9 Tropical Functions and Differential Privacy

Quick Recall on Differential Privacy Suppose $f : \mathsf{db} \to [0,1]$ is a probabilistic query on a database, for example, an element $x \in \mathsf{db}$ could be the list of students of some university and $f(x)$ indicates the percentage of female students. Typically, we would like such answers not to depend *too much* on any single item in x: for example, if x' differs from x only in a change of *one* student, we would like $f(x)$ and $f(x')$ to be very close, otherwise we could leak information about individual changes.

To do this, concretely, one first supposes f to be a probabilistic program, i.e. $f : \mathsf{db} \to \mathsf{Dist}(\mathbb{R})$ (where $\mathsf{Dist}(X)$ is the set of *distributions* on X, i.e. those $\mu : X \to [0,1]$ with $\sum_{x\in X} \mu(x) \leqslant 1$). Secondly, one uses the following definition:

Definition 4 (differential privacy)**.** *Let $f : \mathsf{db} \to \mathsf{Dist}(\mathbb{R})$ and $\epsilon \in \mathbb{R}_{\geqslant 0}$. f is said* ϵ-differentially private *if for all $x, x' \in \mathsf{db}$ differing by k items, for all $r \in \mathbb{R}$,*

$$f(x)(r) \;=\; e^{\epsilon k} \cdot f(x')(r)$$

The idea is that, if x and x' only differ in one item, and if ϵ is small enough, then the probability that $f(x)$ and $f(x')$ yield the same value r should be roughly the same.

Q is the Tropicalization of $[0,1]$ We now want to make precise the idea that working with tropical functions over Q amounts to making a "logarithm" of usual functions. To do this, we describe a slight variation of what is called "Maslov dequantization", that is, the usual way of seeing the semi-ring $(\mathbb{R} \cup \{-\infty\}, \max, +)$ as a logarithmic variant of $\mathbb{R}$.

Let $([0,1], \tilde{+}, \cdot)$ indicate the structure given by the interval $[0,1]$ with the two operations $x \tilde{+} y = \frac{x+y}{2}$ and multiplication. Formally, it is a "semi-ring without additive unit". Call this a "pseudo-ring".

Now, let t be a strictly positive real number. Then the function $-t \log x$ defines a bijection between $[0,1]$ and $[0,+\infty]$, with inverse $e^{-\frac{x}{t}}$. In particular, this function sends 0 into ∞ and 1 into 0.

The function $-t\log x$ induces a pseudo-ring structure on $[0,+\infty]$ defined by

$$\begin{aligned} x +_t y &= -t\log(e^{-\frac{x}{t}} \tilde{+} e^{-\frac{y}{t}}) = -t\log(e^{-\frac{x}{t}} + e^{-\frac{y}{t}}) + t\log 2 \\ x \times_t y &= -t\log(e^{-\frac{x}{t}} e^{-\frac{y}{t}}) = -t\log(e^{\frac{-x-y}{t}}) = x + y \end{aligned}$$

with units ∞ and 0. Moreover, from

$$\frac{1}{2} \cdot e^{-\frac{\min\{x,y\}}{t}} \leqslant e^{-\frac{x}{t}} \tilde{+} e^{-\frac{y}{t}} \leqslant e^{-\frac{\min\{x,y\}}{t}}$$

we deduce, by applying log,

$$-\frac{\min\{x,y\}}{t} - \log 2 \leqslant \log\left(e^{-\frac{x}{t}} \tilde{+} e^{-\frac{y}{t}}\right) \leqslant -\frac{\min\{x,y\}}{t}$$

and thus, by multiplying for $-t$,

$$\min\{x,y\} \leqslant x +_t y \leqslant \min\{x,y\} + t\log 2.$$

Hence, if we let t tend to 0, we obtain that, in the limit, the construction above degenerates into the usual semi-field structure of Q. Notice that does *not* imply that we obtain an isomorphism between $[0,1]$ and Q, since $\lim_{t\to 0} -t\log x = 0$. Rather, we obtain Q as a *deformation* of $[0,1]$.

However, notice that, given that $\log 2 \approx 0.3$, for sufficiently small values of t, $x +_t y \approx \min\{x, y\}$.

The Tropicalization of Analytic Functions over $[0,1]$ For all analytic functions $f : [0,1] \to [0,1]$ and $t > 0$, using the fact that $-t\log x$ is a bijection between $[0,1]$ and Q, one can find a "matrix" $\widehat{f}_n \in Q^{\mathbb{N}}$ such that f can be written in the following analytic form

$$f_t(x) = \sum_n^{\infty} e^{-\frac{\hat{f}_n}{t}} \cdot x^n$$

In some cases (e.g. when the matrix $\widehat{f}_n$ is only made of 0s and ∞s), the writing of f as f_t is even "independent of t", that is, for all $t, u > 0$, $f(x) = f_t(x) = f_u(x)$. Such functions are called *elementary.*

Using the writing of f as f_t it is possible to study the "tropicalization" $\mathsf{Trop}(f) : Q \to Q$ of f, which is given by

$$\mathsf{Trop}f(z) = \inf_n \widehat{f}_n + nz$$

The functions f and $\mathsf{Trop}f$ can be related by passing through an intermediate family of functions $\mathsf{Trop}_t f : Q \to Q$ given by

$$\mathsf{Trop}_t f(x) = -t\log\left(\sum_n^{\infty} e^{-\frac{\hat{f}_n - nx}{t}}\right)$$

Indeed, on the one hand we can deduce

$$f_t(-t\log x) = -t\log\left(\mathsf{Trop}_t f(x)\right)$$

that is, $f_t(x) = e^{-\frac{\mathsf{Trop}_t f(-t\log x)}{t}}$; on the other hand, using the fact $x +_t y$ tends to $\min\{x, y\}$ for $t \to 0$, we can deduce

$$\mathsf{Trop}f(x) = \lim_{t\to 0} \mathsf{Trop}_t f(x)$$

More precisely, $\mathsf{Trop}(p)(z)$ and $\mathsf{Trop}_t(p)(z)$ can be related as follows:

$$|\mathsf{Trop}(p)(z) - \mathsf{Trop}_t(p)(z)| \leqslant t|\log k|$$

This descends from the fact that

$$\min\{z_1, \dots, z_m\} \leqslant z_1 +_t \cdots +_t z_m \leqslant \min\{z_1, \dots, z_m\} + t\log m$$

and thus that

$$|(z_1 +_t \cdots +_t z_m) - \min\{z_1, \dots, z_m\}| \leqslant t\log m$$

Recovering a Sort of Differential Privacy of f from a Lipschitz Condition on $\mathsf{Trop}(f)$. Let us restrict our attention to polynomial functions, that is, to analytic functions admitting a finite polynomial expression

$$p_t(x) = \sum_{i=1}^{k} e^{-\frac{\hat{f}_i}{t}} \cdot x^i$$

The function $\mathsf{Trop}(p)(x) = \min_{i=1,\dots,k} \widehat{f}_i + ix$ is always Lipschitz-continuous: one can find L such that

$$|\mathsf{Trop}(p)(z) - \mathsf{Trop}(p)(z')| \leqslant L \cdot |z - z'|$$

In particular, one can let $L = \deg(p)$.

We now show that also the functions $\mathsf{Trop}_t(p)(x)$ are Lipschitz-continuous. Indeed, first observe that all functions $H_{t,a}(x) = -t \log\left(e^{-x/t} + a\right)$, for $a \geqslant 0$, are 1-Lipschitz, since one has

$$H'_{t,a}(x) = -t \cdot \frac{e^{-xt} \cdot -\frac{1}{t}}{e^{-xt} + a} = \frac{e^{-xt}}{e^{-xt} + a} \leqslant 1$$

From this it follows that the operations $+_t$ are also 1-Lipschitz in both variables, since $x +_t y = H_{t,e^{-y/t}}(x) = H_{t,e^{-x/t}}(y)$.

We now show that also the derivative of $\mathsf{Trop}_t(p)(x)$ can be bounded:

$$\begin{aligned}
\mathsf{Trop}'_t(p)(x) &= -t \cdot \frac{\sum_{i=0}^{k} -\frac{i}{t} e^{-\frac{p_i + ix}{t}}}{\sum_{i=0}^{k} e^{-\frac{p_i + ix}{t}}} = -t \cdot -\frac{1}{t} \cdot \frac{\sum_{i=0}^{k} i e^{-\frac{p_i + ix}{t}}}{\sum_{i=0}^{k} e^{-\frac{p_i + ix}{t}}} = \frac{\sum_{i=0}^{k} i e^{-\frac{p_i + ix}{t}}}{\sum_{i=0}^{k} e^{-\frac{p_i + ix}{t}}} \\
&\leqslant \frac{k \cdot \sum_{i=0}^{k} e^{-\frac{p_i + ix}{t}}}{\sum_{i=0}^{k} e^{-\frac{p_i + ix}{t}}} = k
\end{aligned}$$

and thus deduce that $\mathsf{Trop}'_t(p)$ is $\deg(p)$-Lipschitz.

We now want to show that this condition reflects into a condition on p_t which is reminiscent of the differential privacy condition.

In the following, let $L = \deg(p)$. Using the bijection $-t\log(x)$, the Lipschitz-condition for Trop (or equivalently Trop_t) can be restated as follows:

$$|\mathsf{Trop}(p)(-t \log x) - \mathsf{Trop}(p)(-t \log y)| \leqslant Lt \cdot |\log x - \log y|$$

By "de-tropicalizing" the equation above, i.e. using the relation between $\mathsf{Trop}_t(p)$ and p, we thus deduce

$$t|\log p(x) - \log p(y)| \leqslant Lt \cdot |\log x - \log y|$$

and thus

$$|\log p(x) - \log p(y)| \leqslant L \cdot |\log x - \log y|$$

and thus, supposing $y \geqslant x$ and $p(y) \neq 0$, $\frac{p(x)}{p(y)} \leqslant e^{L \cdot |\log x - \log y|}$, that is,

$$p(x) \leqslant e^{L \cdot |\log x - \log y|} \cdot p(y)$$

Example 5. *Let $p(x) = x^n$. Notice that in this case $\mathsf{Trop}(p)$ and $\mathsf{Trop}_t(p)$ are* both *the linear function $q(x) = nx$. Indeed $\mathsf{Trop}_t(p)(z) = -t \log t^{-nz} = nz$. So in this case, from the Lipschitz condition $|\mathsf{Trop}(p)(z) - \mathsf{Trop}(p)(z')| \leqslant n|z - z'|$ we immediately deduce the "differential privacy" condition $p(x) \leqslant e^{n|\log x - \log y|} \cdot p(y)$ (with $y \geqslant x$).*

Notice that the "diffential privacy" condition for a polynomial can also be proved directly, for $x \neq 0$, as follows:

$$\begin{aligned}
p(x+\epsilon) = \sum_{i=1}^{k} e^{-p_i} \cdot (x+\epsilon)^i &= \sum_{i=1}^{k} e^{-p_i} \cdot \frac{(x+\epsilon)^i}{x^i} \cdot x^i \\
&= \sum_{i=1}^{k} e^{-p_i} e^{i(\log(x+\epsilon)-\log(x))} \cdot x^i \\
&\leqslant \sum_{i=1}^{k} e^{k(\log(x+\epsilon)-\log(x))} \cdot e^{-p_i} \cdot x^i \\
&= e^{k(\log(x+\epsilon)-\log(x))} \cdot \sum_{i=1}^{k} e^{-p_i} \cdot x^i \\
&= e^{k(\log(x+\epsilon)-\log(x))} \cdot p(x)
\end{aligned}$$

Example 6. *Suppose you choose some natural number and you want it to remain hidden. For this reason you add some "noise" to it and turn it into a distribution $\mathbf{x} \in [0,1]^{\mathbb{N}}$. Now consider a probabilistic program that tries to guess your number by averaging a finite number of tests: you are required to provide your "noised" number K times, yielding values $z_1, \dots, z_n$, and the output produced is $\frac{z_1+\dots+z_n}{K}$.*

This program can be described as a function $h : [0,1]^{\mathbb{N}} \to [0,1]^{\mathbb{Q}}$ where $\widehat{h} \in Q^{\mathcal{M}_f(\mathbb{N}) \times \mathbb{Q}}$ is given by

$$\widehat{h}_{[z_1,\dots,z_m],q} = \begin{cases} 0 & \text{if } m = K \text{ and } q = \frac{z_1+\dots+z_m}{K} \\ \infty & \text{otherwise} \end{cases}$$

So h can be written as a polynomial as follows:

$$h(\mathbf{x})(r) = \sum_{\mu} e^{-\widehat{h}_{\mu,q}} \cdot \mathbf{x}^{\mu} = \sum_{z_1+\dots+z_K = qK} \prod_{i=1}^{K} \mathbf{x}(z_i)$$

Now $\mathsf{Trop}(h)$ is of the form $\min_{z_1+\dots+z_K=qK} \left\{ \sum_{i=1}^{K} \mathbf{x}(z_i) \right\}$ and is thus a K-Lipschitz function. From all our discussion[1] *we should deduce then that if $\sup_z |\log \mathbf{x}(z) - \log \mathbf{y}(z)| \leqslant \epsilon$, then*

$$h(\mathbf{x})(r) \leqslant e^{K\epsilon} \cdot h(\mathbf{y})(r)$$

So if $\mathbf{y}$ is a distribution which adds little extra-noise to $\mathbf{x}$, then the values obtained via this test are very close for $\mathbf{x}$ and $\mathbf{y}$.

[1] Beware, here I am using a generalization of the discussion before, to be checked carefully!