Report CS2240: ℓ_p Row Sampling by Lewis Weights

Ayoub El Hanchi, Tyler Kastner

December 2022

1 Problem statement

Given a matrix $F \in \mathbb{R}^{N \times d}$ for $N \gg d$ and $\varepsilon > 0$, we would like to produce a matrix $\widehat{F} \in \mathbb{R}^{n \times d}$ with as small of an n as possible such that:

$$(1-\varepsilon)\|Fw\|_p^p \leq \|\widehat{F}w\|_p^p \leq (1+\varepsilon)\|Fw\|_p^p \quad \forall w \in \mathbb{R}^d \tag{1}$$

If this step can be achieved efficiently, then for many problems, one may replace F with the smaller matrix $\widehat{F}$, leading to significant computational savings . We will assume that the columns of F are linearly independent.

2 An alternative formulation

We treat vectors $f \in R^N$ as elements of $\mathcal{L}([N], 2^{[N]}, U_N)$, the set of functions from the probability space $([N], 2^{[N]}, U_N)$ to $\mathbb{R}$ where U_N is the uniform probability measure, i.e. $U_N(\{i\}) = 1/N$ for all $i \in [N]$. We view the columns of F as d linearly independent functions $\{f_j\}_{j=1}^d$, and their span $\mathcal{F} := \operatorname{span}\{f_1, \ldots, f_d\}$ as a subspace of $\mathcal{L}([N], 2^{[N]}, U_N)$. We denote by $w : \mathcal{F} \to \mathbb{R}^d$ the isormorphism (linear bijection) that sends every $f \in \mathcal{F}$ to the weights $w(f) \in \mathbb{R}^d$ satisfying $f = \sum_{j=1}^d w(f)_j f_j$ (this is well-defined since $\{f_j\}_{j=1}^d$ is a basis of $\mathcal{F}$). The p-norm of $f \in \mathcal{F}$ is given by $\|f\|_p = \mathbb{E}\left[|f(I)|^p\right]^{1/p} = \left(\frac{1}{N}\sum_{i=1}^N |f(i)|^p\right)^{1/p}$. Solving problem (1) is then equivalent to producing functions $\{\hat{f}_j\}_{j=1}^d$ in $\mathcal{L}([n], 2^{[n]}, U_n)$ such that:

$$(1-\varepsilon)\left\|\sum_{j=1}^d w(f)_j f_j\right\|_p^p \leq \left\|\sum_{j=1}^d w(f)_j \hat{f}_j\right\|_p^p \leq (1+\varepsilon)\left\|\sum_{j=1}^d w(f)_j f_j\right\|_p^p \quad \forall f \in \mathcal{F} \tag{2}$$

Given such functions, we can construct $\widehat{F}$ by setting the j^{th} column of $\widehat{F}$ to $\frac{N}{n}\hat{f}_j$.

3 Method

Solving Problem 2 amounts to simultaneously estimating $\|f\|_p^p$ for all $f \in \mathcal{F}$ up to a multiplicative factor. For any $f \in \mathcal{F}$, recall that:

$$\|f\|_p^p = \mathbb{E}\left[|f(I)|^p\right]$$

Since we are interested in estimating $\|f\|_p^p$, we could sample n i.i.d. indices $(I_k)_{k=1}^n$ from the uniform probability measure U_N and form the estimator:

$$\|f\|_p^p \approx \frac{1}{n}\sum_{k=1}^n |f(I_k)|^p$$

Clearly, we have:

$$\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n}|f(I_k)|^p\right] = \|f\|_p^p$$

More generally, one may consider an importance sampling estimator: for a distribution Q on $[N]$, sample $(I_k)_{k=1}^n$ i.i.d. from Q and estimate:

$$\|f\|_p^p \approx \frac{1}{n}\sum_{k=1}^{n}\frac{1}{NQ(\{I_k\})}|f(I_k)|^p$$

which also satisfies:

$$\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n}\frac{1}{NQ(\{I_k\})}|f(I_k)|^p\right] = \|f\|_p^p$$

If we define:

$$\hat{f}_j := \left(\frac{1}{NQ(I_1)}f_j(I_1), \dots, \frac{1}{NQ(I_n)}f_j(I_n)\right)$$

then our estimator can be written as:

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{NQ(\{I_k\})}|f(I_k)|^p = \left\|\sum_{j=1}^{d} w(f)_j \hat{f}_j\right\|_p^p$$

This gives us our candidate for the choice of $\{\hat{f}_j\}_{j=1}^d$. Algorithm 1 summarizes this method. For a matrix F, we use $F(i,:)$ to refer to its i^{th} row. The algorithm takes as input the matrix F, the sampling distribution Q, and the number of samples n, and returns an approximation $\widehat{F}$ of F. The goal of the next sections will be to:

- Given a confidence parameter $\delta \in [0,1]$, an error parameter $\varepsilon > 0$, and a distribution Q, upper bound the number of samples n needed such that $\widehat{F}(F,Q,n)$ is an ε-approximation of F in the sense of (1) with probability at least $1-\delta$.
- Propose computable choices of Q such that the upper bound on n is as small as possible.

Algorithm 1 Matrix approximation by row importance sampling

```
function F̂(F, Q, n)
    F̂ ← 0 ∈ ℝ^{n×d}
    for k ∈ [n] do
        I_k ~ Q
        F̂(k, :) ← 1/(nQ(I_k)) · F(I_k, :)
    end for
    return F̂
end function
```

4 Analysis

Throughout this section, we fix $p \in [1, \infty)$. We start by defining the following random semi-norm on $\mathcal{F}$:

$$\|f\|_n := \left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{NQ(\{I_k\})}|f(I_k)|^p\right)^{1/p}$$

Condition (2) can then be rewritten as:

$$(1-\varepsilon)\|f\|_p^p \leq \|f\|_n^p \leq (1+\varepsilon)\|f\|_p^p \quad \forall f \in \mathcal{F}$$

Denote by $S_p := \{f \in \mathcal{F} \mid \|f\|_p = 1\}$ the unit sphere in $(\mathcal{F}, \|\cdot\|_p)$. The above condition is equivalent to:

$$\inf_{f \in S_p} \|f\|_n^p \geq 1 - \varepsilon$$
$$\sup_{f \in S_p} \|f\|_n^p \leq 1 + \varepsilon$$

We want to bound the probability of the event:

$$\mathcal{B} := \left\{(i_1, \ldots, i_n) \mid \inf_{f \in S_p} \|f\|_n^p < 1 - \varepsilon \vee \sup_{f \in S_p} \|f\|_n^p > 1 + \varepsilon\right\}$$

under the product measure Q^n. By the union bound, we have that:

$$Q^n(\mathcal{B}) \leq Q^n\left(\inf_{f \in S_p} \|f\|_n^p < 1 - \varepsilon\right) + Q^n\left(\sup_{f \in S_p} \|f\|_n^p > 1 + \varepsilon\right)$$

At a high-level, the standard technique to control these tail probabilities is the following:

- Bound the tail probabilities $P(\|f\|_n^p < 1 - \varepsilon)$ and $P(\|f\|_n^p > 1 + \varepsilon)$ for any fixed $f \in S_p$.
- Define a set N_Δ, a discretization of the set S_p, and control the discretization error induced by considering only the set N_Δ instead of S_p.
- Extend the pointwise bound to all of S^p by first extending it to N_Δ using the union bound and to all of S_p using the bound on the discretization error.

This is the approach we will take in the next subsections. See Chapter 5 of [8] for a thorough introduction to the problem of obtaining tail bounds for maxima and minima of random variables.

4.1 Pointwise bound

Fix $f \in S_p$. The random variable $\|f\|_n^p$ is an average of the i.i.d. random variables:

$$X_k^f := \frac{1}{NQ(\{I_k\})}|f(I_k)|^p \quad \text{for all } k \in [n]$$

which satisfy:

$$0 \leq X_k^f \leq \max_{i \in [N]}\left\{\frac{1}{NQ(\{i\})}|f(i)|^p\right\} =: R_p(f, Q) \quad \text{for all } k \in [n]$$

and:

$$\mathbb{E}\left[X_k^f\right] = \|f\|_p^p = 1$$

so that by the Chernoff bound, we have:

$$Q^n(\|f\|_n^p \geq 1 + \varepsilon) \leq \exp\left(-\frac{\varepsilon^2 n}{3R_p(f, Q)}\right) \tag{3}$$

$$Q^n(\|f\|_n^p \leq 1 - \varepsilon) \leq \exp\left(-\frac{\varepsilon^2 n}{3R_p(f, Q)}\right) \tag{4}$$

4.2 Discretization

Recall the definition of a Δ-net.

Definition 1. Let (T, d) be a metric space. Let $S \subset T$. A Δ-net of S in d is a subset N_Δ of S such that for all $x \in S$ there exists an $\pi(x) \in N_\Delta$ such that $d(x, \pi(x)) \leq \Delta$. The Δ-covering number of S in d is defined by:

$$N(S, d, \Delta) := \inf\{|N| \mid N \text{ is } \Delta\text{-net of } S \text{ in } d\}$$

The main result of this section is the following result which discretizes the supremum and infimum we care about.

Proposition 2. *Let $\Delta \in (0, 1)$, and let N_Δ be a Δ-net of S_p in the metric induced by $\|\cdot\|_p$. Then:*

$$\sup_{f \in S_p} \|f\|_n \leq \frac{1}{1-\Delta} \sup_{f \in N_\Delta} \|f\|_n$$
$$\inf_{f \in S_p} \|f\|_n \geq \inf_{f \in N_\Delta} \|f\|_n - \Delta \sup_{f \in S_p} \|f\|_n$$

Proof. Recall that for any $f \in S_p$, $\pi(f) \in N_\Delta$ satisfies $\|f - \pi(f)\|_p \leq \Delta$. Now note that:

$$\begin{aligned}
\sup_{f \in S_p} \{|\|f\|_n - \|\pi(f)\|_n|\} &\leq \sup_{f \in S_p} \|f - \pi(f)\|_n \\
&= \sup_{f \in S_p} \|f - \pi(f)\|_p \left\| \frac{f - \pi(f)}{\|f - \pi(f)\|_p} \right\|_n \\
&\leq \Delta \sup_{f \in S^p} \left\| \frac{f - \pi(f)}{\|f - \pi(f)\|_p} \right\|_n \\
&\leq \Delta \sup_{f \in S_p} \|f\|_n
\end{aligned}$$

where in the last line we used the fact that:

$$\left\{ \frac{f - \pi(f)}{\|f - \pi(f)\|_p} \mid f \in S_p \right\} \subseteq S_p$$

The first inequality then follows from:

$$\begin{aligned}
\sup_{f \in S_p} \|f\|_n &= \sup_{f \in S_p} \{\|f\|_n - \|\pi(f)\|_n + \|\pi(f)\|_n\} \\
&\leq \sup_{f \in S_p} \|\pi(f)\|_n + \sup_{f \in S_p} \{\|f\|_n - \|\pi(f)\|_n\} \\
&\leq \sup_{f \in N_\Delta} \|f\|_n + \Delta \sup_{f \in S_p} \|f\|_n
\end{aligned}$$

For the second inequality, we have similarly:

$$\begin{aligned}
\inf_{f \in S_p} \|f\|_n &= \inf_{f \in S_p} \{\|f\|_n - \|\pi(f)\|_n + \|\pi(f)\|_n\} \\
&\geq \inf_{f \in S_p} \|\pi(f)\|_n - \sup_{f \in S_p} \{\|\pi(f)\|_n - \|f\|_n\} \\
&\geq \inf_{f \in N_\Delta} \|f\|_n - \Delta \sup_{f \in S_p} \|f\|_n
\end{aligned}$$

$\square$

4.3 Uniformization

For our problem, we are interested in covering S_p in the metric induced by the norm $\|\cdot\|_p$. We recall the following standard result (see, e.g., [8], Chapter 5, section 5.2 and exercice 5.5)

Lemma 3. *Let $(V, \|\cdot\|)$ be a d-dimensional normed vector space. Let S denote the unit sphere in $(V, \|\cdot\|)$ and let $d_{\|\cdot\|}$ be the metric induced by $\|\cdot\|$. Then for all $\Delta \in (0,1)$:*

$$N(S, d_{\|\cdot\|}, \Delta) \leq \left(\frac{3}{\Delta}\right)^d$$

We are now ready to state our main theorem:

Theorem 4. *Define:*

$$R_p(Q) := \sup_{f \in S_p} R_p(f, Q)$$

then we have the following tail bounds:

$$Q^n\left(\sup_{f\in S_p} \|f\|_n^p > 1+\varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\right)$$
$$Q^n\left(\inf_{f\in S_p} \|f\|_n^p < 1-\varepsilon\right) \leq 2\exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\right)$$

Proof. We start by proving the first inequality. We have:

$$\begin{aligned}
Q^n\left(\sup_{f\in S_p} \|f\|_n^p > 1+\varepsilon\right) &\leq Q^n\left(\sup_{f\in N_\Delta} \|f\|_n^p > (1-\Delta)^p(1+\varepsilon)\right) \\
&= Q^n\left(\bigcup_{f\in N_\Delta} \{\|f\|_n^p > (1-\Delta)^p(1+\varepsilon)\}\right) \\
&\leq \sum_{f\in N_\Delta} Q^n(\|f\|_n^p > (1-\Delta)^p(1+\varepsilon))
\end{aligned}$$

where the first inequality follows from Proposition 2, and the third by the union bound. Now we set:

$$\Delta := \frac{\varepsilon}{2(1+\varepsilon)p}$$

then we have:

$$(1-\Delta)^p(1+\varepsilon) \geq (1-p\Delta)(1+\varepsilon) = 1 + \frac{\varepsilon}{2}$$

where the first inequality follows from Bernoulli's inequality. Now using (3) we have:

$$\begin{aligned}
Q^n\left(\sup_{f\in S_p}\|f\|_n^p > 1+\varepsilon\right) &\le \sum_{f\in N_\Delta} Q^n(\|f\|_n^p > (1-\Delta)^p(1+\varepsilon)) \\
&\le \sum_{f\in N_\Delta} Q^n\left(\|f\|_n^p > 1+\frac{\varepsilon}{2}\right) \\
&\le \sum_{f\in N_\Delta} \exp\left(-\frac{\varepsilon^2 n}{12R_p(f,Q)}\right) \\
&\le \sum_{f\in N_\Delta} \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)}\right) \\
&\le |N_\Delta| \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)}\right)
\end{aligned}$$

since this bound holds for all Δ-nets of S_p in $d_{\|\cdot\|_p}$, we have:

$$\begin{aligned}
Q^n\left(\sup_{f\in S_p}\|f\|_n^p > 1+\varepsilon\right) &\le N(S_p, d_{\|\cdot\|}, \Delta)\exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)}\right) \\
&\le \left(\frac{3}{\Delta}\right)^d \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)}\right) \\
&= \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{3}{\Delta}\right)\right) \\
&= \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\right)
\end{aligned}$$

This proves the first inequality. For the second, we have by Propositon 2:

$$Q^n\left(\inf_{f\in S_p}\|f\|_n^p < 1-\varepsilon\right) \le Q^n\left(\inf_{f\in N_\Delta}\|f\|_n^p - \Delta\sup_{f\in S_p}\|f\|_n^p < 1-\varepsilon\right)$$

Now define the events:

$$\begin{aligned}
E &:= \left\{\inf_{f\in N_\Delta}\|f\|_n^p - \Delta\sup_{f\in S_p}\|f\|_n^p < 1-\varepsilon\right\} \\
E_1 &:= \left\{\inf_{f\in N_\Delta}\|f\|_n^p - \Delta\sup_{f\in S_p}\|f\|_n^p < 1-\varepsilon\right\}\bigcap\left\{\Delta\sup_{f\in S_p}\|f\|_n^p < \varepsilon/2\right\} \\
E_2 &:= \left\{\inf_{f\in N_\Delta}\|f\|_n^p - \Delta\sup_{f\in S_p}\|f\|_n^p < 1-\varepsilon\right\}\bigcap\left\{\Delta\sup_{f\in S_p}\|f\|_n^p \ge \varepsilon/2\right\}
\end{aligned}$$

notice that $E = E_1 \cup E_2$. Now on E_1 we have:

$$\inf_{f\in N_\Delta}\|f\|_n^p < 1-\varepsilon+\Delta\sup_{f\in S_p}\|f\|_n^p \quad \wedge \quad \Delta\sup_{f\in S_p}\|f\|_n^p < \varepsilon/2$$

which implies:

$$\inf_{f\in N_\Delta}\|f\|_n^p < 1-\frac{\varepsilon}{2}$$

therefore:

$$E_1 \subseteq \left\{\inf_{f\in N_\Delta}\|f\|_n^p < 1-\frac{\varepsilon}{2}\right\}$$

and by definition of E_2, we have:

$$E_2 \subseteq \left\{ \Delta \sup_{f \in S_p} \|f\|_n^p \geq \varepsilon/2 \right\}$$

Hence:

$$\begin{aligned}
Q^n\left(\inf_{f \in S_p} \|f\|_n^p < 1 - \varepsilon\right) &\leq Q^n(E) \\
&= Q^n(E_1 \cup E_2) \\
&\leq Q^n(E_1) + Q^n(E_2) \\
&\leq Q^n\left(\inf_{f \in N_\Delta} \|f\|_n^p < 1 - \frac{\varepsilon}{2}\right) + Q^n\left(\Delta \sup_{f \in S_p} \|f\|_n^p \geq \varepsilon/2\right)
\end{aligned}$$

We now bound each of the two terms separately. For the first, we have:

$$\begin{aligned}
Q^n\left(\inf_{f \in N_\Delta} \|f\|_n^p < 1 - \frac{\varepsilon}{2}\right) &= \left(\bigcup_{f \in N_\Delta} \left\{ \|f\|_n^p < 1 - \frac{\varepsilon}{2}) \right\}\right) \\
&\leq \sum_{f \in N_\Delta} Q^n\left(\|f\|_n^p < 1 - \frac{\varepsilon}{2}\right) \\
&\leq \sum_{f \in N_\Delta} \exp\left(-\frac{\varepsilon^2 n}{12 R_p(f, Q)}\right) \\
&\leq \sum_{f \in N_\Delta} \exp\left(-\frac{\varepsilon^2 n}{12 R_p(Q)}\right) \\
&\leq |N_\Delta| \exp\left(-\frac{\varepsilon^2 n}{12 R_p(Q)}\right)
\end{aligned}$$

since this bound holds for all Δ-nets of S_p in $d_{\|\cdot\|_p}$, we have:

$$\begin{aligned}
Q^n\left(\inf_{f \in N_\Delta} \|f\|_n^p < 1 - \frac{\varepsilon}{2}\right) &\leq N(S_p, d_{\|\cdot\|}, \Delta) \exp\left(-\frac{\varepsilon^2 n}{12 R_p(Q)}\right) \\
&\leq \left(\frac{3}{\Delta}\right)^d \exp\left(-\frac{\varepsilon^2 n}{12 R_p(Q)}\right) \\
&= \exp\left(-\frac{\varepsilon^2 n}{12 R_p(Q)} + d \log\left(\frac{3}{\Delta}\right)\right)
\end{aligned}$$

For the second term, we set:

$$\Delta := \frac{\varepsilon}{2(1 + \varepsilon)}$$

Then:

$$\begin{aligned}
Q^n\left(\Delta \sup_{f \in S_p} \|f\|_n^p \geq \varepsilon/2\right) &= Q^n\left(\sup_{f \in S_p} \|f\|_n^p \geq 1 + \varepsilon\right) \\
&\leq \exp\left(-\frac{\varepsilon^2 n}{12 R_p(Q)} + d \log\left(\frac{6p}{\varepsilon}\right)\right)
\end{aligned}$$

where we used the first inequality we derived in the second line. Putting the bounds together we get:

$$
\begin{aligned}
Q^n\left(\inf_{f\in S_p}\|f\|_n^p < 1-\varepsilon\right) &\le Q^n\left(\inf_{f\in N_\Delta}\|f\|_n^p < 1-\frac{\varepsilon}{2}\right) + Q^n\left(\Delta\sup_{f\in S_p}\|f\|_n^p \ge \varepsilon/2\right)\\
&\le \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6}{\varepsilon}\right)\right) + \exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\right)\\
&\le 2\exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\right)
\end{aligned}
$$

where in the last line we used the fact that $p \ge 1$.

□

Corollary 5. *Let $\delta \in (0,1]$, and let $\varepsilon > 0$. If:*

$$n \ge \frac{12R_p(Q)}{\varepsilon^2}\log\left(\frac{3}{\delta}\right) + \frac{12R_p(Q)d}{\varepsilon^2}\log\left(\frac{6p}{\varepsilon}\right)$$

then with probability $1-\delta$ the output $\widehat{F}(F,Q,n)$ of Algorithm 1 satisfies the inequalities (1).

Proof. The inequalities (1) hold precisely on the complement of the event $\mathcal{B}$. Now:

$$Q^n(\mathcal{B}) \le Q^n\left(\inf_{f\in S_p}\|f\|_n^p < 1-\varepsilon\right) + Q^n\left(\sup_{f\in S_p}\|f\|_n^p > 1+\varepsilon\right) \le 3\exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\right)$$

where the last inequality follows from Theorem 4. Now:

$$
\begin{aligned}
&\delta \ge 3\exp\left(-\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\right)\\
\Leftrightarrow\ &\log\left(\frac{\delta}{3}\right) \ge -\frac{\varepsilon^2 n}{12R_p(Q)} + d\log\left(\frac{6p}{\varepsilon}\right)\\
\Leftrightarrow\ &\log\left(\frac{3}{\delta}\right) \le \frac{\varepsilon^2 n}{12R_p(Q)} - d\log\left(\frac{6p}{\varepsilon}\right)\\
\Leftrightarrow\ &\frac{\varepsilon^2 n}{12R_p(Q)} \ge \log\left(\frac{3}{\delta}\right) + d\log\left(\frac{6p}{\varepsilon}\right)\\
\Leftrightarrow\ &n \ge \frac{12R_p(Q)}{\varepsilon^2}\log\left(\frac{3}{\delta}\right) + \frac{12R_p(Q)d}{\varepsilon^2}\log\left(\frac{6p}{\varepsilon}\right)
\end{aligned}
$$

□

4.4 Optimal sampling distribution Q

For any fixed choice of sampling distribution Q, Corollary 5 upper bounds the number of samples n needed to achieve a fixed level of accuracy and confidence. In particular, the upper bound is proportional to $R_p(Q)$. As we would like to minimize the number of samples required, we would therefore like to pick the sampling distribution Q that minimizes $R_p(Q)$. Our next result gives an explicit expression for this optimal distribution.

Lemma 6. *Define the envelope function of S_p by:*

$$F_p(i) := \sup_{f\in S_p} f(i) = \sup_{f\in S_p}|f(i)|$$

where the second equality follows from the fact that if $f \in S_p$, then $-f \in S_p$. Then the probability measure Q_p^* on $([N], 2^{[N]})$ given by:

$$Q_p^*(\{i\}) \propto F_p^p(i)$$

minimizes $R_p(Q)$, i.e.:

$$Q_p^* \in \operatorname*{argmin}_Q R_p(Q)$$

and:

$$R_p(Q_p^*) = \mathbb{E}\left[F_p^p(I)\right] = \|F_p\|_p^p$$

Proof. We have:

$$\begin{aligned} R_p(Q) &= \sup_{f \in S_p} R_p(f, Q) \\ &= \sup_{f \in S_p} \max_{i \in [N]} \frac{1}{NQ(\{i\})} |f(i)|^p \\ &= \max_{i \in [N]} \frac{1}{NQ(\{i\})} \sup_{f \in S_p} |f(i)|^p \end{aligned}$$

and:

$$R_p(Q_p^*) = \frac{1}{N} \sum_{i=1}^{N} \sup_{f \in S_p} |f(i)|^p = \|F_p\|_p^p$$

Now let Q' be any other distribution on $([N], 2^{[N]})$. Then there exists $i \in [N]$ such that $Q'(\{i\}) \leq Q(\{i\})$. Therefore:

$$R_p(Q') \geq \frac{1}{NQ'(\{i\})} \sup_{f \in S_p} |f(i)|^p \geq \frac{1}{NQ(\{i\})} \sup_{f \in S_p} |f(i)|^p = \frac{1}{N} \sum_{i=1}^{N} \sup_{f \in S_p} |f(i)|^p = R_p(Q_p^*)$$

□

Lemma 6 gives us a closed form solution for the optimal distribution that minimizes the upper bound on the number of samples needed of Corollary 5. However, two remaining questions remain:

- How large is $R_p(Q_p^*) = \|F_p\|_p^p$?
- Can we efficiently sample from the optimal distribution Q_p^* or an approximation of it ?

The goal of the next section is to answer these questions.

5 Sampling distribution

We start our investigation of the optimal distribution Q_p^* and the optimal value $\|F_p\|_p^p$ by first considering the special case $p = 2$. We show that in this case we can explicitly compute $\|F_2\|_2^2 = d$. Furthermore, we show that the orthonormalization of the columns of F is enough to exactly compute Q_2^*. It turns out that the main reason why we can obtain such strong results in the case $p = 2$ is due to the fact that the 2-norm is induced by an inner product.

We next turn our attention to the case $p \neq 2$. The main idea here will be to approximate the p-norm with a norm induced by an inner product. A deep result from the functional analysis literature gives such an inner product, which allows us to:

- Provide upper and lower bounds $d^{\min\{p/2,1\}} \leq \|F_p\|_p^p \leq d^{\max\{p/2,1\}}$.
- Propose an efficiently computable sampling distribution $\hat{Q}_p$ such that $R(\hat{Q}_p) \leq d^{\max\{p/2,1\}}$, matching the upper bound on $\|F_p\|_p^p$.

5.1 Case $p = 2$

Recall that we are working on the space $(\mathcal{F}, \|\cdot\|_2)$ where:

$$\|f\|_2^2 = \frac{1}{N}\sum_{i=1}^{N} f^2(i)$$

Of course, the 2-norm is the norm induced by the standard inner product:

$$\langle f, g\rangle := \frac{1}{N}\sum_{i=1}^{N} f(i)g(i)$$

Now let $\{g_j\}_{j=1}^d$ be an orthonormal basis of $(\mathcal{F}, \langle\cdot,\cdot\rangle)$. The main result of this section is an explicit expression of the envelope F_2.

Lemma 7. *For all $i \in [N]$:*

$$F_2(i) = \left(\sum_{j=1}^{d} g(i)^2\right)^{1/2}$$

Proof. For any $f \in \mathcal{F}$, let $u : \mathcal{F} \to \mathbb{R}^d$ be the isomorphism (linear bijection) that sends f to its coefficients with respect to the orthonormal basis $\{g_i\}_{i=1}^d$ i.e. $f = \sum_{j=1}^d u(f)_j g_j$. We have the explicit expression:

$$u(f) = (\langle f, g_1\rangle, \ldots, \langle f, g_d\rangle)$$

It is clear that u is a linear isometry between the spaces $(\mathcal{F}, \|\cdot\|_2)$ and $(\mathbb{R}^d, \|\cdot\|)$ where $\|\cdot\|$ is the standard Euclidean norm since:

$$\|f\|_2^2 = \langle f, f\rangle = \left\langle \sum_{j=1}^{d} u(f)_j g_j, \sum_{j=1}^{d} u(f)_j g_j \right\rangle = \sum_{j=1}^{d} u(f)_j^2 \|g_j\|_2^2 = \|u(f)\|_2^2$$

Therefore $u(S_2) = S^{d-1}$ where S^{d-1} is the standard Euclidean unit sphere in R^d. Now let $i \in [N]$. Define:

$$\phi(i) := (g_1(i), \ldots, g_d(i))$$

Then

$$\begin{aligned}
F_2(i) &= \sup_{f \in S_2} f(i) \\
&= \sup_{f \in S_2} \sum_{j=1}^{d} u(f)_j g_j(i) \\
&= \sup_{f \in S_2} \langle u(f), \phi(i)\rangle \\
&= \sup_{u \in S^{d-1}} \langle u, \phi(i)\rangle \\
&= \|\phi(i)\|
\end{aligned}$$

where in the penultimate line we used that $u(S_2) = S^{d-1}$. □

Using this explicit expression of F_2, we can immediately compute $\|F_2\|_2^2$.

Corollary 8.

$$\|F_2\|_2^2 = d$$

Proof. Using Lemma 7, we have:

$$\|F_2\|_2^2 = \mathbb{E}\left[F_2^2(I)\right] = \mathbb{E}\left[\sum_{j=1}^{d} g^2(I)\right] = \sum_{j=1}^{d} \mathbb{E}\left[g^2(I)\right] = d$$

□

Similarly, by Lemma 6, we have:

$$Q_2^*(\{i\}) \propto F_2^2(i) = \sum_{j=1}^{d} g_j^2(i)$$

Therefore to compute Q^*, it is enough to orthonormalize the columns $\{f_j\}_{j=1}^d$ of F (using Householder transformations for example).

5.2 Case $p \neq 2$

A quick inspection of the proof of Lemma 7 shows that it relies on the inner product structure of $(\mathcal{F}, \|\cdot\|_2)$. Unfortunately, this structure does not persist for $p \neq 2$. Consequently, the results of the previous section do not generalize in a straightforward way. The idea will be to approximate the p-norm with a norm induced by an inner product. An amazing result of Lewis [4] from the functional analysis literature allows us to accomplish just that. For our setup, the theorem translates to:

Theorem 9. *Fix $p \in [1, \infty)$. There exists a basis $\{g_{p,j}\}_{j=1}^d$ of $\mathcal{F}$ with the following properties. Let $g_p := \left(\sum_{j=1}^d g_{p,j}^2\right)^{1/2}$ and define the inner product, for any $f, h \in \mathcal{F}$:*

$$\langle f, h\rangle_p := \mathbb{E}\left[f(I)h(I)g_p^{p-2}(I)\right] = \frac{1}{N}\sum_{i=1}^{N} f(i)h(i)g_p^{p-2}(i)$$

denote by $n_p(\cdot) := \sqrt{\langle\cdot,\cdot\rangle_p}$ the induced norm, and B_{n_p} be the unit ball of $(\mathcal{F}, n_p)$. Then:

- *$\{g_{p,j}\}_{j=1}^d$ is an orthonormal basis of $(\mathcal{F}, \langle\cdot,\cdot\rangle_p)$.*
- *$\|g_p\|_p = d^{1/p}$*
- *If $p \in [1, 2]$, then for all $f \in \mathcal{F}$:*

$$n_p(f) \leq \|f\|_p \leq d^{1/p-1/2} n_p(f)$$

- *If $p \in [2, \infty)$, then for all $f \in \mathcal{F}$:*

$$d^{1/p-1/2} n_p(f) \leq \|f\|_p \leq n_p(f)$$

The proof of this statement is quite involved and difficult to motivate. We refer the reader to Theorem 2.1 in [5] for a relatively simple but hard to motivate proof. See also section III.B of [9].

5.2.1 Upper and lower bounds on $R_p(Q_p^*)$

The estimates in the last two items of Theorem 9 immediately allows to prove upper and lower bounds on $R_p(Q_p^*) = \|F_p\|_p^p$:

Lemma 10.
$$d^{\min\{p/2,1\}} \leq \|F_p\|_p^p \leq d^{\max\{p/2,1\}}$$

Proof. Fix $p \in [1,2]$. Then by the third item of Theorem 9, we have:

$$n_p(f) \leq \|f\|_p \leq d^{1/p-1/2} n_p(f)$$

which implies:

$$d^{1/2-1/p} B_{n_p} \subseteq B_p \subseteq B_{n_p}$$

therefore on the one hand we have:

$$F_p(i) = \sup_{f \in B_p} f(i) \leq \sup_{f \in B_{n_p}} f(i) = \left(\sum_{j=1}^{d} g_{p,j}^2(i) \right)^{1/2} = g_p(i)$$

where the inequality follows from $B_p \subseteq B_{n_p}$, the second equality follows from the same calculation as in the proof of Lemma 7, and the last equality is by definition of g_p. Similarly, we have:

$$F_p(i) = \sup_{f \in B_p} f(i) \geq \sup_{f \in d^{1/2-1/p} B_{n_p}} f(i) = d^{1/2-1/p} \sup_{f \in B_{n_p}} f(i) = d^{1/2-1/p} g_p(i)$$

combining the upper and lower bounds, taking the p-norm of all sides, and using the second item of Theorem 9, we get:

$$d^{p/2} = d^{p/2-1} \|g_p\|_p^p \leq \|F_p\|_p^p \leq \|g_p\|_p^p = d$$

which finishes the proof of the case $p \in [1,2]$. The case $p \in [2,\infty)$ is similar, with the upper and lower bounds reversed. ◻

5.2.2 A candidate distribution $\hat{Q}_p$

The result of the previous subsection gives us an explicit upper and lower bounds on $R(Q_p^*)$ in terms of the dimension. This in turn gives us an explicit bound on the number of samples n such that the output of $\widehat{F}(F, Q_p^*, n)$ from Algorithm 1 is with high probability a good approximation of F in the sense of (1).

However, it remains unclear how and whether we can compute Q_p^* efficiently (it is likely hard, but I don't have a statement in mind). Here we instead design a distribution $\hat{Q}_p$ with $R(\hat{Q}_p) = d^{\max\{p/2,1\}}$, matching the upper bound on $R(Q_p^*)$ derived in the previous subsection. In the next subsection, we will show that it is efficiently computable (for the case $p \in [1,4)$; we provide a reference for $p \geq 4$).

Denote by $F_{n_p} : [N] \to [0,\infty)$ the envelope of $(\mathcal{F}, n_p)$ which is given by:

$$F_{n_p}(i) := \sup_{f \in B_{n_p}} f(i)$$

In the proof of Lemma 10, we showed that:

$$\min\{1, d^{1/2-1/p}\} F_{n_p}(i) \leq F_p(i) \leq \max\{1, d^{1/2-1/p}\} F_{n_p}(i)$$

In other words, F_{n_p} is a good pointwise approximation of F_p. Now recall from Lemma 6 that:

$$Q_p^*(\{i\}) \propto F_p^p(i)$$

Therefore, it makes sense to consider:

$$\hat{Q}_p(\{i\}) \propto F_{n_p}^p(i)$$

as an approximation of Q_p^*.

Lemma 11.

$$R(\hat{Q}_p) \leq d^{\max\{p/2,1\}}$$

Proof. We have:

$$\begin{aligned}
R_p(\hat{Q}^p) &= \sup_{f \in S_p} R_p(f, \hat{Q}_p) \\
&= \sup_{f \in S_p} \max_{i \in [N]} \frac{1}{N\hat{Q}_p(\{i\})} |f(i)|^p \\
&= \max_{i \in [N]} \frac{1}{N\hat{Q}_p(\{i\})} \sup_{f \in S_p} |f(i)|^p \\
&= \max_{i \in [N]} \left(\frac{1}{N} \sum_{j=1}^{N} F_{n_p}^p(j) \right) \frac{F_p^p(i)}{F_{n_p}^p(i)} \\
&= \|F_{n_p}\|_p^p \cdot \max_{i \in [N]} \left\{ \frac{F_p^p(i)}{F_{n_p}^p(i)} \right\} \\
&\leq \|F_{n_p}\|_p^p \cdot \max\left\{1, d^{p/2-1}\right\} \\
&= d^{\max\{p/2,1\}}
\end{aligned}$$

where the last line follows from the fact that $\|F_{n_p}\|_p^p = d$ as can be seen from the proof of Lemma 10. $\square$

Lemma 10 shows that $\hat{Q}_p$ is a "good" sampling distribution in that we have matching upper bounds for $R(Q_p^*)$ and $R(\hat{Q}_p)$. How can we sample from $\hat{Q}_p$? This is the subject of the next subsection.

5.2.3 Computing $\hat{Q}_p$

How do we go about computing $\hat{Q}_p$? Here we provide an answer for $p \in (1,4)$, which was given by [2]. For $p \geq 4$, we refer to the recent paper [3].

The first key observation we make is that, just like in the case $p = 2$, we have the explicit expression:

$$F_{n_p}(i) = g_p(i) = \left(\sum_{j=1}^{n} g_{p,j}^2(i) \right)^{1/2} \tag{5}$$

Furthermore, note that the two expressions on the right are independent of the choice of orthonormal basis of $(\mathcal{F}, \langle \cdot, \cdot \rangle_p)$. A quick way to see this is that F_{n_p} is a basis-independent quantity that depends only on $(\mathcal{F}, \langle \cdot, \cdot \rangle_p)$.

Now, unfortunately, we are not provided with the inner product $\langle \cdot, \cdot \rangle_p$, and therefore we cannot readily form an orthonormal basis of $(\mathcal{F}, \langle \cdot, \cdot \rangle_p)$ as we did in the case of $p = 2$. Nevertheless, let us think about how we could form such an orthonormal basis assuming we had access to it.

Since we are given a basis $\{f_j\}_{j=1}^d$ of $\mathcal{F}$ through the columns of F, we can orthonormalize them by forming the Gram matrix $G \in \mathbb{R}^{d \times d}$:

$$G_{ij} := \langle f_i, f_j \rangle_p = \frac{1}{N} \sum_{k=1}^{N} f_i(k) f_j(k) F_{n_p}^{p-2}(k) \tag{6}$$

It is then straightforward to verify that the functions $\{h_j\}_{j=1}^d$ given by, for all $i \in [d]$:

$$h_i := \sum_{j=1}^{d} G_{ij}^{-1/2} f_j \tag{7}$$

are orthonormal in $(\mathcal{F}, \langle \cdot, \cdot \rangle_p)$.

Let us rewrite equations (5), (6), and (7) in matrix form. For a vector v, we use $\operatorname{diag}(v)$ to denote the diagonal matrix with v on its diagonal. We overload notation, and for a matrix A, we use $\operatorname{diag}(A)$ to denote the vector obtained by taking the diagonal of the matrix A. Finally, for a real number $a \in \mathbb{R}$ and a vector v, we denote by v^a the vector whose entries are the entries of v raised to the power a. For convenience, define the vector $w_p \in \mathbb{R}^N$ by:

$$w_{p,i} := F_{n_p}^p(i) \propto \hat{Q}_p(\{i\})$$

We call $w_{p,i}$ the p-Lewis weight of the i^{th} row of the matrix F, following the terminology of [2]. Our goal is therefore to compute these p-Lewis weights. We can rewrite equation (6) as:

$$G = F^T \operatorname{diag}\left(w_p^{1-2/p}\right) F \tag{8}$$

Letting $H \in \mathbb{R}^{N \times d}$ be the matrix whose j^{th} column is h_j, we can rewrite equation (7) as:

$$H = FG^{-1/2} \tag{9}$$

Finally using equation (5) with the orthonormal basis $\{h_j\}_{j=1}^d$ and squaring both sides yields:

$$w_p^{2/p} = \operatorname{diag}\left(HH^T\right) \tag{10}$$

Plugging equation (8) and (9) into (10) yields the following equation for w_p in terms of F only:

Proposition 12.

$$w_p^{2/p} = \operatorname{diag}\left(F\left\{F^T \operatorname{diag}\left(w_p^{1-2/p}\right) F\right\} F^T\right)$$

This gives us an equation that the p-Lewis weights must satisfy. However, it is unclear at this point:

- whether the p-Lewis weights are the only solution to this equation.
- how to compute solutions of this equation.

We investigate these questions next. First we make the following trivial observation:

Lemma 13. *Let $i \in [N]$. If $(f_1(i), \dots, f_d(i)) = 0$, then $w_{p,i} = 0$ for all $p \in [1, \infty)$.*

Proof. We have:

$$w_{p,i} = F_{n_p(i)} = \sup_{f \in B_{n_p}} f(i)$$

Now let $f \in B_{n_p}$. Since $\{f_j\}_{j=1}^d$ is a basis of $\mathcal{F}$, there exists $w(f) \in \mathbb{R}^d$ such that:

$$f = \sum_{j=1}^{d} w(f)_j f_j$$

and in particular:

$$f(i) = \sum_{j=1}^{d} w(f)_j f_j(i) = \sum_{j=1}^{d} w(f)_j \cdot 0 = 0$$

therefore $f(i) = 0$ for all $f \in B_{n_p}$, from which it follows that $w_{p,i} = 0$. $\square$

In light of Lemma 13, we will assume without loss of generality that the rows of F are non-zero. (If not, we set the p-Lewis weights of the zero rows to zero, and restrict our attention to the remaining rows.)

In preparation of the main result, we introduce a few definitions. Define the set:

$$\mathbb{R}_{++}^N := \left\{x \in \mathbb{R}^N \mid x_i > 0 \quad \forall i \in [N]\right\}$$

and the function $d : \mathbb{R}_{++}^N \times \mathbb{R}_{++}^N \to [0, \infty)$ given by:

$$d(x, y) := \max_{i \in [N]} \left| \log\left(\frac{x_i}{y_i}\right) \right|$$

Then we have:

Lemma 14. *$(\mathbb{R}_{++}^N, d)$ is a complete metric space.*

Proof. Consider the metric space $(\mathbb{R}^N, d_\infty)$ where d_∞ is the metric induced by the norm $\|\cdot\|_\infty$. This is a complete metric space (by applying completeness of $(\mathbb{R}, |\cdot|)$, which is true by definition of $\mathbb{R}$, on each of the coordinates).

Now note that the coordinate-wise exponential function is a bijection from $\mathbb{R}^N$ to $\mathbb{R}_{++}^N$, whose inverse is the coordinate-wise natural logarithm. Then, for all $x, y \in \mathbb{R}_{++}^N$, we have:

$$d(x, y) = d_\infty(\log(x), \log(y))$$

It is then straightforward to verify that d is a metric using the fact that the logarithm is a bijection.

Finally, let $(x_n)_{n=1}^\infty$ be a Cauchy sequence in $(\mathbb{R}_{++}^N, d)$. Let $\varepsilon > 0$. Then there exists $N \in \mathbb{N}$ such that for all $n, m \geq N$, we have $d(x_n, x_m) \leq \varepsilon$, i.e. $d_\infty(\log(x_n), \log(x_m)) \leq \varepsilon$. Therefore the sequence $(\log(x_n))_{n=1}^\infty$ is Cauchy in $(\mathbb{R}^N, d_\infty)$, and by completeness, there exists $y \in \mathbb{R}^N$ such that $\lim_{n\to\infty} \log(x_n) = y$. Defining $x := \exp(y)$, this last statement is equivalent to $\lim_{n\to\infty} |\log(x_n) - \log(x)| = \lim_{n\to\infty} d(x_n, x) = 0$. Therefore $\lim_{n\to\infty} x_n = x$, so the Cauchy sequence converges in $\mathbb{R}_{++}^N$, showing completeness of $(\mathbb{R}_{++}^N, d)$. □

We introduce some more preliminary results before proving our main result. We define the following notation, for $\alpha \geq 1$:

$$x \approx_\alpha y \Leftrightarrow \frac{1}{\alpha} y \leq x \leq \alpha y \text{ for } x, y \in (0, \infty)$$
$$v \approx_\alpha u \Leftrightarrow \forall i \in [N] \quad u_i \approx_\alpha v_i \text{ for } v, u \in \mathbb{R}_{++}^N$$
$$A \approx_\alpha B \Leftrightarrow \frac{1}{\alpha} B \preceq A \preceq \alpha B \text{ for } A, B \in \mathbb{S}_{++}^d$$

where $\mathbb{S}_{++}^d$ is the set of positive definite matrices. We make the following simple observation:

Lemma 15. *Let $x, y \in \mathbb{R}_{++}^d$. Then:*

$$d(x, y) = \min\left\{\log(\alpha) \mid \alpha \in (1, \infty) \wedge x \approx_\alpha y\right\}$$

Proof. We have the equivalences, for any $i \in [N]$ and $\alpha \geq 1$:

$$\log\left(\frac{x_i}{y_i}\right) \leq \log\alpha \Leftrightarrow \frac{x_i}{y_i} \leq \alpha \Leftrightarrow x_i \leq \alpha y_i$$
$$\log\left(\frac{y_i}{x_i}\right) \leq \log\alpha \Leftrightarrow \frac{y_i}{x_i} \leq \alpha \Leftrightarrow \frac{y_i}{\alpha} \leq x_i$$

which shows:

$$d(x, y) = \max_{i \in [N]} \left| \log\left(\frac{x_i}{y_i}\right) \right| \leq \log\alpha \Leftrightarrow x \approx_\alpha y$$

□

Finally, we have the following facts. The proofs are relatively straightforward and we omit them.

Lemma 16. *Let $\alpha \geq 1$, $x, y \in (0, \infty)$ with $x \approx_\alpha y$, $v, u \in \mathbb{R}^N_{++}$ with $v \approx_\alpha u$, $A, B \in \mathbb{S}^d_{++}$ with $A \approx_\alpha B$. Then the following holds:*

1. $x^p \approx_{\alpha^{|p|}} y^p$ *and* $v^p \approx_{\alpha^{|p|}} u^p$ *for all* $p \in \mathbb{R}$.
2. $\operatorname{diag}(v) \approx_\alpha \operatorname{diag}(u)$.
3. $C^T AC \approx_\alpha C^T BC$ *for all* $C \in \mathbb{R}^{N \times d}$
4. $A^{-1} \approx_\alpha B^{-1}$.
5. $x^T Ax \approx_\alpha x^T Bx$ *for all* $x \in \mathbb{R}^d$

We now state and prove the main result of this section.

Theorem 17. *Fix $p \in [1, \infty)$. The map $T_p : \mathbb{R}^N_{++} \to \mathbb{R}^N_{++}$ given by:*

$$T_p(v) := \operatorname{diag}\left(F\left\{F^T \operatorname{diag}\left(v^{1-2/p}\right)F\right\}F^T\right)^{p/2}$$

satisfies, for all $v, u \in \mathbb{R}^N$:

$$d(T_p(v), T_p(u)) \leq |p/2 - 1| \cdot d(v, u)$$

Remark 18. Before proving this statement, we note that for $p \in [1, 4)$, T_p is a contraction. Since $(\mathbb{R}^N_{++}, d)$ is complete, by the Banach fixed point theorem, it has a unique fixed point. Proposition 12 shows that this unique fixed point are the p-Lewis weights w_p. Therefore, to compute them, it is enough to start from some initial guess $w_p^0 \in \mathbb{R}^N_{++}$ and repeatedly apply the contractive map T_p until the desired accuracy is reached.

Proof. Let $u, v \in \mathbb{R}^N_{++}$, and set $\alpha := \exp(d(u, v))$. Recall that $\phi(i) = (f_1(i), \ldots, f_d(i))$ is the i^{th} row of F. Then we have:

$$\begin{aligned}
v \approx_\alpha u &\Rightarrow v^{1-2/p} \approx_{\alpha^{|1-2/p|}} u^{1-2/p} \\
&\Rightarrow \operatorname{diag}(v^{1-2/p}) \approx_{\alpha^{|1-2/p|}} \operatorname{diag}(u^{1-2/p}) \\
&\Rightarrow F^T \operatorname{diag}(v^{1-2/p})F \approx_{\alpha^{|1-2/p|}} F^T \operatorname{diag}(u^{1-2/p})F \\
&\Rightarrow \left\{F^T \operatorname{diag}(v^{1-2/p})F\right\}^{-1} \approx_{\alpha^{|1-2/p|}} \left\{F^T \operatorname{diag}(u^{1-2/p})F\right\}^{-1} \\
&\Rightarrow \phi(i)^T\left\{F^T \operatorname{diag}(v^{1-2/p})F\right\}^{-1}\phi(i) \approx_{\alpha^{|1-2/p|}} \phi(i)^T\left\{F^T \operatorname{diag}(u^{1-2/p})F\right\}^{-1}\phi(i) \quad \forall i \in [N] \\
&\Rightarrow \left(\phi(i)^T\left\{F^T \operatorname{diag}(v^{1-2/p})F\right\}^{-1}\phi(i)\right)^{p/2} \approx_{\alpha^{|p/2-1|}} \left(\phi(i)^T\left\{F^T \operatorname{diag}(u^{1-2/p})F\right\}^{-1}\phi(i)\right) \quad \forall i \in [N] \\
&\Rightarrow T_p(v)_i \approx_{\alpha^{|p/2-1|}} T_p(u)_i \quad \forall i \in [N] \\
&\Rightarrow T_p(v) \approx_{\alpha^{|p/2-1|}} T_p(u) \\
&\Rightarrow d(T_p(v), T_p(u)) \leq \log\left(\alpha^{|p/2-1|}\right) = |p/2 - 1| \cdot d(u, v)
\end{aligned}$$

The top left statement follows from Lemma 15, the first six implications follow from Lemma 16 using items 1, 2, 3, 4, 5, 1 respectively. The seventh implication is by definition of T_p, the eighth by definition of $\approx_\alpha$ for vectors, and the last again by Lemma 15 and the definition of α. $\square$

Notes

The analysis presented in section 4 is simple, accessible, and new to the best of my knowledge, although likely known to experts. Sadly, it is suboptimal: the second term in Corollary 5 has a factor of d coming from the Δ-covering of the sphere. Perhaps surprisingly, this factor can be removed with a more careful analysis. For the case $p = 2$, we can frame the problem as that of matrix estimation (instead of bounding extrema of random variables over the sphere), and use the matrix Chernoff bound. The resulting bound replaces the factor of d with $\log d$, a significant improvement. For the case $p \neq 2$, the situation is more complicated, but significantly better bounds are known [1, 6, 7]. See [2] for a summary of the best known bounds. The derivation of the sampling distribution proportional to the p-Lewis weights as an approximation of the bound-minimizing optimal distribution Q_p^* is to the best of my knowledge new. The remaining results of section 5 are taken from [2].

References

[1] Jean Bourgain, Joram Lindenstrauss, and Vitali Milman. Approximation of zonoids by zonotopes. *Acta mathematica*, 162:73–141, 1989.

[2] Michael B Cohen and Richard Peng. ℓ_p row sampling by lewis weights. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 183–192, 2015.

[3] Maryam Fazel, Yin Tat Lee, Swati Padmanabhan, and Aaron Sidford. Computing lewis weights to high precision. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2723–2742. SIAM, 2022.

[4] D Lewis. Finite dimensional subspaces of L_p. *Studia Mathematica*, 63(2):207–212, 1978.

[5] Gideon Schechtman and Artem Zvavitch. Embedding subspaces of l_p into ℓ_p^n, $0 < p < 1$. *Mathematische Nachrichten*, 227(1):133–142, 2001.

[6] Michel Talagrand. Embedding subspaces of L_1 into ℓ_1^N. *Proceedings of the American Mathematical Society*, 108(2):363–369, 1990.

[7] Michel Talagrand. Embedding Subspaces of L_p in ℓ_p^N. In *Geometric aspects of functional analysis*, pages 311–326. Springer, 1995.

[8] Ramon Van Handel. Probability in high dimension. Technical report, Princeton University, 2014.

[9] Przemyslaw Wojtaszczyk. *Banach spaces for analysts.* Cambridge University Press, 1996.