Natural Analysts in Adaptive Data Analysis: Supplementary Material

Tijana Zrnic [1] **Moritz Hardt** [1]

1. Preliminaries on Differential Privacy

Here we review two useful properties of differential privacy, which are repeatedly utilized throughout this manuscript. The omitted proofs can be found in the book (Dwork & Roth, 2014).

The first lemma states that the output of a differentially private algorithm remains private under any subsequent post-processing.

Lemma 7 (Post-processing of differential privacy)**.** *Let $\mathcal{F}_2 : \mathcal{X} \to \mathcal{Y}$ be an arbitrary randomized function, and let $\mathcal{F}_1 : \mathcal{D}^n \to \mathcal{X}$ be an (α, β)-differentially private mapping. Then, $\mathcal{F}_2 \circ \mathcal{F}_1$ is also (α, β)-differentially private.*

Differential privacy also has favorable composition properties. Linear composition is easy to show by definition of differential privacy, while for the quadratic improvement stated in Lemma 9 one requires more sophisticated proof techniques, originally outlined in the paper (Dwork et al., 2010).

Lemma 8 (Linear composition of differential privacy)**.** *Let $\mathcal{F}_1 : \mathcal{D}^n \to \mathcal{Y}$ be an (α_1, β_1)-differentially private mapping of a data set $\mathcal{S}$, and let $\mathcal{F}_2 : \mathcal{Y} \times \mathcal{D}^n \to \mathcal{Y}$ be (α_2, β_2)-differentially private for every fixed $y_1 \in \mathcal{Y}$. Then, the composition of $\mathcal{F}_1$ and $\mathcal{F}_2$, obtained as $\mathcal{F}_2(\mathcal{F}_1(\mathcal{S}))$, is $(\alpha_1 + \alpha_2, \beta_1 + \beta_2)$-differentially private.*

Lemma 9 (Strong composition of differential privacy)**.** *Let $\mathcal{F}_1 : \mathcal{D}^n \to \mathcal{Y}$ be an (α, β)-differentially private mapping of a data set $\mathcal{S}$, and for every $i \geq 2$, let $\mathcal{F}_i : \mathcal{Y}^{i-1} \times \mathcal{D}^n \to \mathcal{Y}$ be (α, β)-differentially private for every fixed $y_1, \ldots, y_{i-1} \in \mathcal{Y}^{i-1}$. Then, the composition of $\mathcal{F}_1, \ldots, \mathcal{F}_i$ obtained as $\mathcal{F}_i(\mathcal{F}_{i-1}(\ldots, \mathcal{S})\mathcal{S})$ is $(\sqrt{2i\log(1/\beta')}\alpha + i\alpha(e^\alpha - 1), i\beta + \beta')$-differentially private, for any $\beta' \in (0, 1]$.*

Since $e^\alpha - 1 \leq 2\alpha$ for $\alpha \in [0, 1]$, this paper uses a more convenient composition bound, which states that an i-fold composition of (α, β)-differentially private algorithms is $(\sqrt{2i\log(1/\beta')}\alpha + 2i\alpha^2, i\beta + \beta')$-differentially private.

This work mainly focuses on the Gaussian mechanism for achieving differential privacy, hence we review the privacy properties of this method.

Lemma 10 (Properties of the Gaussian mechanism)**.** *Take any $u, v \in \mathbb{R}^{d_q}$, and let $\xi_1, \xi_2 \sim N(0, \sigma^2 I_{d_q})$ be two independent d_q-dimensional Gaussian noise vectors. Denote $u^\xi = u + \xi_1$ and $v^\xi = v + \xi_2$. Then, it holds that:*

$$\mathbb{P}(u^\xi \in \mathcal{O}) \leq \exp\left(\frac{\sqrt{2\log(1.25/\beta)}\|u - v\|_2}{\sigma}\right)\mathbb{P}(v^\xi \in \mathcal{O}) + \beta,$$

for any $\beta > 0$.

If u and v represent d_q-dimensional empirical answers to the same query on data sets S and S', respectively, where S and S' differ in at most one element, we have $\|u - v\|_2 \leq \frac{\sqrt{d_q}}{n}$. As a result, the Gaussian mechanism with parameter σ is $(\sqrt{2\log(1.25/\beta)}\sqrt{d_q}/n\sigma, \beta)$-differentially private, for any $\beta > 0$.

2. Beyond the Discrete Setting

In this section, we drop the assumption that $\mathcal{H}$ is discrete with Δ-resolution. First we prove that, under no additional regularity, progressive and type A conservative analysts can be arbitrarily adaptive, regardless of their parameters of contraction. We do, however, eliminate the trivial cases: $\lambda = 0$ or $L = 0$ in Definition 1, and $\eta_t = 0$ in Definition 2.

Proposition 3. *Without any assumption on $\mathcal{H}$, progressive and type A conservative analysts can overfit as much as an adaptive analyst with a full view of the transcript, as long as the parameters of contraction are non-trivial.*

Proof. First we design a contraction that will be the main technical idea in the proof for both progressive and conservative analysts. Throughout we assume that answers are one-dimensional, however it is not hard to see that the idea easily extends to higher dimensions. In particular, if $d = d_q$, the following argument should be applied coordinate-wise.

Suppose $h, a \in [0, 1]$. If the real-valued answers happen to lie anywhere in $\mathbb{R}$, first contract them to $[0, 1]$ using, for example, appropriately normalized arctangent.

[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, USA. Correspondence to: Tijana Zrnic <tijana@eecs.berkeley.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

Compute h' by interlacing the decimals of a and h. To be more precise, denote the decimals of a as $a = 0.a^1a^2a^3\ldots$. Similarly let $h = 0.h^1h^2h^3\ldots$. Then, h' is given by $h' = 0.a^1h^1a^2h^2\ldots$; notice that this encoding allows perfect recovery of a and h. Denote this construction by $h' = c(a,h)$.

Now we turn to progressive analysts. Fix any $\lambda \in (0,1)$ and $L > 0$. Let $h_t = \psi_t(h_{t-1}, a_{t-1}) := \min\{\lambda, L\}c(h_{t-1}, a_{t-1})$ for all $t \in \mathbb{N}$ This mapping satisfies the conditions of Definition 1 and hence constitutes a λ-progressive analyst. However, h_t and a_t are compressed with no loss, allowing the whole transcript to be unrolled. Consequently, this analyst can be as adaptive as any data analyst with a full view of the transcript.

A similar argument proves the claim for conservative analysts. Suppose that the step sequence $\{\eta_t\}$ is an arbitrary non-increasing positive sequence. Let $h_t = \psi_t(h_{t-1}, a_{t-1}) := \eta_t c(h_{t-1}, a_{t-1})$ for all $t \in \mathbb{N}$. This update satisfies Definition 2 and, as such, represents a η_t-conservative analyst. As in the previous case, however, the hidden state is a lossless encoding of the transcript, and allows full adaptivity.

This completes the proof of the proposition. □

The previous proposition shows that general contractive maps ψ_t do not ensure better generalization in continuous settings. Under mild regularity, however, linear maps imply that adding noise to the truthful answer is essentially the same as adding noise to the history of the analyst. We exploit this observation in the following theorem, proving a result quantitatively almost identical to that of Theorem 3, although for more general sets $\mathcal{H}$.

Theorem 4. *Let $d_q = d$, and suppose $\{B_t\}$ is a sequence of positive-definite or negative-definite matrices, where $\lambda_{\min} := \min_t \min_i |\lambda_i(B_t)|$. Then, without any discretization assumption on $\mathcal{H}$, there is a computationally efficient mechanism to accurately answers t queries chosen adaptively by a type B λ-conservative analyst, given $n = \tilde{O}(\sqrt{K(\lambda)d\log(t)}/\epsilon^2)$ samples, where $K(\lambda) = O\left(\frac{\log(D/\sqrt{\lambda_{\min}})}{\log(1/\lambda)}\right)$.*

Proof. Fix a number of rounds $t \in \mathbb{N}$. Let the statistical mechanism be the Gaussian mechanism with parameter $\sigma = \frac{\epsilon}{\sqrt{2\log(2td/\epsilon\delta)}}$; that is, the answers are constructed as $a_t = q_t(\mathcal{S}) + \xi_t$, where ξ_t is a d-dimensional vector with entries distributed as $N(0,\sigma^2)$. As stated in Theorem 3, this mechanism is $(\epsilon, \epsilon\delta)$-sample accurate.

We can write $\xi_t = \xi_t^{(1)} + \xi_t^{(2)}$, where $\xi_t^{(1)}, \xi_t^{(2)} \sim N(0, \sigma^2/2)$ are independent Gaussians. Further, we can rewrite the history update as:

$$h_t = \psi_t(h_{t-1}, q_{t-1}(\mathcal{S}) + \xi_{t-1}^{(1)}) + B_t\xi_t^{(2)} := h'_t + B_t\xi_{t-1}^{(2)},$$

where we exploit the linearity of the system. Now consider the variable:

$$h'_t = \psi_t(h_{t-1}, q_{t-1}(\mathcal{S}) + \xi_{t-1}^{(1)}).$$

Let the truncated analyst h_t^k corresponding to h'_t be:

$$h_t^k = \psi_t(h_{t-1}^k, a_{t-1}^k), \;\; h_{t-j}^k = 0, \forall j \geq k,$$
$$\text{where } a_j^k = q_j^k(\mathcal{S}) + \xi_j, \forall j < t, \;\; a_t^k = q_t^k(\mathcal{S}) + \xi_t^{(1)},$$
$$q_t^k = f_t(h_t^k).$$

In all rounds the truncated analyst gets noise with variance at least $\sigma^2/2$, so by Lemma 6, as well as the properties of the Gaussian mechanism, h_t^k is

$$\left(O(\sqrt{kd\log(1/\beta')\log(2td/\epsilon\delta)\log(1.25/\beta)}/\epsilon n), k\beta + \beta'\right)$$

-differentially private, for any $\beta, \beta' > 0$. Recall that $h_t = h'_t + B_t\xi_{t-1}^{(2)}$, and take $h_t^{k,\xi} = h_t^k + B_t\xi_{t-1}^k$, where ξ_{t-1}^k is an independent noise sample identically distributed as $\xi_{t-1}^{(2)}$.

Now we need to compute the parameters of differential privacy of h_t. Since stronger differential privacy only implies better generalization, we consider "simplified" versions of h_t and $h_t^{k,\xi}$, which have less additive noise. One simplification would be to add a noise vector with independent entries, which are distributed as $N(0, \lambda_{\min}^2\sigma^2)$. The reason why this argument works is the following. First, zero-out all non-diagonal entries of B_t; by post-processing, this can only induce weaker differential privacy. Now notice that the diagonal entries of positive-definite or negative-definite matrices are in absolute value lower bounded by $\lambda_{\min}$. Therefore, setting the diagonal entries of B_t to $\lambda_{\min}$ would again worsen the privacy parameters. In conclusion, the privacy parameters of h_t and $h_t^{k,\xi}$ have to be at least as small as those of the simplified histories that would be obtained by adding a noise vector with independent entries that are distributed as $N(0, \lambda_{\min}^2\sigma^2)$, instead of $B_t\xi_{t-1}^{(2)}$ and $B_t\xi_{t-1}^k$, respectively.

By an analogous argument as in Lemma 5, $\|h'_t - h_t^k\| \leq \lambda^k D$. To utilize the Gaussian mechanism, we need to bound $\|h'_t - h_t^k\|_2$. Since ℓ_p-norms are decreasing in p, for $p \geq 1$, we can conclude that $\|h'_t - h_t^k\|_2 \leq \lambda^k D\sqrt{d}$, which follows by assuming contraction happens in ℓ_1-norm, and, subsequently, by applying the Cauchy-Schwarz inequality.

Denote by ν_1, ν_2 two independent d-dimensional vectors whose entries are independent and distributed as $N(0, \lambda_{\min}^2\sigma^2)$. Define $h_{t,s} := h'_t + \nu_1$ and $h_{t,s}^{k,\xi} := h_t^k + \nu_2$. Take the depth of truncation to be $K(\lambda) = O\left(\frac{\log(D\sqrt{d}/\sqrt{\lambda_{\min}}\epsilon)}{\log(1/\lambda)}\right)$. Then, the properties of the Gaussian

mechanism imply that $h_{t,s}$ and $h^{k,\xi}_{t,s}$ are indistinguishable; that is, for some constants ϵ' and δ':

$$\mathbb{P}(h_{t,s} \in \mathcal{O}|\mathcal{S} = S) \leq e^{\epsilon'}\mathbb{P}(h^{k,\xi}_{t,s} \in \mathcal{O}|\mathcal{S} = S) + \delta', \quad (1)$$

and similarly:

$$\mathbb{P}(h^{k,\xi}_{t,s} \in \mathcal{O}|\mathcal{S} = S) \leq e^{\epsilon'}\mathbb{P}(h_{t,s} \in \mathcal{O}|\mathcal{S} = S) + \delta'. \quad (2)$$

By post-processing of differential privacy, we also know that $h^{k,\xi}_{t,s}$ has privacy parameters as least as good as $h^k_{t,s}$; to restate for convenience, $h^{k,\xi}_{t,s}$ is α^ξ, β^ξ)-differentially private, where:

$$\alpha^\xi := O(\sqrt{K(\lambda)d\log(1/\beta')\log(2td/\epsilon\delta)\log(1.25/\beta)}/\epsilon n),$$
$$\beta^\xi := K(\lambda)\beta + \beta'.$$

Recall that there exist constant parameters ϵ' and δ' such that equations (1) and (2) hold. Therefore, we have:

$$\begin{aligned}
&\mathbb{P}(h_{t,s} \in \mathcal{O}|\mathcal{S} = S) \\
\leq &\exp(\epsilon')\mathbb{P}(h^{k,\xi}_{t,s} \in \mathcal{O}|\mathcal{S} = S) + \delta' \\
\leq &\exp(\epsilon' + \alpha^\xi)\mathbb{P}(h^{k,\xi}_{t,s} \in \mathcal{O}|\mathcal{S} = S') + \exp(\epsilon')\beta^\xi + \delta' \\
\leq &\exp(2\epsilon' + \alpha^\xi)\mathbb{P}(h_{t,s} \in \mathcal{O}|\mathcal{S} = S') + \exp(\epsilon' + \alpha^\xi)\delta' \\
&+ \exp(\epsilon')\beta^\xi + \delta'.
\end{aligned}$$

To guarantee $(O(\epsilon), O(\epsilon\delta))$-differential privacy of $h_{t,s}$, the main requirement is to keep α^ξ proportional to ϵ, as all other parameters can be chosen as arbitrarily small constants. This is achieved by having $n = \tilde{O}\left(\frac{\sqrt{K(\lambda)d\log(t)}}{\epsilon^2}\right)$ samples. Recall the main transfer theorem of Bassily et al. (2016): having $(O(\epsilon), O(\epsilon\delta))$-differential privacy together with $(O(\epsilon), O(\epsilon\delta))$-sample accuracy implies ϵ-generalization error with probability at least $1-\delta$. Applying this result completes the proof. □

3. Progressive Analysts: Proofs

3.1. Proof of Lemma 1

The claim follows by applying a union bound together with the Hoeffding concentration bound. In particular, for every *fixed* d_q-dimensional statistical query q and target accuracy ϵ, the following is true by the Hoeffding bound:

$$\mathbb{P}(\|q(\mathcal{S}) - \mathbb{E}_{X\sim\mathcal{P}}[q(X)]\|_\infty > \epsilon) \leq 2d_q\exp(-2n\epsilon^2),$$

where we take a union bound over the coordinates of q. Now notice that, by definition of the truncated analyst, we can write h^k_t, and consequently also q^k_t, as a function of $a^k := (a_{t-k}, \ldots, a_{t-1})$. There are A^{kd_q} possibilities for the value of a^k. With this, we can take the union bound over all possibilities for q^k_t to conclude:

$$\begin{aligned}
\mathbb{P}(\|q^k_t(\mathcal{S}) - \mathbb{E}_{X\sim\mathcal{P}}[q^k_t(X)]\|_\infty > \epsilon) &\leq 2d_qA^{kd_q}\exp(-2n\epsilon^2) \\
&= 2d_q\exp(kd_q\log A - 2n\epsilon^2).
\end{aligned}$$

Since A is polynomial in n, we have that the generalization error scales as $\tilde{O}(\sqrt{kd_q/n})$.

3.2. Proof of Lemma 2

The claimed bound is a consequence of the definition of progressiveness. First, because the truncated and full analyst receive the same answers, we have:

$$\begin{aligned}
\|h_t - h^k_t\| &= \|\psi_t(h_{t-1}, a_{t-1}) - \psi_t(h^k_{t-1}, a_{t-1})\| \\
&\leq \lambda\|h_{t-1} - h^k_{t-1}\| \\
&\leq \lambda^k\|h_{t-k}\|,
\end{aligned}$$

where the last step follows because $h^k_{t-k} = 0$. Now we exploit the Lipschitz properties of the maps $\{\psi_t\}$:

$$\begin{aligned}
\|h_t - h^k_t\| &\leq \lambda^k\|\psi_t(h_{t-k-1}, a_{t-k-1}) - \psi_t(0,0)\| \\
&= \lambda^k\|\psi_t(h_{t-k-1}, a_{t-k-1}) - \psi_t(0,0) \\
&+ \psi_t(0, a_{t-k-1}) - \psi_t(0, a_{t-k-1})\| \\
&\leq \lambda^k(\|\psi_t(h_{t-k-1}, a_{t-k-1}) - \psi_t(0, a_{t-k-1})\| \\
&+ \|\psi_t(0, a_{t-k-1}) - \psi_t(0,0)\|) \\
&\leq \lambda^k(\lambda\|h_{t-k-1}\| + L\|a_{t-k-1}\|) \\
&\leq \frac{\lambda^k LC_{\mathbf{1}}}{1-\lambda},
\end{aligned}$$

where the last step follows by recursively applying the same steps to the term $\|h_{t-k-1}\|$, and due to the fact that $\|a_{t-k-1}\| \leq \|(1, \ldots, 1)\|$.

3.3. Proof of Theorem 1

Take any $h, h' \in \mathcal{H}$, such that $h \neq h'$. Then, h and h' have to differ by at least Δ in norm, assuming that they are equal in all coordinates but one. However, by Lemma 2, we have:

$$\|h_t - h^k_t\| \leq \frac{\lambda^k LC_{\mathbf{1}}}{1-\lambda}.$$

This means that, if $\frac{\lambda^k LC_{\mathbf{1}}}{1-\lambda} < \Delta$, h_t and h^k_t are identical. In other words, a truncated analyst with truncation level $\left\lceil \frac{\log\left(\frac{LC_{\mathbf{1}}}{(1-\lambda)\Delta}\right)}{\log(1/\lambda)} \right\rceil \leq \frac{\log\left(\frac{LC_{\mathbf{1}}}{(1-\lambda)\Delta}\right)}{\log(1/\lambda)} + 1 = \frac{\log\left(\frac{LC_{\mathbf{1}}}{(1-\lambda)\lambda\Delta}\right)}{\log(1/\lambda)} := K(\lambda)$ is identical to the corresponding progressive analyst. Since queries are determined solely by the value of the current history, the queries asked by the full analyst and its truncated version at time t have to be identical. Let each answer be constructed as the projection of the empirical answer to the set $\mathcal{A} = \{0, \frac{\epsilon}{2n}, \frac{\epsilon}{n}, \ldots, 1\}^{d_q}$. Then, by a union bound:

$$\begin{aligned}
&\mathbb{P}(\max_{1\leq i\leq t}\|q_i(\mathcal{S}) - \mathbb{E}_{X\sim\mathcal{P}}[q_i(X)]\|_\infty > \epsilon) \\
\leq &\sum_{i=1}^{t}\mathbb{P}(\|q_i(\mathcal{S}) - \mathbb{E}_{X\sim\mathcal{P}}[q_i(X)]\|_\infty > \epsilon) \\
\leq &\ 2td_q\exp(K(\lambda)d_q\log(2n/\epsilon + 1) - 2n\epsilon^2),
\end{aligned}$$

where the last step applies Lemma 1. Since $\|q_i(\mathcal{S}) - a_i\|_\infty \leq O\left(\frac{1}{n}\right)$, we can conclude that the generalization error scales as $\tilde{O}(\sqrt{K(\lambda)d_q \log(t)/n})$.

3.4. Proof of Claim 1

Suppose that the data samples are supported on $\mathbb{R}$, with no atoms. Let d_q and d_a be the dimensions of the queries and query results, respectively. Typically $d_q = d_a$, but this assumption is not necessary for the current counterexample. We will prove the claim for $K = 1$, which will immediately imply the claim for all $K \in \mathbb{N}$. Let g_1 be any bijection between $\mathbb{R}^{d_a}$ and $\mathbb{R}$, and g_2 be any bijection between $\mathbb{R}^2$ and $\mathbb{R}$; the existence of such functions is a standard set-theoretic result. Pick a "reserved value" $r \in [0,1]^{d_q}$. All queries the analyst wishes to ask while interacting with the response mechanism must have the inverse image of r to be a singleton; this does not effectively limit the scope of queries, since r can have infinite precision. After the first round, set $q_2(g_1(a_1)) = r$. In all higher rounds $t \geq 3$, set $q_t(g_2(g_1(a_{t-1}), q_{t-1}^{-1}(r))) = r$. Since g_1 and g_2 are bijections, at any round t one can recover a_{t-1}, as well as the previous encoding of the transcript $q_{t-1}^{-1}(r)$, which allows recursive recovery of all answers $a_1, a_2, \ldots, a_{t-1}$. Since the queries are constructed deterministically based on the current transcript, knowing all answers encodes all query-answer pairs in a lossless fashion. Therefore, despite only having access to q_{t-1} and a_{t-1} at time t, the analyst is familiar with the full transcript. Consequently, the analyst can be arbitrarily adaptive, which completes the proof.

4. Conservative Analysts, Type A: Proofs

4.1. Proof of Lemma 3

Fix h_{t-1}, and take two different answers $a_{t-1}, a'_{t-1} \in \mathcal{A}$. Denote the histories resulting from evolving h_{t-1} using a_{t-1} and a'_{t-1} by h_t and h'_t, respectively. Then, by definition of type A conservative analysts:

$$\begin{aligned}\|h_t - h'_t\| &= \|\psi_t(h_{t-1}, a_{t-1}) - \psi_t(h_{t-1}, a'_{t-1})\| \\ &\leq \eta_t \|a_{t-1} - a'_{t-1}\| \\ &\leq \eta_t C_{\mathbf{1}},\end{aligned}$$

where $C_{\mathbf{1}} := \|(1, \ldots, 1)\|$, which follows from the assumption that the answers are bounded to $[0,1]^{d_q}$. Since the set $\mathcal{H}$ has Δ-resolution, if $h_t \neq h'_t$, it has to hold that $\|h_t - h'_t\| \geq \Delta$. Therefore, if $\eta_t C_{\mathbf{1}} < \Delta$, the history is determined solely depending on h_{t-1} and ψ_t, with no dependence on a_{t-1} and a'_{t-1}. Denote $K(\eta_t) := \min\{t : \eta_t C_{\mathbf{1}} < \Delta\}$. Since η_t is a non-increasing sequence, using recursive reasoning one can conclude that the history at all times after $K(\eta_t)$ does not depend on the value of the current answer. As a result, we can set all answers after time $K(\eta_t)$ to be equal to 0, with no change on the analyst's history sequence. This exactly means that $h_t^{K(\eta_t)} = h_t$, for $K(\eta_t) := \{\min t : \eta_t < \Delta/C_{\mathbf{1}}\}$, which completes the proof.

4.2. Proof of Lemma 4

By the strong composition of differential privacy, h_k^k is $(\sqrt{2k\log(1/\beta')}\alpha + 2k\alpha^2, k\beta + \beta')$-differentially private. For all rounds after the k-th one, the answers are constant and independent of the data set $\mathcal{S}$, meaning they are $(0,0)$-differentially private. Hence, by the linear composition of differential privacy, for all $t \geq k$ the history remains $(\sqrt{2k\log(1/\beta')}\alpha + 2k\alpha^2, k\beta + \beta')$-differentially private. Both composition results are stated in the supplementary material.

4.3. Proof of Proposition 1

The proof follows directly from Lemma 3 and Lemma 4.

4.4. Proof of Theorem 2

Let the statistical mechanism be the truncated Gaussian mechanism with parameter $\sigma = \frac{\epsilon}{\sqrt{2\log(2td_q/\epsilon\delta)}}$; that is, the answers are constructed as $a_t = [q_t(\mathcal{S}) + \xi_t]_{[0,1]^{d_q}}$, where ξ_t is a d_q-dimensional vector with entries distributed as $N(0, \sigma^2)$. Then, by the sub-gaussian tail bound (Boucheron et al., 2013), as well as a union bound:

$$\begin{aligned}\mathbb{P}(\max_{1\leq i\leq t} \|a_i - q_i(\mathcal{S})\|_\infty \geq \epsilon) &\leq \mathbb{P}(\max_{1\leq i\leq t} \|\xi_i\|_\infty \geq \epsilon) \\ &\leq 2td_q \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) \\ &= \epsilon\delta,\end{aligned}$$

where in the last step we plug in the choice of σ. Therefore, such a mechanism is $(\epsilon, \epsilon\delta)$-sample accurate. By properties of the Gaussian mechanism, this mechanism is also

$$\left(\frac{\sqrt{4d_q\log(2td_q/\epsilon\delta)\log(1.25K(\eta_t)/\epsilon\delta)}}{n\epsilon}, \epsilon\delta/K(\eta_t)\right)$$

-differentially private. By Proposition 1, for an *arbitrarily large* t, the history h_t is

$$\left(\sqrt{2K(\eta_t)\log(1/\beta')}\frac{\sqrt{4d_q\log(2td_q/\epsilon\delta)\log(1.25K(\eta_t)/\epsilon\delta)}}{n\epsilon} + 2K(\eta_t)\frac{4d_q\log(2td_q/\epsilon\delta)\log(1.25K(\eta_t)/\epsilon\delta)}{n^2\epsilon^2}, \epsilon\delta + \beta'\right)$$

-differentially private, for any $\beta' > 0$. Given $n \geq \tilde{O}\left(\frac{\sqrt{K(\eta_t)d_q\log(t)}}{\epsilon^2}\right)$, this composition is $(O(\epsilon), O(\epsilon\delta))$-differentially private. By the main transfer theorem of Bassily et al. (2016), having $(O(\epsilon), O(\epsilon\delta))$-differential

privacy, as well as $(O(\epsilon), O(\epsilon\delta))$-sample accuracy, implies ϵ-generalization error with probability at least $1-\delta$. Therefore, the generalization error scales as $\tilde{O}((K(\eta_t)d_q \log(t))^{1/4}/\sqrt{n})$.

5. Conservative Analysts, Type B: Proofs

5.1. Proof of Lemma 5

To prove the claim, we exploit the fact that the truncated analyst and the full analyst receive the same noise variables at any given round. This implies:

$$\begin{aligned}\|h_t - h_t^k\| = &\ \|A_t h_{t-1} + B_t q_{t-1}(\mathcal{S}) + B_t\xi_{t-1} - A_t h_{t-1}^k \\ &- B_t q_{t-1}^k(\mathcal{S}) - B_t\xi_{t-1}\| \\ = &\ \|\psi_t(h_{t-1}, q_{t-1}(\mathcal{S})) - \psi_t(h_{t-1}^k, q_{t-1}^k(\mathcal{S}))\| \\ \leq &\ \lambda\|h_{t-1} - h_{t-1}^k\| \\ \leq &\ \lambda^k\|h_{t-k}\| \\ \leq &\ \lambda^k D,\end{aligned}$$

where the first equality follows by canceling the noise term, the first inequality uses the fact that the analyst is λ-conservative, the second inequality applies the previous argument recursively, and uses the fact that $h_{t-k}^k = 0$, and the last inequality follows by the assumption of $\mathcal{H}$ being bounded.

5.2. Proof of Lemma 6

Since $h_{t-k}^k = 0$, it is independent of the data set $\mathcal{S}$ and, as such, it must be $(0,0)$-differentially private. By linear composition of differential privacy, h_{t-k+1} is then (α,β)-differentially private. Moreover, in the last k rounds, all indivudual answers are (α,β)-differentially private, so by the strong composition of differential privacy, the history is $(\sqrt{2k\log(1/\beta')}\alpha + 2k\alpha^2, k\beta+\beta')$-differentially private. The composition results for differential privacy are stated in the supplementary material.

5.3. Proof of Proposition 2

For any $h, h' \in \mathcal{H}$, such that $h \neq h'$, it has to hold that $\|h - h'\| \geq \Delta$. However, as shown in Lemma 5, for every k, $\|h_t^k - h_t\| \leq \lambda^k D$. Therefore, for truncation level $\left\lceil \frac{\log(D/\Delta)}{\log(1/\lambda)} \right\rceil \leq \frac{\log(D/\Delta\lambda)}{\log(1/\lambda)} := K(\lambda)$, $h_t^{K(\lambda)} = h_t$. Consequently, since $h_t^{K(\lambda)}$ is $(\sqrt{2K(\lambda)\log(1/\beta')}\alpha + 2K(\lambda)\alpha^2, K(\lambda)\beta+\beta')$-differentially private by Lemma 6, then so is h_t.

5.4. Proof of Theorem 3

Let the statistical mechanism be the Gaussian mechanism with parameter $\sigma = \frac{\epsilon}{\sqrt{2\log(2td_q/\epsilon\delta)}}$; that is, the answers are constructed as $a_t = q_t(\mathcal{S}) + \xi_t$, where ξ_t is a d_q-dimensional vector with entries distributed as $N(0,\sigma^2)$. Then, by the sub-gaussian tail bound (Boucheron et al., 2013), as well as a union bound:

$$\begin{aligned}\mathbb{P}(\max_{1\leq i\leq t}\|a_i - q_i(\mathcal{S})\|_\infty \geq \epsilon) &= \mathbb{P}(\max_{1\leq i\leq t}\|\xi_i\|_\infty \geq \epsilon) \\ &\leq 2td_q \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right) \\ &= \epsilon\delta,\end{aligned}$$

where in the last step we plug in the choice of σ. Therefore, such a mechanism is $(\epsilon, \epsilon\delta)$-sample accurate. By properties of the Gaussian mechanism, this mechanism is also

$$\left(\frac{\sqrt{4d_q\log(2td_q/\epsilon\delta)\log(1.25K(\lambda)/\epsilon\delta)}}{n\epsilon}, \epsilon\delta/K(\lambda)\right)$$

-differentially private. By Proposition 1, for an *arbitrarily large* t, the history h_t is

$$\left(\sqrt{2K(\lambda)\log(1/\beta')}\frac{\sqrt{4d_q\log(2td_q/\epsilon\delta)\log(1.25K(\lambda)/\epsilon\delta)}}{n\epsilon} + 2K(\lambda)\frac{4d_q\log(2td_q/\epsilon\delta)\log(1.25K(\lambda)/\epsilon\delta)}{n^2\epsilon^2}, \epsilon\delta+\beta'\right)$$

-differentially private, for any $\beta' > 0$. Given $n \geq \tilde{O}\left(\frac{\sqrt{K(\lambda)d_q\log(t)}}{\epsilon^2}\right)$, this composition is $(O(\epsilon), O(\epsilon\delta))$-differentially private. By the main transfer theorem of Bassily et al. (2016), having $(O(\epsilon), O(\epsilon\delta))$-differential privacy, as well as $(O(\epsilon), O(\epsilon\delta))$-sample accuracy, implies ϵ-generalization error with probability at least $1-\delta$. Therefore, the generalization error scales as $\tilde{O}((K(\lambda)d_q\log(t))^{1/4}/\sqrt{n})$.

References

Bassily, R., Nissim, K., Smith, A., Steinke, T., Stemmer, U., and Ullman, J. Algorithmic stability for adaptive data analysis. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pp. 1046–1059. ACM, 2016.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Dwork, C., Rothblum, G. N., and Vadhan, S. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pp. 51–60. IEEE, 2010.