecture could have several parts based on how large $\|x - y\|_p$ is, but the primary conjecture is as follows.

Conjecture 1.0.2. *With F and μ modified as above, and ρ, ϵ, and k as in theorem 1.0.1, the following bound holds*

$$\mu\left(\frac{\|x - y\|_p}{1 + \epsilon}\right) \leq \rho(F(x), F(y)) \leq \mu\big((1 + \epsilon) \|x - y\|_p\big)$$

for all $x, y \in P$ with $\|x - y\|_p = \Omega(1)$.

The setup for the proof would be the same as for theorem 1.0.1, relying on 1st and 2nd moment estimates for $\xi(\lambda |W|)$; however, because the density for a p-stable random variable W is only implicitly defined, the needed 1st and 2nd moment estimates are not so straightforward, but could be empirically found on the computer using methods such as [3] to draw the p-stable random variables. This approach, in which we directly project the points from $\mathbb{R}^D$, may be contrasted to embedding $\ell_p^D \hookrightarrow \ell_1^n$ and applying theorem 1.0.1 there. Pisier [17] (see also [15, chapter 8] and [9, chapter 9]) shows that such embeddings exist with distortion $(1 + \epsilon)$, with n proportional to D and depending on p and ϵ.

2. Overview of the proof

In this section, we explain the choices for the function ξ and the metric ρ, as well as the use of Cauchy random variables, outlining the proof along the way.

Consider a point $v \in \mathbb{R}^D$. The 1-stability of the Cauchy distribution dictates that the coordinates of the projected point $F(v)$ are Cauchy distributed: $F(v)_j \sim \|v\|_1 X_j$ with $X_j \stackrel{\text{i.i.d.}}{\sim}$ Cauchy (1). The metric ρ is then an empirical mean:

$$\rho(F(v), 0) = \frac{1}{k} \sum_{j=1}^{k} \xi(\|v\|_1 X_j),$$

and if we marginalize out the Cauchy dependence, we recover the deterministic function μ of $\|v\|_1$:

$$\mathbb{E}\rho(F(v), 0) = \mathbb{E}\xi(\|v\|_1 X) =: \mu(\|v\|_1) \quad \text{for} \quad X \sim \text{Cauchy}(1).$$

We can now outline the proof as follows: let $x - y = v \in \mathbb{R}^D$. The projection map $F : \mathbb{R}^D \to \mathbb{R}^k$ is linear and the metric ρ is translation invariant, so our goal is to show $\mu(\|v\|_1) \approx \rho(F(v), 0)$ or upon setting $\|v\|_1 = \lambda$,

$$\mu(\lambda) \approx \frac{1}{k} \sum_{j=1}^{k} \xi(\lambda X_j)$$

with high probability. As usual, we use the exponential Markov inequality and the i.i.d. assumption to estimate

$$\mathbb{P}\left\{ \frac{1}{k} \sum_{j=1}^{k} \xi(\lambda |X_j|) - \mu(\lambda) > t \right\} \leq \left(e^{-st} \mathbb{E} e^{s\left(\xi(\lambda |X|) - \mu(\lambda)\right)} \right)^k$$

with a similar setup for the lower tail. However, Cauchy random variables X only have finite *fractional* moments,

$$\mathbb{E}|X|^b < \infty \quad \text{only for} \quad |b| < 1,$$

so the presence of $\xi(\lambda |X|)$ in the exponential requires $\xi(\lambda) = c\ln(o(\lambda))$ when λ is large. Our choice of ξ ensures this behavior:

$$\xi(\lambda) = \ln(1 + \sqrt{\lambda}) + \frac{1}{2}\ln(1 + \lambda) \leq 2\ln(1 + \sqrt{\lambda}),$$

while the presence of the "1+" in the logarithms ensures ξ is nonnegative, increasing, and sends 0 to 0. The function ξ is thus subadditive and preserves the triangle inequality:

$$\xi(|x_i - y_i|) \leq \xi(|x_i - z_i|) + \xi(|z_i - y_i|),$$

ensuring ρ is a metric on $\mathbb{R}^k$. Because μ is the expectation of ξ, it inherits these properties, so that $\mu \circ \|\|_1$ induces a metric on the original space $\mathbb{R}^D$.

We show in sections 4 and 5 that our tail bounds take the following form: To be concrete, here is the upper tail case, but the other lower tail cases are similar

$$\min_{s} e^{-s\Delta} \mathbb{E} e^{s\left(\xi(\lambda|X|) - \mu(\lambda)\right)} \le \exp\left(-\frac{\Delta^2}{4(V^2 + A)}\right) \qquad (\diamond)$$

with Δ depending on $\mu(\lambda)$, the function V^2 giving an upper bound for the 2nd moment or the variance of $\xi(\lambda|X|)$, and the auxiliary function $A(\lambda)$, derived from tail estimates for $\xi(\lambda|X|)$. The particular form of ξ was chosen to give explicit control over all these quantities as λ varies, allowing us to obtain bounds on equation $(\diamond)$ that only weakly depend on λ.

We arrive at the particular form $(\diamond)$ for the tail bounds by estimating the moment generating function as follows, taking the upper tail as an example: with $Y = \xi(\lambda|X|) - \mu(\lambda)$, we split $\mathbb{E}\exp(sY)$ into two terms and desire each to be bounded by something quadratic in s: for the 1st term, using a 2nd order Taylor expansion for the exponential,

$$\mathbb{E}\exp(sY)\mathbb{I}\{sY \le 1\} \le 1 + s^2 \mathbb{E} Y^2 \le 1 + s^2 V^2$$

while for the 2nd term, we use integration by parts, eventually showing

$$\mathbb{E}\exp(sY)\mathbb{I}\{sY > 1\} = e\mathbb{P}\{Y > 1/s\} + \int_1^\infty e^t \mathbb{P}\{Y > t/s\}\, dt \le s^2 A(\lambda).$$

We can show the integrand decays exponentially in t using our choice of ξ and the explicit density for the Cauchy distribution:

$$\begin{aligned}
\mathbb{P}\{Y > t/s\} &= \mathbb{P}\{\xi(\lambda|X|) > \mu(\lambda) + t/s\} \\
&\le \mathbb{P}\left\{2\ln(1 + \sqrt{\lambda|X|}) > \mu(\lambda) + t/s\right\} \le C(\lambda)e^{-t/s}
\end{aligned}$$

with C depending on λ and $\mu(\lambda)$. We can then combine these estimates and optimize in s:

$$\min_{s} e^{-s\Delta} \mathbb{E}\exp(sY) \le \min_{s} e^{-s\Delta}\exp(s^2 V^2 + s^2 A(\lambda)) = \exp\left(\frac{-\Delta^2}{4(V^2 + A(\lambda))}\right),$$

using $s = \Delta/(2(V^2 + A(\lambda)))$.

The tail probabilities now have the form

$$\mathbb{P}\left\{\left|\frac{1}{k}\sum_{j=1}^{k}\xi(\lambda|X_j|) - \mu(\lambda)\right| > \Delta\right\} \le 2\exp\left(-k\frac{\Delta^2}{4(V^2 + A(\lambda))}\right)$$

for a single λ corresponding to a single vector $v = x - y \in \mathbb{R}^D$. There are at most $\binom{N}{2}$ pairs of points from P, so we would want to choose the target dimension as

$$k = (c + 2)\ln(N)\frac{4(V^2 + A(\lambda))}{\Delta^2}$$

to ensure with probability at least $1 - N^{-c}$,

$$\left| \frac{1}{k} \sum_{j=1}^{k} \xi(\lambda \, |X_j|) - \mu(\lambda) \right| \leq \Delta$$

for all pairs of points simultaneously. However, the error Δ and the target dimension k both depend on λ, so we require uniform estimates for these quantities. We find these by breaking up the possible values for λ into three regimes: small, medium, and big
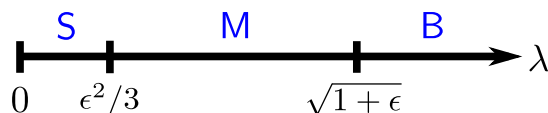
Our choice of ξ provides an explicit function for $\mu := \mathbb{E}\xi(\lambda \, |X|)$, (lemma A.1.1)

$$\mu(\lambda) := \frac{1}{2} \ln(1 + \lambda^2) + \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1 + \lambda} \right) \quad \text{with} \quad \operatorname{atanh}(x) := \sum_{j=0}^{\infty} \frac{x^{2j+1}}{2j+1}$$

for $|x| < 1$. The big regime has μ behaving like the log term, while the medium and small regimes have it behaving like $\Theta(\sqrt{2\lambda})$. The choice of ξ also gives us a bound on the variance (corollary A.3.2)

$$\operatorname{Var}(\xi(\lambda \, |X|)) \leq \min\left\{ \frac{\pi^2}{2}, 2\mathbb{E}\ln(1 + \lambda \, |X|) \right\}.$$

The constant bound, independent of λ, is used for the big regime, while the expectation bound provides finer control on the variance when λ is small, via another explicit function, lemma A.3.3, of λ.

For the big regime, taking Δ as

$$\mu((1 + \epsilon)\lambda) - \mu(\lambda) \quad \text{and} \quad \mu(\lambda) - \mu((1 + \epsilon)^{-1}\lambda),$$

both of which are bounded by lemma A.2.1, together with the constant bound for the variance give theorem 3.0.1, as $A(\lambda)$ is bounded here.

For the medium and small regimes, we take $\Delta = \epsilon\mu(\lambda)$ and use corollary A.3.4 to bound $V^2/\mu(\lambda)^2$. The split between medium and small regimes occurs because of the $\ln(\lambda)$ term in that ratio: the target dimension has $\Delta^2 \sim \lambda$ on the bottom, while the upper tail bound 4.0.1 required s^* to only have Δ on top

$$s^* = \Delta/(2(V^2 + A(\lambda))) < 1/2.$$

This mismatch in powers of Δ forces us to choose a cutoff λ; because $A(\lambda)$ (and V^2) have terms proportional to λ, the above inequality can only hold for λ not too small. This gives the $\ln(\epsilon^2/3)$ term in theorem 3.0.3 for the medium regime.

For the small regime, there is no such restriction on s for the lower tail bound 5.0.2, but the target dimension still grows like $\ln(1/\lambda)$ as λ decreases

(See lemma 3.0.5.). We stop that growth by fixing a particular λ_0, showing that for all smaller λ, the $(1 - \epsilon)$ error has a suitable replacement in theorem 3.0.9. The key is lemma 3.0.7: we choose λ_0 so that $\lambda_0 \max_i |X_i| < 1/6$ with high probability, making *both* $\xi(\lambda |X_i|)$ and $\mu(\lambda)$ behave like $\Theta(\sqrt{\lambda}) = \sqrt{\eta}\Theta(\sqrt{\lambda_0})$ for $\lambda = \eta\lambda_0$ with $\eta \in (0, 1)$. Because λ_0 turns out to be $\Theta(1/(N^{c+2}k))$, the $-\ln(\lambda_0)$ in the target dimension forces $k = \Theta(\epsilon^{-2}\ln^2(N^{c+2}))$, a quadratic dependence on $\ln(N)$.

We finish the proofs in the next section, while the upper and lower tail estimates are provided in sections 4 and 5. We collect the estimates on the 1st and 2nd moments in appendix A, and ancillary identities for those estimates in appendix B.

3. Finishing the proof

We now tie down the target dimension k. Recall P is a set of N points in $\mathbb{R}^D$, and $F : \mathbb{R}^D \to \mathbb{R}^k$ is a matrix of i.i.d. Cauchy (1) entries. In what follows, the estimates are not sharp.

Theorem 3.0.1 (Big Regime). *For $\epsilon \in (0, 1)$ and $\|x - y\|_1 \geq \sqrt{1 + \epsilon}$,*

$$\mu\left(\frac{\|x - y\|_1}{1 + \epsilon}\right) \leq \rho\big(F(x), F(y)\big) \leq \mu\big((1 + \epsilon)\|x - y\|_1\big)$$

for all $x, y \in P$ with probability at least $1 - N^{-c}$ provided

$$k \geq \frac{C}{\epsilon^2(1 - \epsilon)^2}\ln(N^{c+2}) \quad with \quad C = 64\left(\frac{\pi^2}{2} + \frac{16\sqrt{2}}{e\pi}e^{\operatorname{atanh} 1/\sqrt{2}}\right).$$

Remark 3.0.2. The constants are not expected to be sharp; C is computed so that k is uniformly bounded with respect to $\|x - y\|_1 \geq \sqrt{1 + \epsilon}$.

Proof. Let $\lambda = \|x - y\|_1$. We want to use the lower and upper tail estimates from lemmas 5.0.1 and 4.0.1, so it remains to verify

$$s^* = \frac{\Delta}{2(V^2 + A(\lambda))} \leq 1/2$$

with Δ either

$$\mu(\lambda) - \mu\big((1 + \epsilon)^{-1}\lambda\big) \quad \text{or} \quad \mu\big((1 + \epsilon)\lambda\big) - \mu(\lambda).$$

By lemma A.2.1, the differences Δ are at most ϵ for $\lambda \geq \sqrt{1 + \epsilon}$, while the upper bound for the variance of $\xi(\lambda |X|)$ is $V^2 = \pi^2/2$ by corollary A.3.2. Because $\epsilon < 1$, we then certainly have $s^* < 1/2$.

As explained in section 2, the target dimension k is chosen to ensure the union bound is at most N^{-c} for both tails combined. The choice of C comes from the lower bound for the Δ's from lemma A.2.1 and the larger of the two $A(\lambda)$ functions in lemmas 5.0.1 and 4.0.1. $\square$

Theorem 3.0.3 (Medium Regime). *For $\|x - y\|_1 \in [\epsilon^2/3, \sqrt{1+\epsilon}]$ and $\epsilon \in (0,1)$,*

$$(1 - \epsilon)\mu\big(\|x - y\|_1\big) \le \rho\big(F(x), F(y)\big) \le (1 + \epsilon)\mu\big(\|x - y\|_1\big)$$

for all $x, y \in P$ with probability at least $1 - N^{-c}$ provided

$$k = \frac{4\ln(N^{c+2})}{\epsilon^2}\left(C + \frac{8}{\pi}\big(1 - \ln(\epsilon^2/3)\big)\right) \quad \text{with } C \text{ bounded.}$$

Remark 3.0.4. We have not been able to establish an upper bound result

$$\rho(F(x), F(y)) \le (1 + \epsilon)\mu(\|x - y\|_1)$$

with high probability when $\|x - y\| < \epsilon^2/3$. Our proofs break down or require a much higher estimate for the target dimension k. We conjecture that $k = O(\ln^2(N^c)/\epsilon^2)$ still suffices, in light of theorem 3.0.9 for the small regime.

Proof. With $\lambda = \|x - y\|_1$, we take $\Delta = \epsilon\mu(\lambda)$. By lemma 3.0.5, the lower bound for $\rho(F(x), F(y))$ requires an initial estimate for the target dimension of $\tilde{k} = 2\ln(N^{c+2})\epsilon^{-2}C(1 - \ln(\lambda^*))$ with λ^* the smallest λ we wish to consider. The upper bound will force our choice of λ^*.

We now want to use the upper tail estimate from lemma 4.0.1. It remains to check

$$s^* = \frac{\epsilon\mu(\lambda)}{2(V^2 + A(\lambda))} \le 1/2,$$

and it suffices to show $\epsilon\mu(\lambda)/A(\lambda) \le 1$. With $b = 16e/(\pi(e-1)^2) \approx 4.7$ from $A(\lambda)$, we use lemma A.1.2 for the upper bound for $\mu(\lambda)$ to find, after some estimation,

$$\frac{\epsilon\mu(\lambda)}{A(\lambda)} \le \epsilon\frac{\sqrt{1+\lambda^2}}{b}\left(\frac{3}{\sqrt{2}}\frac{1}{(1+\lambda)\sqrt{\lambda}} + \frac{\lambda}{2}\right) \le \begin{cases} .7\epsilon & \text{if } \lambda \in [1, \sqrt{1+\epsilon}] \\ 1 & \text{if } \lambda \in [\epsilon^2/3, 1] \end{cases},$$

recalling $\epsilon \in (0,1)$.

Using the expression for $A(\lambda)$, we now have the following estimate for the target dimension. Because V^2 is an estimate on the variance now, we can remove the 1+'s from corollary A.3.4 to find

$$k = \frac{4\ln(N^{c+2})}{\epsilon^2}\left(\frac{9}{2}(\sqrt{1+\epsilon} + 4/\pi) + 2 + \frac{8}{\pi}\big(1 - \ln(\epsilon^2/3)\big) + \frac{b}{2}(1 + \sqrt{1+\epsilon})^2\right)$$

using $\mu(\lambda) \ge \sqrt{2\lambda}/(1 + \lambda)$ from remark A.1.3. The b dependent term here is enough to ensure $k \ge \tilde{k}$, so both sides of the inequality for $\rho(F(x), F(y))$ hold with high probability and this dimension k. □

The following two lemmas lead to theorem 3.0.9, which shows that a lower bound for $\rho(F(x), F(x))$ continues to hold for *all* $\|x - y\|_1 < 2$.

Lemma 3.0.5. *For $\epsilon \in (0,1)$ and $0 < \|x - y\|_1 \in [\lambda^*, 2)$,*

$$(1 - \epsilon)\mu\big(\|x - y\|_1\big) \leq \rho\big(F(x), F(y)\big)$$

for all $x, y \in P$ with probability at least $1 - N^{-c}$ provided

$$k = \frac{2\ln(N^{c+2})}{\epsilon^2}\left(C + \frac{8}{\pi}\big(1 - \ln(\lambda^*)\big)\right) \quad \text{with } C \text{ bounded.}$$

Remark 3.0.6. The estimates are not sharp.

Proof. With $\lambda = \|x - y\|_1$, we take $\Delta = \epsilon\mu(\lambda)$. Using corollary A.3.4 and the lower tail esimate from lemma 5.0.2, the target dimension is

$$k = \frac{2\ln(N^{c+2})}{\epsilon^2} \max\big\{1 + (9/2)(2 + 4/\pi),\, 3 + (8/\pi)\big(1 - \ln(\lambda^*)\big)\big\}$$

to ensure the bound holds with probability at least $1 - N^{-c}$ for all pairs of points. $\qquad\square$

Lemma 3.0.7. *For $1 \leq i \leq k$, let $X_i \overset{i.i.d.}{\sim} Cauchy(1)$. For $0 < \epsilon < 1$ and $0 < \lambda_0 \leq 1$, suppose*

$$(1 - \epsilon)\mu(\lambda_0) \leq \frac{1}{k}\sum_{i=1}^{k} \xi(\lambda_0\,|X_i|)$$

and $\lambda_0 \max_i |X_i| \leq c_0 \leq 1/6$.
   Then if $0 < \eta < 1$, the same X_i also satisfy

$$(1 - \epsilon')\mu(\eta\lambda_0) \leq \frac{1}{k}\sum_{i=1}^{k} \xi(\eta\lambda_0\,|X_i|)$$

with ϵ' depending on ϵ, c_0, and λ_0. If $\lambda_0 \leq \epsilon^2$, then we can have

$$1 - \epsilon' = (1 - \epsilon)\frac{1 - \epsilon^2/\sqrt{2}}{1 + 3\epsilon^2}.$$

Remark 3.0.8. Analogous upper bounds are also possible, with a similar proof.

Proof. A fourth order Taylor expansion with Lagrange remainder shows

$$\sqrt{a} \leq \xi(a) \leq \sqrt{a}(1 + a/2) \quad \text{for} \quad 0 \leq a \leq 1/6.$$

Because $\max_i |X_i| \leq c_0 \leq 1/6$ and $0 < \eta < 1$, we invoke the above inequality twice to find

$$\frac{\sqrt{\eta}}{1 + \lambda_0 c_0/2}\xi(\lambda_0\,|X_i|) \leq \frac{\sqrt{\eta}}{1 + \lambda_0\,|X_i|\,/2}\xi(\lambda_0\,|X_i|) \leq \xi(\eta\lambda_0\,|X_i|).$$

By assumption, summing over i and dividing by k yields

$$(1 - \epsilon)\frac{\sqrt{\eta}}{1 + \lambda_0 c_0/2}\mu(\lambda_0) \leq \frac{1}{k}\sum_{i=1}^{k}\xi(\eta\lambda_0 |X_i|)$$

We finish by using remark A.1.3 (twice) to "absorb" $\sqrt{\eta}$ into μ,

$$\sqrt{\eta}\mu(\lambda_0) \geq \frac{\sqrt{2\eta\lambda_0}}{1 + \lambda_0} \geq \frac{\mu(\eta\lambda_0)}{1 + \lambda_0}\frac{1 + \eta\lambda_0}{1 + \eta\lambda_0(1 + 1/\sqrt{2})} \geq \frac{\mu(\eta\lambda_0)}{1 + \lambda_0}(1 - \eta\lambda_0/\sqrt{2}) \quad \square$$

Theorem 3.0.9 (Small Regime). *For $\epsilon \in [N^{-(c+2)/2}, 1]$ and all $\|x - y\|_1 < 2$, the following bound holds:*

$$(1 - \epsilon)\frac{1 - \epsilon^2/\sqrt{2}}{1 + 3\epsilon^2}\mu(\|x - y\|_1) \leq \rho(F(x), F(y)),$$

with probability at least $1 - N^{-c}$, provided

$$k = \frac{C\ln^2(N^{c+2})}{\epsilon^2}.$$

Proof. We can use lemma 3.0.5 with $\lambda^* = \lambda_0$ to cover all distances $\|x - y\|_1$ down to λ_0. We then choose λ_0 in order to extend the lower bound to distances smaller than λ_0, using lemma 3.0.7.

Concretely, recall from section 2 that because F is a linear map of i.i.d. Cauchy entries,

$$\rho(F(x/\eta), F(y/\eta)) \sim \frac{1}{k}\sum_{i=1}^{k}\xi(\lambda_0 |X_i|) \quad \text{and} \quad \rho(F(x), F(y)) \sim \frac{1}{k}\sum_{i=1}^{k}\xi(\eta\lambda_0 |X_i|)$$

with *the same $X_i \overset{\text{i.i.d.}}{\sim}$* Cauchy (1). Let $Z = \max_{1 \leq i \leq k}|X_i|$. To use lemma 3.0.7, we just need to ensure $\lambda_0 Z \leq 1/6$ with high probability. By the independence of the X_i,

$$\mathbb{P}\{\lambda_0 Z \leq 1/6\} = \left(1 - (2/\pi)\arctan(6\lambda_0)\right)^k \geq 1 - k(12/\pi)\lambda_0.$$

So set $\lambda_0 = \pi/(12N^{c+2}k)$. Choosing k according to lemma 3.0.5 with $\lambda_0 = \lambda^*$, we have the following inequality for the target dimension

$$k \geq \frac{2\ln(N^{c+2})}{\epsilon^2}\left(C + \frac{8}{\pi}\left(1 - \ln(\pi/(12N^{c+2}k))\right)\right).$$

Taking $k = C\ln^2(N^{c+2})/\epsilon^2$ satisfies the above, provided $\epsilon^2 > N^{-c-2}$, say. We now have the conditions of lemma 3.0.7 satisfied for all $\binom{N}{2}$ pairs of points, with probability at least $1 - N^{-c}$, and $\lambda_0 < \epsilon^2$. $\square$

4. Upper tails

In the following lemmas, the estimates are not sharp.

Lemma 4.0.1 (General Upper Tail). *With $Y = \mathbb{E}\xi(\lambda\,|X|) - \mu(\lambda)$ and $V^2 \geq \mathrm{Var}(\xi(\lambda\,|X|))$,*

$$\min_s e^{-s\Delta} \mathbb{E}\exp(sY) \leq \exp\left(\frac{-\Delta^2}{4(V^2 + A(\lambda))}\right)$$

and is minimized at s^* with

$$A(\lambda) = \frac{16e}{\pi(e-1)^2}\sqrt{\frac{\lambda^2}{1+\lambda^2}} \quad provided \quad s^* = \frac{\Delta}{2(V^2 + A(\lambda))} \leq 1/2.$$

Proof. From the discussion in section 2, we just need to establish the $A(\lambda)$ function for the integration by parts terms. To ensure $\mathbb{E}\exp(sY)$ is finite, we require $s < 1$. For $1 > s > 0$ and $t > 1$, we then estimate, with $w = \mu + t/s$,

$$\mathbb{P}\{\xi(\lambda\,|X|) > w\} \leq \mathbb{P}\left\{2\ln(1 + \sqrt{\lambda\,|X|}) > w\right\} \leq \frac{2}{\pi}\arctan\left(\frac{\lambda}{(\exp(w/2)-1)^2}\right),$$

and for all $w \geq w_0$ may be bounded by

$$\leq \frac{2}{\pi}\frac{\lambda}{(\exp(w/2)-1)^2} \leq \frac{2\lambda}{\pi}\frac{e^{-w}}{(1 - \exp(-w_0/2))^2} =: C(\lambda)e^{-t/s}.$$

We shall choose w_0 and hence $C(\lambda)$ a bit later; note that $C(\lambda)$ contains the factor $e^{-\mu}$.

With $Y = \xi(\lambda\,|X|) - \mu(\lambda)$, we can now estimate the integration by parts terms

$$\mathbb{E}\exp(sY)\mathbb{I}\{sY > 1\} = e\mathbb{P}\{Y > 1/s\} + \int_1^\infty e^t \mathbb{P}\{Y > t/s\}\,dt$$

as at most

$$eC(\lambda)e^{-1/s} + C(\lambda)\int_1^\infty e^{t(1-1/s)}\,dt = eC(\lambda)\frac{e^{-1/s}}{1-s} \leq \frac{4}{e}C(\lambda)\frac{s^2}{1-s}$$

using $e^{-1/s} \leq (2/e)^2 s^2$ and $s \in (0,1)$. Assuming $s \leq 1/2$, we can now write, for a suitable upper bound $A(\lambda)$,

$$\mathbb{E}\exp(sY) \leq 1 + V^2 s^2 + \frac{8}{e}C(\lambda)s^2 \leq 1 + s^2(V^2 + A(\lambda)),$$

and we may optimize in s

$$\min_s e^{-s\Delta}\mathbb{E}\exp(sY) \leq \min_s e^{-s\Delta + s^2(V^2 + A)} = \exp\left(\frac{-\Delta^2}{4(V^2 + A)}\right)$$

at

$$s^* = \frac{\Delta}{2(V^2 + A)}.$$

It remains to choose w_0 and hence $A(\lambda)$. Recalling the formula for μ either from section 2 or directly from lemma A.1.1, we can lower bound $w = \mu(\lambda) + t/s \geq (1/2)\ln(1 + \lambda^2) + 2 \geq 2$ provided $s \leq 1/2$. Choosing $w_0 = 2$,

$$\frac{8}{e}C(\lambda) \leq \frac{16}{\pi e}\frac{\lambda}{(1 + \lambda^2)^{1/2}}\frac{1}{(1 - 1/e)^2} = \frac{16e}{\pi(e - 1)^2}\sqrt{\frac{\lambda^2}{1 + \lambda^2}} =: A(\lambda). \qquad \square$$

5. Lower tails

Unlike for the upper tails, we can control the lower tails for the full range of λ. We address λ bounded away from 0 using the same techniques as for the upper tail. The lower tail proof for small λ simplifies because $\xi(\lambda|X|)$ is nonnegative, so that there is no restriction on optimizing s in the moment generating function.

In the following lemmas, the estimates are not sharp.

Lemma 5.0.1 (Lower Tail, Big Regime). *With* $Y = \mathbb{E}\xi(\lambda|X|) - \mu(\lambda)$ *and* $V^2 \geq \mathrm{Var}(\xi(\lambda|X|))$,

$$\min_s e^{-s\Delta}\mathbb{E}\exp(-sY) \leq \exp\left(\frac{-\Delta^2}{4(V^2 + A(\lambda))}\right)$$

and is minimized at s^* *with*

$$A(\lambda) = \frac{16}{e\pi}e^{\mathrm{atanh}(1/\sqrt{2})}\sqrt{\frac{1 + \lambda^2}{\lambda^2}} \quad provided \quad s^* = \frac{\Delta}{2(V^2 + A(\lambda))} \leq 1/2.$$

Proof. Just as in the upper tail computations,

$$\mathbb{E}\exp(-sY)\mathbb{I}\{-sY \leq 1\} \leq \mathbb{E}(1 - sY + (-sY)^2) = 1 + V^2 s^2.$$

and

$$\mathbb{E}\exp(-sY)\mathbb{I}\{-sY > 1\} \leq e\mathbb{P}\{-Y > 1/s\} + \int_1^\infty \exp(t)\mathbb{P}\{-Y > t/s\}\,dt.$$

We shall again determine the function $A(\lambda)$ by estimating a tail, but now it is the lower tail

$$\mathbb{P}\{-Y > t/s\} = \mathbb{P}\{\xi(\lambda|X|) < \mu - t/s\}.$$

By subadditivity of $\sqrt{a}$,

$$\xi(a) = \ln(1 + \sqrt{a}) + \frac{1}{2}\ln(1 + a) \geq \ln(\sqrt{1 + a}) + \frac{1}{2}\ln(1 + a) = \ln(1 + a).$$

We now can estimate

$$\mathbb{P}\{-Y > t/s\} < \frac{2}{\pi}\arctan\left(e^{-t/s}\frac{e^\mu}{\lambda}\right) < \frac{2}{\pi}\frac{e^\mu}{\lambda}e^{-t/s} =: C(\lambda)e^{-t/s}.$$

We can then upper bound the integration by parts terms just like in the proof for the upper tail lemma 4.0.1. Assuming $s \le 1/2$, we choose an upper bound $A(\lambda)$ for $(8/e)C(\lambda)$ and arrive at

$$\min_s e^{-s\Delta}\mathbb{E}\exp(-sY) \le \exp\left(\frac{-\Delta^2}{4(V^2 + A(\lambda))}\right)$$

To find $A(\lambda)$, note that $\mu(\lambda) \le \mathrm{atanh}(1/\sqrt{2}) + (1/2)\ln(1 + \lambda^2)$, so that

$$\frac{8}{e}C(\lambda) \le \frac{16}{e\pi}e^{\mathrm{atanh}(1/\sqrt{2})}\sqrt{\frac{1 + \lambda^2}{\lambda^2}} =: A(\lambda)$$

which is bounded for λ away from 0. □

Lemma 5.0.2 (Lower Tail, Small Regimes). *With $Y = \mathbb{E}\xi(\lambda\,|X|) - \mu(\lambda)$ and $V^2 \ge \mathbb{E}\xi(\lambda\,|X|)^2$,*

$$\min_s e^{-s\Delta}\mathbb{E}\exp(-sY) \le \exp\left(\frac{-\Delta^2}{2V^2}\right)$$

Proof. Because $\xi(\lambda\,|X|)$ is nonnegative, we can use the 2nd order Taylor expansion of $\exp(-x)$ to write

$$e^{-s\Delta+s\mu}\mathbb{E}\exp(-s\xi(\lambda\,|X|)) \le e^{-s\Delta+s\mu}(1 - s\mu + \frac{s^2}{2}V^2) = e^{-s\Delta+s^2V^2/2}$$

and we can then optimize s in the usual way. □

Acknowledgments

Appendix A: The first and second moments

Here we derive the explicit formula for $\mu(\lambda) = \mathbb{E}\xi(\lambda|X|)$ in lemma A.1.1 and the upper bounds V^2 for $\mathrm{Var}(\xi(\lambda|X|))$ in corollary A.3.2. Some of this work is

a bit tedious, but it will allow us to give explicit upper bounds on the target dimension

$$k = C \frac{V^2 + A}{\Delta^2}.$$

We need bounds for $\mu(\lambda)$ when λ is small (remark A.1.3) as well as lower bounds for the Δ's

$$\mu((1 + \epsilon)\lambda) - \mu(\lambda) \quad \text{and} \quad \mu(\lambda) - \mu((1 + \epsilon)^{-1}\lambda)$$

when λ is "large" (lemma A.2.1).

Because the Cauchy density has particularly simple behavior when extended to the complex plane, we heavily rely on complex analysis techniques. We chose ξ to be the linear combination

$$\xi(\lambda \, |X|) = \ln(1 + \sqrt{\lambda \, |X|}) + \frac{1}{2} \ln(1 + \lambda \, |X|)$$

as it will simplify the estimates as well as be easy to compute using a pair of contour integrals. For both moments, the contour integral setup below will greatly facilitate computations; in particular, it will allow us to avoid estimating

$$\mathbb{E} \ln^2(1 + \sqrt{\lambda \, |X|}) \quad \text{and} \quad \mathbb{E} \ln^2(1 + \lambda \, |X|)$$

individually, which while possible, is not necessary for our results.

Proposition A.0.1 (Contour Integral Setup)**.** *For $\lambda > 0$, $b > 0$, and $X \sim$ Cauchy(1),*

$$\mathbb{E} \ln^b(1 + \sqrt{\lambda \, |X|}) = \ln^b(1 + \sqrt{i\lambda}) + \ln^b(1 + \sqrt{-i\lambda})$$
$$- \frac{1}{2} \mathbb{E} \ln^b(1 + i\sqrt{\lambda \, |X|}) - \frac{1}{2} \mathbb{E} \ln^b(1 - i\sqrt{\lambda \, |X|}).$$

Remark A.0.2. The task is then to simplify the complex logarithms on the right hand side when particular values of b are chosen. We shall choose $b = 1$ and $b = 2$ in the next sections.

Proof. We want to compute

$$I(\lambda) := \mathbb{E} \ln^b(1 + \sqrt{\lambda \, |X|}) = \frac{2}{\pi} \int_0^\infty \frac{\ln^b(1 + \sqrt{\lambda x})}{1 + x^2} \, dx$$

via contour integration. Extending to $z \in \mathbb{C} - (-\infty, 0]$, let

$$f(z) := \frac{2}{\pi} \frac{\ln^b(1 + \sqrt{\lambda z})}{1 + z^2}$$

which has simple poles at $z = \pm i$.

We shall compute $I(\lambda)$ be using two different contours that both traverse the interval $[0, R]$ in the positive direction. Specifically, $\mathcal{C}^+$ is oriented counterclockwise, while $\mathcal{C}^-$ is oriented clockwise, setting

$$\mathcal{C}^+ := [0, R] \cup \mathcal{C}^+(R) \cup \mathcal{C}_\epsilon^+(R) \quad \text{and} \quad \mathcal{C}^- := [0, R] \cup \mathcal{C}^-(R) \cup \mathcal{C}_\epsilon^-(R)$$

with "large" arcs
$$\mathcal{C}^{\pm}(R) := \left\{ Re^{\pm i\theta} \mid 0 \le \theta \le \pi - \epsilon \right\}$$
and segments rotating as $\epsilon \to 0$ to the negative real axis
$$\mathcal{C}_{\epsilon}^{\pm}(R) := \left\{ re^{\pm i(\pi - \epsilon)} \mid R \ge r \ge 0 \right\}.$$

Check that
$$\lim_{R \to \infty} \int_{\mathcal{C}^{\pm}(R)} f(z) \, dz = 0.$$

Keeping in mind the orientations of the contours, the residue theorem dictates for $R > 1$,

$$\int_{\mathcal{C}^{+}} f(z) \, dz = 2\pi i \operatorname{res}_{z=i} f(z)$$

$$= 2\pi i \lim_{z \to i} (z - i) \frac{2}{\pi} \frac{\ln^b(1 + \sqrt{\lambda z})}{(z - i)(z + i)} = 2 \ln^b(1 + \sqrt{\lambda i})$$

and similarly

$$\int_{\mathcal{C}^{-}} f(z) \, dz = -2\pi i \lim_{z \to -i} (z - (-i)) \frac{2}{\pi} \frac{\ln^b(1 + \sqrt{\lambda z})}{(z - i)(z + i)} = 2 \ln^b(1 + \sqrt{-i\lambda}).$$

It remains to show that

$$\lim_{R \to \infty} \lim_{\epsilon \to 0} \left(\int_{\mathcal{C}_{\epsilon}^{+}(R)} f(z) \, dz + \int_{\mathcal{C}_{\epsilon}^{-}(R)} f(z) \, dz \right)$$
$$= \mathbb{E} \ln^b(1 + i\sqrt{\lambda |X|}) + \mathbb{E} \ln^b(1 - i\sqrt{\lambda |X|}).$$

For these $\mathcal{C}_{\epsilon}^{\pm}(R)$ integrals, note that

$$\sqrt{re^{\pm i(\pi - \epsilon)}} = \sqrt{r} e^{\mp i\epsilon/2} e^{\pm i\pi/2} = \pm i \sqrt{r} e^{\mp i\epsilon/2},$$

which approaches $\pm i\sqrt{r}$ when $\epsilon \to 0$. Consequently, when $z = re^{i(\pi - \epsilon)} = -re^{-i\epsilon}$, we can use the dominated convergence theorem to conclude

$$\lim_{\epsilon \to 0} \int_{\mathcal{C}_{\epsilon}^{+}(R)} f(z) \, dz = \lim_{\epsilon \to 0} \int_{R}^{0} f(-re^{-i\epsilon}) \left(-e^{-i\epsilon} \right) dr$$

$$= \lim_{\epsilon \to 0} \int_{0}^{R} \frac{2}{\pi} \frac{e^{-i\epsilon} \ln^b(1 + i\sqrt{\lambda r} e^{-i\epsilon/2})}{1 + r^2 e^{-2i\epsilon}} \, dr$$

$$= \int_{0}^{R} \frac{2}{\pi} \frac{\ln^b(1 + i\sqrt{\lambda r})}{1 + r^2} \, dr,$$

checking that the integrand is bounded by a summable one when $\epsilon < \pi/8$ say. Sending $R \to \infty$ recovers

$$\lim_{R \to \infty} \lim_{\epsilon \to 0} \int_{\mathcal{C}_{\epsilon}^{+}(R)} f(z) \, dz = \mathbb{E} \ln^b(1 + i\sqrt{\lambda |X|}).$$

Similar reasoning applies to the $\mathcal{C}_\epsilon^-(R)$ integral to yield

$$\lim_{R\to\infty}\lim_{\epsilon\to 0}\int_{\mathcal{C}_\epsilon^-(R)} f(z)\,dz = \mathbb{E}\ln^b(1 - i\sqrt{\lambda\,|X|})$$

Putting everything together, we have

$$2\ln^b(1 + \sqrt{i\lambda}) + 2\ln^b(1 + \sqrt{-i\lambda})$$
$$= 2I(\lambda) + \mathbb{E}\ln^b(1 + i\sqrt{\lambda\,|X|}) + \mathbb{E}\ln^b(1 - i\sqrt{\lambda\,|X|})$$

that is

$$\ln^b(1 + \sqrt{i\lambda}) + \ln^b(1 + \sqrt{-i\lambda})$$
$$= I(\lambda) + \frac{1}{2}\mathbb{E}\ln^b(1 + i\sqrt{\lambda\,|X|}) + \frac{1}{2}\mathbb{E}\ln^b(1 - i\sqrt{\lambda\,|X|})$$

as claimed. $\qquad\square$

A.1. 1st moment

Recall from definition B.1.6 that $\mathrm{atanh}(x)$ may be defined by the power series

$$\mathrm{atanh}(x) = \sum_{j=0}^{\infty} \frac{x^{2j+1}}{2j+1} \quad \text{for} \quad |x| < 1.$$

Lemma A.1.1. *If $\lambda > 0$ and $X \sim Cauchy(1)$, then*

$$\mathbb{E}\ln(1 + \sqrt{\lambda\,|X|}) = \mathrm{atanh}\left(\frac{\sqrt{2\lambda}}{1+\lambda}\right) + \frac{1}{2}\ln(1+\lambda^2) - \frac{1}{2}\mathbb{E}\ln(1+\lambda\,|X|)$$

that is,

$$\mu(\lambda) := \mathbb{E}\xi(\lambda\,|X|) = \mathrm{atanh}\left(\frac{\sqrt{2\lambda}}{1+\lambda}\right) + \frac{1}{2}\ln(1+\lambda^2).$$

Proof. Starting from proposition A.0.1 with $b = 1$,

$$\mathbb{E}\ln(1 + \sqrt{\lambda\,|X|})$$
$$= \ln(1 + \sqrt{i\lambda}) + \ln(1 + \sqrt{-i\lambda}) - \frac{1}{2}\mathbb{E}\ln(1 + i\sqrt{\lambda\,|X|}) - \frac{1}{2}\mathbb{E}\ln(1 - i\sqrt{\lambda\,|X|}).$$

By lemma B.1.7 and the atanh addition formula B.1.8,

$$\ln(1 + \sqrt{\lambda i}) + \ln(1 + \sqrt{-\lambda i})$$
$$= \mathrm{atanh}(\sqrt{\lambda i}) + \mathrm{atanh}(\sqrt{-\lambda i}) + \frac{1}{2}\ln(1 - (\sqrt{\lambda i})^2) + \frac{1}{2}\ln(1 - (\sqrt{-\lambda i})^2)$$

$$= \operatorname{atanh}\left(\frac{\sqrt{i\lambda} + \sqrt{-i\lambda}}{1 + \sqrt{i\lambda}\sqrt{-i\lambda}}\right) + \frac{1}{2}\ln(1 - i\lambda) + \frac{1}{2}\ln(1 - (-i\lambda))$$

$$= \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1 + \lambda}\right) + \frac{1}{2}\ln(1 - (i\lambda)^2)$$

$$= \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1 + \lambda}\right) + \frac{1}{2}\ln(1 + \lambda^2).$$

By remark B.0.10,

$$\ln(1 + i\sqrt{\lambda\,|X|}) + \ln(1 - i\sqrt{\lambda\,|X|}) = \ln(1 - (i\sqrt{\lambda\,|X|})^2) = \ln(1 + \lambda\,|X|).$$

Consequently,

$$\mathbb{E}\ln(1 + \sqrt{\lambda\,|X|}) + \frac{1}{2}\mathbb{E}\ln(1 + \lambda\,|X|) = \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1 + \lambda}\right) + \frac{1}{2}\ln(1 + \lambda^2)$$

as claimed. $\qquad\square$

We use the following lemma to show that $\mu(\lambda) = \Theta(\sqrt{\lambda})$ as well when λ is small.

Lemma A.1.2. *For $\lambda > 0$,*

$$\frac{\sqrt{2\lambda}}{1 + \lambda} < \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1 + \lambda}\right) < \frac{\sqrt{2\lambda}}{1 + \lambda}\left(1 + \frac{1}{2}\ln\left(1 + \frac{2\lambda}{1 + \lambda^2}\right)\right) < \frac{3}{\sqrt{2}}\frac{\sqrt{\lambda}}{1 + \lambda}$$

and approaches 0 as $\lambda \to \infty$. Further, for any $\lambda \le \lambda_0 \le 1$,

$$\frac{\sqrt{2\lambda}}{1 + \lambda} < \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1 + \lambda}\right) < \frac{\sqrt{2\lambda}}{1 + \lambda}\left(1 + \frac{1}{2}\ln\left(1 + \frac{2\lambda_0}{1 + \lambda_0^2}\right)\right).$$

Remark A.1.3. By lemma A.1.1, we now also have the bound

$$\frac{\sqrt{2\lambda}}{1 + \lambda} \le \mu(\lambda) \le \frac{\sqrt{2\lambda}}{1 + \lambda}\left(1 + \frac{\lambda_0}{1 + \lambda_0^2}\right) + \frac{\lambda^2}{2} \le \frac{3}{\sqrt{2}}\frac{\sqrt{\lambda}}{1 + \lambda} + \frac{\lambda^2}{2}.$$

using $\ln(1 + x) \le x$ twice.

Proof. The limit for large λ is immediate. From the power series for atanh, conclude $\operatorname{atanh}(x) > x$ for $x > 0$. We can also give the upper bound

$$\operatorname{atanh}(x) = \sum_{j=0}^{\infty}\frac{x^{2j+1}}{2j+1} = x\sum_{j=0}^{\infty}\frac{(x^2)^j}{2j+1} \le x\left(1 + \frac{1}{2}\sum_{j=1}^{\infty}\frac{(x^2)^j}{j}\right)$$

$$= x\left(1 - \frac{1}{2}\ln(1 - x^2)\right).$$

So,

$$\operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1+\lambda}\right) \leq \frac{\sqrt{2\lambda}}{1+\lambda}\left(1 + \frac{1}{2}\ln\left(1 + \frac{2\lambda}{1+\lambda^2}\right)\right)$$

Noting that $\lambda/(1+\lambda^2)$ is strictly increasing for $\lambda \in (0,1)$, we can fix the λ^2 term at a particular constant. $\square$

A.2. Estimating deviations of the mean

We derive the estimates used in the large scale concencentration proofs given above. Both differences

$$\mu((1+\epsilon)\lambda) - \mu(\lambda) \quad \text{and} \quad \mu(\lambda) - \mu((1+\epsilon)^{-1}\lambda)$$

are controlled by lemma A.2.1 by requiring $\lambda \geq \sqrt{1+\epsilon}$. Because

$$\mu(\lambda) = \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1+\lambda}\right) + \frac{1}{2}\ln(1+\lambda^2)$$

both deviations will be sums of two terms, an atanh term and a ln term.

Lemma A.2.1. *For $1 \leq a$ and $1/\sqrt{a} \leq \lambda$,*

$$a - 1 > \mu(a\lambda) - \mu(\lambda) \geq \frac{a-1}{4}(1 - (a-1))$$

Proof. We shall show that for $\lambda \geq 1/\sqrt{a}$, the difference in the atanh terms is nonpositive. We then immediately have the upper bound

$$\mu(a\lambda) - \mu(\lambda) \leq \frac{1}{2}\ln\left(1 + \frac{(a^2-1)\lambda^2}{1+\lambda^2}\right) < \ln(a) \leq a - 1.$$

On the other hand, because $\lambda \geq 1/\sqrt{a}$, the ln contribution also has the lower bound

$$\frac{1}{2}\ln\left(1 + \frac{(a^2-1)\lambda^2}{1+\lambda^2}\right) \geq \frac{1}{2}\ln\left(1 + \frac{(a^2-1)(1/a)}{1+1/a}\right) = \frac{1}{2}\ln\left(1 + (a-1)\right)$$

$$\geq \frac{(a-1)}{2}\left(1 - \frac{a-1}{2}\right)$$

using a 2nd order Taylor series with Lagrange remainder in the last line, recalling $a \geq 1$ here.

For the lower bound for $\mu(a\lambda) - \mu(\lambda)$, it remains to control how negative the atanh contribution is. With

$$u = \frac{\sqrt{2a\lambda}}{1+a\lambda} \quad \text{and} \quad v = \frac{\sqrt{2\lambda}}{1+\lambda},$$

we can use the atanh addition formula B.1.8,

$$\operatorname{atanh}(u) - \operatorname{atanh}(v) = \operatorname{atanh}(u) + \operatorname{atanh}(-v) = \operatorname{atanh}\left(\frac{u + (-v)}{1 + u(-v)}\right)$$

for $u, v \in (-1, 1)$, which is the case for us here. After some simplification, we recover

$$\frac{u - v}{1 - uv} = (\sqrt{a} - 1)\sqrt{2\lambda}\frac{1 - \lambda\sqrt{a}}{(1 - \lambda\sqrt{a})^2 + \lambda(1 + a)}$$

which is negative for $\lambda \geq 1/\sqrt{a}$. Because atanh is an odd function, taking it of the above gives a negative contribution for such λ. Use the AM-GM inequality to upper bound

$$-\frac{u - v}{1 - uv} \leq \frac{\sqrt{a} - 1}{\sqrt{2}\sqrt{1 + a}} =: w,$$

then use the estimate

$$\operatorname{atanh}(w) \leq \frac{w}{1 - w^2} = \frac{(\sqrt{a} - 1)\sqrt{2}\sqrt{1 + a}}{(1 + \sqrt{a})^2} \leq \frac{\sqrt{a} - 1}{2}$$

as the remaining factor is seen to be decreasing for $a \geq 1$ upon taking logarithms. Using $\sqrt{a} \leq 1 + (a - 1)/2$, we finally have.

$$\mu(a\lambda) - \mu(\lambda) \geq \frac{(a - 1)}{2}\left(1 - \frac{a - 1}{2}\right) - \frac{\sqrt{a} - 1}{2} \geq \frac{a - 1}{4}(1 - (a - 1)). \quad \square$$

A.3. 2nd moment

To estimate the 2nd moment $\mathbb{E}\xi^2(\lambda | X|)$, note that for any $a, b > 0$, the AM-GM inequality gives $(a + b)^2 \leq 2(a^2 + b^2)$, so that

$$\mathbb{E}\xi^2(\lambda | X|) = \mathbb{E}\left(\ln(1 + \sqrt{\lambda | X|}) + \frac{1}{2}\ln(1 + \lambda | X|)\right)^2$$

$$\leq \mathbb{E}\left(2\ln^2(1 + \sqrt{\lambda | X|}) + \frac{1}{2}\ln^2(1 + \lambda | X|)\right).$$

It turns out this last expression also arises from a contour integral.

Lemma A.3.1. *If $\lambda > 0$ and $X \sim Cauchy(1)$, then*

$$\mathbb{E}\left(2\ln^2(1 + \sqrt{\lambda | X|}) + \frac{1}{2}\ln^2(1 + \lambda | X|)\right)$$

$$= 2\mathbb{E}\arctan^2(\sqrt{\lambda | X|}) + \mu^2(\lambda) - \left(\arctan(\lambda) - h(\sqrt{\lambda})\right)^2$$

with

$$h(\sqrt{\lambda}) = \frac{\pi}{2} + \arctan\left(\frac{\sqrt{\lambda}}{\sqrt{2}} - \frac{1}{\sqrt{2\lambda}}\right).$$

Proof. The computations will be a bit more involved than those for the first moment. Starting from proposition A.0.1 with $b = 2$,

$$\mathbb{E} \ln^2(1 + \sqrt{\lambda |X|})$$
$$= \ln^2(1 + \sqrt{i\lambda}) + \ln^2(1 + \sqrt{-i\lambda})$$
$$- \frac{1}{2}\mathbb{E} \ln^2(1 + i\sqrt{\lambda |X|}) - \frac{1}{2}\mathbb{E} \ln^2(1 - i\sqrt{\lambda |X|}),$$

that is,

$$\mathbb{E} 2\ln^2(1 + \sqrt{\lambda |X|}) + \mathbb{E} \ln^2(1 + i\sqrt{\lambda |X|}) + \mathbb{E} \ln^2(1 - i\sqrt{\lambda |X|})$$
$$= 2\ln^2(1 + \sqrt{i\lambda}) + 2\ln^2(1 + \sqrt{-i\lambda}).$$

By lemma A.3.5,

$$\mathbb{E} \ln^2(1 + i\sqrt{\lambda |X|}) + \mathbb{E} \ln^2(1 - i\sqrt{\lambda |X|})$$
$$= \mathbb{E} \frac{1}{2} \ln^2(1 + (\sqrt{\lambda |X|})^2) - 2\mathbb{E} \arctan^2(\sqrt{\lambda |X|})$$
$$= \mathbb{E} \frac{1}{2} \ln^2(1 + \lambda |X|) - 2\mathbb{E} \arctan^2(\sqrt{\lambda |X|}).$$

For the residue terms, we use lemma A.3.6:

$$2\ln^2(1 + \sqrt{i\lambda}) + 2\ln^2(1 + \sqrt{-i\lambda})$$
$$= \frac{1}{4} \ln^2(1 + \lambda^2) - \arctan^2(\lambda)$$
$$+ \ln(1 + \lambda^2)g(\sqrt{\lambda}) + 2\arctan(\lambda)h(\sqrt{\lambda}) + g^2(\sqrt{\lambda}) - h^2(\sqrt{\lambda})$$

with

$$g(\sqrt{\lambda}) = \operatorname{atanh}\left(\frac{\sqrt{2\lambda}}{1 + \lambda}\right) \quad \text{and} \quad h(\sqrt{\lambda}) = \frac{\pi}{2} + \arctan\left(\frac{\sqrt{\lambda}}{\sqrt{2}} - \frac{1}{\sqrt{2\lambda}}\right).$$

Recalling our computation of $\mu(\lambda)$ in lemma A.1.1, we can further simplify:

$$2\ln^2(1 + \sqrt{i\lambda}) + 2\ln^2(1 + \sqrt{-i\lambda})$$
$$= \left(\frac{1}{4} \ln^2(1 + \lambda^2) + \ln(1 + \lambda^2)g(\sqrt{\lambda}) + g^2(\sqrt{\lambda})\right)$$
$$- \arctan^2(\lambda) + 2\arctan(\lambda)h(\sqrt{\lambda}) - h^2(\sqrt{\lambda})$$
$$= \mu^2(\lambda) - \arctan^2(\lambda) + 2\arctan(\lambda)h(\sqrt{\lambda}) - h^2(\sqrt{\lambda})$$

Putting everything together we may conclude

$$\mathbb{E}\left(2\ln^2(1 + \sqrt{\lambda |X|}) + \frac{1}{2} \ln^2(1 + \lambda |X|)\right)$$
$$= 2\mathbb{E} \arctan^2(\sqrt{\lambda |X|}) + \mu^2(\lambda) - \left(\arctan(\lambda) - h(\sqrt{\lambda})\right)^2. \qquad \square$$

Corollary A.3.2 (The Variance Is Bounded). *For $\lambda > 0$ and $X \sim Cauchy(1)$,*

$$\text{Var}(\xi(\lambda \,|X|)) \leq \min\left\{ 2\mathbb{E}\ln(1 + \lambda \,|X|), \frac{\pi^2}{2} \right\}.$$

Proof. Just note that for $\nu > 0$,

$$\arctan^2(\sqrt{\nu}) \leq \min\left\{ \ln(1 + \nu), \frac{\pi^2}{4} \right\}.$$

The constant follows from $\arctan(x) \leq \pi/2$ for all $x \in \mathbb{R}$, while the $\ln(1 + \nu)$ bound follows from comparing derivatives, noting that both functions take 0 when $\nu = 0$. $\qquad\square$

For quantitative estimates for the 2nd moment and the variance, we make the $\mathbb{E}\ln(1 + \lambda \,|X|)$ term explicit in the above bound.

Lemma A.3.3. *For $\lambda \geq 0$ and $X \sim Cauchy(1)$,*

$$\mathbb{E}\ln(1 + \lambda \,|X|) = -\frac{2}{\pi}\ln(\lambda)\arctan(\lambda) + \frac{1}{2}\ln(1 + \lambda^2) + \frac{2}{\pi}\text{Ti}_2(\lambda).$$

Proof. From lemma B.0.1

$$\mathbb{E}\ln(1 + \lambda \,|X|) = \frac{2}{\pi}\frac{1}{2i}(\text{Li}_2(1 + i\lambda) - \text{Li}_2(1 - i\lambda))$$

We use the reflection formula B.2.1 to expand the dilogarithm terms.

Recall from lemma B.2.1, for $z \in (\mathbb{C} - \mathbb{R}) \cup (0, 1)$,

$$\text{Li}_2(z) + \text{Li}_2(1 - z) - \text{Li}_2(1) = -\ln(z)\ln(1 - z).$$

Consequently, using definition B.1.3 for Ti_2,

$$\frac{1}{2i}(\text{Li}_2(1 + i\lambda) - \text{Li}_2(1 - i\lambda))$$

$$= \frac{1}{2i}\big(-\ln(-i\lambda)\ln(1 + i\lambda) - \text{Li}_2(-i\lambda) + \text{Li}_2(1) \big)$$

$$-\frac{1}{2i}\big(-\ln(i\lambda)\ln(1 - i\lambda) - \text{Li}_2(i\lambda) + \text{Li}_2(1) \big)$$

$$= \frac{1}{2i}\ln(\lambda)(\ln(1 - i\lambda) - \ln(1 + i\lambda)) + \frac{\pi}{4}(\ln(1 - i\lambda) + \ln(1 + i\lambda)) + \text{Ti}_2(\lambda)$$

By lemma B.0.9 (really the remark there) and the definition of arctan,

$$\frac{1}{2i}(\text{Li}_2(1 + i\lambda) - \text{Li}_2(1 - i\lambda)) = -\ln(\lambda)\arctan(\lambda) + \frac{\pi}{4}\ln(1 + \lambda^2) + \text{Ti}_2(\lambda).$$

Thus,

$$\mathbb{E}\ln(1 + \lambda \,|X|) = \frac{2}{\pi}\frac{1}{2i}(\text{Li}_2(1 + i\lambda) - \text{Li}_2(1 - i\lambda))$$

$$= -\frac{2}{\pi}\ln(\lambda)\arctan(\lambda) + \frac{1}{2}\ln(1 + \lambda^2) + \frac{2}{\pi}\text{Ti}_2(\lambda). \qquad\square$$

Corollary A.3.4. *For* $0 < \lambda < 2$

$$\frac{\mathbb{E}\xi^2(\lambda \,|X|)}{\mu(\lambda)^2} \leq \begin{cases} 1 + 2\left(\lambda + \frac{4}{\pi}(1 - \ln(\lambda))\right) & \text{for } \lambda \in (0, 1] \\ 1 + \frac{9}{2}\left(\lambda + \frac{4}{\pi}\right) & \text{for } \lambda \in (1, 2) \end{cases}.$$

Proof. By corollary A.3.2 and lemma A.3.3, we have

$$\mathbb{E}\xi^2(\lambda \,|X|) \leq \ln(1 + \lambda^2) + \frac{4}{\pi}\,\mathrm{Ti}_2(\lambda) - \frac{4}{\pi}\ln(\lambda)\arctan(\lambda) + \mu^2(\lambda)$$

$$\leq \lambda^2 + \frac{4}{\pi}\lambda - \frac{4}{\pi}\lambda\ln(\lambda) + \mu^2(\lambda)$$

because $\mathrm{Ti}_2(\lambda)$ is an alternating series with terms of decreasing magnitude for $\lambda < 2$ and that for $\lambda \leq 1$, $\ln(\lambda)$ is nonnegative. For $\lambda \in (1, 2)$, we can drop the $\ln(\lambda)$ term for an upper bound. Consequently, using $\mu(\lambda) \geq \sqrt{2\lambda}/(1 + \lambda)$ from remark A.1.3,

$$\frac{\mathbb{E}\xi^2(\lambda \,|X|)}{\mu(\lambda)^2} \leq 1 + 2\left(\lambda + \frac{4}{\pi}(1 - \ln(\lambda))\right)$$

for $\lambda \leq 1$, and

$$\frac{\mathbb{E}\xi^2(\lambda \,|X|)}{\mu(\lambda)^2} \leq 1 + \frac{9}{2}\left(\lambda + \frac{4}{\pi}\right)$$

for $\lambda \in (1, 2)$. $\qquad\square$

Lemma A.3.5. *For* $r > 0$,

$$\ln^2(1 + ir) + \ln^2(1 - ir) = \frac{1}{2}\ln^2(1 + r^2) - 2\arctan^2(r).$$

Proof. We are adding complex conjugates, so the left hand side is

$$2\Re \ln^2(1 + ir) = 2\Re\left(\frac{1}{2}\ln(1 + r^2) + i\arctan(r)\right)^2$$

$$= 2\left(\frac{1}{4}\ln^2(1 + r^2) - \arctan^2(r)\right) = \frac{1}{2}\ln^2(1 + r^2) - 2\arctan^2(r).$$

$\qquad\square$

Lemma A.3.6. *For* $\nu > 0$,

$$\ln^2(1 + \nu\sqrt{i}) + \ln^2(1 + \nu\sqrt{-i})$$

$$= \frac{1}{8}\ln^2(1 + \nu^4) - \frac{1}{2}\arctan^2(\nu^2)$$

$$+ \frac{1}{2}\ln(1 + \nu^4)g(\nu) + \arctan(\nu^2)h(\nu) + \frac{1}{2}(g^2(\nu) - h^2(\nu))$$

with

$$g(\nu) = \mathrm{atanh}\left(\frac{\nu\sqrt{2}}{1 + \nu^2}\right) \quad \text{and} \quad h(\nu) = \frac{\pi}{2} + \arctan\left(\frac{\nu}{\sqrt{2}} - \frac{1}{\nu\sqrt{2}}\right).$$

Proof. Using lemma B.1.7,

$$\ln^2(1 + \nu\sqrt{i}) = \left(\operatorname{atanh}(\nu\sqrt{i}) + \frac{1}{2}\ln(1 - i\nu^2)\right)^2$$

$$= \operatorname{atanh}^2(\nu\sqrt{i}) + \operatorname{atanh}(\nu\sqrt{i})\ln(1 - i\nu^2) + \frac{1}{4}\ln^2(1 - i\nu^2)$$

and similarly

$$\ln^2(1 + \nu\sqrt{-i}) = \left(\operatorname{atanh}(\nu\sqrt{-i}) + \frac{1}{2}\ln(1 + i\nu^2)\right)^2$$

$$= \operatorname{atanh}^2(\nu\sqrt{-i}) + \operatorname{atanh}(\nu\sqrt{-i})\ln(1 + i\nu^2) + \frac{1}{4}\ln^2(1 + i\nu^2)$$

Adding yields several terms:

$$\boxed{1}(\nu) := \frac{1}{4}\ln^2(1 + i\nu^2) + \frac{1}{4}\ln^2(1 - i\nu^2)$$

$$\boxed{2}(\nu) := \operatorname{atanh}(\nu\sqrt{i})\ln(1 - i\nu^2) + \operatorname{atanh}(\nu\sqrt{-i})\ln(1 + i\nu^2)$$

$$\boxed{3}(\nu) := \operatorname{atanh}^2(\nu\sqrt{i}) + \operatorname{atanh}^2(\nu\sqrt{-i})$$

From lemma A.3.5,

$$\boxed{1}(\nu) = \frac{1}{4}\left(\frac{1}{2}\ln^2(1 + \nu^4) - 2\arctan^2(\nu^2)\right) = \frac{1}{8}\ln^2(1 + \nu^4) - \frac{1}{2}\arctan^2(\nu^2).$$

We also have

$$\boxed{2}(\nu) = \operatorname{atanh}(\nu\sqrt{i})\left(\frac{1}{2}\ln(1 + \nu^4) - i\arctan(\nu^2)\right)$$

$$+ \operatorname{atanh}(\nu\sqrt{-i})\left(\frac{1}{2}\ln(1 + \nu^4) + i\arctan(\nu^2)\right)$$

$$= \frac{1}{2}\ln(1 + \nu^4)\left(\operatorname{atanh}(\nu\sqrt{i}) + \operatorname{atanh}(\nu\sqrt{-i})\right)$$

$$- i\arctan(\nu^2)\left(\operatorname{atanh}(\nu\sqrt{i}) - \operatorname{atanh}(\nu\sqrt{-i})\right)$$

$$= \frac{1}{2}\ln(1 + \nu^4)g(\nu) + \arctan(\nu^2)h(\nu).$$

Let

$$g(\nu) := \operatorname{atanh}(\nu\sqrt{i}) + \operatorname{atanh}(\nu\sqrt{-i})$$

$$= \operatorname{atanh}\left(\frac{\nu(\sqrt{i} + \sqrt{-i})}{1 + \nu^2\sqrt{-i^2}}\right) = \operatorname{atanh}\left(\frac{\nu\sqrt{2}}{1 + \nu^2}\right)$$

by the atanh addition formula B.1.8, as $\sqrt{\pm i} = (1 \pm i)/\sqrt{2}$ are conjugates of each other.

Let

$$h(\nu) := -i\big(\operatorname{atanh}(\nu\sqrt{i}) - \operatorname{atanh}(\nu\sqrt{-i})\big).$$

Then

$$
\begin{aligned}
g^2&(\nu) - h^2(\nu) \\
&= \operatorname{atanh}^2(\nu\sqrt{i}) + \operatorname{atanh}^2(\nu\sqrt{-i}) + 2\operatorname{atanh}(\nu\sqrt{i})\operatorname{atanh}(\nu\sqrt{-i}) \\
&\quad + \big(\operatorname{atanh}(\nu\sqrt{i}) - \operatorname{atanh}(\nu\sqrt{-i})\big)^2 \\
&= 2\big(\operatorname{atanh}^2(\nu\sqrt{i}) + \operatorname{atanh}^2(\nu\sqrt{-i})\big) = 2\boxed{3}(\nu).
\end{aligned}
$$

So we are left to understand $h(\nu)$. By lemma A.3.7, it is

$$h(\nu) = \frac{\pi}{2} + \arctan\left(\frac{\nu}{\sqrt{2}} - \frac{1}{\nu\sqrt{2}}\right). \qquad \square$$

Lemma A.3.7. *For $\nu > 0$,*

$$h(\nu) := -i\big(\operatorname{atanh}(\nu\sqrt{i}) - \operatorname{atanh}(\nu\sqrt{-i})\big) = \frac{\pi}{2} + \arctan\left(\frac{\nu}{\sqrt{2}} - \frac{1}{\nu\sqrt{2}}\right).$$

Remark A.3.8. For $\nu < 1$, we can rewrite the above as

$$\frac{\pi}{2} - \arctan\left(\frac{1-\nu^2}{\nu\sqrt{2}}\right) = \arctan\left(\frac{\nu\sqrt{2}}{1-\nu^2}\right).$$

Proof. We cannot directly use the atanh addition formula because there is a singularity when ν crosses 1. However, by definition of atanh B.1.6, we can convert $h(\nu)$ as follows, using $\sqrt{-i} = -i\sqrt{i}$

$$
\begin{aligned}
h(\nu) &:= -i\big(\operatorname{atanh}(\nu\sqrt{i}) - \operatorname{atanh}(\nu\sqrt{-i})\big) \\
&= -i\big(-i\arctan(i\nu\sqrt{i}) - (-i)\arctan(i\nu\sqrt{-i})\big) \\
&= -\arctan(i\nu\sqrt{i}) + \arctan(\nu\sqrt{i}).
\end{aligned}
$$

We now use the inversion formula B.3.1 for arctan.

$$
\begin{aligned}
h(\nu) &= -\arctan(i\nu\sqrt{i}) + \frac{\pi}{2} - \arctan(1/(\nu\sqrt{i})) \\
&= \frac{\pi}{2} - \big(\arctan(i\nu\sqrt{i}) + \arctan(-i\sqrt{i}/\nu)\big)
\end{aligned}
$$

The following identity holds

$$\big(\arctan(i\nu\sqrt{i}) + \arctan(-i\sqrt{i}/\nu)\big) = -\arctan\left(\frac{\nu}{\sqrt{2}} - \frac{1}{\nu\sqrt{2}}\right),$$

because both analytic expressions are 0 at $\nu = 1$, and their derivatives match for $\nu > 0$. $\qquad \square$

Appendix B: Polylogarithms and their friends

The polylogarithms $\mathrm{Li}_b(z)$ arise when we compute or estimate the first and second moments of the coordinate projections; they will help us give quantitative bounds which are needed in some of the proofs. References for polylogarithms are [11] and [14].

As initial motivation for studying such functions, we have the following lemma.

Lemma B.0.1. *Let $X \sim Cauchy(1)$ and $\nu > 0$. Then for $b > -1$,*

$$\mathbb{E}\ln^b(1 + \nu|X|) = \frac{\Gamma(b+1)}{i\pi}(\mathrm{Li}_{b+1}(1+i\nu) - \mathrm{Li}_{b+1}(1-i\nu)).$$

Proof. We have

$$I_b(\nu) := \mathbb{E}\ln^b(1+\nu|X|) = \frac{2}{\pi}\int_0^\infty \frac{\ln^b(1+\nu x)}{1+x^2}\,dx$$

Change variables $u = 1 + \nu x$ and then $t = \ln(u)$ to find

$$I_b(\nu) = \frac{2\nu}{\pi}\int_0^\infty \frac{t^b e^t}{(e^t - (1+i\nu))(e^t - (1-i\nu))}\,dt.$$

Using partial fractions, we may write

$$\begin{aligned}
I_b(\nu) &= \frac{1}{i\pi}\int_0^\infty t^b \frac{2i\nu e^t}{(e^t - (1+i\nu))(e^t - (1-i\nu))}\,dt \\
&= \frac{1}{i\pi}\int_0^\infty \frac{t^b(1+i\nu)}{e^t - (1+i\nu)} - \frac{t^b(1-i\nu)}{e^t - (1-i\nu)}\,dt \\
&= \frac{\Gamma(b+1)}{i\pi}(\mathrm{Li}_{b+1}(1+i\nu) - \mathrm{Li}_{b+1}(1-i\nu)).
\end{aligned}$$

by definition B.0.7. The polylogarithms are defined because $\nu > 0$, and if $b > 0$, the value at $\nu = 0$ is also defined. ☐

General references for complex analysis are [18] for proofs and [16] for intuition. If $z = x + iy \in \mathbb{C}$ with $x, y \in \mathbb{R}$, then $\Re(z) := x$ and $\Im(z) := y$. If $z = re^{i\theta} = x + iy \in \mathbb{C}$, denote $z^* = re^{-i\theta} = x - iy$ for the complex conjugate. Further $|z|^2 = zz^* = x^2 + y^2$. Thus, if $w = se^{i\phi}$, we have

$$(zw)^* = (rse^{i(\theta+\phi)})^* = rse^{-i(\theta+\phi)} = z^*w^*.$$

Further, if $w \neq 0$,

$$\left|\frac{z}{w}\right|^2 = \frac{zz^*}{ww^*} = \frac{r^2}{s^2} = \frac{|z|^2}{|w|^2}.$$

For us, analytic functions are synonymous with holomorphic ones. We shall be using two theorems from complex analysis repeatedly. Cf. [18, page 52, 96].

Theorem B.0.2 (Analytic Continuation). *Let f and g be analytic functions in a connected open subset Ω of $\mathbb{C}$. If $f(z) = g(z)$ for all z in a non-empty open subset of Ω, then $f(z) = g(z)$ throughout Ω.*

Theorem B.0.3 (Primitives). *Let f be an analytic function in a simply connected subset Ω of $\mathbb{C}$. Then for $z_0, z \in \Omega$, the function*

$$F(z) := \int_{z_0}^{z} f(w)\, dw = \int_{\gamma} f(w)\, dw$$

is analytic too, with γ any path from z_0 to z lying in Ω.

Definition B.0.4 (The Logarithm). For $z = re^{i\theta} \in \mathbb{C} - (-\infty, 0]$, define (the principle branch of) the logarithm of z, $\ln(z)$ as

$$\ln(z) := \ln(r) + i\theta = \int_{1}^{z} \frac{dw}{w}$$

for any path from 1 to z in $\mathbb{C} - (-\infty, 0]$.

Remark B.0.5. Note that $\ln(z^*) = \ln(r) - i\theta = \ln(z)^*$. The map $w \mapsto 1/w$ takes $\mathbb{C} - (-\infty, 0]$ to itself; for if $w = se^{i\phi}$, with $|\phi| < \pi$, then $1/w = (1/s)e^{-i\phi}$ which also lives in $\mathbb{C} - (-\infty, 0]$. With this choice of principle branch, the logarithm still satisfies $-\ln(1/w) = \ln(w)$ via

$$-\ln(1/w) = -(\ln(1/s) + i(-\phi)) = \ln(s) + i\phi = \ln(w).$$

Similarly, note that if $\Re(z), \Re(w) > 0$, then $zw = rse^{i(\theta+\phi)}$ with $|\theta + \phi| < \pi$ so $\arg(zw) = \theta + \phi$ and

$$\ln(zw) = \ln(rs) + i(\theta + \phi) = \ln(z) + \ln(w)$$

in this case. However, the general identity $\ln(z_1 z_2) = \ln(z_1) + \ln(z_2)$ does not hold.

Definition B.0.6 (The Polylogarithm of Order 1). Define the polylogarithm of order 1, $\mathrm{Li}_1(z)$ as

$$\mathrm{Li}_1(z) := \sum_{j=1}^{\infty} \frac{z^j}{j} \quad \text{for} \quad |z| < 1$$

and

$$\mathrm{Li}_1(z) := -\ln(1-z) = \ln\left(\frac{1}{1-z}\right) \quad \text{for} \quad z \in \mathbb{C} - [1, \infty).$$

For general z, the domain makes sense, as $1 - z = -(z-1) \in \mathbb{C} - (-\infty, 0]$ for the z in question. Recall when $|z| < 1$,

$$-\ln(1-z) = \sum_{j=1}^{\infty} \frac{z^j}{j},$$

noting that both sides agree when $z = 0$, and upon differentiating,

$$\frac{d}{dz} \sum_{j=1}^{\infty} \frac{z^j}{j} = \sum_{j=0}^{\infty} z^j = \frac{1}{1-z} = \frac{d}{dz}(-\ln(1-z))$$

which means $-\ln(1-z)$ and the sum differ by a constant, namely 0.

The order of the polylogarithms may be extended; the general integral form below will be useful for some of the computations later.

Definition B.0.7. For $b > 0$, define the *polylogarithm of order b* as

$$\mathrm{Li}_b(z) := \sum_{j=1}^{\infty} \frac{z^j}{j^b} \quad \text{for} \quad |z| < 1$$

and

$$\mathrm{Li}_b(z) := \frac{1}{\Gamma(b)} \int_0^{\infty} \frac{z t^{b-1}}{e^t - z} \, dt = \frac{1}{\Gamma(b)} \int_0^{\infty} \frac{z t^{b-1} e^{-t}}{1 - e^{-t} z} \, dt.$$

for $z \in \mathbb{C} - [1, \infty)$.

The nonintegral order polylogarithms also extend to the unit circle when the order is greater than 1.

Lemma B.0.8. *For $b > 1$ and $z \in \mathbb{C}$ with $|z| = 1$,*

$$\mathrm{Li}_b(z) < b.$$

Proof. By definition,

$$\mathrm{Li}_b(z) = \sum_{j=1}^{\infty} \frac{z^j}{j^b} \quad \text{so that when } |z| = 1, \quad |\mathrm{Li}_b(z)| \le \sum_{j=1}^{\infty} \frac{|z|^j}{j^b} = \sum_{j=1}^{\infty} \frac{1}{j^b}$$

The series is finite because $b > 1$; concretely, by the integral test (because $1/x^b$ is convex),

$$\sum_{j=1}^{\infty} \frac{1}{j^b} = 1 + \sum_{j=2}^{\infty} \frac{1}{j^b} \le 1 + \int_1^{\infty} \frac{1}{x^b} \, dx = 1 + (b-1) \frac{-1}{x^{b-1}} \Big|_1^{\infty} = b < \infty. \quad \square$$

Lemma B.0.9. *For $z \in (\mathbb{C} - \mathbb{R}) \cup (-1, 1)$ and $b > 0$,*

$$\mathrm{Li}_b(z) + \mathrm{Li}_b(-z) = \frac{1}{2^{b-1}} \mathrm{Li}_b(z^2).$$

If $b > 1$, the equality also holds when $z = \pm 1$.

Remark B.0.10. When $b = 1$, recover

$$\ln(1-z) + \ln(1+z) = -\big(\mathrm{Li}_1(z) + \mathrm{Li}_1(-z)\big) = -\mathrm{Li}_1(z^2) = \ln(1 - z^2).$$

Proof. First assume $|z| < 1$. From the power series,

$$\mathrm{Li}_b(z) + \mathrm{Li}_b(-z) = \sum_{j=1}^{\infty} \frac{z^j + (-z)^j}{j^b} = \sum_{j=1}^{\infty} z^j \frac{1 + (-1)^j}{j^b}$$

$$= 2\sum_{j=1}^{\infty} \frac{z^{2j}}{(2j)^b} = \frac{1}{2^{b-1}} \sum_{j=1}^{\infty} \frac{(z^2)^j}{j^b} = \frac{1}{2^{b-1}} \mathrm{Li}_b(z^2).$$

Both sides are analytic functions on $(\mathbb{C}-\mathbb{R})\cup(-1,1)$, so by analytic continuation, the identity continues to hold there. If $b > 1$, the power series are also defined at $z = \pm 1$. $\square$

A useful property of the polylogarithms and the logarithm that we shall use repeatedly in computations is that they are all symmetric about the real axis, that is, $\mathrm{Li}_b(z^*)^* = \mathrm{Li}_b(z)$ or concretely

$$\Re\,\mathrm{Li}_b(z^*) = \Re\,\mathrm{Li}_b(z) \quad \text{and} \quad \Im\,\mathrm{Li}_b(z^*) = -\Im\,\mathrm{Li}_b(z).$$

Powers and polynomials of such functions also have this property. Intuitively this symmetry follows from the real coefficents in their power series expansions, so that $\mathrm{Li}(x) \in \mathbb{R}$ when $x < 1$. Rigorously, we use the Schwarz reflection principle; because $\mathrm{Li}_b(z)$ is analytic in $\mathbb{C} - [1,\infty)$ when $0 \leq \arg(z) < \pi$ and real valued on $(-\infty, 1)$, $\mathrm{Li}_b(z)$ may be extended to the rest of $\mathbb{C} - [1, \infty)$ in an analytic fashion. Analytic continuation then dictates that this extension coincides with the original definition of $\mathrm{Li}_b(z)$. See [18] pages 57–59 for the Schwarz reflection principle, page 56 for showing the integral definitions of $\mathrm{Li}_b(z)$ are analytic, and page 52 for the principle of analytic continuation.

B.1. Arctan and the inverse tangent integrals

The function $t \mapsto \arctan(t)$ is proportional to the distribution function of $|X|$ with $X \sim \mathrm{Cauchy}\,(1)$. It is then perhaps not surprising that arctan and its relatives arise in working with functions of Cauchy random variables. We outline the properties we shall be using here.

The following definition is opaque but most useful to us.

Definition B.1.1. Define $\arctan(z)$ as

$$\arctan(z) := \sum_{j=0}^{\infty} (-1)^j \frac{z^{2j+1}}{2j+1} \quad \text{for} \quad |z| < 1,$$

and

$$\arctan(z) = \int_0^z \frac{1}{1+w^2}\,dw \quad \text{for} \quad z \in (\mathbb{C} - i\mathbb{R}) \cup (-i, i).$$

Equivalently,

$$\arctan(z) := \frac{1}{2i}(\ln(1 + iz) - \ln(1 - iz)) = \frac{1}{2i}(\mathrm{Li}_1(iz) - \mathrm{Li}_1(-iz)).$$

Remark B.1.2. From the integral formulation, we also immediately have, with $v = -w$,

$$\arctan(-z) = \int_0^{-z} \frac{1}{1+w^2}\, dw = -\int_0^z \frac{1}{1+(-v)^2}\, dv = -\arctan(z).$$

The last definition for $\arctan(z)$ follows from

$$\frac{d}{dz} \frac{1}{2i}(\ln(1+iz) - \ln(1-iz)) = \frac{1}{2i}\left(\frac{i}{1+iz} - \frac{(-i)}{1-iz}\right)$$

$$= \frac{1}{2}\left(\frac{1}{1+iz} + \frac{1}{1-iz}\right) = \frac{1}{1+z^2} = \frac{d}{dz}\arctan(z)$$

and that $\arctan(0) = 0$.

We can generalize.

Definition B.1.3. For $z \in \mathbb{C} - i\mathbb{R} \cup (-i, i)$ and $b > 0$, define the *inverse tangent integral of order b* as

$$\mathrm{Ti}_b(z) := \sum_{j=0}^{\infty} (-1)^j \frac{z^{2j+1}}{(2j+1)^b} \quad \text{for} \quad |z| < 1$$

and

$$\mathrm{Ti}_b(z) = \frac{\mathrm{Li}_b(iz) - \mathrm{Li}_b(-iz)}{2i} \quad \text{for} \quad z \in \mathbb{C} - i\mathbb{R} \cup (-i, i).$$

Remark B.1.4. Note if $|y| < 1$, we find

$$\mathrm{Li}_b(iy) - \mathrm{Li}_b(-iy) = \sum_{j=1}^{\infty} \frac{(iy)^j - (-iy)^j}{j^b} = \sum_{j=1}^{\infty} i^j \frac{y^j}{j^b}(1 - (-1)^j)$$

$$= 2\sum_{j=0}^{\infty} i^{2j+1} \frac{y^{2j+1}}{(2j+1)^b} = 2i \sum_{j=0}^{\infty} (-1)^j \frac{y^{2j+1}}{(2j+1)^b} =: 2i\,\mathrm{Ti}_b(y) \in i\mathbb{R}.$$

Hence,

$$\mathrm{Ti}_b(y) = \frac{\mathrm{Li}_b(iy) - \mathrm{Li}_b(-iy)}{2i}$$

when $|y| < 1$ and $b > 0$. The right hand side continues to make sense for $y \in (\mathbb{C} - i\mathbb{R}) \cup (-i, i)$, so we may define

$$\mathrm{Ti}_b(z) := \frac{\mathrm{Li}_b(iz) - \mathrm{Li}_b(-iz)}{2i}$$

as an analytic function on $z \in (\mathbb{C} - i\mathbb{R}) \cup (-i, i)$ that agrees with the power series on the interior of the unit circle.

Remark B.1.5. In particular, we have $\mathrm{Ti}_1(z) = \arctan(z)$.

To focus on the behavior of arctan on $(-i, i)$ which was not addressed in the inversion formula B.3.1, we change points of view through a rotation of the complex plane.

Definition B.1.6. Define the function atanh as

$$\mathrm{atanh}(x) = \sum_{j=0}^{\infty} \frac{x^{2j+1}}{2j+1} \quad \text{for} \quad |x| < 1,$$

and as

$$\mathrm{atanh}(z) = \int_0^z \frac{1}{1-w^2}\, dw = -i\arctan(iz) \quad \text{for} \quad z \in (\mathbb{C} - \mathbb{R}) \cup (-1,1).$$

or equivalently as

$$\mathrm{atanh}(z) = \frac{1}{2}(\ln(1+z) - \ln(1-z)) = \frac{1}{2}(\mathrm{Li}_1(z) - \mathrm{Li}_1(-z)).$$

To see that the definitions are consistent, note first from the power series, $\mathrm{atanh}(0) = 0 = \arctan(0)$, while on the other hand,

$$\frac{d}{dz}(-i)\arctan(iz) = \frac{(-i)}{1+(iz)^2}(i) = \frac{1}{1-z^2} = \frac{d}{dz}\,\mathrm{atanh}(z).$$

Lemma B.1.7. *Let* $z \in (\mathbb{C} - \mathbb{R}) \cup (-1,1)$ *then*

$$\ln(1+z) = \mathrm{atanh}(z) + \frac{1}{2}\ln(1-z^2).$$

Proof. Just split into even and odd degree terms.

$$\mathrm{Li}_1(z) = \sum_{j=1}^{\infty} \frac{z^j}{j} = \sum_{j=0}^{\infty} \frac{z^{2j+1}}{(2j+1)} + \sum_{j=1}^{\infty} \frac{z^{2j}}{(2j)} = \mathrm{atanh}(z) + \frac{1}{2}\sum_{j=1}^{\infty} \frac{(z^2)^j}{j}$$

$$= \mathrm{atanh}(z) + \frac{1}{2}\,\mathrm{Li}_1(z^2).$$

The equality extends to $(\mathbb{C} - \mathbb{R}) \cup (-1,1)$ as both sides are analytic there. We now have

$$\ln(1+z) = -\mathrm{Li}_b(-z) = -\mathrm{atanh}(-z) + \frac{1}{2}\ln(1-(-z)^2) = \mathrm{atanh}(z) + \frac{1}{2}\ln(1+z^2)$$

as desired.   □

Here is the addition formula.

Lemma B.1.8 (Atanh Addition Formula). *If* $-1 < x,y < 1$,

$$\mathrm{atanh}(x) + \mathrm{atanh}(y) = \mathrm{atanh}\left(\frac{x+y}{1+xy}\right).$$

If $z \in \mathbb{C} - \mathbb{R}$,

$$\mathrm{atanh}(z) + \mathrm{atanh}(z^*) = \mathrm{atanh}\left(\frac{2\Re(z)}{1+|z|^2}\right)$$

Proof. Because atanh is odd, the addition formula also covers subtraction too. Check that

$$\frac{d}{dz}\operatorname{atanh}\left(\frac{z+w}{1+zw}\right) = \frac{1}{1-z^2} = \frac{d}{dz}\operatorname{atanh}(z).$$

So

$$\operatorname{atanh}\left(\frac{z+w}{1+zw}\right) = \operatorname{atanh}(z) + c$$

with c a constant. Taking $z = 0$ forces $c = \operatorname{atanh}(w)$ as desired.

For $z, w \in (\mathbb{C} - \mathbb{R}) \cup (-1, 1)$, let

$$f(z, w) := \frac{z+w}{1+zw}.$$

We want to know when $f(z, w)$ also lies in the domain of atanh. When $w = z^*$,

$$\left|\frac{z+z^*}{1+zz^*}\right| \le \frac{2\,|\Re(z)/\,|z||}{\frac{1}{|z|} + |z|} \le \frac{2}{\frac{1}{|z|} + |z|} \le \frac{1}{\sqrt{|z|/\,|z|}} = 1.$$

by the AM-GM inequality. The equality case occurs just if $|z| = 1$, but in that case, $|\Re(z)|/\,|z| < 1$ as $z = \pm 1$ is not allowed for atanh. We are thus ok for all $z \in (\mathbb{C} - \mathbb{R}) \cup (-1, 1)$ in this $w = z^*$ case.

When $x, y \in (-1, 1)$, we may consider

$$\partial_x f(x, y) = \frac{1}{1+xy} - \frac{(x+y)}{(1+xy)^2}y = \frac{1}{(1+xy)^2}(1+xy-xy-y^2) = \frac{1-y^2}{(1+xy)^2} > 0$$

and by symmetry, $\partial_y f(x, y) > 0$. So f is increasing in each of the individual coordinates. In particular, when $-1 < x < y < 1$,

$$\frac{2x}{1+x^2} = f(x, x) < f(x, y) < f(y, y) = \frac{2y}{1+y^2}.$$

For each permutation of x, y, and 0, check that

$$|f(x, y)| < \frac{2t}{1+t^2} < 1 \quad \text{with} \quad t = \max\{|x|, |y|\},$$

by the AM-GM inequality, with strict inequality because $|x|, |y| < 1$. $\qquad\square$

B.2. Dilogarithm properties

The dilogarithm is the polylogarithm of order 2.

Lemma B.2.1 (Reflection Formula). *For $z \in (\mathbb{C} - \mathbb{R}) \cup (0, 1)$,*

$$\operatorname{Li}_2(z) + \operatorname{Li}_2(1-z) - \operatorname{Li}_2(1) = -\ln(z)\ln(1-z).$$

Proof. (Compare to [11, page 5].) Consider

$$\frac{d}{dz}(\text{Li}_2(z) + \text{Li}_2(1-z)) = \frac{\text{Li}_1(z)}{z} + \frac{\text{Li}_1(1-z)}{1-z}(-1) = \frac{-\ln(1-z)}{z} + \frac{\ln(z)}{1-z}.$$

On the other hand,

$$\frac{d}{dz}(-\ln(z)\ln(1-z)) = \frac{-\ln(1-z)}{z} + \frac{\ln(z)}{1-z}.$$

Because the domain $(\mathbb{C} - \mathbb{R}) \cup (0,1)$ is simply connected and the derivative above is analytic there, we have

$$-\ln(z)\ln(1-z) + \ln(z_0)\ln(1-z_0)$$
$$= \text{Li}_2(z) + \text{Li}_2(1-z) - (\text{Li}_2(z_0) + \text{Li}_2(1-z_0))$$

for some z_0 which we may take to lie on $(0,1)$. Taking the limit as $z_0 \to 0$ is safe, as the Taylor series for $\ln(1-z_0)$ ensures $\ln(z_0)\ln(1-z_0) \to 0$, while the dilogarithm is continuous on $(-\infty, 1]$. Hence,

$$-\ln(z)\ln(1-z) = \text{Li}_2(z) + \text{Li}_2(1-z) - \text{Li}_2(1)$$

as desired. Note that proving the identity via integration by parts has to make this same limiting argument. □

B.3. Inversion formulas

The following lemma allows us to describe the survival function of $|X|$ with $X \sim \text{Cauchy}(1)$ in a convenient way. Note that the survival function for $|X|$ will only consider $z = x > 0$.

Lemma B.3.1. *For $z \in \mathbb{C} - i\mathbb{R}$,*

$$\arctan(z) + \arctan\left(\frac{1}{z}\right) = \begin{cases} \pi/2 & \text{if } \Re(z) > 0 \\ -\pi/2 & \text{if } \Re(z) < 0. \end{cases}$$

Remark B.3.2. On the imaginary axis, $\arctan(ir) = i\operatorname{atanh}(r)$ and atanh is only defined for $r \in (-1, 1)$ so $1/r$ does not make sense there. Consequently the domain in question has two connected components, so different constants should not be unexpected.

Proof. First note that the left hand side is a constant

$$\frac{d}{dz}\left(\arctan(z) + \arctan\left(\frac{1}{z}\right)\right) = \frac{1}{1+z^2} + \frac{1}{1+z^{-2}}\frac{-1}{z^2} = 0.$$

The constant is determined by representative points $z = \pm 1$ in the right and left hand planes respectively. □

References

[1] AILON, N. and CHAZELLE, B. (2009). The Fast Johnson-Lindenstrauss Transform and Approximate Nearest Neighbors*. *SIAM Journal on Computing* **39** 302–322. MR2506527

[2] BRINKMAN, B. and CHARIKAR, M. (2005). On the Impossibility of Dimension Reduction in L_1. *J. ACM* **52** 766–788. MR2176562

[3] CHAMBERS, J. M., MALLOWS, C. L. and STUCK, B. W. (1976). A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association* **71** 340–344. MR0415982

[4] CORAZZA, P. (1999). Introduction to Metric-Preserving Functions. *The American Mathematical Monthly* **106** 309–323. MR1682385

[5] DRINEAS, P., MAGDON-ISMAIL, M., MAHONEY, M. W. and WOODRUFF, D. P. (2012). Fast Approximation of Matrix Coherence and Statistical Leverage. *Journal of Machine Learning Research* **13** 32. arXiv: 1109.3843. MR3033372

[6] DRINEAS, P. and MAHONEY, M. W. (2016). RandNLA: Randomized Numerical Linear Algebra. *Communications of the ACM* **59** 80–90.

[7] INDYK, P. (2006). Stable Distributions, Pseudorandom Generators, Embeddings, and Data Stream Computation. *J. ACM* **53** 307–323. MR2238947

[8] JOHNSON, W. B. and LINDENSTRAUSS, J. (1984). Extensions of Lipschitz Mappings into a Hilbert Space. *Contemporary Mathematics* **26** 189–206. MR0737400

[9] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* **23**. Springer-Verlag, Berlin. MR1102015

[10] LEE, J. R. and NAOR, A. (2004). Embedding the Diamond Graph in L_p and Dimension Reduction in L_1. *Geometric & Functional Analysis GAFA* **14** 745–747. arXiv: math/0407520. MR2084978

[11] LEWIN, L. (1981). *Polylogarithms and Associated Functions*. North-Holland Publishing Co., New York-Amsterdam. MR0618278

[12] LI, P., HASTIE, T. J. and CHURCH, K. W. (2007). Nonlinear Estimators and Tail Bounds for Dimension Reduction in l_1 Using Cauchy Random Projections. *Journal of Machine Learning Research (JMLR)* **8** 2497–2532. arXiv: cs/0610155. MR2353840

[13] MATOUŠEK, J. (2008). On Variants of the Johnson–Lindenstrauss Lemma. *Random Structures and Algorithms* **33** 142–156. MR2436844

[14] MAXIMON, L. C. (2003). The Dilogarithm Function for Complex Argument. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **459** 2807–2819. MR2015991

[15] MILMAN, V. D. and SCHECHTMAN, G. (1986). *Asymptotic Theory of Finite-Dimensional Normed Spaces. Lecture Notes in Mathematics* **1200**. Springer-Verlag, Berlin. With an appendix by M. Gromov. MR0856576

[16] NEEDHAM, T. (1997). *Visual Complex Analysis*. The Clarendon Press, Oxford University Press, New York. MR1446490

[17] PISIER, G. (1983). On the Dimension of the l_p^n-Subspaces of Banach Spaces, for $1 \leq p < 2$. *Transactions of the American Mathematical Society* **276** 201–211. MR0684503

[18] STEIN, E. M. and SHAKARCHI, R. (2003). *Complex analysis. Princeton Lectures in Analysis* **2**. Princeton University Press, Princeton, NJ. MR1976398

[19] VEMPALA, S. S. (2004). *The Random Projection Method. DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **65**. American Mathematical Society, Providence, RI. MR2073630