Linear Dimension Reduction Approximately Preserving Level-Sets of the 1-Norm

by

Michael P. Casey

Department of Mathematics
Duke University

Date: ________________________________
Approved:

____
Sayan Mukherjee, Supervisor

____
Galen Reeves

____
Paul Bendich

____
Jonathan Mattingly

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Mathematics
in the Graduate School of
Duke University

2019

ABSTRACT

Linear Dimension Reduction Approximately Preserving Level-Sets of the 1-Norm

by

Michael P. Casey

Department of Mathematics
Duke University

Date: ____________________
Approved:

____
Sayan Mukherjee, Supervisor

____
Galen Reeves

____
Paul Bendich

____
Jonathan Mattingly

An abstract of a dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Mathematics
in the Graduate School of
Duke University

2019

Abstract

We choose a family of matrices $F : \mathbb{R}^D \to \mathbb{R}^k$ and a metric ρ on $\mathbb{R}^k$ such that with high probability, $\rho(F(x), F(y))$ is a strictly concave increasing function of $\|x - y\|_1 > 8\epsilon^2$ for $x, y \in \mathbb{R}^D$, up to a multiplicative error of $1 \pm \epsilon$. In particular, if X is a set of N points in $\mathbb{R}^D$, the target dimension k may be chosen as $C \ln^2(N^{c+2})/(\epsilon^2(1-\epsilon)^2)$, with C a constant and $\epsilon > N^{-c}$, to ensure all pairs of points of X of distance at least $8\epsilon^2$ are treated this way, with failure probability at most N^{-c} for $c > 1$. In some cases, distances smaller than $8\epsilon^2$ can also be addressed. For distances larger than $\sqrt{1+\epsilon}$, the target dimension can be reduced to $C \ln(N^{c+2})/(\epsilon^2(1-\epsilon)^2)$.

Acknowledgements

I should like to thank Professor Mukherjee for willing to be my advisor. I should also like to thank Professors Reeves, Bendich, and Mattingly for willing to be on my committee.

I should like to thank Mom, Dad, Katie, and everyone who has been praying for me throughout my time at Duke.

I should finally like to thank Mary, Joseph, and the Holy Trinity for helping me be patient during this project.

For the Blessed Virgin Mary

&

For Grandpa Art

Contents

Abstract **iv**

Acknowledgements **v**

1 Introduction **1**

1.1 Motivation 1

1.2 Main Theorem 2

2 Metric-Preserving Functions **4**

2.1 Metric-Preserving Functions 7

3 Setup for Proving Concentration Behavior **12**

3.1 Reduction to Studying a Single Coordinate 12

3.1.1 Cauchy Distribution 14

3.1.2 Concentration and the Moment Generating Function 17

3.1.3 Common Lemmas for Estimating the MGF 20

4 Proving Concentration **23**

4.1 Estimating the Moment Generating Function 23

4.2 Large Scales 26

4.2.1 Estimating Deviations of the Mean 33

4.3 Small Scales 37

4.4 Really Small Scales 42

4.4.1 Bounds on Maxima 47

5 The First and Second Moments **50**

5.1 1st Moment 55

5.2 An Auxiliary Mean 58

5.3 2nd Moment 60

6 Conclusions **68**

6.1 Conclusions 68

6.2 Conjectures 69

A Polylogarithms and Their Friends **71**

A.1 Arctan and the Inverse Tangent Integrals 80

A.2 Dilogarithm Properties 87

A.3 Inversion Formulas 88

B Basic Inequalities **90**

Bibliography **95**

Chapter 1

Introduction

1.1 Motivation

The Johnson-Lindenstrauss lemma [JL84] states that given a finite pointset $P \in \mathbb{R}^D$ and $0 < \epsilon < 1$, there are 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 $k = \Theta(\epsilon^{-2} \ln |P|)$. There are also results stating that analogous results for using the 1-norm do not hold. For example, in [BC05] and [LN04], 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, [BC05] shows k must be $N^{1/2 - O(\epsilon \ln(1/\epsilon))}$ if one wants $c = 1 + \epsilon$.

In light of these negative results, people have instead tried estimating the median of the coordinates of $F(x) - F(y)$ under suitable distributions of F, or using 1-homogeneous functions on those coordinates that allow estimating $\|x - y\|_1$. See [LHC07] for example. For k-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.

We relax 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 $k = \Theta(\epsilon^{-c} \ln^c |P|)$, 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", 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

$$\left\{x \in \mathbb{R}^D \mid \|x-y\| = t\right\}$$

look like scaled versions of spheres for the 1-norm 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. We shall say a bit more about these functions in chapter 2.

1.2 Main Theorem

Throughout, $\ln^a(x) := (\ln(x))^a$, and for $p \geq 1$, $\ell_p^k = \mathbb{R}^k$ with metric induced by the p-norm.

Theorem 1.2.0.1. *Let $P \subset \mathbb{R}^D$ be a set of N points. For $a \geq 0$, set*

$$\xi(a) := \ln(1+\sqrt{a}) + \frac{1}{2}\ln(1+a) \quad \textit{and} \quad \mu(a) := \operatorname{atanh}\left(\frac{\sqrt{2a}}{1+a}\right) + \frac{1}{2}\ln(1+a^2).$$

Equip $\mathbb{R}^k$ with the metric

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

For $1 \leq j \leq D$ and $1 \leq i \leq k$, let $F_{ij} \overset{i.i.d.}{\sim} \mathit{Cauchy}(1)$ be the entries of $F : \mathbb{R}^D \to \mathbb{R}^k$. If

$$c \geq 3, \quad 1/4 \geq \epsilon \geq N^{-c}, \quad \textit{and} \quad k \geq \frac{C}{\epsilon^2(1-\epsilon)^2}(\ln^2(N^c)),$$

then if $\|x-y\|_1 \geq \sqrt{1+\epsilon}$,

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

while if $\sqrt{1+\epsilon} \geq \|x-y\|_1 \geq 8\epsilon^2$,

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

and finally if $\|x-y\|_1 \leq 8\epsilon^2$,

$$(1-\epsilon)(1-4\epsilon^2)\mu\left(\|x-y\|_1\right) \leq \rho(F(x), F(y))$$

for all $x, y \in P$ *with probability at least* $1 - N^{-c-2}$.

Remark 1.2.0.2. We have not been able to establish an upper bound result

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

with high probability when $\|x-y\| < 8\epsilon^2$. 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.

The final bound for k is proved in corollary 4.4.1.2. The main idea is using the Cauchy random variables F_{ij} to encode $\|x-y\|_1$ in the coordinates of $F(x-y)$. These coordinates are still random, but applying the ρ metric yields a sum of i.i.d. random variables that concentrate about their mean, which necessarily depends on $\|x-y\|_1$. We are thus able to recover a function of $\|x-y\|_1$ this way. We had to choose the function ξ to grow logarithmically as Cauchy random variables only have fractional moments, while concentration phenomena usually require moments of all orders. We say more about this particular choice for ξ in chapter 5.

Chapter 2

Metric-Preserving Functions

In this chapter, we provide some intuition for the metric ideas we use. Denote (X, ρ) for a metrizable space X equipped with a metric ρ that induces its topology. As a function, $\rho : X \times X \to \mathbb{R}_+$ satisfies, for any $x, y, z \in X$,

- $\rho(x, y) = \rho(y, x)$
- $\rho(x, y) = 0$ just if $x = y$
- $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$.

In k-nearest neighbor applications, one is given a point set $P \subset \mathbb{R}^D$ and a query point q; the point q induces an ordering on the points of P via the ambient metric:

$$\|q - p_1\|_1 \leq \|q - p_2\|_1 \leq \|q - p_3\|_1 \dots$$

and one desires to recover p_1 through p_k or points that approximately realize such distances. Note that any monotone increasing function μ preserves this order:

$$\mu(\|q - p_1\|_1) \leq \mu(\|q - p_2\|_1) \leq \mu(\|q - p_3\|_1) \dots$$

If μ is a strictly concave, increasing function that takes 0 at 0, then $\mu \circ \|\cdot\|_1$ is a metric as well, as shown in section 2.1. Our theorem 1.2.0.1 shows that such orders above can be approximately preserved while projecting the points of P down to a lower dimensional space. However the metric we use in the target space has to be changed. Because k-nearest neighbor methods often rely on properties of the metric that is actually being used, we study such metrics ρ_f in this chapter.

Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be strictly concave, increasing, and satisfy $f(0) = 0$. Define the metric ρ_f on $\mathbb{R}^k$ as

$$\rho_f(x, y) := \sum_{j=1}^{k} f(|x_j - y_j|).$$

Note that by construction, the metric ρ_f is translation invariant on $\mathbb{R}^k$. We still call the sublevel set $B(t)$ the (closed) ball about 0 of radius t and $S(t) = \partial B(t)$ the corresponding sphere:

$$B(t) := \left\{x \in \mathbb{R}^k \mid \rho_f(x,0) \le t\right\} \quad \text{and} \quad \partial B(t) = S(t) := \left\{x \in \mathbb{R}^k \mid \rho_f(x,0) = t\right\}.$$

This notation is justified as $B(t)$ is star-shaped: for any $0 < \alpha < 1$, if $x \in S(t)$, then $\alpha x \in B(t)$ because f is strictly monotone increasing:

$$\rho_f(\alpha x, 0) = \sum_{j=1}^{k} f(\alpha |x_i|) < \sum_{j=1}^{k} f(|x_i|) = t.$$

To better understand the surface of $S(t)$, we focuse on the positive quadrant $\mathbb{R}^k_+ = [0,\infty)^k$. The surface of $S(t)$ in all other quadrants is just a reflection of the one in $\mathbb{R}^k_+$. Compare to Schur's Majorization Inequality [Ste04, chapter 13].

Lemma 2.0.0.1. *On $\mathbb{R}^k$, let ρ_f be as above and let $t > 0$. For all distinct $x, y \in S(t) \cap \mathbb{R}^k_+$, the open line segment (x, y) does not lie in $B(t)$. Concretely, if $0 < \alpha < 1$,*

$$\alpha x + (1-\alpha) y \notin B(t).$$

Remark 2.0.0.2. Contrast this behavior to the p-norms with $p \ge 1$. For the 1-norm, the associated balls are crosspolytopes: all such (x, y) lie on $S(t) \subset B(t)$. For $p > 1$, the norms are "strictly convex": all such (x, y) lie in the interior of $B(t)$, that is, in $B(t) - S(t)$.

Proof. We just need to show

$$\rho_f(\alpha x + (1-\alpha)y, 0) > t.$$

Because $x, y \in \mathbb{R}^k_+$ and $0 < \alpha < 1$,

$$\begin{aligned}
&\rho_f(\alpha x + (1-\alpha)y, 0) := \sum_{j=1}^{k} f(|\alpha x_j + (1-\alpha)y_j - 0|) = \sum_{j=1}^{k} f(\alpha x_j + (1-\alpha)y_j) \\
&> \sum_{j=1}^{k} \alpha f(x_j) + (1-\alpha) f(y_j) \\
&= \alpha \rho_f(x,0) + (1-\alpha)\rho_f(y,0) = \alpha t + (1-\alpha) t \\
&= t
\end{aligned}$$

because $x, y \in S(t)$. The strict inequality in the above is because f is strictly concave. $\square$

For further intuition on the shape of the sphere $S(t)$, consider comparing $S(t) \cap \mathbb{R}^k_+$ to dilations of the crosspolytope. Denote $y \in H(x)$ if y lies in the convex hull of x and all permutations of its coordinates:

$$x_\sigma := (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(k)})$$

with σ an element of the permutation group on k letters. Different σ may map to the same point if some of the coordinates of x are identical; however, if the coordinates of x are distinct, there are $k!$ such x_σ. Note that $\|y\|_1 = \|x\|_1 = \|x_\sigma\|_1$ for all such σ, so they all lie on the same codim-1 simplex of the corresponding crosspolytope. We shall show that when $x \in S(t)$, *only* the vertices of $H(x)$, that is only x and the x_σ lie on $S(t)$; all other points of $H(x)$ lie outside $B(t)$.

Lemma 2.0.0.3. *Let $x \in S(t) \cap \mathbb{R}^k_+$ for $t > 0$. If $y \in H(x) \cap S(t) \cap \mathbb{R}^k_+$, then*

$$y = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(k)})$$

for some permutation σ on k letters.

Proof. We just need to show $\rho_f(y, 0) > \rho_f(x, 0) = t$ for all $y \in H(x)$ that is not one of the x_σ. By definition,

$$y = (y_1, \ldots, y_k) = \sum_\sigma \alpha_\sigma x_\sigma = \sum_\sigma \alpha_\sigma (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(k)}).$$

with $\sum_\sigma \alpha_\sigma = 1$ and $0 \le \alpha_\sigma \le 1$. Focusing on a single coordinate,

$$y_j = \sum_\sigma \alpha_\sigma x_{\sigma(j)} = \sum_{l=1}^k \left(\sum_{\sigma \mid \sigma(j)=l} \alpha_\sigma \right) x_l = \sum_{l=1}^k d_{jl} x_l \quad \text{with} \quad d_{jl} := \sum_{\sigma \mid \sigma(j)=l} \alpha_\sigma.$$

Note that $\sum_l d_{jl} = 1 = \sum_j d_{jl}$ because $\sum_\sigma \alpha_\sigma = 1$. As a matrix, $\{d_{jl}\}$ is said to be *doubly-stochastic*. We now have

$$\rho_f(y, 0) = \sum_{j=1}^k f(y_j) = \sum_{j=1}^k f\left(\sum_{l=1}^k d_{jl} x_l \right) \ge \sum_{j=1}^k \sum_{l=1}^k d_{jl} f(x_l) = \sum_{l=1}^k f(x_l) = t.$$

Again, the inequality is strict as soon as one of the $d_{jl} < 1$, that is, when $y \neq x_\sigma$ for some permuation σ, because f is strictly convex. $\square$

2.1 Metric-Preserving Functions

There is a nice review on such functions in [Cor99].

Definition 2.1.0.1. A function $f : [0, \infty) \to [0, \infty)$ is *metric-preserving* if for any metric space (X, ρ), the function $f \circ \rho : X \times X \to [0, \infty)$ is also a metric on X. The function f is said to be *strongly metric-preserving* if $f \circ \rho$ induces the same topology on X as ρ does.

The embeddings we have in mind use nonlinear functions $f : [0, \infty) \to [0, \infty)$ which are monotone increasing and *subadditive*, that is,

$$f(x) \leq f(y) \quad \text{and} \quad f(x + y) \leq f(x) + f(y) \quad \text{for} \quad 0 \leq x \leq y \in \mathbb{R}.$$

Upon taking $x = 0$ in the second inequality, we note any subadditive $f : [0, \infty) \to \mathbb{R}$, must have $f(0) \geq 0$.

Lemma 2.1.0.2. *Any concave function $f : [0, \infty) \to \mathbb{R}$ satisfying $f(0) \geq 0$ is subadditive.*

Remark 2.1.0.3. Note that f need not be nonnegative or increasing either. Note also that

$$-f(x + y) \geq -f(x) + (-f(y)),$$

that is, convex functions that are 0 at 0 are *superadditive*.

Proof. We need to show

$$f(x + y) \leq f(x) + f(y)$$

for $0 \leq x, y \in \mathbb{R}$. If x or y is 0, there is nothing to show because $f(0) \geq 0$. We may thus assume $x, y > 0$. Let

$$\lambda = \frac{x}{x + y} \quad \text{so that} \quad 1 - \lambda = \frac{y}{x + y}.$$

By concavity,

$$f(x) = f(\lambda(x + y) + (1 - \lambda)0) \geq \lambda f(x + y) + (1 - \lambda)f(0) \geq \lambda f(x + y).$$

Similarly,

$$f(y) = f((1-\lambda)(x+y) + \lambda 0) \geq (1-\lambda)f(x+y) + \lambda f(0) \geq (1-\lambda)f(x+y).$$

Add to conclude

$$f(x) + f(y) \geq \lambda f(x+y) + (1-\lambda)f(x+y) = f(x+y).$$

□

Corollary 2.1.0.4. *Any increasing concave function $f : [0, \infty) \to [0, \infty)$ satisfying $f(0) = 0$ is strongly metric-preserving.*

Remark 2.1.0.5. In particular, f preserves the ordering of the points. If $x, y, z \in (X, \rho)$, then

$$\rho(x, y) \leq \rho(x, z) \Rightarrow f(\rho(x, y)) \leq f(\rho(x, z)).$$

Proof. Let (X, ρ) be a metric space. To check the triangle inequality, just note for $x, y, z \in X$,

$$f(\rho(x, y)) \leq f(\rho(x, z) + \rho(z, y)) \leq f(\rho(x, z)) + f(\rho(z, y))$$

the first inequality follows because f is monotone increasing, while the second follows from the previous lemma 2.1.0.2 because f is subadditive.

The composition $f(\rho(x, y))$ is symmetric in x and y because ρ is, and $f(\rho(x, y)) = 0$ just if $\rho(x, y) = 0$, because f is strictly increasing. Because ρ is a metric, $\rho(x, y) = 0$ just if $x = y$.

To see that $f \circ \rho$ generates the same topology on X as ρ, recall that on a metric space, the topology induced by the metric is generated by the open balls in that metric. We shall show there is a one-to-one correspondence between the $f \circ \rho$ balls and the ρ balls. Specifically, let

$$B_f(y, t) := \{x \in X \mid f(\rho(x, y)) < t\} \quad \text{and} \quad B_\rho(y, t) := \{x \in X \mid \rho(x, y) < t\}$$

Because f is strictly increasing, f is invertible and preserves the order of $\mathbb{R}$. Consequently,

$$B_f(y,t) = \left\{x \in X \mid \rho(x,y) < f^{-1}(t)\right\} = B_\rho(y, f^{-1}(t))$$

and similarly

$$B_\rho(y,r) = \{x \in X \mid f(\rho(x,y)) < f(r)\} = B_f(y, f(r)).$$

$\square$

We record some properties of concave increasing functions that we use.

Lemma 2.1.0.6. *If $f, g : \mathbb{R}_+ \to \mathbb{R}_+$ are concave increasing functions, then $f \circ g : \mathbb{R}_+ \to \mathbb{R}_+$ is concave and increasing as well.*

Proof. Let $0 < \lambda < 1$ and $x, y \in \mathbb{R}_+$. By the concavity of f,

$$\lambda f(g(x)) + (1-\lambda) f(g(y)) \leq f(\lambda g(x) + (1-\lambda) g(y)).$$

By the concavity of g,

$$\lambda g(x) + (1-\lambda) g(y) \leq g(\lambda x + (1-\lambda) y).$$

Because f is increasing, we consequently have

$$f(\lambda g(x) + (1-\lambda) g(y)) \leq f(g(\lambda x + (1-\lambda) y))$$

as desired. $\square$

Lemma 2.1.0.7. *Let $f : [0, \infty) \to \mathbb{R}$ be concave with $f(0) \geq 0$. Then, for $s > 1$,*

$$f(sx) \leq s f(x).$$

Similarly, for $0 < \epsilon < 1$,

$$f(\epsilon x) \geq \epsilon f(x).$$

Remark 2.1.0.8. So if f is increasing,

$$(1-\epsilon) f(x) \leq f((1-\epsilon)x) \leq f(x) \leq f((1+\epsilon)x) \leq (1+\epsilon) f(x).$$

Proof. Consider $x = (sx)/s$. If f is concave,

$$\frac{f(sx)}{s} \leq \left(1 - \frac{1}{s}\right) f(0) + \frac{1}{s} f(sx) \leq f\left(\left(1 - \frac{1}{s}\right) 0 + \frac{1}{s}(sx)\right) = f(x).$$

Setting $\epsilon = 1/s \in (0, 1)$ and $y = x/\epsilon$, we also recover from the above

$$\epsilon f(y) \leq f(\epsilon y).$$

□

By corollary 2.1.0.4, the following function is strictly metric-preserving.

Example 2.1.1. The function

$$\frac{x}{1+x}$$

on $\mathbb{R}_+$ is strictly increasing and strictly concave.

Proof. Let $g(x) = x/(1+x)$. Then for $x \geq 0$,

$$\frac{d}{dx}\frac{x}{1+x} = \frac{(1+x) - x}{(1+x)^2} = \frac{1}{(1+x)^2} > 0 \quad \text{and} \quad \frac{d^2}{dx^2}\frac{x}{1+x} = \frac{-2}{(1+x)^3} < 0.$$

□

The choice of μ in theorem 1.2.0.1 is due to the following lemma; we shall see in the next chapter, particularly lemma 3.1.2.1 why it is a good choice.

Lemma 2.1.0.9. *Let W be a nonnegative random variable with a continuous density, and suppose $\xi : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing concave function with $\xi(0) = 0$. Then the function $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ defined as*

$$\mu(\lambda) := \mathbb{E}\xi(\lambda W)$$

is also strictly increasing, concave, and satisfies $\mu(0) = 0$.

Proof. The condition $\mu(0) = 0$ is clear. Let $1 > s > 0$ and $\lambda_1, \lambda_2 > 0$. By assumption, ξ satisfies

$$s\xi(\lambda_1 w) + (1-s)\xi(\lambda_2 w) < \xi((s\lambda_1 + (1-s)\lambda_2)w) \quad \text{for any } w > 0.$$

If f is the density for W, we may integrate the above to conclude

$$\begin{aligned}
&s\mu(\lambda_1) + (1-s)\mu(\lambda_2) = s\mathbb{E}\xi(\lambda_1 W) + (1-s)\mathbb{E}\xi(\lambda_2 W) \\
&= \int_0^\infty \big(s\xi(\lambda_1 w) + (1-s)\xi(\lambda_2 w)\big) f(w)\,dw \\
&< \int_0^\infty \xi((s\lambda_1 + (1-s)\lambda_2)w) f(w)\,dw \\
&= \mathbb{E}\xi((s\lambda_1 + (1-s)\lambda_2)W) = \mu(s\lambda_1 + (1-s)\lambda_2)
\end{aligned}$$

as desired. □

Chapter 3

Setup for Proving Concentration Behavior

3.1 Reduction to Studying a Single Coordinate

Let $F : \mathbb{R}^D \to \mathbb{R}^k$ be a matrix (F_{ij}). If $v \in \mathbb{R}^D$, then the ith coordinate of $F(v)$ looks like

$$(F(v))_i = \sum_{j=1}^{D} F_{ij} v_j.$$

We show below that if the F_{ij} are independent identically distributed p-stable random variables, then by remark 3.1.0.4,

$$(F(v))_i \sim \|v\|_p F_{iD}.$$

Consequently,

$$\rho(F(v), 0) = \frac{1}{k} \sum_{i=1}^{k} \xi(\|v\|_p F_{iD})$$

and our goal is to show that this sum concentrates about its mean when k is large enough.

Definition 3.1.0.1 (p-Norms)**.** For $1 \leq p < \infty$ and $v \in \mathbb{R}^D$, the p-norm of v is

$$\|v\|_p := \left(\sum_{j=1}^{D} |v_j|^p \right)^{1/p}.$$

The associated metric induced by the p-norm on $\mathbb{R}^D$ is

$$\rho_p(x, y) := \|x - y\|_p = \left(\sum_{j=1}^{D} |x_j - y_j|^p \right)^{1/p}$$

These norms are convenient in part because they are "positively" 1-homogeneous,

$$\|Cv\|_p = C \|v\|_p \quad \text{for} \quad C \geq 0.$$

In particular, when $\|v\|_p > 0$, $v/\|v\|_p$ has p-norm 1. Given the nonexistence results in [BC05] and [LN04], the metric ρ that we choose will not have this scaling property. It will still be translation invariant though.

The following definition may be found in [MS86, chapter 8].

Definition 3.1.0.2 (Symmetric p-Stable Random Variables)**.** For $0 < p \leq 2$, a random variable W is drawn from the standard symmetric p-stable distribution if

$$\mathbb{E}\exp(itW) = \exp\left(-\frac{|t|^p}{p}\right).$$

Such random variables have the following useful property.

Lemma 3.1.0.3. *For $1 \leq j \leq D$, let $W_j \overset{i.i.d.}{\sim} W$ with W standard symmetric and p-stable with $2 \geq p \geq 1$. Then if $v \in \mathbb{R}^D$,*

$$\mathbb{E}\exp\left(it\sum_{j=1}^{D} W_j v_j\right) = \exp\left(-\frac{|t|^p}{p}\|v\|_p^p\right).$$

Remark 3.1.0.4. So if $\|v\|_p = 1$, we have a new standard symmetric p-stable random variable $\sum W_j v_j$, and if $x \neq 0 \in \mathbb{R}^D$,

$$\sum_{j=1}^{D} W_j x_j = \|x\|_p \sum_{j=1}^{D} W_j \frac{x_j}{\|x\|_p} \sim \|x\|_p W \left\|\frac{x_j}{\|x\|_p}\right\|_p = \|x\|_p W.$$

That is, the distribution of the sum carries the p-norm information of x. We shall show in lemma 3.1.1.5 that Cauchy random variables are 1-stable.

Proof. By independence,

$$\mathbb{E}\exp\left(it\sum_{j=1}^{D} W_j v_j\right) = \prod_{j=1}^{D}\mathbb{E}\exp(itv_jW_j) = \prod_{j=1}^{D}\exp(-|tv_j|^p/p)$$
$$= \exp\left(-\frac{|t|^p}{p}\sum_{j=1}^{D}|v_j|^p\right) = \exp\left(-\frac{|t|^p}{p}\|v\|_p^p\right).$$

□

3.1.1 Cauchy Distribution

Definition 3.1.1.1. The symmetric Cauchy distribution with parameter $\lambda > 0$, denoted Cauchy (λ), has probability density function

$$f_\lambda(x) := \frac{\lambda}{\pi(\lambda^2 + x^2)}.$$

Remark 3.1.1.2. Check

$$\int_{-\infty}^{\infty} f_\lambda(x)\,dx = \frac{2\lambda}{\pi}\int_0^\infty \frac{1}{\lambda^2+x^2}\,dx = \frac{2}{\lambda\pi}\int_0^\infty \left(1+\frac{x^2}{\lambda^2}\right)^{-1}\,dx$$

with $u = x/\lambda$,

$$= \frac{2}{\pi}\int_0^\infty (1+u^2)^{-1}\,du$$

with $u = \tan(v)$,

$$= \frac{2}{\pi}\int_0^{\frac{\pi}{2}} (1+\tan(v)^2)^{-1}\sec(v)^2\,dv = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} dv = 1$$

as desired.

Lemma 3.1.1.3. *Let $X \sim$ Cauchy(1). Then the distribution function of $|X|$ is*

$$\mathbb{P}\{|X| \le t\} = \frac{2}{\pi}\arctan(t).$$

Remark 3.1.1.4. So, by the inversion formula for arctan A.3.0.1,

$$\mathbb{P}\{|X| > t\} = 1 - \frac{2}{\pi}\arctan(t) = \frac{2}{\pi}\arctan\left(\frac{1}{t}\right).$$

Proof. For $t \ge 0$, compute

$$\mathbb{P}\{|X| \le t\} = \mathbb{P}\{-t \le X \le t\} = \frac{1}{\pi}\int_{-t}^{t}\frac{1}{1+x^2}\,dx = \frac{2}{\pi}\int_0^t \frac{1}{1+x^2}\,dx$$

with $x = \tan(v)$

$$= \frac{2}{\pi}\int_0^{\arctan(t)} (1+\tan^2(v))^{-1}\sec^2(v)\,dv$$
$$= \frac{2}{\pi}(\arctan(t) - \arctan(0)) = \frac{2}{\pi}\arctan(t).$$

□

We verify that $X \sim \text{Cauchy}\,(1)$ is 1-stable.

Lemma 3.1.1.5. *Let $X \sim$ Cauchy$\,(1)$, then for $t \in \mathbb{R}$,*

$$\mathbb{E}\exp(itX) = \exp(-|t|).$$

Proof. We consider the contour integral

$$\frac{1}{\pi}\int_C \frac{\exp(itz)}{1+z^2}\,dz$$

with the contour C the half-circle of radius R in the upper half plane, together with the interval $[-R, R]$ for $R > 0$. The contour is oriented counter clockwise. We intend to take $R \to \infty$, and as soon as $R > 1$, the contour encloses $z = i$. Because

$$\frac{1}{1+z^2} = \frac{1}{(z-i)(z+i)}$$

the integrand contains an isolated simple pole at $z = i$, so by the residue formula [SS03, page 75-76 chapter 3],

$$\begin{aligned}
&\frac{1}{\pi}\int_C \frac{\exp(itz)}{1+z^2}\,dz = \frac{1}{\pi}\frac{2i\pi}{\operatorname{res}_{z=i}}\frac{\exp(itz)}{1+z^2}\\
&= 2i\lim_{z\to i}(z-i)\frac{\exp(itz)}{1+z^2}\\
&= 2i\lim_{z\to i}\exp(itz)(z-i)\left(\frac{1}{(z-i)(z+i)}\right)\\
&= 2i\lim_{z\to i}\exp(itz)\frac{1}{z+i}\\
&= \exp(-t).
\end{aligned}$$

For $z = Re^{i\theta}$ with $0 < \theta < \pi$,

$$\exp(itz) = \exp(itR(\cos\theta + i\sin\theta)) = \exp(itR\cos\theta)\exp(-tR\sin\theta)$$

which goes to 0 as $R \to \infty$ because $t > 0$. We may also assume $R > \sqrt{2}$ in order to use lemma 3.1.1.6:

$$\left|\frac{1}{1+z^2}\right| \leq \frac{1}{R^2}\frac{1}{R^{-4} + (1-2R^{-2})}.$$

Consequently, with C_+ the semicircle of radius R,

$$\left|\int_{C_+} \frac{\exp(itz)}{1+z^2}\, dz\right| \leq \int_{C_+} \left|\frac{\exp(itz)}{1+z^2}\right| \, dz \leq \frac{\exp(-tR\sin\theta)}{R^2} \frac{\pi R}{R^{-4}+(1-2R^{-2})} \to 0$$

for $t > 0$ and $R \to \infty$. We can conclude

$$\exp(-t) = \lim_{R\to\infty} \frac{1}{\pi} \int_{-R}^{R} \frac{\exp(itx)}{1+x^2}\, dx + \lim_{R\to\infty} \frac{1}{\pi} \int_{C_+} \frac{\exp(itz)}{1+z^2}\, dz = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\exp(itx)}{1+x^2}\, dx.$$

When $t < 0$, we have to use the opposite semicircle, with the closed contour now oriented clockwise. The same bounds now hold, as $-\pi < \theta < 0$ makes

$$\exp(itz) = \exp(itR\cos\theta)\exp(-tR\sin\theta)$$

have magnitude at most 1, while the residue is now taken at $z = -i$:

$$\begin{aligned}
&\frac{1}{\pi} \int_{C} \frac{\exp(itz)}{1+z^2}\, dz = -\frac{1}{\pi} \frac{2i\pi}{\operatorname{res}_{z=-i}} \frac{\exp(itz)}{1+z^2} \\
&= -2i \lim_{z\to -i} (z-(-i)) \frac{\exp(itz)}{1+z^2} \\
&= -2i \lim_{z\to -i} \exp(itz)(z+i) \left(\frac{1}{(z+i)(z-i)}\right) \\
&= -2i \lim_{z\to -i} \exp(itz) \frac{1}{z-i} \\
&= \exp(t)
\end{aligned}$$

with the initial minus sign because the contour is clockwise. □

Lemma 3.1.1.6. *Let $z = re^{i\theta} \in \mathbb{C}$ with $r > 0$. Then if $r > \sqrt{2}$,*

$$\left|\frac{1}{1+z^2}\right| \leq \frac{1}{r^2} \frac{1}{\sqrt{r^{-4}+(1-2r^{-2})}},$$

while if $|\theta| \leq \pi/4$,

$$\left|\frac{1}{1+z^2}\right| \leq \frac{1}{\sqrt{1+r^4}}.$$

Proof. With $z = re^{i\theta}$,

$$\begin{aligned}
&1 + z^2 = 1 + r^2 e^{2i\theta} = 1 + r^2(\cos(2\theta) + i\sin(2\theta)) \\
&= 1 + r^2\cos(2\theta) + ir^2\sin(2\theta)
\end{aligned}$$

so

$$\begin{aligned}
\left|1+z^2\right| &= (1+r^4\cos^2(2\theta)+2r^2\cos(2\theta)+r^4\sin^2(2\theta))^{1/2} \\
&= (1+r^4+2r^2\cos(2\theta))^{1/2} \\
&\geq (1+r^4(1-2/r^2))^{1/2} = r^2(1/r^4+(1-2/r^2))^{1/2}
\end{aligned}$$

If $r=|z|>\sqrt{2}$, all terms in the lower bound are positive. Hence,

$$\left|\frac{1}{1+z^2}\right| \leq \frac{1}{r^2}\frac{1}{\sqrt{r^{-4}+(1-2r^{-2})}}.$$

On the other hand, if $|\theta|\leq \pi/4$,

$$\left|1+z^2\right| = (1+r^4+2r^2\cos(2\theta))^{1/2} \geq (1+r^4)^{1/2} = r^2(1+1/r^4)^{1/2}$$

so that

$$\left|\frac{1}{1+z^2}\right| \leq \frac{1}{\sqrt{1+r^4}}.$$

□

3.1.2 Concentration and the Moment Generating Function

From our initial discussion of p-stable random variables and remark 3.1.0.4, taking each entry F_{ij} of the matrix $F:\mathbb{R}^D\to\mathbb{R}^k$ as a standard symmetric p-stable random variable $W_i\sim W$ makes each of the k coordinates $F(v)_i$ have a distribution like $\|v\|_p W$. These k coordinates are still random though, so more work has to be done to recover information related to $\|v\|_p$. If $\xi:\mathbb{R}_+\to\mathbb{R}_+$ is strictly increasing, and hence invertible, one would hope that the empirical average

$$\frac{1}{k}\sum_{i=1}^{k}\xi(\|v\|_p|W_i|) \quad \text{deviates little from its mean} \quad \mathbb{E}\frac{1}{k}\sum_{i=1}^{k}\xi(\|v\|_p|W_i|)=\mathbb{E}\xi(\|v\|_p|W|).$$

If that were the case, the function $\lambda\mapsto\mathbb{E}\xi(\lambda|W|)$, which is invertible, could be inverted to recover $\|v\|_p$ up to some error. When the emperical average behaves this way, we say it *concentrates* about its mean. The following lemma, which bounds the probabilities that the empirical average can be far from the mean, is a standard first step in showing concentration. The lemma will allow us to transition from considering sums of independent random variables to just the behavior of a single random variable.

Lemma 3.1.2.1. *For $1 \le i \le k$, let $W_i \overset{i.i.d.}{\sim} W$. Let $\xi : \mathbb{R}_+ \to \mathbb{R}_+$ and*

$$\mu(\lambda) := \mathbb{E}\xi(\lambda\,|W|).$$

Then for $s > 0$, and $\lambda_+ > \lambda > \lambda_-$,

$$\mathbb{P}\left\{\frac{1}{k}\sum_{i=1}^{k}\xi(\lambda\,|W_i|) > \mu(\lambda_+)\right\} \le \Big(\exp(-s\mu(\lambda_+))\mathbb{E}\exp(s\xi(\lambda\,|W|))\Big)^k,$$

and

$$\mathbb{P}\left\{\frac{1}{k}\sum_{i=1}^{k}\xi(\lambda\,|W_i|) < \mu(\lambda_-)\right\} \le \Big(\exp(s\mu(\lambda_-))\mathbb{E}\exp(-s\xi(\lambda\,|W|))\Big)^k.$$

Remark 3.1.2.2. Alternatively with

$$\Delta_+ := \mu(\lambda_+) - \mu(\lambda) \quad \text{and} \quad \Delta_- = \mu(\lambda) - \mu(\lambda_-),$$

the linearity of the expectation allows us to rewrite the above bounds as

$$\mathbb{P}\left\{\frac{1}{k}\sum_{i=1}^{k}\xi(\lambda\,|W_i|) > \mu(\lambda_+)\right\} \le \Big(\exp(-s\Delta_+)\mathbb{E}\exp(s(\xi(\lambda\,|W|) - \mu(\lambda)))\Big)^k,$$

and

$$\mathbb{P}\left\{\frac{1}{k}\sum_{i=1}^{k}\xi(\lambda\,|W_i|) < \mu(\lambda_-)\right\} \le \Big(\exp(-s\Delta_-)\mathbb{E}\exp(-s(\xi(\lambda\,|W|) - \mu(\lambda)))\Big)^k.$$

This formulation allows knowledge of the variance to come into play, but makes the lower tail proof less straightforward.

Proof. We use Markov's inequality B.0.0.2 for nonnegative random variables. With $s > 0$,

$$\mathbb{P}\left\{\frac{1}{k}\sum_{i=1}^{k}\xi(\lambda\,|W_i|) > \mu(\lambda_+)\right\} = \mathbb{P}\left\{s\sum_{i=1}^{k}\xi(\lambda\,|W_i|) > sk\mu(\lambda_+)\right\}$$

$$= \mathbb{P}\left\{\exp\left(s\sum_{i=1}^{k}\xi(\lambda\,|W_i|)\right) > \exp(sk\mu(\lambda_+))\right\}$$

$$\leq \exp(-sk\mu(\lambda_+))\mathbb{E}\exp\left(s\sum_{i=1}^{k}\xi(\lambda\,|W_i|)\right)$$

$$= \exp(-sk\mu(\lambda_+))\prod_{i=1}^{k}\mathbb{E}\exp(s\xi(\lambda\,|W_i|)) = \Big(\exp(-s\mu(\lambda_+))\mathbb{E}\exp(s\xi(\lambda\,|W|))\Big)^k$$

using independence of the W_i and then that $W_i \sim W$ in the last line.

Similarly,

$$\mathbb{P}\left\{\frac{1}{k}\sum_{i=1}^{k}\xi(\lambda\,|W_i|) < \mu(\lambda_-)\right\} = \mathbb{P}\left\{-s\sum_{i=1}^{k}\xi(\lambda\,|W_i|) > -sk\mu(\lambda_-)\right\}$$

$$= \mathbb{P}\left\{\exp\left(-s\sum_{i=1}^{k}\xi(\lambda\,|W_i|)\right) > \exp(-sk\mu(\lambda_-))\right\}$$

$$\leq \exp(sk\mu(\lambda_-))\mathbb{E}\exp\left(-s\sum_{i=1}^{k}\xi(\lambda\,|W_i|)\right)$$

$$= \exp(sk\mu(\lambda_-))\prod_{i=1}^{k}\mathbb{E}\exp(-s\xi(\lambda\,|W_i|)) = \Big(\exp(s\mu(\lambda_-))\mathbb{E}\exp(-s\xi(\lambda\,|W|))\Big)^k$$

□

The plan is then to minimize the right hand sides over s, which usually requires finding good upper bounds for the moment generating function

$$\mathbb{E}\exp(\pm sY) \quad \text{with} \quad Y = \xi(\lambda\,|W|) \quad \text{or} \quad Y = \xi(\lambda\,|W|) - \mu(\lambda)$$

as a function of s. Even in cases where the moment generating function $\mathbb{E}\exp(s\xi(\lambda\,|W|))$ is explicitly known, such minimization might not be easy to do, sometimes because the derivatives in s are functions which are difficult to bound well. Often however, having a good upper bound on the moment generating function for which s can be optimized is sufficient, as will be the case here. In the next chapter, we shall derive the actual estimates for the Cauchy case, and show how they dictate the choice of the target dimension k. The following lemmas will be used there.

3.1.3 Common Lemmas for Estimating the MGF

Lemma 3.1.3.1. *Let Y be a random variable with distribution function F and density f continuous on $[a, b]$. Let $g : \mathbb{R} \to \mathbb{R}$ a continuously differentiable function. Then, if $\mathbb{E}g(Y) < \infty$,*

$$\mathbb{E}g(Y)\mathbb{I}\{a \leq Y \leq b\} = \int_a^b g'(t)\mathbb{P}\{Y > t\}\, dt + g(a)(1 - F(a)) - g(b)(1 - F(b))$$

for $a \leq b \in \mathbb{R}$.

Proof. The proof is via integration by parts. If

$$F(t) := \mathbb{P}\{Y \leq t\}$$

is the distribution function for Y, then $1 - F(t)$ goes to 0 as $t \to \infty$. If $F' = f$ with f continuous on $[a, b]$,

$$\begin{aligned}
&\int_a^b g(y)\, dF(y) = \int_a^b g(y) f(y)\, dy = \int_a^b g(y) \frac{d}{dy}(-(1 - F(y)))\, dy \\
&= g(y)(-(1 - F(y)))|_a^b - \int_a^b g'(y)(-(1 - F(y)))\, dy \\
&= -g(y)\mathbb{P}\{Y > y\}|_a^b + \int_a^b g'(y)\mathbb{P}\{Y > y\}\, dy \\
&= g(a)\mathbb{P}\{Y > a\} - g(b)\mathbb{P}\{Y > b\} + \int_a^b g'(y)\mathbb{P}\{Y > y\}\, dy.
\end{aligned}$$

□

Lemma 3.1.3.2. *For $0 < s$ and $0 < u$,*

$$\exp(-s/u) \leq \left(\frac{2}{es}\right)^2 u^2$$

The statement is only useful for small u, say $0 < u \leq 1$.

Proof. We want to compare $\exp(-s/u)$ to c^2u^2 with c depending on s. Taking logs,

$$\begin{aligned}
&-\frac{s}{u} \leq 2\ln(c) + 2\ln(u) \\
&-s \leq 2u\ln(c) + 2u\ln(u)
\end{aligned}$$

Minimize the right-hand side in u

$$0 = 2\ln(c) + 2\ln(u) + 2 \Rightarrow -\ln(c) - 1 = \ln(u) \Rightarrow \frac{1}{ce} = u$$

and at this value of u,

$$2u^* \ln(c) + 2u^* \ln(u^*) = \frac{2\ln(c)}{ce} - \frac{2}{ce}\ln(ce)$$
$$= \frac{2}{ce}(\ln(c) - \ln(c) - \ln(e)) = \frac{-2}{ce}$$

So we require c to be

$$-s \leq -2/(ce) \Rightarrow -c \leq -2/(es) \Rightarrow c \geq 2/(es)$$

We take equality. □

Lemma 3.1.3.3. *For* $0 \leq t \leq 1$*,*

$$\exp(t) \leq 1 + t + \frac{e-1}{2}t^2.$$

For $t \leq 0$*,*

$$\exp(t) \leq 1 + t + \frac{t^2}{2}.$$

In particular, for all $t \leq 1$*,*

$$\exp(t) \leq 1 + t + \frac{e-1}{2}t^2 \leq 1 + t + t^2.$$

I doubt the smaller constants are tremendously useful. The bound for $t \leq 0$ is used for the lower tail with no need to split the MGF integral if one uses the 2nd moment by itself. If one uses the variance, t no longer has a sign and the more general bound needs to be used.

Proof. Because $\exp(u)$ is convex, if $0 \leq u \leq 1$ we may write $\exp(u)$ as

$$\exp(u \cdot 1 + (1-u) \cdot 0) \leq u\exp(1) + (1-u)\exp(0) = 1 + (e-1)u \leq 1 + 2u.$$

Consequently,

$$\exp(t) - 1 = \int_0^t \exp(u)\,du \le \int_0^t (1 + (e-1)u)\,du = t + \frac{e-1}{2}t^2$$

that is,

$$\exp(t) \le 1 + t + \frac{e-1}{2}t^2 \le 1 + t + t^2$$

For the $t \le 0$ case, Taylor's theorem with remainder (about $t = 0$) gives

$$\exp(t) = 1 + t + \frac{t^2}{2}\exp(\xi)$$

for some $\xi \le 0$. Because $\exp(\xi)$ is monotone increasing, we have

$$\exp(t) \le 1 + t + \frac{t^2}{2}\exp(0) = 1 + t + \frac{t^2}{2} \le 1 + t + t^2.$$

Taylor's theorem with Lagrange remainder about $t = 0$ also shows $\exp(t) \ge 1 + t$ for all $t \in \mathbb{R}$ as the remainder term is always nonnegative. $\square$

Chapter 4

Proving Concentration

In this chapter, we shall prove bounds of the form

$$\exp(s\mu(\lambda_-))\mathbb{E}\exp(-s\xi(\lambda\,|W|)) \leq \exp\left(-\frac{\Delta_-^2}{4(V^2+A_-)}\right)$$

and

$$\exp(-s\mu(\lambda_+))\mathbb{E}\exp(s\xi(\lambda\,|W|)) \leq \exp\left(-\frac{\Delta_+^2}{4(V^2+A_+)}\right)$$

for special choices of s, with $A_\pm$ functions of λ and V^2 an upper bound on either the second moment or the variance for $\xi(\lambda\,|W|)$. We provide estimates for the reciprocals of the exponential rates in order to estimate the target dimension k. By lemma 3.1.2.1, taking k as

$$\ln(2/\delta)\max\left\{\frac{4(V^2+A_+)}{\Delta_+^2},\frac{4(V^2+A_-)}{\Delta_-^2}\right\}$$

ensures

$$\mu(\lambda_-) \leq \frac{1}{k}\sum_{j=1}^{k}\xi(\lambda\,|W_i|) \leq \mu(\lambda_+)$$

with probability at least $1-\delta$. Taking $\delta < N^c$ with $c \geq 3$ ensures that the above bound holds for all $\binom{N}{2} < N^2$ pairs of points, with total probability at least

$$1-\delta N^2 > 1-\frac{N^2}{N^c} \geq 1-\frac{1}{N}.$$

4.1 Estimating the Moment Generating Function

We modify an argument from [Mat08], which will allow us to focus on estimating $\mathbb{P}\{Y > t\}$ for Y the desired random variable. The next lemma is the crux of that argument.

Lemma 4.1.0.1. *Let $0 < u < 1$ and Y a random variable with continuous distribution function F and continuous density f on $(0, \infty)$. Then if $\mathbb{E}\exp(uY) < \infty$,*

$$\mathbb{E}\exp(uY)\mathbb{I}\{Y \leq 1/u\} \leq 1 + u\mathbb{E}Y + u^2\mathbb{E}Y^2$$

and

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

Proof. For the first integral, let F be the distribution function for Y, that is,

$$F(t) := \mathbb{P}\{Y \leq t\}.$$

By lemma 3.1.3.3,

$$\mathbb{E}\exp(uY)\mathbb{I}\{Y \leq 1/u\} = \int_{-\infty}^{1/u} \exp(uy)\, dF(y) \leq \int_{-\infty}^{1/u} (1 + uy + u^2y^2)\, dF(y)$$
$$\leq \int_{-\infty}^{\infty} (1 + uy + u^2y^2)\, dF(y) = 1 + u\mathbb{E}Y + u^2\mathbb{E}Y^2$$

as $uy \leq 1$ here.

For the second integral, use lemma 3.1.3.1.

$$\mathbb{E}\exp(uY)\mathbb{I}\{Y > 1/u\} = \int_{1/u}^{\infty} \exp(uy)\, dF(y)$$
$$= e\mathbb{P}\{Y > 1/u\} + \int_{1/u}^{\infty} u\exp(uy)\mathbb{P}\{Y > y\}\, dy$$

in which we have assumed the survival function $\mathbb{P}\{Y > y\}$ decays faster than e^{-y} in order to address the boundary term. Having assumed this function is also continuous, the usual change of variables $t = uy$ yields

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

□

To estimate the survival functions, we first establish what they are for us.

Lemma 4.1.0.2. *For $0 < \alpha \leq 1$, $0 < \lambda, t$, and $X \sim$ Cauchy(1),*

$$\mathbb{P}\{\ln(1 + \lambda^\alpha |X|^\alpha) > t\} = \frac{2}{\pi}\arctan\left(\frac{\lambda}{(\exp(t) - 1)^{1/\alpha}}\right).$$

and is differentiable for $t > 0$.

If there is a constant in front of the logarithm, just rescale t in the final result.

Proof. We have by the arctan inversion formula A.3.0.1

$$\mathbb{P}\left\{\ln(1+\lambda^\alpha |X|^\alpha) > t\right\} = \mathbb{P}\left\{\lambda^\alpha |X|^\alpha > \exp(t) - 1\right\}$$
$$= \mathbb{P}\left\{|X| > \frac{1}{\lambda}(\exp(t)-1)^{1/\alpha}\right\}$$
$$= \frac{2}{\pi}\arctan\left(\frac{\lambda}{(\exp(t)-1)^{1/\alpha}}\right) = 1 - \frac{2}{\pi}\arctan\left(\frac{(\exp(t)-1)^{1/\alpha}}{\lambda}\right).$$

As a composition of differentiable functions, the survival function above is differentiable for $t > 0$. Because $0 < \alpha \leq 1$, the derivative is continuous too with a finite limit as t goes to 0. □

We specialize lemma 4.1.0.2 to $\alpha = 1/2$ for more workable estimates.

Lemma 4.1.0.3. *Let $X \sim$ Cauchy(1) and $\lambda > 0$. Then if $t \geq 2$,*

$$\mathbb{P}\left\{\xi(\lambda|X|) > t\right\} \leq C_1(\lambda)\exp(-t) \quad \text{with} \quad C_1(\lambda) := \frac{2}{\pi}\frac{\lambda}{(1-1/e)^2}$$

While if $2\ln(1+\sqrt{\lambda}) \leq t$,

$$\mathbb{P}\left\{\xi(\lambda|X|) > t\right\} \leq C_2(\lambda)\exp(-t/2) \quad \text{with} \quad C_2(\lambda) := \frac{2}{\pi}(1+\sqrt{\lambda}).$$

Proof. First, because $\xi(a) \leq 2\ln(1+\sqrt{a})$ for all $a \geq 0$,

$$\left\{\xi(\lambda|X|) > t\right\} \subseteq \left\{2\ln(1+\sqrt{\lambda|X|}) > t\right\}$$

so that, using lemma 4.1.0.2 with $\alpha = 1/2$,

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

Now, for all $\lambda > 0$,

$$\frac{2}{\pi}\arctan\left(\frac{\lambda}{(\exp(t/2)-1)^2}\right) \leq \frac{2}{\pi}\frac{\lambda}{(\exp(t/2)-1)^2} = \frac{2\lambda}{\pi}\frac{\exp(-t)}{(1-\exp(-t/2))^2}$$
$$\leq \frac{2\lambda}{\pi(1-\exp(-t_0/2))^2}\exp(-t)$$

for all $t \geq t_0 > 0$ because

$$\frac{d}{dt}(1 - \exp(-t/2)) = \frac{1}{2}\exp(-t) > 0.$$

On the other hand, if $t \geq 2\ln(1+\sqrt{\lambda})$, we then have

$$\begin{aligned}
&\frac{2}{\pi}\arctan\left(\frac{\lambda}{(\exp(t/2)-1)^2}\right) \\
&= \frac{2}{\pi}\exp(-t/2)\frac{\lambda}{(\exp(t/2)-1)(1-\exp(-t/2))} \\
&\leq \frac{2}{\pi}\exp(-t/2)\frac{\lambda}{\sqrt{\lambda}(1-1/(1+\sqrt{\lambda}))} \\
&= \frac{2}{\pi}\exp(-t/2)\frac{\sqrt{\lambda}(1+\sqrt{\lambda})}{\sqrt{\lambda}} \\
&= \frac{2}{\pi}\exp(-t/2)(1+\sqrt{\lambda})
\end{aligned}$$

which is bounded above by $2/\pi$ for the t in question. □

4.2 Large Scales

Lemma 4.2.0.1 (Upper Tail, Large Scales)**.** *For $1/2 > u > 0$, $X \sim$ Cauchy(1), $Y = \xi(\lambda\,|X|) - \mu(\lambda)$, and $V^2 \geq \mathbb{E}Y^2$,*

$$\exp(-u\Delta_+)\mathbb{E}\exp(uY)$$

can be minimized to

$$\exp\left(-\frac{\Delta_+^2}{4(V^2+A_+)}\right) \quad \text{at} \quad u = \frac{\Delta_+}{2(V^2+A_+)},$$

with A_+ a bounded nonnegative function of $\lambda \geq 1$.

In particular, for $\epsilon \leq 1/4$, $\lambda > 1/\sqrt{1+\epsilon}$, and

$$\Delta_+ = \mu((1+\epsilon)\lambda) - \mu(\lambda),$$

we have the bound

$$\frac{4(V^2+A_+)}{\Delta_+^2} \leq \frac{64}{\epsilon^2(1-\epsilon)^2}\left(\frac{\pi^2}{2} + \frac{64\pi}{e(\pi^2-1/2)}\right).$$

Remark 4.2.0.2. This bound is not tight; I believe there are better ways to estimate the A_+ term, possibly by iterating the argument at the end of the proof.

Proof. We break up $\mathbb{E}\exp(uY)$ into two integrals using lemma 4.1.0.1. The first integral is

$$\mathbb{E}\exp(uY)\mathbb{I}\{uY \leq 1\} \leq \mathbb{E}(1+uY+u^2Y^2)\mathbb{I}\{uY \leq 1\} \leq \mathbb{E}(1+uY+u^2Y^2) \leq 1+u^2V^2$$

with $V^2 \geq \mathbb{E}Y^2$.

The second integral is

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

We thus need an upper bound on

$$\mathbb{P}\{Y > t/u\} = \mathbb{P}\{\xi(\lambda\,|X|) > \mu(\lambda) + t/u\}$$

for $t \geq 1$.

We want to use lemma 4.1.0.3 to estimate these tail probabilities, so we compare

$$\mu(\lambda) + t/u \overset{?}{\geq} 2\ln(1+\sqrt{\lambda})$$

Using the exact formula for $\mu(\lambda)$ from lemma 5.1.0.1 and noting the atanh contribution is nonnegative by lemma 5.1.0.2,

$$\mu(\lambda) + t/u \geq \frac{1}{2}\ln(1+\lambda^2) + t/u \geq \frac{1}{2}\ln(1+\lambda^2) + 2$$

because $1/2 > u > 0$. For $\lambda \leq 1$, we certainly have

$$2\ln(1+\sqrt{\lambda}) \leq 2\sqrt{\lambda} \leq 2 \leq \frac{1}{2}\ln(1+\lambda^2) + 2.$$

For $\lambda \geq 1$, we have

$$2\ln(1+\sqrt{\lambda}) = 2\ln(\sqrt{\lambda}) + 2\ln(1+1/\sqrt{\lambda}) = \ln(\lambda) + 2\ln(1+1/\sqrt{\lambda}) \leq \ln(\lambda) + 2\ln(2)$$

while

$$\frac{1}{2}\ln(1+\lambda^2) + 2 = 2 + \ln(\lambda) + \frac{1}{2}\ln(1+1/\lambda^2) > 2 + \ln(\lambda).$$

Because $\ln(2) < 1$, we are ok here too.

With $C_2(\lambda)$ the function in lemma 4.1.0.3,

$$e\mathbb{P}\{Y > 1/u\} = e\mathbb{P}\{\xi(\lambda|X|) > \mu + 1/u\} \le eC_2 \exp\left(-\frac{\mu + 1/u}{2}\right)$$

We also have from that lemma

$$\begin{aligned}
&\int_1^\infty \exp(t)\mathbb{P}\{Y > t/u\}\, dt = \int_1^\infty \exp(t)\mathbb{P}\{\xi(\lambda|X|) > \mu + t/u\}\, dt \\
&\le C_2 \int_1^\infty \exp(t)\exp\left(-\frac{\mu + t/u}{2}\right) \\
&= C_2 \exp(-\mu/2) \int_1^\infty \exp(t(1 - 1/(2u)))\, dt
\end{aligned}$$

The integral makes sense only for $1 - 1/(2u) < 0$, that is $2u < 1$.

$$\begin{aligned}
&= \frac{C_2 \exp(-\mu/2)}{1 - 1/(2u)} \int_1^\infty (\exp(t(1 - 1/(2u))))'\, dt \\
&= \frac{C_2 \exp(-\mu/2)}{1 - 1/(2u)} \exp(t(1 - 1/(2u)))|_1^\infty \\
&= (-1)\frac{eC_2 \exp(-\mu/2)\exp(-1/(2u))}{1 - 1/(2u)} \\
&= eC_2 \exp(-\mu/2)\exp(-1/(2u))\frac{2u}{1 - 2u}
\end{aligned}$$

By lemma 3.1.3.2,

$$\exp(-1/(2u)) \le \left(\frac{2}{e/2}\right)^2 u^2 = \frac{16}{e^2}u^2,$$

so we can estimate everything together as

$$\begin{aligned}
&\mathbb{E}\exp(uY)\mathbb{I}\{uY > 1\} \le e\mathbb{P}\{Y > 1/u\} + \int_1^\infty \exp(t)\mathbb{P}\{Y > t/u\}\, dt \\
&\le eC_2 \exp(-\mu/2)\exp(-1/(2u))\left(1 + \frac{2u}{1 - 2u}\right) \\
&= eC_2 \exp(-\mu/2)\exp(-1/(2u))\frac{1}{1 - 2u} \\
&\le C_2 \exp(-\mu/2)\frac{16}{e}\frac{u^2}{1 - 2u}.
\end{aligned}$$

Note that

$$C_2(\lambda)\exp(-\mu/2) \le \frac{2}{\pi}\frac{(1 + \sqrt{\lambda})}{(1 + \lambda^2)^{1/4}}.$$

By subadditivity,

$$\frac{2}{\pi}\frac{(1+\sqrt{\lambda})}{(1+\lambda^2)^{1/4}} \geq \frac{2}{\pi}\frac{(1+\sqrt{\lambda})}{1^{1/4}+(\lambda^2)^{1/4}} = \frac{2}{\pi}.$$

For an upper bound, if $\lambda \leq 1$,

$$\frac{2}{\pi}\frac{(1+\sqrt{\lambda})}{(1+\lambda^2)^{1/4}} \leq \frac{2}{\pi}(1+\sqrt{\lambda}) \leq \frac{4}{\pi}.$$

On the other hand, if $\lambda \geq 1$,

$$\frac{2}{\pi}\frac{(1+\sqrt{\lambda})}{(1+\lambda^2)^{1/4}} \leq \frac{4}{\pi}\frac{\sqrt{\lambda}}{(1+\lambda^2)^{1/4}} < \frac{4}{\pi}\frac{\sqrt{\lambda}}{\lambda^{2/4}} = \frac{4}{\pi}.$$

We thus have

$$\mathbb{E}\exp(uY)\mathbb{I}\left\{uY > 1\right\} \leq C_2\exp(-\mu/2)\frac{16}{e}\frac{u^2}{1-2u} \leq \frac{4}{\pi}\frac{16}{e}\frac{u^2}{1-2u} =: A_+u^2$$

if we choose an upper bound on $u \leq u_0 < 1/2$.

We then want to optimize u for

$$\begin{aligned}
&\exp(-u\Delta_+)\mathbb{E}\exp(uY) \leq \exp(-u\Delta_+)\left(1+V^2u^2+A_+u^2\right) \\
&\leq \exp(-u\Delta_+ + u^2(V^2+A_+))
\end{aligned}$$

If

$$k(u) := -u\Delta_+ + u^2(V^2+A_+)$$

which is convex, then

$$0 = k'(u^*) = -\Delta_+ + 2(u^*)(V^2+A_+) \Rightarrow \frac{\Delta_+}{2(V^2+A_+)} = u^*$$

and at u^*,

$$\begin{aligned}
&k(u^*) = -\frac{\Delta_+}{2(V^2+A_+)}\Delta_+ + \left(\frac{\Delta_+}{2(V^2+A_+)}\right)^2(V^2+A_+) \\
&= -\frac{\Delta_+^2}{2(V^2+A_+)} + \frac{\Delta_+^2}{4(V^2+A_+)} = -\frac{\Delta_+^2}{4(V^2+A_+)}.
\end{aligned}$$

We need to make sure $u^* < 1/2$. We have a lower bound on A_+ of

$$\frac{4}{\pi}\frac{16}{e}\frac{1}{1-2u_0} \geq \frac{8}{\pi}\frac{16}{e}$$

if we choose $u_0 = 1/4$. We have to verify then that $u^* \leq 1/4$. In this case,

$$u^* \leq \frac{\Delta_+}{2A_+} \leq \frac{\Delta_+}{(16/\pi)(16/e)} < 0.034\Delta_+.$$

If we choose $\Delta_+ = \mu((1+\epsilon)\lambda) - \mu(\lambda)$, then $\Delta_+ < \epsilon$ for $\lambda \geq 1/\sqrt{1+\epsilon}$ by lemma 4.2.1.5.

We can now estimate A_+ and u^* a bit better. For $\lambda \geq 1/\sqrt{1+\epsilon}$, we take $V^2 = \pi^2/2$ as our upper bound for the variance by remark 5.3.0.2. Consequently, $u^* < \Delta_+/\pi^2 < 0.102\epsilon$ as A_+ is positive and $\Delta < \epsilon$. We can now estimate A_+ as

$$A_+ = \frac{4}{\pi}\frac{16}{e}\frac{1}{1-2u} \leq \frac{4}{\pi}\frac{16}{e}\frac{1}{1-2\epsilon/\pi^2} = \frac{64\pi}{e(\pi^2-2\epsilon)} \leq \frac{64\pi}{e(\pi^2-1/2)}$$

if $\epsilon \leq 1/4$. We then have, using lemma 4.2.1.5 again

$$\frac{4(V^2+A_+)}{\Delta_+^2} \leq \frac{16}{\epsilon^2(1-\epsilon)^2}4(V^2+A_+) \leq \frac{64}{\epsilon^2(1-\epsilon)^2}\left(\frac{\pi^2}{2} + \frac{64\pi}{e(\pi^2-1/2)}\right).$$

□

Lemma 4.2.0.3 (Lower Tail, Large Scales)**.** *For $1 > u > 0$, $X \sim$ Cauchy(1), $Y = \xi(\lambda\,|X|) - \mu(\lambda)$, and $V^2 \geq \mathbb{E}Y^2$,*

$$\exp(-u\Delta_-)\mathbb{E}\exp(-uY)$$

can be minimized to

$$\exp\left(-\frac{\Delta_-^2}{4(V^2+A_-)}\right) \quad at \quad u = \frac{\Delta_-}{2(V^2+A_-)},$$

with A_- a bounded nonnegative function of λ for $\lambda \geq 1$.

In particular, for $\epsilon \leq 1/4$, $\lambda > \sqrt{1+\epsilon}$, and

$$\Delta_- = \mu(\lambda) - \mu((1+\epsilon)^{-1}\lambda),$$

we have the bound

$$\frac{4(V^2+A_-)}{\Delta_-^2} \leq \frac{64}{\epsilon^2(1-\epsilon)^2}\left(\frac{\pi^2}{2} + \frac{8\pi^2}{e\pi(\pi^2-1/4)}\sqrt{2}\right).$$

Remark 4.2.0.4. Again, the bound is not sharp, as there should be better ways to estimate A_-, possibly by iterating the argument found in the proof.

Proof. Note that Y does not have a sign, so we try breaking up the corresponding integral again.

$$\mathbb{E}\exp(-uY) = \mathbb{E}\exp(-uY)\mathbb{I}\{-uY \le 1\} + \mathbb{E}\exp(-uY)\mathbb{I}\{-uY > 1\}.$$

We can still use lemma 4.1.0.1 applied to $-Y$. Just as in the upper tail computations,

$$\mathbb{E}\exp(-uY)\mathbb{I}\{-uY \le 1\} \le \mathbb{E}(1 - uY + (-uY)^2) = 1 + V^2u^2.$$

and

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

Now,

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

By subadditivity of $\sqrt{a}$,

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

So

$$\mathbb{P}\{-Y > t/u\} = \mathbb{P}\{\xi(\lambda|X|) < \mu - t/u\}$$
$$\le \mathbb{P}\{\ln(1+\lambda|X|) < \mu - t/u\}$$
$$= \mathbb{P}\{1+\lambda|X| < \exp(\mu - t/u)\}$$
$$= \mathbb{P}\left\{|X| < \frac{\exp(\mu - t/u) - 1}{\lambda}\right\}$$
$$= \frac{2}{\pi}\arctan\left(\frac{\exp(\mu - t/u) - 1}{\lambda}\right)$$
$$\le \frac{2}{\pi}\arctan\left(\frac{\exp(\mu - t/u) - \exp(-t/u)}{\lambda}\right)$$
$$= \frac{2}{\pi}\arctan\left(\exp(-t/u)\frac{(\exp(\mu) - 1)}{\lambda}\right)$$

Finally, using the basic upper bound for arctan,

$$\mathbb{P}\{-Y > t/u\} \leq \frac{2}{\pi}\frac{\exp(\mu - t/u) - 1}{\lambda} < \frac{2}{\pi\lambda}\exp(\mu - t/u) =: C_1(\lambda)\exp(-t/u)$$

Note that the unbounded contribution to $\mu(\lambda)$ is

$$\frac{1}{2}\ln(1+\lambda^2)$$

so that

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

which is bounded for $\lambda \geq \lambda_0 > 0$. We thus have *provided* $1 - 1/u < 0$ that is, $u < 1$,

$$\begin{aligned}
&\int_1^\infty \exp(t)\mathbb{P}\{-Y > t/u\}\, dt < C_1\int_1^\infty \exp(t - t/u)\, dt \\
&= \frac{C_1}{1-1/u}\int_1^\infty (\exp(t(1-1/u)))'\, dt \\
&= \frac{C_1}{1-1/u}\exp(t(1-1/u))|_1^\infty \\
&= -\frac{C_1 e}{1-1/u}\exp(-1/u) = \frac{C_1 e u}{1-u}\exp(-1/u).
\end{aligned}$$

Putting things together

$$\begin{aligned}
&\mathbb{E}\exp(-uY)\mathbb{I}\{-uY > 1\} \leq e\mathbb{P}\{-Y > 1/u\} + \int_1^\infty \exp(t)\mathbb{P}\{-Y > t/u\}\, dt \\
&< eC_1\exp(-1/u) + \frac{C_1 e u}{1-u}\exp(-1/u) \\
&= \frac{eC_1}{1-u}\exp(-1/u) \\
&\leq \frac{eC_1}{1-u}\frac{4}{e^2}u^2 = u^2\frac{4}{e(1-u)}\frac{2}{\pi}\sqrt{1+\frac{1}{\lambda^2}} \leq u^2\frac{8}{e\pi(1-u)}\sqrt{1+1/(1+\epsilon)} \leq A_- u^2
\end{aligned}$$

if we have a bound $1 > u_0 \geq u$ and assuming $\lambda \geq \sqrt{1+\epsilon}$.

So

$$\exp(-u\Delta_-)\mathbb{E}\exp(-uY) < \exp(-u\Delta_-)(1+(V^2+A_-)u^2) \leq \exp(-u\Delta_- + u^2(V^2+A_-)).$$

Because

$$k(u) := -u\Delta_- + u^2(V^2+A_-)$$

is convex, we can find the global minimizer u^* at

$$0 = k'(u^*) = -\Delta_- + 2u^*(V^2 + A_-) \Rightarrow \frac{\Delta_-}{2(V^2 + A_-)} = u^*$$

so that

$$k(u^*) = -\frac{\Delta_-^2}{2(V^2 + A_-)} + \frac{\Delta_-^2}{4(V^2 + A_-)^2}(V^2 + A_-) = -\frac{\Delta_-^2}{4(V^2 + A_-)}$$

We need to ensure $u^* < 1$. By remark 5.3.0.2, we take $V^2 = \pi^2/2$ as our upper bound for the variance. Consequently, $u^* < \Delta_-/\pi^2$ as A_- is positive. By lemma 4.2.1.5 and the discussion following, $\Delta_- < \epsilon$, when $\lambda \geq \sqrt{1+\epsilon}$ making $u^* < \epsilon/9 < 1/2$ for $\epsilon < 1$ certainly.

We can now estimate A_- as

$$A_- \leq \frac{8}{e\pi(1 - \epsilon/\pi^2)}\sqrt{1 + 1/(1+\epsilon)} \leq \frac{8\pi^2}{e\pi(\pi^2 - \epsilon)}\sqrt{1 + 1/(1+\epsilon)} \leq \frac{8\pi^2}{e\pi(\pi^2 - 1/4)}\sqrt{2}$$

for $\epsilon \leq 1/4$. Finally, using lemma 4.2.1.5 again

$$\frac{4(V^2 + A_-)}{\Delta_-^2} \leq \frac{16}{\epsilon^2(1-\epsilon)^2}4(V^2 + A_-) \leq \frac{64}{\epsilon^2(1-\epsilon)^2}\left(\frac{\pi^2}{2} + \frac{8\pi^2}{e\pi(\pi^2 - 1/4)}\sqrt{2}\right).$$

□

4.2.1 Estimating Deviations of the Mean

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

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

when λ is not too small. 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. The first evidence that this deviations are bounded in λ is the following.

Lemma 4.2.1.1. *For $\lambda > 0$ and $a > 1$,*

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

Remark 4.2.1.2. Note this is bounded above by

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

for all $\lambda > 0$ and by

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

for all $1 \geq \lambda \geq 0$.

Proof. We have

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

□

Lemma 4.2.1.3. *For $\lambda > 0$ and $a > 1$,*

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

Remark 4.2.1.4. Note the change in sign when λ crosses $1/\sqrt{a}$. We shall need it in some of the later bounds.

Proof. By the atanh addition formula A.1.0.11,

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

for $u, v \in (-1,1)$, which is the case for us here. With

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

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

while

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

so that

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

Because atanh is an odd function, taking atanh of the above will give negative numbers when $\lambda\sqrt{a} > 1$. □

Lemma 4.2.1.5. *For $1 \le a$ and $1/\sqrt{a} \le \lambda$,*

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

Proof. The atanh contribution is negative for $\lambda \ge 1/\sqrt{a}$ with input

$$(\sqrt{a}-1)\sqrt{2\lambda}\frac{1-\lambda\sqrt{a}}{(1-\lambda\sqrt{a})^2+\lambda(1+a)}$$

Let

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

We estimate

$$v = (\sqrt{a}-1)\sqrt{\lambda/2}(\lambda\sqrt{a}-1)\frac{2}{(\lambda\sqrt{a}-1)^2+\lambda(1+a)}$$
$$\le (\sqrt{a}-1)\sqrt{\lambda/2}(\lambda\sqrt{a}-1)\frac{1}{(\lambda\sqrt{a}-1)\sqrt{\lambda(1+a)}}$$
$$= \frac{\sqrt{a}-1}{\sqrt{2}\sqrt{1+a}} < \frac{\sqrt{a}}{\sqrt{2}\sqrt{1+a}} < \frac{1}{\sqrt{2}}.$$

So

$$w := \frac{\sqrt{a}-1}{\sqrt{2}\sqrt{1+a}} < 1$$

We may write

$$\begin{aligned}
&\operatorname{atanh}(v) \leq \operatorname{atanh}(w) \leq \frac{w}{1-w^2} \\
&= \frac{\sqrt{a}-1}{\sqrt{2}\sqrt{1+a}}\left(1 - \frac{(\sqrt{a}-1)^2}{2(1+a)}\right)^{-1} = \frac{(\sqrt{a}-1)\sqrt{2}\sqrt{1+a}}{2(1+a)-(\sqrt{a}-1)^2} \\
&= \frac{(\sqrt{a}-1)\sqrt{2}\sqrt{1+a}}{2+2a-(a+1-2\sqrt{a})} = \frac{(\sqrt{a}-1)\sqrt{2}\sqrt{1+a}}{1+a+2\sqrt{a}} = \frac{(\sqrt{a}-1)\sqrt{2}\sqrt{1+a}}{(1+\sqrt{a})^2} \\
&\leq \frac{\sqrt{a}-1}{2}
\end{aligned}$$

as the remaining factor is decreasing in a:

$$\begin{aligned}
&\frac{d}{da}\frac{(1+a)^{1/2}}{(1+\sqrt{a})^2} \\
&= \frac{1}{(1+\sqrt{a})^4}\left((1+\sqrt{a})^2\frac{1}{2(1+a)^{1/2}} - 2(1+a)^{1/2}(1+\sqrt{a})\frac{1}{2\sqrt{a}}\right) \\
&= \frac{1}{2(1+\sqrt{a})^2}\left(\frac{1}{(1+a)^{1/2}} - 2\frac{(1+a)^{1/2}}{1+\sqrt{a}}\frac{1}{\sqrt{a}}\right) \\
&= \frac{1}{2(1+\sqrt{a})^2(1+a)^{1/2}}\left(1 - 2\frac{(1+a)}{\sqrt{a}+a}\right) \\
&= \frac{1}{2(1+\sqrt{a})^3(1+a)^{1/2}\sqrt{a}}(a+\sqrt{a}-2-2a) \\
&= \frac{1}{2(1+\sqrt{a})^3(1+a)^{1/2}\sqrt{a}}(\sqrt{a}-a-2) \\
&< 0 \quad \text{certainly for} \quad a \geq 1.
\end{aligned}$$

On the other hand, because $\lambda \geq 1/\sqrt{a}$, the ln contribution is now

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

using lemma 4.2.1.6 to lower bound the logarithm.

Hence, for $\lambda \geq 1/\sqrt{a}$, and using lemma 4.4.0.5 to approximate the square root in the

2nd line,

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

$$\geq \frac{(a-1)}{2}\left(1 - \frac{a-1}{2}\right) - \frac{a-1}{4} = \frac{(a-1)}{2} - \frac{a-1}{4} - \frac{(a-1)^2}{4}$$

$$= \frac{a-1}{4} - \frac{(a-1)^2}{4} = \frac{a-1}{4}(1-(a-1)).$$

If we drop the negative atanh contribution, we can find an upper bound

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

Note this need not be an upper bound if the atanh contribution were positive. □

Lemma 4.2.1.6. *For $-1 < t$,*

$$\ln(1+t) \geq t(1 - t/2).$$

Remark 4.2.1.7. One may also use the alternating series test if $t < 1$.

Proof. Let $f(t) = \ln(1+t)$. Then

$$f'(t) = \frac{1}{1+t}, \quad f''(t) = -\frac{1}{(1+t)^2}, \quad \text{and} \quad f^{(3)}(t) = \frac{2}{(1+t)^3} > 0.$$

Because $t > -1$, all these terms have a sign. By Taylor's theorem with Lagrange remainder, with $x \in (0, t)$,

$$f(t) = f(0) + f'(0)t + f''(0)\frac{t^2}{2} + f^{(3)}(x)\frac{t^3}{3!} \geq t - \frac{t^2}{2} = t(1 - t/2)$$

□

4.3 Small Scales

In the last section dividing by $\Delta_\pm$ was ok as these quantities were bounded around ϵ and away from 0. In this section, we shall have $\Delta_\pm = \pm\epsilon\mu(\lambda) \to 0$ when $\lambda \to 0$, which will need slightly different arguments. Even here, we shall not take λ too small, as the target dimension k will then grow accordingly.

Lemma 4.3.0.1 (Upper Tail, Small Scales). *For* $1/2 > u > 0$*,* $X \sim \mathit{Cauchy}(1)$*,* $1 \geq \lambda > 8\epsilon^2$*, and* $Y = \xi(\lambda\,|X|)$*, if* $V^2 \geq \mathbb{E}Y^2$*,*

$$\exp(-u(1+\epsilon)\mu(\lambda))\mathbb{E}\exp(uY)$$

can be minimized to

$$\exp\left(-\frac{\epsilon^2\mu^2(\lambda)}{4(V^2+A_+)}\right) \quad \mathit{at} \quad u = \frac{\epsilon\mu(\lambda)}{2(V^2+A_+)},$$

with A_+ *a bounded nonnegative function of* $\lambda \leq 1$*.*

In particular, for $\epsilon \leq 1/4$*,*

$$\frac{4(V^2+A_+)}{\epsilon^2\mu^2(\lambda)} \leq \frac{8}{\epsilon^2}\left(3.126 + 1 + \frac{4}{\pi} - \frac{4}{\pi}\ln(\lambda) + 8 + 2\sqrt{2} + \frac{1}{4}\right)$$

Remark 4.3.0.2. Note how the logarithmic term blows up when λ becomes small. We shall discuss this in section 4.4. The restriction that $\lambda \geq 8\epsilon^2$ prevents us from using this lemma to show concentration at moderately small scales. There may be a different proof technique that could do so, possibly using a particular moment instead of the full moment generating function; compare [PN95].

Proof. We break up $\mathbb{E}\exp(uY)$ into two integrals using lemma 4.1.0.1. With $V^2 \geq \mathbb{E}Y^2$, the first integral is

$$\mathbb{E}\exp(uY)\mathbb{I}\{uY \leq 1\} \leq \mathbb{E}(1+uY+u^2Y^2)\mathbb{I}\{uY \leq 1\}$$
$$\leq \mathbb{E}(1+uY+u^2Y^2) \leq 1 + u\mu + u^2V^2.$$

The second integral is

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

We thus need an upper bound on

$$\mathbb{P}\{Y > t/u\} = \mathbb{P}\{\xi(\lambda\,|X|) > t/u\}$$

for $t \geq 1$.

We want to use lemma 4.1.0.3 with $C_1(\lambda)$ to estimate these tail probabilities as we are assuming λ is bounded here. If we assume $u < 1/2$, then $t/u > 2$. In this case, the lemma says

$$e\mathbb{P}\{Y > 1/u\} \leq eC_1(\lambda)\exp(-1/u)$$

while (noting that $1 - 1/u < 0$ for us)

$$\begin{aligned}
&\int_1^\infty \exp(t)\mathbb{P}\{Y > t/u\}\, dt \leq C_1(\lambda)\int_1^\infty \exp(t - t/u)\, dt \\
&= \frac{C_1(\lambda)}{1 - 1/u}\int_1^\infty (\exp(t(1 - 1/u)))'\, dt \\
&= \frac{C_1(\lambda)}{1 - 1/u}\exp(t(1 - 1/u))|_1^\infty \\
&= (-1)\frac{C_1(\lambda)}{1 - 1/u}\exp(1 - 1/u) = \frac{C_1(\lambda)eu\exp(-1/u)}{1 - u}.
\end{aligned}$$

By lemma 3.1.3.2,

$$\exp(-1/u) \leq \left(\frac{2}{e}\right)^2 u^2 = \frac{4}{e^2}u^2,$$

so we can estimate everything together as

$$\begin{aligned}
&\mathbb{E}\exp(uY)\mathbb{I}\{uY > 1\} \leq e\mathbb{P}\{Y > 1/u\} + \int_1^\infty \exp(t)\mathbb{P}\{Y > t/u\}\, dt \\
&\leq eC_1(\lambda)\exp(-1/u) + \frac{C_1(\lambda)eu\exp(-1/u)}{1 - u} \\
&= \frac{C_1(\lambda)}{1 - u}e\exp(-1/u) \leq \frac{C_1(\lambda)}{1 - u}e\frac{4}{e^2}u^2 = \frac{C_1(\lambda)}{1 - u}\frac{4}{e}u^2 \leq \frac{C_1(\lambda)}{1 - u_0}\frac{4}{e}u^2 \\
&= \frac{2}{\pi}\frac{\lambda}{(1 - 1/e)^2}\frac{4}{e}\frac{1}{1 - u_0}u^2 \\
&= \frac{8e}{\pi(e - 1)^2(1 - u_0)}\lambda u^2
\end{aligned}$$

assuming an upper bound on $u < u_0 \leq 1/2$. Choosing $u_0 = 1/2$, we set

$$A_+(\lambda) := \frac{8e}{\pi(e - 1)^2(1 - u_0)}\lambda = \frac{16e}{\pi(e - 1)^2}\lambda$$

Consequently,

$$\begin{aligned}
&\exp(-u(1 + \epsilon)\mu(\lambda))\mathbb{E}\exp(uY) \leq \exp(-u(1 + \epsilon)\mu(\lambda))\big(1 + u\mu(\lambda) + u^2(V^2 + A_+(\lambda))\big) \\
&\leq \exp\big(- u(1 + \epsilon)\mu(\lambda) + u\mu(\lambda) + u^2(V^2 + A_+(\lambda))\big) \\
&= \exp\big(- u\epsilon\mu(\lambda) + u^2(V^2 + A_+(\lambda))\big)
\end{aligned}$$

and we want to minimize this last quantity in u. Let

$$k(u) := -u\epsilon\mu(\lambda) + u^2(V^2 + A_+(\lambda))$$

Then, setting the derivative to 0 yields

$$0 = -\epsilon\mu(\lambda) + 2u^*(V^2 + A_+(\lambda)) \Rightarrow u^* = \frac{\epsilon\mu(\lambda)}{2(V^2 + A_+(\lambda))}.$$

Because $k(u)$ is a convex function, u^* is a global minimizer at which

$$k(u^*) = -\epsilon\mu(\lambda)\frac{\epsilon\mu(\lambda)}{2(V^2 + A_+(\lambda))} + \frac{\epsilon^2\mu^2(\lambda)}{4(V^2 + A_+(\lambda))^2}(V^2 + A_+(\lambda))$$
$$= -\frac{\epsilon^2\mu^2(\lambda)}{4(V^2 + A_+(\lambda))}.$$

We need to check $u^* < 1/2$. Because $V^2 \geq \mathbb{E}Y^2 \geq (\mathbb{E}Y)^2 = \mu^2(\lambda)$ by Jensen's inequality, we always have for $0 < \lambda \leq 1$

$$u^* = \frac{\epsilon\mu(\lambda)}{2(V^2 + A_+(\lambda))} < \frac{\epsilon\mu(\lambda)}{2V^2} \leq \frac{\epsilon}{2\mu(\lambda)} \leq \frac{\epsilon}{2}\frac{(1+\lambda)}{\sqrt{2\lambda}} \leq \frac{\epsilon}{\sqrt{2\lambda}} < 1/4 < 1/2$$

provided $4\epsilon < \sqrt{2\lambda}$, that is, $8\epsilon^2 < \lambda$. In this case, we can estimate $A_+(\lambda)$ as

$$A_+(\lambda) \leq \frac{C_1(\lambda)}{1 - 1/4}\frac{4}{e} = \frac{16}{3e}\frac{2\lambda}{\pi}\frac{1}{(1-1/e)^2}$$
$$= \lambda\frac{32e}{3\pi(e-1)^2} < 3.126\lambda.$$

Consequently using lemma 5.3.0.3 for the bound V^2, we can give the bound

$$\frac{4(V^2 + A_+)}{\epsilon^2\mu^2(\lambda)} \leq \frac{(1+\lambda)^2}{2\epsilon^2\lambda}4\left(V^2 + A_+\right) \leq \frac{2(1+\lambda)^2}{\epsilon^2}\left(\frac{V^2}{\lambda} + 3.126\right)$$
$$= \frac{2(1+\lambda)^2}{\epsilon^2}\left(\frac{V^2}{\lambda} + 3.126\right)$$

When $\lambda \leq 1$, this is

$$\frac{2(1+\lambda)^2}{\epsilon^2}\left(3.126 + 1 + \frac{4}{\pi} - \frac{4}{\pi}\ln(\lambda) + \frac{8}{(1+\lambda)^2} + 2\lambda\frac{\sqrt{2\lambda}}{1+\lambda} + \frac{\lambda^3}{4}\right)$$
$$\leq \frac{8}{\epsilon^2}\left(3.126 + 1 + \frac{4}{\pi} - \frac{4}{\pi}\ln(\lambda) + 8 + 2\sqrt{2} + \frac{1}{4}\right).$$

□

Lemma 4.3.0.3 (Lower Tail, Small Scales)**.** *Let $t < \mu(\lambda)$, $X \sim$ Cauchy(1), and $Y = \xi(\lambda\,|X|)$ with $\mathbb{E}Y^2 \le V^2$. Then at $u = (\mu(\lambda) - t)/V^2$,*

$$\exp(tu)\mathbb{E}\exp(-uY) \le \exp\left(-\frac{(t-\mu)^2}{2V^2}\right)$$

In particular, for $t = (1-\epsilon)\mu(\lambda)$, $\epsilon \le 1/4$, and $0 \le \lambda \le 1$

$$\frac{2V^2}{(t-\mu)^2} \le \frac{4}{\epsilon^2}\left(1 + \frac{4}{\pi} - \frac{4}{\pi}\ln(\lambda) + 8 + 2\sqrt{2} + \frac{1}{4}\right)$$

and for $1 \le \lambda \le 2$,

$$\frac{2V^2}{(t-\mu)^2} \le \frac{9}{\epsilon^2}\left(\frac{\pi^2}{2} + 4 + 2\sqrt{2}\right).$$

Remark 4.3.0.4. I do not think the bound is tight. Again, note how the bound deteriorates when λ becomes small. We shall discuss this in section 4.4.

Proof. By lemma 3.1.3.3,

$$\exp(tu)\mathbb{E}\exp(-uY) \le \exp(tu)\mathbb{E}\left(1 - uY + \frac{u^2}{2}Y^2\right)$$
$$\le \exp(tu)(1 - u\mu + \frac{u^2}{2}V^2) \le \exp\left(u(t-\mu) + \frac{u^2}{2}V^2\right).$$

We want to minimize

$$k(u) := u(t-\mu) + \frac{u^2}{2}V^2.$$

Setting the derivative to 0 yields

$$0 = (t-\mu) + uV^2 \Rightarrow \frac{\mu - t}{V^2} = u^* > 0.$$

The minimizer u^* is a global minimizer because $k(u)$ is convex. At u^*,

$$k(u^*) = -\frac{(\mu-t)^2}{V^2} + \frac{V^2}{2}\left(\frac{\mu-t}{V^2}\right)^2 = -\frac{(\mu-t)^2}{2V^2}$$

so that

$$\exp(tu^*)\mathbb{E}\exp(-u^*Y) \le \exp\left(-\frac{(t-\mu)^2}{2V^2}\right).$$

Using lemma 5.3.0.3 for the bound V^2, we can give the bound for $\lambda \le 1$

$$\begin{aligned}
\frac{2V^2}{(\mu - t)^2} &= \frac{2}{\epsilon^2 \mu^2(\lambda)} V^2 \le \frac{(1+\lambda)^2}{\epsilon^2} \frac{V^2}{\lambda} \\
&\le \frac{4}{\epsilon^2} \left(1 + \frac{4}{\pi} - \frac{4}{\pi} \ln(\lambda) + \frac{8}{(1+\lambda)^2} + 2\lambda \frac{\sqrt{2\lambda}}{1+\lambda} + \frac{\lambda^3}{4} \right) \\
&\le \frac{4}{\epsilon^2} \left(1 + \frac{4}{\pi} - \frac{4}{\pi} \ln(8\epsilon^2) + 8 + 2\sqrt{2} + \frac{1}{4} \right)
\end{aligned}$$

and for $1 \le \lambda \le 2$,

$$\frac{2V^2}{(\mu - t)^2} \le \frac{(1+\lambda)^2}{\epsilon^2} \frac{V^2}{\lambda} \le \frac{9}{\epsilon^2} \left(\frac{\pi^2}{2} + 2 + 2\sqrt{2} + 2 \right) = \frac{9}{\epsilon^2} \left(\frac{\pi^2}{2} + 4 + 2\sqrt{2} \right).$$

□

4.4 Really Small Scales

In the last section 4.3, we saw the reciprocals of the concentration rates blow up like $\ln(1/\lambda)$ as $\lambda \to 0$. In this section, we show that we can stop that blow up at a particular $\lambda_0 = \Theta(\delta)$ with $\delta > 0$ the failure probability.

We shall show in lemma 4.4.0.3 that $\xi(a) \approx \sqrt{a}$ for small a which will play well with $\mu(\lambda) = \Theta(\sqrt{\lambda})$ for $\lambda \le 1$, as seen in remark 5.1.0.3. The following lemma 4.4.0.1 shows how this approximate homogeneity could be used. Specifically, let $X_i \overset{\text{i.i.d.}}{\sim} \text{Cauchy}(1)$ for $1 \le i \le k$. We show in section 4.4.1 that with high probability, $\max_i |X_i| \le C_k$ for some increasing function C_k of k. The hope would be to invoke the approximate homogeneity above to conclude that if the concentration results hold for $\lambda \approx \epsilon/C_k$, it holds for all $\lambda \le \epsilon/C_k$ too.

Unfortunately, at least for Cauchy random variables, C_k grows quickly with k, so that one already needs a concentration result for moderately small λ. We were able to give a lower tail concentration result in lemma 4.3.0.3 with no restriction on how small λ can be, but the upper tail concentration result in lemma 4.3.0.1 required $\lambda \ge 8\epsilon^2$.

Lemma 4.4.0.1. *For $1 \leq i \leq k$, let $X_i \overset{i.i.d.}{\sim}$ Cauchy(1). For $0 < \epsilon < 1/4$ and $0 < \lambda_0 \leq 1$, suppose*

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

and $\lambda_0 \max_i |X_i| \leq c_0 \leq 1/6$.

Then if $0 < \eta < 1$,

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

with ϵ' depending on ϵ, c_0, and λ_0. If $\lambda_0 = O(\epsilon^2)$, then ϵ' may be made to be $O(\epsilon)$ for ϵ small enough.

Remark 4.4.0.2. We shall see in section 4.4.1 that λ_0 must be taken very small in order for $\lambda_0 \max_i |X_i| \leq c_0$ with high probability.

Proof. Because $\max_i |X_i| \leq c_0 \leq .16$ and $0 < \eta < 1$, we can invoke lemma 4.4.0.3 once to say

$$\sqrt{\eta\lambda_0 |X_i|} \leq \xi(\eta\lambda_0 |X_i|) \leq \sqrt{\eta\lambda_0 |X_i|}\left(1 + \frac{\eta\lambda_0 |X_i|}{2}\right)$$

and then again, writing $\sqrt{\eta\lambda_0 |X_i|} = \sqrt{\eta}\sqrt{\lambda_0 |X_i|}$

$$\frac{\sqrt{\eta}}{1+\lambda_0 |X_i|/2}\xi(\lambda_0 |X_i|) \leq \xi(\eta\lambda_0 |X_i|) \leq \sqrt{\eta}\xi(\lambda_0 |X_i|)\left(1 + \frac{\eta\lambda_0 |X_i|}{2}\right)$$

leaving a bound of

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

By assumption, summing over i and dividing by k yields

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

We shall now use remark 5.1.0.3 (twice) to "absorb" $\sqrt{\eta}$ into μ, recalling

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

It will help to rewrite the last bound as

$$\sqrt{2\lambda}\frac{1+\lambda}{1+\lambda^2}+\frac{\lambda^2}{2}=\sqrt{2\lambda}\left(\frac{1+\lambda}{1+\lambda^2}+(\lambda/2)^{3/2}\right).$$

We look at the bounds individually.

For the lower bound,

$$\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k}\xi(\eta\lambda_0\,|X_i|)\geq(1-\epsilon)\frac{\sqrt{\eta}}{1+\lambda_0c_0/2}\mu(\lambda_0)\\
&\geq(1-\epsilon)\frac{\sqrt{\eta}}{1+\lambda_0c_0/2}\frac{\sqrt{2\lambda_0}}{1+\lambda_0}\\
&=(1-\epsilon)\sqrt{2\eta\lambda_0}\frac{1}{(1+\lambda_0c_0/2)(1+\lambda_0)}\\
&\geq(1-\epsilon)\mu(\eta\lambda_0)\left(\frac{1+\eta\lambda_0}{1+\eta\lambda_0^2}+(\eta\lambda_0/2)^{3/2}\right)^{-1}\frac{1}{(1+\lambda_0c_0/2)(1+\lambda_0)}
\end{aligned}$$

We need to control the multiplier above. Its inverse is

$$\left(\frac{1+\eta\lambda_0}{1+\eta\lambda_0^2}+(\eta\lambda_0/2)^{3/2}\right)(1+\lambda_0c_0/2)(1+\lambda_0)$$

If $\lambda_0=O(\epsilon^2)$, then this multiplier is $1+O(\epsilon^2)$ for ϵ small enough.

For the upper bound,

$$\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k}\xi(\eta\lambda_0\,|X_i|)\leq(1+\epsilon)\sqrt{\eta}\mu(\lambda_0)\left(1+\frac{\eta\lambda_0c_0}{2}\right)\\
&\leq(1+\epsilon)\sqrt{2\eta\lambda_0}\left(\frac{1+\lambda_0}{1+\lambda_0^2}+(\lambda_0/2)^{3/2}\right)\left(1+\frac{\eta\lambda_0c_0}{2}\right)\\
&\leq(1+\epsilon)\mu(\eta\lambda_0)(1+\eta\lambda_0)\left(\frac{1+\lambda_0}{1+\lambda_0^2}+(\lambda_0/2)^{3/2}\right)\left(1+\frac{\eta\lambda_0c_0}{2}\right)
\end{aligned}$$

The multiplier

$$(1+\eta\lambda_0)\left(\frac{1+\lambda_0}{1+\lambda_0^2}+(\lambda_0/2)^{3/2}\right)\left(1+\frac{\eta\lambda_0c_0}{2}\right)$$

is $1+O(\epsilon^2)$ as soon as $\lambda_0=O(\epsilon^2)$ when ϵ is small enough. □

We now show $\xi(a)\approx\sqrt{a}$ for small a which will play well with $\mu(\lambda)=\Theta(\sqrt{\lambda})$ for $\lambda\leq 1$, as seen in remark 5.1.0.3.

Lemma 4.4.0.3. *For $0 < a < 1/6 \approx .16$,*

$$\sqrt{a} \le \xi(a) = \ln(1+\sqrt{a}) + \frac{1}{2}\ln(1+a) \le \sqrt{a}\left(1+\frac{a}{2}\right)$$

Remark 4.4.0.4. Note the upper bound is better than the upper bound of

$$\sqrt{a} + \frac{a}{2} = \sqrt{a}\left(1+\frac{\sqrt{a}}{2}\right).$$

Proof. We focus on

$$f(x) = \ln(1+x) + \frac{1}{2}\ln(1+x^2),$$

using Taylor's theorem with Lagrange remainder about $x = 0$.

We have

$$f'(x) = \frac{1}{1+x} + \frac{x}{1+x^2}$$

which is strictly positive for all $x > 0$.

We also have

$$\begin{aligned}
f''(x) &= -\frac{1}{(1+x)^2} + \frac{1}{(1+x^2)^2}(1+x^2-2x^2) = -\frac{1}{(1+x)^2} + \frac{1}{(1+x^2)^2}(1-x^2)\\
&= \frac{1}{(1+x)^2(1+x^2)^2}(-1-2x^2-x^4+1+2x+x^2-x^2(1+2x+x^2))\\
&= \frac{1}{(1+x)^2(1+x^2)^2}(-2x^2-x^4+2x+x^2-x^2-2x^3-x^4)\\
&= \frac{1}{(1+x)^2(1+x^2)^2}(-2x^2-2x^4+2x-2x^3)\\
&= \frac{2x}{(1+x)^2(1+x^2)^2}(1-x-x^2-x^3)
\end{aligned}$$

For $0 \le x \le 1/2$, the 2nd derivative is positive, so for some $z \in (0, 1/2)$,

$$f(x) = f(0) + f'(0)x + f''(z)\frac{x^2}{2} \ge f'(0)x = x$$

We then have (from the first simplification of $f''(x)$ above)

$$
\begin{aligned}
f'''(x) &= \left(-\frac{1}{(1+x)^2} + \frac{1}{(1+x^2)^2}(1-x^2)\right)' \\
&= \left(-\frac{1}{(1+x)^2} + \frac{1}{(1+x^2)^2} - \frac{x^2}{(1+x^2)^2}\right)' \\
&= \frac{2}{(1+x)^3} - \frac{2(2x)}{(1+x^2)^3} - \frac{1}{(1+x^2)^4}((1+x^2)^2(2x) - x^2 2(1+x^2)2x) \\
&= \frac{2}{(1+x)^3} - \frac{4x}{(1+x^2)^3} - \frac{1}{(1+x^2)^3}(2x(1+x^2) - 4x^3)) \\
&= \frac{2}{(1+x)^3} - \frac{6x}{(1+x^2)^3} + \frac{2x^3}{(1+x^2)^3} \\
&= \frac{2}{(1+x)^3} - \frac{(6x-2x^3)}{(1+x^2)^3}
\end{aligned}
$$

We finally have

$$
\begin{aligned}
f^{(4)}(x) &= -\frac{6}{(1+x)^4} - \frac{1}{(1+x^2)^6}((1+x^2)^3(6-6x^2) - (6x-2x^3)3(1+x^2)^2 2x) \\
&= -\frac{6}{(1+x)^4} - \frac{6}{(1+x^2)^4}((1+x^2)(1-x^2) - x(6x-2x^3)) \\
&= -\frac{6}{(1+x)^4} - \frac{6}{(1+x^2)^4}(1-x^4-6x^2+2x^4) \\
&= -\frac{6}{(1+x)^4} - \frac{6}{(1+x^2)^4}(1-6x^2+x^4)
\end{aligned}
$$

Certainly for $|x| < 1/\sqrt{6}$, all terms are negative, so for some $z \in (0, 1/\sqrt{6})$,

$$
f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} + \frac{x^4}{4!}f^{(4)}(z) = x + \frac{x^3}{3} + \frac{x^4}{4!}f^{(4)}(z) \le x + \frac{x^3}{3}.
$$

Putting both bounds together, as $1/2 > 1/\sqrt{6}$, we have for all $0 \le x < 1/\sqrt{6}$,

$$
x \le f(x) \le x\left(1 + \frac{x^2}{2}\right)
$$

Setting $x = \sqrt{a}$, we have our result. □

Lemma 4.4.0.5. *For* $0 < \epsilon$

$$
1 + \frac{\epsilon}{2}\left(1 - \frac{\epsilon}{4}\right) \le \sqrt{1+\epsilon} \le 1 + \frac{\epsilon}{2}.
$$

Proof. For the upper bound,

$$
\sqrt{1+\epsilon} \le \sqrt{1+\epsilon+\epsilon^2/4} = 1 + \epsilon/2.
$$

Let $f(x) = (1+x)^{1/2}$.

$$f'(x) = \frac{1}{2(1+x)^{1/2}}, \quad f''(x) = -\frac{1}{4(1+x)^{3/2}}, \quad \text{and} \quad f^{(3)}(x) = \frac{3}{8(1+x)^{5/2}} > 0.$$

So for some $x \in (0, \epsilon)$ and all $\epsilon \geq 0$,

$$f(\epsilon) = f(0) + f'(0)\epsilon + f''(0)\frac{\epsilon^2}{2} + f^{(3)}(x)\frac{\epsilon^3}{3!} = 1 + \frac{\epsilon}{2} - \frac{\epsilon^2}{8} + f^{(3)}(x)\frac{\epsilon^3}{3!} \geq 1 + \frac{\epsilon}{2} - \frac{\epsilon^2}{8}.$$

Thus,

$$1 + \frac{\epsilon}{2}\left(1 - \frac{\epsilon}{4}\right) \leq \sqrt{1+\epsilon} \leq 1 + \frac{\epsilon}{2}.$$

□

4.4.1 Bounds on Maxima

Lemma 4.4.1.1. *Let X_i for $1 \leq i \leq k$ be independent identically distributed random variables. Let Z be the largest of $|X_i|$. Then provided*

$$\alpha = \frac{1}{kp_t} > 1, \quad \textit{with} \quad p_t = \mathbb{P}\{|X_i| > t\},$$

there is the bound

$$\mathbb{P}\{Z > t\} \leq \exp(-H(\alpha)kp_t) \quad \textit{with} \quad H(x) := x\ln(x) + 1 - x.$$

Proof. Let $Y_i(t)$ be the indicator function $\mathbb{I}(|X_i| > t)$ which is a Bern (p_t) random variable with

$$p_t = \mathbb{P}\{|X_i| > t\}.$$

If $Z > t$, then *at least* one of the X_i is greater than t:

$$\mathbb{P}\{Z_j > t\} = \mathbb{P}\left\{\sum_{i=1}^{k} Y_i(t) > 1\right\}, \quad \text{while} \quad \mathbb{E}\sum_{i=1}^{k} Y_i(t) = kp_t.$$

If $\alpha = 1/(kp_t) > 1$, the Chernoff-Hoeffding bounds for the binomial distribution apply, using lemma B.0.0.5

$$\mathbb{P}\{Z > t\} = \mathbb{P}\left\{\sum_{i=1}^{k} Y_i(t) > \alpha kp_t\right\} \leq \exp\left(-H(\alpha)kp_t\right)$$

with

$$H(\alpha) := \alpha \ln(\alpha) + 1 - \alpha.$$

□

For the above to be useful, we link α to k as follows.

Let $p_t = 1/(kC_k)$ with $C_k > 1$ possibly depending on k so that $\alpha = C_k$ and

$$H(\alpha)kp_t = H(C_k)\frac{1}{C_k} = (C_k \ln(C_k) + 1 - C_k)\,\frac{1}{C_k} = \ln(C_k) + \frac{1}{C_k} - 1$$

which is nonnegative and increasing in C_k for $C_k \geq 1$ because

$$\frac{d}{dc}(\ln(c) + \frac{1}{c} - 1) = \frac{1}{c} - \frac{1}{c^2} = \frac{c-1}{c^2} > 0$$

for $c > 1$.

If the desired failure probability is at most $\delta \in (0, 1)$, taking $C_k = e/\delta$ makes

$$\exp(-H(C_k)kp_t) = \exp(-\ln(e/\delta) - (\delta/e) + 1) = e^{-\delta/e}\delta < \delta.$$

Note that none of the above calculations use the actual behavior of p_t with respect to t.

We now specialize to Cauchy random variables. If $X_i \sim \text{Cauchy}\,(1)$,

$$p_t = \mathbb{P}\left\{|\lambda X_i| > t\right\} = \frac{2}{\pi}\arctan(\lambda/t) \leq \frac{2\lambda}{\pi t}$$

Consequently,

$$t \leq \frac{2\lambda}{\pi p_t} = \frac{2\lambda}{\pi}\frac{ke}{\delta}.$$

Typically, we want $\delta = N^{-c}$ with $c \geq 3$ say in order for the dimension reduction guarantee to hold for all pairs of points. Picking a larger value for the failure probability δ would make t smaller though. The alternative is to take λ small. We can now use lemma 4.4.0.1.

Corollary 4.4.1.2 (Lower Tail)**.** *For all* $0 < \eta < 1$*,* $N^{-c} \leq \epsilon < 1/4$*,* $X_j \overset{i.i.d.}{\sim}$ *Cauchy*$\,(1)$*, and*

$$\lambda_0 = \epsilon^2 \frac{\pi}{8keN^c},$$

the following bound holds:

$$(1-\epsilon)(1-4\epsilon^2)\mu(\eta\lambda_0) \leq \frac{1}{k}\sum_{j=1}^{k}\xi(\eta\lambda_0\,|X_j|)(1+\epsilon)(1+4\epsilon^2)\mu(\eta\lambda_0)$$

with failure probability at most $2/N^c$.

Proof. By lemma 4.4.1.1,

$$\max_i \{\lambda_0\,|X_i|\} \leq \frac{\epsilon^2}{4} < 1/6 \quad \text{with failure probability at most } 1/N^c.$$

On the other hand, we want to use lemma 4.3.0.3 to say

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

with failure probability at most $1/N^c$, noting there is no restriction on the size of λ_0. Following the discussion at the beginning of this chapter, we have to choose k as

$$\begin{aligned} &k \geq \ln(N^c)\frac{4}{\epsilon^2}\left(1+\frac{4}{\pi}-\frac{4}{\pi}\ln(\lambda_0)+8+2\sqrt{2}+\frac{1}{4}\right) \\ &= \frac{C}{\epsilon^2}\ln^2(N^c) \end{aligned}$$

for some constant C, having used $\ln(N^c) > \ln(1/\epsilon)$ and our choice of λ_0. We are now free to use lemma 4.4.0.1 to conclude

$$(1-\epsilon)(1-4\epsilon^2)\mu(\eta\lambda_0) \leq \frac{1}{k}\sum_{j=1}^{k}\xi(\eta\lambda_0\,|X_j|)(1+\epsilon)(1+4\epsilon^2)\mu(\eta\lambda_0)$$

with failure probability at most $2/N^c$. □

Chapter 5

The First and Second Moments

In the definition of metric ρ on $\mathbb{R}^k$, we could have used

$$\ln(1+\lambda\,|X|) \quad \text{or} \quad \ln(1+\sqrt{\lambda\,|X|}),$$

as the function applied to each coordinate, and we know the exact moments of such functions in terms of polylogarithms, using lemmas A.0.0.1 and A.0.0.13. However, it turns out using the linear combination

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

greatly simplifies the first moment and estimates of the second moment in terms of known functions. This first moment is also approximately 1/2-homogeneous at small scales (that is, for small λ), which will allow us to recover concentration properties there too. This homogeneity is lost if we use either of the logarithms individually, as a $-\lambda\ln(\lambda)$ term appears in those cases, as can already be seen in computing $\mathbb{E}\ln(1+\lambda\,|X|)$ in lemma 5.2.0.1. That term instead will appear in our estimates for the second moment and will become important when proving concentration at small scales.

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.

Remark 5.0.0.1. The computation of the first moment $\mu(\lambda)$ can also be done by using the reflection formula A.2.0.1 and the exact expressions for $\mathbb{E}\ln(1+\lambda\,|X|)$ and $\mathbb{E}\ln(1+\sqrt{\lambda\,|X|})$ in terms of dilogarithms. It would have a similar flavor to the proof of lemma 5.2.0.1 for computing $\mathbb{E}\ln(1+\lambda\,|X|)$. Such proofs may be debatably shorter, but the contour integral setup below hopefully makes the form of $\mu(\lambda)$ less mysterious. This proof method also appears to allow further generalizations without so much trouble, especially if we replace $\sqrt{x}$ by x^α with $\alpha = 1/2^k$ for integer $k \in \mathbb{N}$. Fractional powers of the logarithms may also be more approachable this way.

Lemma 5.0.0.2 (Contour Integral Setup)**.** *For $\lambda > 0$, $b > 0$, and $X \sim$ Cauchy(1),*

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

Remark 5.0.0.3. The task is then to simplify the complex logarithms on the right hand side when particular values of b are chosen. We shall do so 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 given by

$$\mathcal{C}^+ := \partial\left\{re^{i\theta} \mid 0 \le r \le R \text{ and } 0 \le \theta \le \pi - \epsilon\right\} \quad \text{oriented counterclockwise}$$

and

$$\mathcal{C}^- := \partial\left\{re^{i\theta} \mid 0 \le r \le R \text{ and } -\pi + \epsilon \le \theta \le 0\right\} \quad \text{oriented clockwise.}$$

Setting

$$\mathcal{C}^\pm(R) := \left\{Re^{\pm i\theta} \mid 0 \le \theta \le \pi - \epsilon\right\} \quad \text{and} \quad \mathcal{C}^\pm_\epsilon(R) := \left\{re^{\pm i(\pi-\epsilon)} \mid 0 \le r \le R\right\},$$

we have

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

We shall show that

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

and

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

On the other hand, keeping in mind the orientations of the contours, the residue theorem dictates for $R > 1$,

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

and similarly

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

For the $\mathcal{C}^{\pm}(R)$ integrals, using lemmas 5.0.0.5 and 3.1.1.6, we have when $|z| = R > \sqrt{2}$

$$|f(z)| \le \frac{\left(\ln(1+\sqrt{R}) + (\pi+\ln(2))/2\right)^b}{R^2\sqrt{R^{-4}+(1-2R^{-2})}}$$

Consequently, with $z = Re^{i\theta}$ so that $dz = Rie^{i\theta}\,d\theta$,

$$\left|\int_{\mathcal{C}^{+}(R)} f(z)\,dz\right| \le \int_0^{\pi-\epsilon}\left|f(Re^{i\theta})Rie^{i\theta}\right|\,d\theta$$
$$\le \frac{\left(\ln(1+\sqrt{\lambda R}) + (\pi+\ln(2))/2\right)^b}{R\sqrt{R^{-4}+(1-2R^{-2})}}(\pi-\epsilon)$$
$$\to 0$$

when $R\to\infty$ using lemma 5.0.0.4. Similar reasoning applies to the $\mathcal{C}^{-}(R)$ integral.

For the $\mathcal{C}_{\epsilon}^{\pm}(R)$ 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 $\pm ir$ when $\epsilon \to 0$. Consequently, when $z = re^{i(\pi-\epsilon)} = -re^{-i\epsilon}$,

$$\begin{aligned}
&\lim_{\epsilon\to 0}\int_{\mathcal{C}_\epsilon^+(R)} f(z)\,dz = \lim_{\epsilon\to 0}\int_R^0 f(-re^{-i\epsilon})\,(-e^{-i\epsilon})dr \\
&= \lim_{\epsilon\to 0}\int_0^R \frac{e^{-i\epsilon}\ln^b(1+i\sqrt{\lambda r}e^{-i\epsilon/2})}{1+r^2e^{-2i\epsilon}}\,dr
\end{aligned}$$

We want to use the dominated convergence theorem to take the limit inside the integral. Using lemmas 5.0.0.5 and 3.1.1.6 again, now assuming $\epsilon < \pi/8$,

$$\left|\frac{e^{-i\epsilon}\ln^b(1+i\sqrt{\lambda r}e^{-i\epsilon/2})}{1+r^2e^{-2i\epsilon}}\right| \le \frac{\big(\ln(1+\sqrt{\lambda r}) + (\pi+\ln(2))/2\big)^b}{\sqrt{1+r^4}}$$

which is not only bounded for $0 \le r \le R$, but also stays integrable when $r \to \infty$, again by lemma 5.0.0.4. The dominated convergence theorem now can say

$$\lim_{\epsilon\to 0}\int_{\mathcal{C}_\epsilon^+(R)} f(z)\,dz = \int_0^R \lim_{\epsilon\to 0}\frac{e^{-i\epsilon}\ln^b(1+i\sqrt{\lambda r}e^{-i\epsilon/2})}{1+r^2e^{-2i\epsilon}}\,dr = \int_0^R \frac{\ln^b(1+i\sqrt{\lambda r})}{1+r^2}\,dr.$$

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

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

that is

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

as claimed. □

Lemma 5.0.0.4. *For $b > 0$, $\lambda > 0$, and $c > 0$,*

$$\frac{\big(\ln(1+\sqrt{\lambda r})+c\big)^b}{r} \to 0 \quad \textit{as} \quad r \to \infty.$$

Proof. For $b < 2$, we can just use

$$\frac{\big(\ln(1+\sqrt{\lambda r})+c\big)^b}{r} \le \frac{\big(\sqrt{\lambda r}+c\big)^b}{r} = \frac{(\lambda r)^{b/2}}{r}\big(1+c/\sqrt{\lambda r}\big)^b \to 0$$

when $r \to \infty$ as $r^{1-\epsilon}/r^2 \to 0$ in that case.

For the larger b, the proof is by induction. Because we are sending $r \to \infty$, we may assume $r > 1/\lambda$ so that $\sqrt{\lambda r} < \lambda r$. In this case, similar reasoning to the above shows

$$\lim_{r\to\infty} \frac{\big(\ln(1+\lambda r)+c\big)^b}{r} = 0$$

for $b < 1$, which we take as the base case.

For the induction step, L'Hospital's rule dictates

$$\begin{aligned}
&0 \le \lim_{r\to\infty} \frac{\big(\ln(1+\lambda r)+c\big)^b}{r} = \lim_{r\to\infty} \frac{b\big(\ln(1+\lambda r)+c\big)^{b-1}}{1}\frac{\lambda}{1+\lambda r} \\
&< \lim_{r\to\infty} \frac{b\big(\ln(1+\lambda r)+c\big)^{b-1}}{1}\frac{\lambda}{\lambda r} = \lim_{r\to\infty} \frac{b\big(\ln(1+\lambda r)+c\big)^{b-1}}{r} \\
&= b\lim_{r\to\infty} \frac{\big(\ln(1+\lambda r)+c\big)^{b-1}}{r}
\end{aligned}$$

so if the limit is 0 for $b' \le b-1$, it is 0 for $b' \le b$ as well. □

Lemma 5.0.0.5. *Let $z \in \mathbb{C} - (-\infty, 0]$ and $\alpha \le 1/2$. Then, with $r = |z|$,*

$$|\ln(1+z^\alpha)| \le \ln(1+r^\alpha) + \frac{\pi + \ln(2)}{2}.$$

Proof. Let $z = re^{i\theta}$ with $|\theta| < \pi$. Then

$$z^\alpha = \exp(\alpha \ln(z)) = \exp(\alpha(\ln(r) + i\theta)) = r^\alpha e^{i\alpha\theta}.$$

Consequently,

$$\begin{aligned}
&\ln(1+z^\alpha) = \ln(1 + r^\alpha\cos(\alpha\theta) + ir^\alpha\sin(\alpha\theta)) \\
&= \frac{1}{2}\ln((1+r^\alpha\cos(\alpha\theta))^2 + r^{2\alpha}\sin^2(\alpha\theta)) + i\arctan\left(\frac{r^\alpha\sin(\alpha\theta)}{1+r^\alpha\cos(\alpha\theta)}\right) \\
&= \frac{1}{2}\ln(1 + 2r^\alpha\cos(\alpha\theta) + r^{2\alpha}) + i\arctan\left(\frac{r^\alpha\sin(\alpha\theta)}{1+r^\alpha\cos(\alpha\theta)}\right).
\end{aligned}$$

The cosine term is nonnegative as $|\alpha\theta| < \alpha\pi \leq \pi/2$ by our assumption on α.

By the AM-GM inequality B.0.0.1,

$$|\ln(1+z^\alpha)|^2 \leq \frac{1}{4}\ln^2(1+r^{2\alpha}+\cos^2(\alpha\theta)+r^{2\alpha}) + \arctan^2\left(\frac{r^\alpha \sin(\alpha\theta)}{1+r^\alpha\cos(\alpha\theta)}\right)$$
$$\leq \frac{1}{4}\ln^2(2+2r^{2\alpha}) + \frac{\pi^2}{4}$$

as $|\arctan(x)| \leq \pi/2$ for $x \in \mathbb{R}$.

Because the square root function is subadditive, we finally have

$$|\ln(1+z^\alpha)| \leq \left(\frac{1}{4}\ln^2(2+2r^{2\alpha}) + \frac{\pi^2}{4}\right)^{1/2}$$
$$\leq \frac{1}{2}\ln(2(1+r^{2\alpha})) + \frac{\pi}{2}$$
$$= \frac{\ln(2)}{2} + \ln(\sqrt{1+r^{2\alpha}}) + \frac{\pi}{2} \leq \frac{\ln(2)}{2} + \ln(1+r^\alpha) + \frac{\pi}{2}.$$

□

In the following sections, we specialize to the case $b = 1$ and $b = 2$ in order to compute the 1st and 2nd moments respectively. The complex integrals and residues then simplify to more identifiable functions.

5.1 1st Moment

Lemma 5.1.0.1. *If $\lambda > 0$ and $X \sim$ Cauchy(1), then*

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

that is,

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

Proof. Starting from lemma 5.0.0.2 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 A.1.0.9 and the atanh addition formula A.1.0.11,

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

By lemma A.0.0.12,

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

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. □

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

Lemma 5.1.0.2. *For $\lambda>0$,*

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

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

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

Remark 5.1.0.3. Numerically, the upper bound overestimates by a factor of $\sqrt{2}$ when $\lambda=1$, but the estimate gets much better for larger λ.

By lemma 5.1.0.1, we now also have the bound

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

using

$$0 < \frac{1}{2}\ln(1+\lambda^2) \le \frac{\lambda^2}{2}$$

for $\lambda > 0$.

Proof. The limit for large λ is immediate. We first show that the input has a unique maximum at $\lambda = 1$. It will be easier to view it as a function of $\nu = \sqrt{\lambda}$ as ν then has a positive derivative with respect to λ.

$$\frac{d}{d\nu}\frac{\nu\sqrt{2}}{1+\nu^2} = \frac{\sqrt{2}}{(1+\nu^2)^2}((1+\nu^2) - \nu(2\nu)) = \frac{\sqrt{2}}{(1+\nu^2)^2}(1-\nu^2)$$

which is positive for $\sqrt{\lambda} = \nu < 1$ and negative for $\sqrt{\lambda} = \nu > 1$. Because atanh is monotone increasing on $\mathbb{R}_+$, we have a unique maximum at $\nu = 1 = \lambda$, at which point the input is $1/\sqrt{2}$.

For the lower bound, note from the power series for atanh,

$$\text{atanh}(x) = \sum_{j=0}^{\infty} \frac{x^{2j+1}}{2j+1}$$

all terms are nonnegative when $x > 0$, so $\text{atanh}(x) > x$ in this case.

For the upper bound, use lemma A.1.0.10:

$$\text{atanh}(u) \le \frac{u}{1-u^2}$$

for $|u| < 1$. Consequently,

$$\text{atanh}\left(\frac{\sqrt{2\lambda}}{1+\lambda}\right) \le \frac{\sqrt{2\lambda}}{1+\lambda}\left(1 - \frac{2\lambda}{1+2\lambda+\lambda^2}\right)^{-1} = \frac{\sqrt{2\lambda}}{1+\lambda}\left(\frac{1+\lambda^2}{1+2\lambda+\lambda^2}\right)^{-1}$$
$$= \frac{\sqrt{2\lambda}}{1+\lambda}\frac{1+2\lambda+\lambda^2}{1+\lambda^2} = \frac{\sqrt{2\lambda}}{1+\lambda}\left(1 + \frac{2\lambda}{1+\lambda^2}\right).$$

At the beginning of the proof, we showed $\lambda/(1+\lambda^2)$ is strictly increasing for $\lambda \le 1$, so for any $\lambda \le \lambda_0 \le 1$,

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

□

5.2 An Auxiliary Mean

The following mean will be useful in some of the later estimates for the second moment.

Lemma 5.2.0.1. *For* $\lambda \geq 0$ *and* $X \sim \textit{Cauchy}(1)$*,*

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

Remark 5.2.0.2. It is clear that the expecation is nonnegative when $0 < \lambda \leq 1$. When $\lambda > 1$, see lemma 5.2.0.3.

Proof. From lemma A.0.0.1

$$\begin{aligned} &\mathbb{E} \ln(1 + \lambda |X|) = \frac{1}{i\pi}(\operatorname{Li}_2(1 + i\lambda) - \operatorname{Li}_2(1 - i\lambda)) \\ &= \frac{2}{\pi}\frac{1}{2i}(\operatorname{Li}_2(1 + i\lambda) - \operatorname{Li}_2(1 - i\lambda)) \end{aligned}$$

We use the reflection formula A.2.0.1 to expand the dilogarithm terms.

Recall from lemma A.2.0.1, for $z \in (\mathbb{C} - \mathbb{R}) \cup (0, 1)$,

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

So we have

$$\operatorname{Li}_2(1 - i\lambda) = -\ln(i\lambda) \ln(1 - i\lambda) - \operatorname{Li}_2(i\lambda) + \operatorname{Li}_2(1)$$

and

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

Consequently,

$$\begin{aligned} &\frac{1}{2i}(\operatorname{Li}_2(1 + i\lambda) - \operatorname{Li}_2(1 - i\lambda)) \\ &= \frac{1}{2i}\big(-\ln(-i\lambda) \ln(1 + i\lambda) - \operatorname{Li}_2(-i\lambda) + \operatorname{Li}_2(1)\big) \\ &\quad -\frac{1}{2i}\big(-\ln(i\lambda) \ln(1 - i\lambda) - \operatorname{Li}_2(i\lambda) + \operatorname{Li}_2(1)\big) \\ &= \frac{1}{2i}\big(-\ln(-i\lambda) \ln(1 + i\lambda) + \ln(i\lambda) \ln(1 - i\lambda) + \operatorname{Li}_2(i\lambda) - \operatorname{Li}_2(-i\lambda)\big) \\ &= \frac{1}{2i}\left(-\left(\ln(\lambda) - \frac{i\pi}{2}\right) \ln(1 + i\lambda) + \left(\ln(\lambda) + \frac{i\pi}{2}\right) \ln(1 - i\lambda)\right) + \operatorname{Ti}_2(\lambda) \\ &= \frac{1}{2i} \ln(\lambda)(\ln(1 - i\lambda) - \ln(1 + i\lambda)) + \frac{\pi}{4}(\ln(1 - i\lambda) + \ln(1 + i\lambda)) + \operatorname{Ti}_2(\lambda) \end{aligned}$$

By lemma A.0.0.11 (really the remark there) and the definition of arctan,

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

Thus,

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

□

We take some time to better understand how $\mathbb{E}\ln(1+\lambda|X|)$ behaves as a function of λ. We know it is increasing from its definition as an expectation of increasing functions of λ, but depending on the size of λ, certain terms contribute much more than others.

Lemma 5.2.0.3. *For $\lambda > 0$, let*

$$f(\lambda) := \mathrm{Ti}_2(\lambda) - \ln(\lambda)\arctan(\lambda).$$

Then

$$0 < f(\lambda) = f\left(\frac{1}{\lambda}\right) \le f(1) = \mathrm{Ti}_2(1) < 1.$$

and goes to 0 as $\lambda \to \infty$ or $\lambda \to 0$.

Proof. Take the derivative

$$\frac{d}{d\lambda}(\mathrm{Ti}_2(\lambda) - \ln(\lambda)\arctan(\lambda)) = \frac{\arctan(\lambda)}{\lambda} - \frac{\arctan(\lambda)}{\lambda} - \frac{\ln(\lambda)}{1+\lambda^2} = -\frac{\ln(\lambda)}{1+\lambda^2}$$

which is positive for $\lambda < 1$ and negative for $\lambda > 1$ and hence $\lambda = 1$ is the unique maximizer.

For showing the equality, consider the inversion formula A.3.0.3 for $\mathrm{Ti}_2(\lambda)$,

$$\mathrm{Ti}_2(\lambda) = \mathrm{Ti}_2(1/\lambda) + \frac{\pi}{2}\ln(\lambda)$$

so that, using the inversion formula A.3.0.1 for $\arctan(\lambda)$,

$$f(\lambda) = \mathrm{Ti}_2(\lambda) - \ln(\lambda)\arctan(\lambda) = \mathrm{Ti}_2(1/\lambda) + \ln(\lambda)\left(\frac{\pi}{2} - \arctan(\lambda)\right)$$
$$= \mathrm{Ti}_2(1/\lambda) + \ln(\lambda)\arctan(1/\lambda) = \mathrm{Ti}_2(1/\lambda) - \ln(1/\lambda)\arctan(1/\lambda)$$
$$= f(1/\lambda)$$

For the lower bound, all terms of $f(\lambda)$ are nonnegative for $0 < \lambda < 1$, and hence $f(1/\lambda)$ must be nonnegative too. Alternatively, the 3rd line above has all terms nonnegative for $\lambda \geq 1$: the power series for $\mathrm{Ti}_2(1/\lambda)$ is alternating with terms of strictly decreasing magnitude:

$$\mathrm{Ti}_2(1/\lambda) = \sum_{j=0}^{\infty} (-1)^j \frac{\lambda^{-(2j+1)}}{(2j+1)^2}.$$

This series also shows

$$\frac{1}{\lambda} > \mathrm{Ti}_2(1/\lambda) > \frac{1}{\lambda} - \frac{1}{9\lambda^2} > 0.$$

The same reasoning yields $1 > \mathrm{Ti}_2(1) > 1 - 1/9$. □

5.3 2nd Moment

To estimate the 2nd moment $\mathbb{E}\xi^2(\lambda\,|X|)$, note that for any $a, b > 0$,

$$(a+b)^2 = a^2 + b^2 + 2ab = a^2 + b^2 + 2\sqrt{a^2 b^2} \leq a^2 + b^2 + 2\left(\frac{a^2}{2} + \frac{b^2}{2}\right) = 2a^2 + 2b^2$$

by the AM-GM inequality B.0.0.1. Consequently,

$$\begin{aligned}
&\mathbb{E}\xi^2(\lambda\,|X|) = \mathbb{E}\left(\ln(1+\sqrt{\lambda\,|X|}) + \frac{1}{2}\ln(1+\lambda\,|X|)\right)^2 \\
&\leq \mathbb{E}\left(2\ln^2(1+\sqrt{\lambda\,|X|}) + 2\frac{1}{4}\ln^2(1+\lambda\,|X|)\right) \\
&= \mathbb{E}\left(2\ln^2(1+\sqrt{\lambda\,|X|}) + \frac{1}{2}\ln^2(1+\lambda\,|X|)\right).
\end{aligned}$$

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

Lemma 5.3.0.1. *If $\lambda > 0$ and $X \sim$ Cauchy(1), then*

$$\begin{aligned}
&\mathbb{E}\left(2\ln^2(1+\sqrt{\lambda\,|X|}) + \frac{1}{2}\ln^2(1+\lambda\,|X|)\right) \\
&= 2\mathbb{E}\arctan^2(\sqrt{\lambda\,|X|}) + \mu^2(\lambda) - \arctan^2(\lambda) + 2\arctan(\lambda)h(\sqrt{\lambda}) - h^2(\sqrt{\lambda})
\end{aligned}$$

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 lemma 5.0.0.2 with $b = 2$,

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

that is,

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

By lemma 5.3.0.4,

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

For the residue terms, we use lemma 5.3.0.7:

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

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 5.1.0.1, we can further simplify:

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

Putting everything together we may conclude

$$
\begin{aligned}
&\mathbb{E}\left(2\ln^2(1+\sqrt{\lambda\,|X|})+\frac{1}{2}\ln^2(1+\lambda\,|X|)\right)\\
&=2\mathbb{E}\arctan^2(\sqrt{\lambda\,|X|})+\mu^2(\lambda)-\arctan^2(\lambda)+2\arctan(\lambda)h(\sqrt{\lambda})-h^2(\sqrt{\lambda}).
\end{aligned}
$$

□

Remark 5.3.0.2. Using lemma 5.3.0.5, we thus have the upper bound

$$
\begin{aligned}
&\mathbb{E}\xi^2(\lambda\,|X|)\\
&\leq \min\left\{2\mathbb{E}\ln(1+\lambda\,|X|),\frac{\pi^2}{2}\right\}+\mu^2(\lambda)-\arctan^2(\lambda)+2\arctan(\lambda)h(\sqrt{\lambda})-h^2(\sqrt{\lambda})\\
&=\min\left\{2\mathbb{E}\ln(1+\lambda\,|X|),\frac{\pi^2}{2}\right\}+\mu^2(\lambda)-\big(\arctan(\lambda)-h(\sqrt{\lambda})\big)^2\\
&\leq \min\left\{2\mathbb{E}\ln(1+\lambda\,|X|),\frac{\pi^2}{2}\right\}+\mu^2(\lambda).
\end{aligned}
$$

In particular, the variance is bounded from above by $\pi^2/2$ for *all* $\lambda > 0$.

Lemma 5.3.0.3. *For* $0 < \lambda \leq 1$

$$
\frac{\mathbb{E}\xi^2(\lambda\,|X|)}{\lambda}\leq \lambda+\frac{4}{\pi}-\frac{4}{\pi}\ln(\lambda)+\frac{8}{(1+\lambda)^2}+2\lambda\frac{\sqrt{2\lambda}}{1+\lambda}+\frac{\lambda^3}{4}
$$

while for $\lambda \geq 1$*,*

$$
\frac{\mathbb{E}\xi^2(\lambda\,|X|)}{\lambda}\leq \frac{\pi^2}{2}+2+\lambda\sqrt{2}+\frac{\lambda^3}{4}.
$$

Proof. By remark 5.3.0.2 and lemma 5.2.0.1, we have

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

using that $\operatorname{Ti}_2(\lambda)$ is an alternating series with terms of decreasing magnitude for $\lambda < 2$ and that for $\lambda \leq 1$, $\ln(\lambda)$ is nonnegative. By remark 5.1.0.3,

$$
\mu^2(\lambda)\leq\left(2\frac{\sqrt{2\lambda}}{1+\lambda}+\frac{\lambda^2}{2}\right)^2=\frac{8\lambda}{(1+\lambda)^2}+2\lambda^2\frac{\sqrt{2\lambda}}{1+\lambda}+\frac{\lambda^4}{4}.
$$

Consequently, for $\lambda \leq 1$,

$$\frac{\mathbb{E}\xi^2(\lambda\,|X|)}{\lambda} \leq \lambda + \frac{4}{\pi} - \frac{4}{\pi}\ln(\lambda) + \frac{8}{(1+\lambda)^2} + 2\lambda\frac{\sqrt{2\lambda}}{1+\lambda} + \frac{\lambda^3}{4}$$

When $\lambda \geq 1$, we can instead use the other upper bound for the atanh term, using lemma 5.1.0.2.

$$\frac{\mathbb{E}\xi^2(\lambda\,|X|)}{\lambda} \leq \frac{\pi^2}{2\lambda} + \frac{1}{\lambda}\left(\sqrt{2} + \frac{\lambda^2}{2}\right)^2 \leq \frac{\pi^2}{2} + 2 + \lambda\sqrt{2} + \frac{\lambda^3}{4}.$$

□

Lemma 5.3.0.4. *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

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

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

□

Lemma 5.3.0.5. *For $\nu > 0$,*

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

is strictly concave. Consequently, for any $\lambda > 0$ and $X \sim$ Cauchy(1),

$$\mathbb{E}\arctan^2(\sqrt{\lambda\,|X|}) \leq \min\left\{\mathbb{E}\ln(1+\lambda\,|X|), \frac{\pi^2}{4}\right\}.$$

Remark 5.3.0.6. The function is also strictly increasing and 0 when $\nu = 0$, so the function is subadditive too. This bound contains a $-\lambda\ln(\lambda)$ term for small λ.

Proof. Upon taking the derivative,

$$\frac{d}{d\nu}\arctan^2(\sqrt{\nu}) = \frac{2\arctan(\sqrt{\nu})}{1+(\sqrt{\nu})^2}\frac{1}{2\sqrt{\nu}} = \frac{\arctan(\sqrt{\nu})}{\sqrt{\nu}}\frac{1}{1+\nu}$$

which is positive for $\nu > 0$. We shall show the derivative is decreasing as well, as a product of decreasing functions. We focus on the arctan fraction, as $1/(1+\nu)$ is decreasing.

$$\frac{d}{d\nu}\frac{\arctan(\sqrt{\nu})}{\sqrt{\nu}} = \frac{1}{\nu}\left(\frac{\sqrt{\nu}}{1+\nu}\frac{1}{2\sqrt{\nu}} - \frac{\arctan(\sqrt{\nu})}{2\sqrt{\nu}}\right)$$
$$= \frac{1}{2\nu^{3/2}}\left(\frac{\sqrt{\nu}}{1+\nu} - \arctan(\sqrt{\nu})\right)$$

so we just need to show the bracketed term is nonpositive. It is 0 when $\nu = 0$, and we show it is decreasing:

$$\frac{d}{d\nu}\left(\frac{\sqrt{\nu}}{1+\nu} - \arctan(\sqrt{\nu})\right) = \frac{1}{2\sqrt{\nu}(1+\nu)} - \frac{\sqrt{\nu}}{(1+\nu)^2} - \frac{1}{1+\nu}\frac{1}{2\sqrt{\nu}} = -\frac{\sqrt{\nu}}{(1+\nu)^2} < 0$$

as desired.

For the upper bounds, 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 are 0 when $\nu = 0$.

$$\frac{d}{d\nu}\arctan^2(\sqrt{\nu}) = \frac{\arctan(\sqrt{\nu})}{\sqrt{\nu}}\frac{1}{1+\nu} \leq \frac{\sqrt{\nu}}{\sqrt{\nu}}\frac{1}{1+\nu} = \frac{1}{1+\nu} = \frac{d}{d\nu}\ln(1+\nu).$$

□

Lemma 5.3.0.7. *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

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

Proof. Using lemma A.1.0.9,

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

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 5.3.0.4,

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

We also have

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

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 A.1.0.11, 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

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

So we are left to understand $h(\nu)$. By lemma 5.3.0.8, it is

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

□

Lemma 5.3.0.8. *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 5.3.0.9. 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 A.1.0.6, we can convert $h(\nu)$ as follows

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

using $\sqrt{-i} = -i\sqrt{i}$. We now use the inversion formula A.3.0.1 for arctan.

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

We claim

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

Both expressions are 0 at $\nu = 1$, so we just need to show the derivatives match for $\nu > 0$.

On one hand,

$$\begin{aligned}
&\frac{d}{d\nu}(-1)\arctan\left(\frac{\nu}{\sqrt{2}} - \frac{1}{\nu\sqrt{2}}\right) = -\left(1 + \left(\frac{\nu}{\sqrt{2}} - \frac{1}{\nu\sqrt{2}}\right)^2\right)^{-1}\frac{d}{d\nu}\left(\frac{\nu}{\sqrt{2}} - \frac{1}{\nu\sqrt{2}}\right)\\
&= -\left(1 + \frac{1}{2}\left(\frac{\nu^2 - 1}{\nu}\right)^2\right)^{-1}\left(\frac{1}{\sqrt{2}} + \frac{1}{\nu^2\sqrt{2}}\right)\\
&= -\frac{1}{\sqrt{2}}\left(\frac{2\nu^2 + \nu^4 - 2\nu^2 + 1}{2\nu^2}\right)^{-1}\left(\frac{\nu^2 + 1}{\nu^2}\right) = -\frac{1}{\sqrt{2}}\frac{2\nu^2}{1 + \nu^4}\frac{\nu^2 + 1}{\nu^2}\\
&= -\frac{\sqrt{2}(1 + \nu^2)}{1 + \nu^4}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\frac{d}{d\nu}\big(\arctan(i\nu\sqrt{i}) + \arctan(-i\sqrt{i}/\nu)\big)\\
&= \frac{i\sqrt{i}}{1 - i\nu^2} + \frac{-i\sqrt{i}}{1 - i/\nu^2}\frac{(-1)}{\nu^2} = \frac{i\sqrt{i}}{1 - i\nu^2} + \frac{i\sqrt{i}}{\nu^2 - i} = \frac{i\sqrt{i}}{1 - i\nu^2} + \frac{-\sqrt{i}}{i\nu^2 + 1}\\
&= \frac{\sqrt{i}}{1 + \nu^4}\big(i(1 + i\nu^2) - (1 - i\nu^2)\big) = \frac{\sqrt{i}}{1 + \nu^4}\big(i - \nu^2 - 1 + i\nu^2\big)\\
&= \frac{\sqrt{i}}{1 + \nu^4}(i - 1)(1 + \nu^2) = \frac{-\sqrt{2}\sqrt{-i}\sqrt{i}}{1 + \nu^4}(1 + \nu^2)\\
&= -\frac{\sqrt{2}(1 + \nu^2)}{1 + \nu^4}
\end{aligned}$$

Because the derivatives match, we are done. □

Chapter 6

Conclusions

6.1 Conclusions

One of the main motivations for this work was to better understand which properties of a Banach space can be dropped while still keeping a dimension reduction result. We have seen that translation invariance can be kept, and the (round) ℓ_2 metric is not essential, as the results we have approximately preserve level sets of the 1-norm. We have also seen that dropping the 1-homogeneity of the metric makes proofs much more long-winded: we had to have separate proofs based on the magnitude of λ, and the bounded variance property was crucial for the proofs at large scales. Normally, the 1-homogeneity allows us to fix a scale λ_0 and prove concentration only there, as concentration at all other scales follows from it.

Many of the standard tools for proving concentration inequalities rely on the boundedness of the random variables, for which knowledge of the second moment is sufficient; we could not use those at all, due to the dependence on λ and the unboundedness of the Cauchy random variables. In many cases, trying to truncate the random variables or bound λ gave poor estimates.

A second observation is

$$\frac{1}{1+\epsilon}\mu(\lambda) \leq \mu\left(\frac{\lambda}{1+\epsilon}\right) < \mu((1+\epsilon)\lambda) \leq (1+\epsilon)\mu(\lambda)$$

because μ is concave, strictly increasing, and satisfies $\mu(0) = 0$, with the last inequality following from

$$\mu(\lambda) = \mu\left(\frac{(1+\epsilon)\lambda}{1+\epsilon}\right) \geq \frac{1}{1+\epsilon}\mu((1+\epsilon)\lambda).$$

Thus, our theorem also gives a bi-Lipschitz embedding result at all distances mentioned. We preferred the current formulation of the theorem because $\mu(\lambda) \sim \ln(\lambda)$ for large λ, so the bi-Lipschitz result would tell us less about the possible values of $\|x-y\|_1$ than the original formulation when the distances are large. There does not seem to be a standard name for keeping the error inside μ this way. However, for $a \geq 0$, let

$$\mu_-(a) := \mu\left(\frac{a}{1+\epsilon}\right) \quad \text{and} \quad \mu_+(a) := \mu((1+\epsilon)a).$$

In this case, John Roe [Roe03, page 151] would call F in the theorem a "coarse embedding" of $P \subset \ell_1^D$ into (R^k, ρ).

A third observation is that while special functions can be useful for testing ideas, the contour integral approach led to much simpler proofs in many cases. Finding the appropriate contour took some time though and many of the calculations were more approachable from the practice gleaned from working with those special functions.

Finally, throughout this work we have found how useful computer algebra software, such as Mathematica, can be when making conjectures. Realizing that μ could have a much simpler form by taking ξ as a linear combination of logarithms was only discovered after staring at power series for the latter for a spell. We have also seen how much patience can be needed when verifying such conjectures by hand.

6.2 Conjectures

There is a theorem of Pisier [Pis83] which says that there are subspaces of ℓ_1^d $(1 \pm \epsilon)$ isomorphic to ℓ_p^n with n proportional to d. See also [MS86, Chapter 8]. The references mentioned say there *exist* linear embeddings H of $P \subset \ell_p^{D'}$ into ℓ_1^D, but constructing such an H is not immediately clear. The proofs there substitute a finite sequence of p-stable random variables with an infinite sequence that is "more tractable" for the proofs, but it is not clear how to go in the other direction. If one had such an $H : \ell_p^{D'} \overset{1+\epsilon}{\hookrightarrow} \ell_1^D$ available,

Corollary 6.2.0.1. *With F, μ, ρ, ϵ, and k as in theorem 1.2.0.1, the bounds*

$$(1-\epsilon)\mu\left((1-\epsilon)\|x-y\|_p\right) \leq \rho(FH(x), FH(y)) \leq (1+\epsilon)\mu((1+\epsilon)\|x-y\|_p)$$

hold with probability at least $1 - N^{-c-2}$ for all $x, y \in P$ satisfying $\|x-y\|_p > 8\epsilon^2(1+\epsilon)$.

We're using that k did not depend on D in theorem 1.2.0.1. The following conjecture concerns directly constructing an embedding matrix from $\mathbb{R}^D \to \mathbb{R}^k$ satisfying the required bounds. Specifically,

Conjecture 6.2.0.2. For $1 \leq j \leq D$ and $1 \leq i \leq k$, draw the entries of the matrix $F : \mathbb{R}^D \to \mathbb{R}^k$, F_{ij}, as i.i.d. copies of a standard p-stable random variable. Then with ρ, ϵ, and k as in theorem 1.2.0.1, and with $\mu : \mathbb{R}_+ \to \mathbb{R}_+$ defined as

$$\mu(\lambda) := \mathbb{E}\xi(\lambda F_{11}),$$

the following bounds hold

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

for all $x, y \in P$ with probability at least $1 - N^{-c-2}$.

Appendix A

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 [Lew81] and [Max03].

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

Lemma A.0.0.1. *Let $X \sim \mathit{Cauchy}(1)$ 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

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

With $u = 1 + \nu x$,

$$\begin{aligned}\frac{2}{\pi}\int_0^\infty \frac{\ln^b(1+\nu x)}{1+x^2}\,dx &= \frac{2}{\pi}\frac{\nu}{\nu}\int_1^\infty \frac{\ln^b(u)}{1+\nu^{-2}(u-1)^2}\,\frac{du}{\nu} \\ &= \frac{2\nu}{\pi}\int_1^\infty \frac{\ln^b(u)}{\nu^2 + (u-1)^2}\,du\end{aligned}$$

and with $t = \ln(u)$ so that $dt = du/u \Rightarrow e^t\,dt = du$,

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

Because

$$\begin{aligned}
&\frac{1+i\nu}{e^t-(1+i\nu)}-\frac{1-i\nu}{e^t-(1-i\nu)} \\
&=\frac{2i\nu e^t-(1+\nu^2)+(1+\nu^2)}{(e^t-(1+i\nu))(e^t-(1-i\nu))} \\
&=\frac{2i\nu e^t}{(e^t-(1+i\nu))(e^t-(1-i\nu))},
\end{aligned}$$

we may write

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

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 [SS03] for proofs and [Nee97] 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. [SS03, page 52,96].

Theorem A.0.0.2 (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 A.0.0.3 (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 A.0.0.4 (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 A.0.0.5*.* 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

$$zw = rse^{i(\theta+\phi)} \quad \text{with} \quad |\theta + \phi| < \pi$$

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 A.0.0.6 (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.

Unlike $\mathrm{Li}_1(z)$, the higher order polylogarithms extend to the unit circle.

Definition A.0.0.7 (Higher Integral Order Polylogarithms)**.** For $2 \leq n \in \mathbb{N}$, the *polylogarithm of order* n, $\mathrm{Li}_n(z)$ is defined as

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

and by

$$\mathrm{Li}_n(z) := \int_0^z \frac{\mathrm{Li}_{n-1}(w)}{w}\, dw \quad \text{for} \quad z \in \mathbb{C} - (1, \infty).$$

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

Definition A.0.0.8. 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

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

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

To check that the definitions are consistent, note that if $|z| < 1$, then $\left|e^{-t}z\right| < 1$ too, and we may use the geometric series to rewrite:

$$\mathrm{Li}_b(z) = \frac{1}{\Gamma(b)} \int_0^\infty zt^{b-1}e^{-t} \sum_{j=0}^\infty e^{-tj} z^j\, dt$$

and if we can exchange the integral and the sum,

$$= \sum_{j=0}^\infty \frac{z^{j+1}}{\Gamma(b)} \int_0^\infty t^{b-1}e^{-t(j+1)}\, dt = \sum_{j=1}^\infty \frac{z^j}{\Gamma(b)} \int_0^\infty t^{b-1}e^{-tj}\, dt$$

Now, with $s/j = t$,

$$\frac{1}{\Gamma(b)} \int_0^\infty t^{b-1}e^{-tj}\, dt \;= \frac{1}{\Gamma(b)} \int_0^\infty (s/j)^{b-1}e^{-s}\, \frac{ds}{j} \;= \frac{1}{j^b}\frac{1}{\Gamma(b)} \int_0^\infty (s)^{b-1}e^{-s}\, ds \;= \frac{1}{j^b}$$

so we recover, when $|z| < 1$,

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

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

Lemma A.0.0.9. *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),

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

□

For all $b > 1$, we also have

$$\mathrm{Li}_b(z) = \int_0^z \frac{\mathrm{Li}_{b-1}(w)}{w}\, dw$$

just note from the power series

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

Because $1/z$ and $\mathrm{Li}_{b-1}(z)$ analytic on $\mathbb{C} - [1, \infty)$ away from 0, $\mathrm{Li}_{b-1}(z)/z$ is too. Still from the power series, all terms are degree 0 or higher, so $\mathrm{Li}_{b-1}(z)/z$ is also analytic for $|z| < 1$ and hence on all of $\mathbb{C} - [1, \infty)$. Consequently,

$$\int_0^z \frac{\mathrm{Li}_{b-1}(w)}{w}\, dw$$

is analytic there too. Because this integral agrees with $\mathrm{Li}_b(z)$ for $|z| < 1$, analytic continuation dictates it agrees with $\mathrm{Li}_b(z)$ on the full domain $\mathbb{C} - [1, \infty)$.

Lemma A.0.0.10. *For $b > 1$, and $0 < x < 1$,*

$$\mathrm{Li}_b(x) \leq x\, \mathrm{Li}_b(1).$$

Proof. From the power series,

$$\mathrm{Li}_b(x) = \sum_{j=1}^{\infty} \frac{x^j}{j^b} = x \sum_{j=1}^{\infty} \frac{x^{j-1}}{j^b} \leq x \sum_{j=1}^{\infty} \frac{1}{j^b} = x\, \mathrm{Li}_b(1)$$

having used $x^k \leq 1$ for $0 < x < 1$ and $k \geq 0$. □

Lemma A.0.0.11. *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 A.0.0.12. 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 $z = \pm 1$. □

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,

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

Powers and polynomials of such functions also have this property. Intuitively this symmetry follows from the real coeffecients 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 [SS03] 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.

Lemma A.0.0.13. *Let $\lambda > 0$, $X \sim \mathit{Cauchy}(1)$ and $b > -1$. Then*

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

Proof. With $\lambda y = x$, we have

$$I_b(\lambda) := \mathbb{E} \ln^b(1 + \sqrt{\lambda X}) := \frac{2}{\pi} \int_0^{\infty} \frac{\ln^b(1 + \sqrt{\lambda y})}{1 + y^2} \, dy = \frac{2}{\pi} \frac{\lambda}{\lambda} \int_0^{\infty} \frac{\ln^b(1 + \sqrt{x})}{1 + \lambda^{-2} x^2} \frac{dx}{\lambda}$$
$$= \frac{2\lambda}{\pi} \int_0^{\infty} \frac{\ln^b(1 + \sqrt{x})}{\lambda^2 + x^2} \, dx.$$

If $u^2 = x$ so that $2u\,du = dx$,

$$\begin{aligned}
I_b(\lambda) &= \frac{2\lambda}{\pi}\int_0^\infty \frac{\ln^b(1+u)}{\lambda^2+u^4}2u\,du \\
&= \frac{2\lambda}{\pi}\int_0^\infty \frac{\ln^b(1+u)}{(u^2+i\lambda)}\frac{2u}{u^2-i\lambda}\,du \\
&= \frac{2\lambda}{\pi}\int_0^\infty \frac{\ln^b(1+u)}{(u^2+i\lambda)}\left(\frac{1}{(u-\sqrt{i\lambda})}+\frac{1}{(u+\sqrt{i\lambda})}\right)\,du
\end{aligned}$$

Now

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

So we can write

$$I_b(\lambda) = \frac{\sqrt{\lambda}}{i\pi\sqrt{i}}\int_0^\infty \ln^b(1+u)\left(\frac{1}{u-i\sqrt{i\lambda}}-\frac{1}{u+i\sqrt{i\lambda}}\right)\left(\frac{1}{(u-\sqrt{i\lambda})}+\frac{1}{(u+\sqrt{i\lambda})}\right)\,du,$$

and with $v = u+1$,

$$\begin{aligned}
I_b(\lambda) &= \frac{\sqrt{\lambda}}{i\pi\sqrt{i}}\int_1^\infty \ln^b(v)\left(\frac{1}{v-1-i\sqrt{i\lambda}}-\frac{1}{v-1+i\sqrt{i\lambda}}\right) \\
&\quad \left(\frac{1}{(v-1-\sqrt{i\lambda})}+\frac{1}{(v-1+\sqrt{i\lambda})}\right)\,dv \\
&= \frac{\sqrt{\lambda}}{i\pi\sqrt{i}}\int_1^\infty \ln^b(v)\left(\frac{1}{v-a_+}-\frac{1}{v-a_-}\right)\left(\frac{1}{v-c_+}+\frac{1}{v-c_-}\right)\,dv
\end{aligned}$$

upon setting

$$a_\pm := 1 \pm i\sqrt{i\lambda} \quad \text{and} \quad c_\pm := 1 \pm \sqrt{i\lambda}.$$

So

$$I_b(\lambda) = \mathbb{E}\ln^b(1+\sqrt{\lambda X}) = \boxed{a_+, c_+} - \boxed{a_-, c_+} + \boxed{a_+, c_-} - \boxed{a_-, c_-}$$

with

$$\boxed{z_1, z_2} = \frac{\sqrt{\lambda}}{i\pi\sqrt{i}}\int_1^\infty \frac{\ln^b(v)}{(v-z_1)(v-z_2)}\,dv = \frac{\sqrt{\lambda}\Gamma(b+1)}{(z_1-z_2)i\pi\sqrt{i}}(\mathrm{Li}_{b+1}(z_1)-\mathrm{Li}_{b+1}(z_2)).$$

by lemma A.0.0.14.

Compute,

$$\begin{aligned}
&\boxed{a_+, c_+} - \boxed{a_-, c_+}\\
&= \frac{\Gamma(b+1)\sqrt{\lambda}}{i\pi\sqrt{i}}\left(\frac{\mathrm{Li}_{b+1}(a_+) - \mathrm{Li}_{b+1}(c_+)}{(i-1)\sqrt{i\lambda}} - \frac{\mathrm{Li}_{b+1}(a_-) - \mathrm{Li}_{b+1}(c_+)}{-(i+1)\sqrt{i\lambda}}\right)\\
&= \frac{\Gamma(b+1)}{i^2\pi}\left(\frac{\mathrm{Li}_{b+1}(a_+) - \mathrm{Li}_{b+1}(c_+)}{(i-1)} - \frac{\mathrm{Li}_{b+1}(a_-) - \mathrm{Li}_{b+1}(c_+)}{-(i+1)}\right)\\
&= \frac{\Gamma(b+1)}{i^2\pi}\left(\frac{\mathrm{Li}_{b+1}(a_+) - \mathrm{Li}_{b+1}(c_+)}{(i-1)} + \frac{\mathrm{Li}_{b+1}(a_-) - \mathrm{Li}_{b+1}(c_+)}{(i+1)}\right)\\
&= \frac{\Gamma(b+1)}{i^2\pi(-2)}((i+1)\,\mathrm{Li}_{b+1}(a_+) + (i-1)\,\mathrm{Li}_{b+1}(a_-) - 2i\,\mathrm{Li}_{b+1}(c_+))
\end{aligned}$$

and similarly

$$\begin{aligned}
&\boxed{a_+, c_-} - \boxed{a_-, c_-}\\
&= \frac{\Gamma(b+1)\sqrt{\lambda}}{i\pi\sqrt{i}}\left(\frac{\mathrm{Li}_{b+1}(a_+) - \mathrm{Li}_{b+1}(c_-)}{(i+1)\sqrt{i\lambda}} - \frac{\mathrm{Li}_{b+1}(a_-) - \mathrm{Li}_{b+1}(c_-)}{-(i-1)\sqrt{i\lambda}}\right)\\
&= \frac{\Gamma(b+1)}{i^2\pi}\left(\frac{\mathrm{Li}_{b+1}(a_+) - \mathrm{Li}_{b+1}(c_-)}{(i+1)} + \frac{\mathrm{Li}_{b+1}(a_-) - \mathrm{Li}_{b+1}(c_-)}{(i-1)}\right)\\
&= \frac{\Gamma(b+1)}{i^2\pi(-2)}((i-1)\,\mathrm{Li}_{b+1}(a_+) + (i+1)\,\mathrm{Li}_{b+1}(a_-) - 2i\,\mathrm{Li}_{b+1}(c_-)).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathbb{E}\ln^b(1+\sqrt{\lambda X})\\
&= \frac{\Gamma(b+1)(2i)}{i^2\pi(-2)}(\mathrm{Li}_{b+1}(a_+) + \mathrm{Li}_{b+1}(a_-) - \mathrm{Li}_{b+1}(c_+) - \mathrm{Li}_{b+1}(c_-))\\
&= \frac{\Gamma(b+1)}{i\pi}(\mathrm{Li}_{b+1}(c_+) + \mathrm{Li}_{b+1}(c_-) - \mathrm{Li}_{b+1}(a_+) - \mathrm{Li}_{b+1}(a_-))
\end{aligned}$$

as desired. □

Lemma A.0.0.14. *Let $z_1, z_2 \in \mathbb{C} - [1,\infty)$ with $z_1 \neq z_2$, and $b > -1$. Then*

$$\int_1^\infty \frac{\ln^b(v)}{(v-z_1)(v-z_2)}\,dv = \frac{\Gamma(b+1)}{z_1-z_2}(\mathrm{Li}_{b+1}(z_1) - \mathrm{Li}_{b+1}(z_2)).$$

Proof. Because

$$\frac{z_1}{v-z_1} - \frac{z_2}{v-z_2} = \frac{z_1 v - z_1 z_2 - z_2 v + z_1 z_2}{(v-z_1)(v-z_2)} = \frac{(z_1-z_2)v}{(v-z_1)(v-z_2)},$$

we can write, with $w = \ln(v)$ so that $dw = dv/v$,

$$\int_1^\infty \frac{\ln^b(v)}{(v-z_1)(v-z_2)}\,dv = \frac{1}{z_1-z_2}\int_1^\infty \ln^b(v)\left(\frac{z_1}{v-z_1} - \frac{z_2}{v-z_2}\right)\frac{dv}{v}$$
$$= \frac{1}{z_1-z_2}\int_0^\infty \left(\frac{z_1 w^b}{e^w - z_1} - \frac{z_2 w^b}{e^w - z_2}\right) dw$$
$$= \frac{\Gamma(b+1)}{z_1-z_2}(\mathrm{Li}_{b+1}(z_1) - \mathrm{Li}_{b+1}(z_2)).$$

□

A.1 Arctan and the Inverse Tangent Integrals

From lemma 3.1.1.3, we have seen $\arctan(t)$ is proportional to the distribution function for the standard Cauchy distribution. 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 A.1.0.1. 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 A.1.0.2. The function arctan is related to the usual tangent function as follows. On $(-\pi/2, \pi/2)$, recall $\tan(\theta)$ is strictly monotone increasing, so its inverse function is well-defined:

$$\frac{d}{d\theta}\tan(\theta) = \frac{d}{d\theta}\frac{\sin(\theta)}{\cos(\theta)}$$
$$= \frac{1}{\cos^2(\theta)}(\cos^2(\theta) - (-\sin^2(\theta))) = \sec^2(\theta) = 1 + \tan^2(\theta) \geq 1.$$

Note $|\tan(\theta)| \to \infty$ as $|\theta| \to \pi/2$. We can thus define $\arctan : \mathbb{R} \to (-\pi/2, \pi/2)$ by

$$\arctan(\tan(\theta)) = \theta.$$

Take the derivative to find

$$1 = \arctan'(\tan(\theta))(1 + \tan^2(\theta))$$
$$\frac{1}{1 + \tan^2(\theta)} = \arctan'(\tan(\theta))$$

so that with $x = \tan(\theta)$,

$$\frac{1}{1 + x^2} = \arctan'(x) \quad \text{or} \quad \int_0^r \frac{1}{1 + x^2} = \arctan(r)$$

by the fundamental theorem of calculus, noting that $\arctan(0) = 0$. The definition A.1.0.1 here, thus agrees with what one would expect for arctan on $[0, \infty)$. It also shows arctan is an analytic function on $(\mathbb{C} - i\mathbb{R}) \cup (-i, i)$ as this domain is simply connected and $(1 + w^2)^{-1}$ is analytic there. The power series for $\arctan(z)$ then follows by considering any path from 0 to z contained in the interior of the unit disk. The integrand $(1 + w^2)^{-1}$ may then be expressed as a geometric series:

$$\arctan(z) = \int_0^z \frac{1}{1 + w^2}\, dw = \int_0^z \sum_{j=0}^{\infty} (-w^2)^j\, dw = \sum_{j=0}^{\infty} (-1)^j \int_0^z w^{2j}\, dw$$
$$= \sum_{j=0}^{\infty} (-1)^j \frac{z^{2j+1}}{2j + 1}.$$

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 A.1.0.3. 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 A.1.0.4. 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 A.1.0.5. In particular, we have

$$\mathrm{Ti}_1(z) = \arctan(z).$$

and for $b > 1$,

$$\mathrm{Ti}_b(z) := \int_0^z \frac{\mathrm{Ti}_{b-1}(w)}{w}\, dw.$$

To see the latter, note that $\mathrm{Ti}_b(0) = 0$ from the power series, while differentiating yields

$$\frac{d}{dz}\mathrm{Ti}_b(z) = \frac{d}{dz}\frac{\mathrm{Li}_b(iz) - \mathrm{Li}_b(-iz)}{2i}$$
$$= \frac{1}{2i}\left(\frac{\mathrm{Li}_{b-1}(iz)}{iz} i - \frac{\mathrm{Li}_{b-1}(-iz)}{-iz}(-i)\right) = \frac{1}{2i}\left(\frac{\mathrm{Li}_{b-1}(iz)}{z} - \frac{\mathrm{Li}_{b-1}(-iz)}{z}\right)$$
$$= \frac{1}{z}\mathrm{Ti}_{b-1}(z).$$

These formulas also follow from lemma A.1.0.8, noting that

$$\mathrm{Ti}_b(z) = \frac{1}{i}\chi_b(iz)$$

which we introduce further below.

To focus on the behavior of arctan on $(-i, i)$ which was not addressed in the inversion formula A.3.0.1, we change points of view through a rotation of the complex plane.

Definition A.1.0.6. Define the function atanh as

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

and as

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

or equivalently as

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

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

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

Of course, we can generalize,

Definition A.1.0.7. Define χ_b, *the Legendre χ function of order $b > 0$*, as

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

and

$$\chi_b(z) = \frac{1}{2}(\mathrm{Li}_b(z) - \mathrm{Li}_b(-z)) \quad \text{for} \quad z \in (\mathbb{C} - \mathbb{R}) \cup (-1, 1)$$

In particular, $\chi_1(z) = \operatorname{atanh}(z)$.

Lemma A.1.0.8. *For $b > 1$ and $z \in (\mathbb{C} - \mathbb{R}) \cup (-1, 1)$,*

$$\chi_b(z) = \int_0^z \frac{\chi_{b-1}(w)}{w}\, dw.$$

Proof. By definition,

$$\int_0^z \frac{\chi_{b-1}(w)}{w}\, dw = \frac{1}{2} \int_0^z \frac{\mathrm{Li}_{b-1}(w) - \mathrm{Li}_{b-1}(-w)}{w}\, dw$$
$$= \frac{1}{2} \int_0^z \frac{\mathrm{Li}_{b-1}(w)}{w}\, dw - \frac{1}{2} \int_0^z \frac{\mathrm{Li}_{b-1}(-w)}{w}\, dw$$

and with $v = -w$,

$$= \frac{1}{2} \mathrm{Li}_b(z) - \frac{1}{2} \int_0^{-z} \frac{\mathrm{Li}_{b-1}(v)}{-v}\, (-dv) = \frac{1}{2} \mathrm{Li}_b(z) - \frac{1}{2} \int_0^{-z} \frac{\mathrm{Li}_{b-1}(v)}{v}\, dv$$
$$= \frac{1}{2} \mathrm{Li}_b(z) - \frac{1}{2} \mathrm{Li}_b(-z) = \chi_b(z).$$

□

Lemma A.1.0.9. *Let $z \in (\mathbb{C} - \mathbb{R}) \cup (-1, 1)$ then*

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

and for $b > 0$,

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

Proof. Just split into even and odd degree terms.

$$\mathrm{Li}_b(z) = \sum_{j=1}^{\infty} \frac{z^j}{j^b} = \sum_{j=0}^{\infty} \frac{z^{2j+1}}{(2j+1)^b} + \sum_{j=1}^{\infty} \frac{z^{2j}}{(2j)^b} = \chi_b(z) + \frac{1}{2^b} \sum_{j=1}^{\infty} \frac{(z^2)^j}{j^b}$$
$$= \chi_b(z) + \frac{1}{2^b} \mathrm{Li}_b(z^2).$$

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

When $b = 1$, we have

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

as desired. □

A useful property of atanh is

Lemma A.1.0.10. *For* $0 \le u < 1$,

$$\operatorname{atanh}(u) \le \frac{u}{1-u^2}$$

Proof. From the power series,

$$\operatorname{atanh}(u) = \sum_{j=0}^{\infty} \frac{u^{2j+1}}{2j+1} = u \sum_{j=0}^{\infty} \frac{(u^2)^j}{2j+1} \le u \sum_{j=0}^{\infty} (u^2)^j = \frac{u}{1-u^2}.$$

□

Here is the addition formula.

Lemma A.1.0.11 (Atanh Addition Formula)**.** *If* $-1 < x, y < 1$,

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

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

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

Proof. Because atanh is odd, the addition formula also covers subtraction by $-\operatorname{atanh}(y) = \operatorname{atanh}(-y)$. Recall

$$\frac{d}{dz}\operatorname{atanh}(z) = \frac{1}{1-z^2},$$

while

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

$$= \left((1+zw)^2 - (z+w)^2\right)^{-1}(1-w^2)$$

$$= (1 + 2zw + (zw)^2 - z^2 - w^2 - 2zw)^{-1}(1-w^2)$$

$$= (1 - z^2 - w^2(1-z^2))^{-1}(1-w^2)$$

$$= \frac{(1-w^2)}{(1-z^2)(1-w^2)} = \frac{1}{1-z^2}.$$

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 $w = z^*$,

$$\frac{z+z^*}{1+zz^*} = \frac{2\Re(z)}{1+|z|^2} = \frac{2\Re(z)/|z|}{\frac{1}{|z|}+|z|}$$

so that

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

by the AM-GM inequality B.0.0.1. The equality case occurs just if $|z| = 1$, but in that case,

$$\left|\frac{\Re(z)}{|z|}\right| < 1$$

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,

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

So f is increasing in each of the individual coordinates. In particular, when $-1 < x < y < 1$,

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

If $0 \leq x < y$, then we have

$$0 \leq f(x, y) < \frac{2y}{1+y^2} < \frac{y}{y} = 1$$

and if $x < y \leq 0$, we have

$$0 \leq |f(x,y)| < \frac{2|x|}{1+x^2} < \frac{|x|}{|x|} = 1$$

and finally if $x < 0 < y$, with $t = \max\{|x|, |y|\}$,

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

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

□

A.2 Dilogarithm Properties

The dilogarithm is the polylogarithm of order 2.

Lemma A.2.0.1 (Reflection Formula)**.** *For* $z \in (\mathbb{C} - \mathbb{R}) \cup (0,1)$,

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

Proof. (Compare to [Lew81, page 5].) Consider

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

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

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

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,

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

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

A.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 A.3.0.1. *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 A.3.0.2. On the imaginary axis, $\arctan(ir) = i\,\text{atanh}(r)$ 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 in the right and left hand planes. From the case $z = 1$,

$$\arctan(1) + \arctan\left(\frac{1}{1}\right) = 2\arctan(1) = 2\frac{\pi}{4} = \frac{\pi}{2}.$$

and similarly from the case $z = -1$,

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

□

In this section similar "inversion formulas" will be derived for some of arctan's relatives. As previously remarked, the interval $(-i, i)$ was omitted in the above. An addition formula happens there instead; see lemma A.1.0.11.

Lemma A.3.0.3. *Let $z \in \mathbb{C} - i\mathbb{R}$. If $\Re(z) > 0$,*

$$\text{Ti}_2(z) = \text{Ti}_2\left(\frac{1}{z}\right) + \frac{\pi}{2}\ln(z)$$

and if $\Re(z) < 0$,

$$\mathrm{Ti}_2(z) = \mathrm{Ti}_2\left(\frac{1}{z}\right) - \frac{\pi}{2}\ln(-z).$$

Proof. By definition,

$$\frac{d}{dz}\,\mathrm{Ti}_2(z) = \frac{\arctan(z)}{z}.$$

On the other hand,

$$\frac{d}{dz}\left(\mathrm{Ti}_2\left(\frac{1}{z}\right) \pm \frac{\pi}{2}\ln(\pm z)\right) = \frac{\arctan(z)}{(1/z)}\frac{(-1)}{z^2} \pm \frac{\pi}{2}\frac{1}{\pm z}(\pm) = \frac{1}{z}\left(\pm\frac{\pi}{2} - \arctan\left(\frac{1}{z}\right)\right).$$

Now use lemma A.3.0.1. If $\Re(z) > 0$,

$$\frac{d}{dz}\left(\mathrm{Ti}_2\left(\frac{1}{z}\right) + \frac{\pi}{2}\ln(z)\right) = \frac{1}{z}\left(\frac{\pi}{2} - \arctan\left(\frac{1}{z}\right)\right) = \frac{1}{z}\arctan(z)$$

and if $\Re(z) < 0$,

$$\frac{d}{dz}\left(\mathrm{Ti}_2\left(\frac{1}{z}\right) - \frac{\pi}{2}\ln(z)\right) = \frac{1}{z}\left(-\frac{\pi}{2} - \arctan\left(\frac{1}{z}\right)\right) = \frac{1}{z}\arctan(z).$$

So in both cases,

$$\mathrm{Ti}_2(z) = c_\pm + \mathrm{Ti}_2\left(\frac{1}{z}\right) \pm \frac{\pi}{2}\ln(\pm z)$$

Taking $z = 1$ in the "+" case and $z = -1$ in the "-" case shows that $c_\pm = 0$. □

Appendix B

Basic Inequalities

We shall be using the arithmetic mean-geometric mean inequality in many places; the following is from [Ste04, chapter 2].

Lemma B.0.0.1 (AM-GM Inequality)**.** *For any nonnegative reals* $(a_1, \ldots, a_D)$ *and* $(p_1, \ldots, p_D)$ *such that* $\sum p_i = 1$*,*

$$\prod_{i=1}^{D} a_i^{p_i} \leq \sum_{i=1}^{D} p_i a_i.$$

Proof. Let

$$\bar{a} = \sum_{i=1}^{D} p_i a_i$$

so that upon replacing a_i by $\hat{a}_i = a_i/\bar{a}$, we just have to show $\prod_{i=1}^{D} \hat{a}_i^{p_i} \leq 1$.

Recall for any $x \in \mathbb{R}$,

$$1 + x \leq \exp(x) \quad \text{or replacing } y = 1 + x, \quad y \leq \exp(y - 1)$$

with $y \geq 1$ if $x \geq 0$. In particular, for any $p_i \geq 0$,

$$y^{p_i} \leq \exp(p_i y - p_i).$$

Applying the above to $y = \hat{a}_i$, we have

$$\prod_{i=1}^{D} \hat{a}_i^{p_i} \leq \prod_{i=1}^{D} \exp(p_i \hat{a}_i - p_i) = \exp\left(\sum_{i=1}^{D} p_i \hat{a}_i - p_i\right) = \exp(1 - 1) = 0 = 1$$

as desired. $\square$

All of our high probability results rely on the following inequality.

Lemma B.0.0.2 (General Markov Inequality)**.** *Let X be a real-valued random variable, and let $\phi : \mathbb{R} \to \mathbb{R}_+$ be a nonnegative increasing function. Then*

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}\phi(X)}{\phi(t)}.$$

Remark B.0.0.3. The theorem is only useful when the expectation exists; we typically use

$$\phi(t) = \exp(st)$$

for s in some domain of $\mathbb{R}$, but even relatively simple functions like

$$\phi(t) = \max\left\{0, t^\beta\right\} \quad \text{with} \quad \beta > 0$$

can be used. In [PN95], the latter functions are shown to give better estimates; however, properties of the exponential greatly outway these benefits for us in most cases.

Proof. The proof here may be found on page 1 of [PN95]. If μ is the probability measure for X,

$$\mathbb{E}\phi(X) = \int_{-\infty}^{\infty} \phi(x)\, d\mu(x) \geq \int_t^{\infty} \phi(x)\, d\mu(x) \geq \phi(t) \int_t^{\infty} d\mu(x) = \phi(t)\mathbb{P}\{X \geq t\}.$$

The penultimate inequality holds because $\phi \geq 0$ is increasing. Rearrange to find the result. □

The following function appears in the Chernoff-Hoeffding bounds further below.

Lemma B.0.0.4. *For $1 \neq x > 0$, the function*

$$H(x) := x\ln(x) + 1 - x > 0.$$

It has a unique minimum of 0 at $x = 1$.

Proof. The critical point occurs when

$$0 = H'(x) = \ln(x) - 1 + \frac{x}{x} = \ln(x) \Rightarrow x = 1.$$

On the other hand,

$$H''(x) = \frac{1}{x} \geq 0 \quad \text{for all} \quad x \geq 0.$$

So $H(x)$ is a convex function on $[0, \infty)$ with a unique minimum of 0 at $x = 1$. Consequently, $H(x)$ is strictly positive for $0 < x \neq 1$. $\square$

Denote by $\mathrm{Bern}\,(p)$ the distribution for a Bernoulli random variable taking values in $\{0, 1\}$, taking the value 1 with probability p. Denote by $\mathrm{Binom}\,(n, p)$ the distribution of the sum of n $\mathrm{Bern}\,(p)$ random variables. We use the following Chernoff-Hoeffding bounds for the binomial distribution when we are considering really small scales; here is a useful formulation [BD15, page 255-56]:

Lemma B.0.0.5. *Let $Z \sim Binom\,(n, p)$, that is,*

$$Z = \sum_{i=1}^{n} X_i \quad \text{with} \quad X_i \sim Bern\,(p)\,.$$

Then

$$\mathbb{P}\left\{Z \geq t\right\} \leq \left(\frac{t}{np}\right)^{-t} e^{t-np} \quad \text{for} \quad t \geq np$$

and

$$\mathbb{P}\left\{Z \leq t\right\} \leq \left(\frac{t}{np}\right)^{-t} e^{t-np} \quad \text{for} \quad 0 \leq t \leq np.$$

In particular,

$$\mathbb{P}\left\{Z \geq cnp\right\} \leq e^{-H(c)np} \quad \text{and} \quad \mathbb{P}\left\{Z \leq \frac{np}{c}\right\} \leq \exp\left(-H\left(\frac{1}{c}\right)np\right).$$

Proof. For a free parameter $s > 0$, compute

$$\mathbb{P}\left\{Z \geq t\right\} = \mathbb{P}\left\{sZ \geq st\right\} = \mathbb{P}\left\{e^{sZ} \geq e^{st}\right\} \leq e^{-st}\mathbb{E}e^{sZ} \leq e^{-st}\mathbb{E}e^{s\sum_i X_i}.$$

By independence,

$$\mathbb{P}\left\{Z \geq t\right\} \leq e^{-st}\prod_{i=1}^{n}\mathbb{E}e^{sX_i} = e^{-st}\left(pe^s + q\right)^n = \exp\left(-st + n\ln\left(pe^s + q\right)\right).$$

To optimize s, minimize

$$f(s) := -st + n\ln\left(pe^s + q\right).$$

Setting the derivative to 0 yields

$$
\begin{aligned}
f'(s) &= -t + \frac{npe^s}{pe^s + q} = 0 \\
\frac{t}{np} &= \frac{1}{p + qe^{-s}} \\
\frac{np}{t} &= p + qe^{-s} \\
\frac{np}{t} - p &= qe^{-s} \\
\frac{p}{q}\left(\frac{n}{t} - 1\right) &= e^{-s}.
\end{aligned}
$$

which is ok, as $t > np$ here:

$$
e^{-s} = \frac{np}{q}\frac{n-t}{nt} = \frac{np}{t}\frac{n-t}{n-np}.
$$

At this value of s,

$$
\begin{aligned}
e^s &= \frac{q}{p}\frac{t}{n-t} \\
pe^s + q &= q\left(1 + \frac{t}{n-t}\right) = q\frac{n}{n-t}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\exp\left(-st + n\ln\left(pe^s + q\right)\right) = \left(\frac{p}{q}\left(\frac{n}{t} - 1\right)\right)^t \left(q\frac{n}{n-t}\right)^n \\
&= \left(\frac{p}{q}\left(\frac{n-t}{t}\right)\right)^t \left(q\frac{n}{n-t}\right)^n = \left(\frac{p}{t}\left(\frac{n-t}{q}\right)\right)^t \left(\frac{n(1-p)}{n-t}\right)^n \\
&= \left(\frac{np}{t}\left(\frac{n-t}{n(1-p)}\right)\right)^t \left(\frac{n-np}{n-t}\right)^n = \left(\frac{np}{t}\right)^t \left(\frac{n-np}{n-t}\right)^{n-t}.
\end{aligned}
$$

As $n \geq t \geq np$, the last factor is at least 1, making

$$
= \left(\frac{np}{t}\right)^t \left(1 + \frac{t-np}{n-t}\right)^{n-t} \leq \left(\frac{np}{t}\right)^t e^{t-np}.
$$

For the lower tail, with $s < 0$,

$$
\mathbb{P}\left\{Z \leq t\right\} = \mathbb{P}\left\{sZ \geq st\right\} = \mathbb{P}\left\{e^{sZ} \geq e^{st}\right\} \leq e^{-st}\mathbb{E}e^{sZ}
$$

just as before. Because the same optimal value for e^{-s} holds, the previous calculation gives

$$
\exp\left(-st + n\ln\left(pe^s + q\right)\right) = \left(\frac{np}{t}\right)^t \left(\frac{n-np}{n-t}\right)^{n-t}
$$

and as $n \geq t$ and $np \geq t$, the last factor is now at most 1:

$$= \left(\frac{np}{t}\right)^t \left(1 - \frac{np - t}{n - t}\right)^{n-t} \leq \left(\frac{np}{t}\right)^t e^{t-np}.$$

With $t = \alpha np$, the bounds simplify:

$$\left(\frac{np}{t}\right)^t e^{t-np} = \exp\left(\alpha np \ln \frac{1}{\alpha} + (\alpha - 1)np\right) = \exp(-H(\alpha)np).$$

With $t = np/\alpha$,

$$\left(\frac{np}{t}\right)^t e^{t-np} = \exp\left(\frac{np}{\alpha} \ln \alpha + \frac{np}{\alpha} - np\right) = \exp\left(-H\left(\frac{1}{\alpha}\right) np\right).$$

□

Bibliography

[BC05] Bo Brinkman and Moses Charikar, *On the Impossibility of Dimension Reduction in L_1*, J. ACM **52** (2005), no. 5, 766–788.

[BD15] Grard Biau and Luc Devroye, *Lectures on the Nearest Neighbor Method*, Springer Series in the Data Sciences, Springer International Publishing, Cham, 2015, DOI: 10.1007/978-3-319-25388-6.

[Cor99] Paul Corazza, *Introduction to Metric-Preserving Functions*, The American Mathematical Monthly **106** (1999), no. 4, 309–323.

[JL84] William B. Johnson and Joram Lindenstrauss, *Extensions of Lipschitz Mappings into a Hilbert Space*, Contemporary Mathematics **26** (1984), 189–206.

[Lew81] Leonard Lewin, *Polylogarithms and Associated Functions*, North-Holland Publishing Co., New York-Amsterdam, 1981. MR 618278

[LHC07] Ping Li, Trevor J. Hastie, and Kenneth W. Church, *Nonlinear Estimators and Tail Bounds for Dimension Reduction in l_1 Using Cauchy Random Projections*, Journal of Machine Learning Research (JMLR) **8** (2007), 2497–2532. MR 2353840

[LN04] James R. Lee and Assaf Naor, *Embedding the Diamond Graph in L_p and Dimension Reduction in L_1*, Geometric & Functional Analysis GAFA **14** (2004), no. 4, 745–747 (en).

[Mat08] Jiří Matoušek, *On Variants of the Johnson–Lindenstrauss Lemma*, Random Structures and Algorithms **33** (2008), no. 2, 142–156 (en).

[Max03] Leonard C. Maximon, *The Dilogarithm Function for Complex Argument*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **459** (2003), no. 2039, 2807–2819 (en).

[MS86] Vitali D. Milman and Gideon Schechtman, *Asymptotic Theory of Finite-Dimensional Normed Spaces*, Lecture Notes in Mathematics, vol. 1200, Springer-Verlag, Berlin, 1986, With an appendix by M. Gromov. MR 856576

[Nee97] Tristan Needham, *Visual Complex Analysis*, The Clarendon Press, Oxford University Press, New York, 1997. MR 1446490

[Pis83] Gilles Pisier, *On the Dimension of the l_p^n-Subspaces of Banach Spaces, for* $1 \leq p < 2$, Transactions of the American Mathematical Society **276** (1983), no. 1, 201–211. MR 684503

[PN95] Thomas K. Philips and Randolph Nelson, *The Moment Bound is Tighter than Chernoff's Bound for Positive Tail Probabilities*, The American Statistician **49** (1995), no. 2, 175–178. MR 1347726

[Roe03] John Roe, *Lectures on Coarse Geometry*, University Lecture Series, vol. 31, American Mathematical Society, Providence, RI, 2003. MR 2007488

[SS03] Elias M. Stein and Rami Shakarchi, *Complex analysis*, Princeton Lectures in Analysis, vol. 2, Princeton University Press, Princeton, NJ, 2003. MR 1976398

[Ste04] J. Michael Steele, *The Cauchy-Schwarz Master Class*, MAA Problem Books Series, Mathematical Association of America, Washington, DC; Cambridge University Press, Cambridge, 2004, DOI: 10.1017/CBO9780511817106. MR 2062704