Proceedings of Machine Learning Research vol xxx:1–27, 2022

Fully Adaptive Composition in Differential Privacy

Justin Whitehouse JWHITEHO@ANDREW.CMU.EDU
Carnegie Mellon University

Aaditya Ramdas ARAMDAS@CMU.EDU
Carnegie Mellon University

Ryan Rogers RROGERS@LINKEDIN.COM
Data Science Applied Research, LinkedIn

Zhiwei Steven Wu ZSTEVENWU@CMU.EDU
Carnegie Mellon University

Abstract

Composition is a key feature of differential privacy. Well-known advanced composition theorems allow one to query a private database quadratically more times than basic privacy composition would permit. However, these results require that the privacy parameters of all algorithms be fixed *before* interacting with the data. To address this, Rogers et al. (2016) introduced *fully adaptive* composition, wherein both algorithms and their privacy parameters can be selected adaptively. The authors introduce two probabilistic objects to measure privacy in adaptive composition: *privacy filters*, which provide differential privacy guarantees for composed interactions, and *privacy odometers*, time-uniform bounds on privacy loss. There are substantial gaps between advanced composition and existing filters and odometers. First, existing filters place stronger assumptions on the algorithms being composed. Second, these odometers and filters suffer from large constants, making them impractical. We construct filters that match the tightness of advanced composition, including constants, despite allowing for adaptively chosen privacy parameters. We also construct several general families of odometers. These odometers can match the tightness of advanced composition at an arbitrary, preselected point in time, or at all points in time simultaneously, up to a doubly-logarithmic factor. We obtain our results by leveraging recent advances in time-uniform martingale concentration. In sum, we show that fully adaptive privacy is obtainable at almost no loss, and conjecture that our results are essentially unimprovable (even in constants) in general.

1. Introduction

Differential privacy (Dwork et al., 2006b) is a rigorous algorithmic criterion that provides meaningful guarantees of individual privacy for conducting analysis on sensitive data. Intuitively, an algorithm is differentially private if similar inputs induce similar distributions on outputs. More formally, a randomized algorithm $A : \mathcal{X} \to \mathcal{Y}$ is differentially private if, for any set of outcomes $G \subset \mathcal{Y}$ and any *neighboring* inputs $x, x' \in \mathcal{X}$, we have

$$\mathbb{P}(A(x) \in G) \leq e^{\epsilon}\mathbb{P}(A(x') \in G) + \delta, \tag{1}$$

where ϵ and δ are the privacy parameters of the algorithm.

A fundamental property of differential privacy is graceful composition. Suppose $A_1, \dots, A_n$ are n algorithms such that each A_m is (ϵ_m, δ_m)-differentially private. Advanced composition (Dwork et al., 2010; Kairouz et al., 2015; Dwork and Roth, 2014) states that, for any $\delta' > 0$, the *composed* sequence of algorithms is (ϵ, δ)-differentially private, where

$$\epsilon = \sqrt{2\log\left(\frac{1}{\delta'}\right)\sum_{m\leq n}\epsilon_m^2} + \sum_{m\leq n}\epsilon_m\left(\frac{e^{\epsilon_m}-1}{e^{\epsilon_m}+1}\right), \qquad \delta = \delta' + \sum_{m\leq n}\delta_m. \tag{2}$$

When all privacy parameters are the same and small, we roughly have $\epsilon = O(\sqrt{n}\epsilon_m)$. This means analysts can make extended use of sensitive datasets with a slow degradation of privacy.

However, there is a major disconnect between most existing results on privacy composition and modern data analysis. As analysts view the outputs of algorithms, the future manner in which they interact with the data changes. Advanced composition allows analysts to adaptively select algorithms, but not privacy parameters. In many cases, analysts may wish to choose the subsequent privacy parameters based on the outcomes of the previous private algorithms. For example, if an analyst learns, from past computations, that they only need to run one more computation, they should be able to use the remainder of their privacy budget in the final round. Likewise, if an analyst is having a hard time deriving conclusions, they should be allowed to adjust privacy parameters to extend the allowable number of computations.

This desideratum has motivated the study of *fully adaptive* composition, wherein one is allowed to adaptively select the privacy parameters of the algorithms. Rogers et al. (2016) define two probabilistic objects which can be used to ensure privacy guarantees in fully adaptive composition. The first, called a *privacy filter*, is an adaptive stopping condition that ensures an entire interaction between an analyst and a dataset retains a pre-specified target privacy level, even when the privacy parameters are chosen adaptively. The second, called a *privacy odometer*, provides a sequence of high-probability upper bounds on how much privacy has been lost up to any point in time. While this work took the first steps towards fully adaptive composition, their filters and odometers suffered from large constants and the latter suffered from sub-optimal asymptotic rates.

We show that, as long as a target privacy level is pre-specified, one can obtain almost the same rate as advanced composition, including constants. We also construct families of privacy odometers that are not only tighter than the originals, but can be optimized for various target levels of privacy. Overall, we show that full adaptivity is not a cost—but rather a feature—of differential privacy.

1.1. Related Work

Privacy composition. There is a long line of work on privacy composition. The "basic composition" theorem states that, when composing private algorithms, the privacy parameters (both ϵ and δ) add up linearly (Dwork et al., 2006b,a; Dwork and Lei, 2009). The "advanced composition" theorem allows the total ϵ to grow sublinearly with a small degradation on δ (Dwork et al., 2010). Later work (Kairouz et al., 2015; Murtagh and Vadhan, 2016) studies "optimal" composition, a computationally intractable formula that tightly characterizes the overall privacy of composed mechanisms.

More recently, several variants of privacy have been studied including (zero)-concentrated differential privacy (zCDP) (Bun and Steinke, 2016; Dwork and Rothblum, 2016), Renyi differential privacy (RDP) (Mironov, 2017), and f-differential privacy (f-DP) (Dong et al., 2021). These all exhibit tighter composition results than differential privacy, but for restricted classes of mechanisms. Unlike our work, these results do not allow adaptive choices of privacy parameters.

Privacy filters and odometers. Rogers et al. (2016) originally introduced privacy filters and odometers, which allow privacy composition with adaptively selected privacy parameters. While their contributions provide a decent approximation of advanced composition, their bounds suffer from large constants, which prevents practical usage. Our work directly improves over these initial results. First, we construct privacy filters essentially matching advanced composition. We also provide flexible families of privacy odometers that outperform those of Rogers et al. (2016). Lastly, our filters and odometers can allow algorithms being composed to satisfy zCDP, yielding greater control over the privacy loss of certain private mechanisms.

Feldman and Zrnic (2021) leverage RDP to construct Rényi filters, which match the rate of advanced composition when the algorithms being composed satisfy *probabilistic* (i.e. point-wise) differential privacy (Kasiviswanathan and Smith, 2014). We obtain the same rate as Rényi filters under the less restrictive assumption that the algorithms being composed satisfy conditional differential privacy (see Definition 1). Since converting from differential privacy to probabilistic differential privacy can be costly (see Lemma 1), our filters demonstrate an improvement. One advantage of Rényi filters is that they can obtain a tighter control over privacy loss when the algorithms being composed satisfy RDP, albeit not always in closed form.

Feldman and Zrnic (2021) and Lécuyer (2021) construct RDP odometers. The former work sequentially composes Rényi filters and the latter work simultaneously runs multiple Rényi filters and takes a union bound. Neither odometer provides high probability, time-uniform bounds on privacy loss, making these results incomparable to our own. We believe our notion of odometers, which aligns with that of Rogers et al. (2016), is more natural.

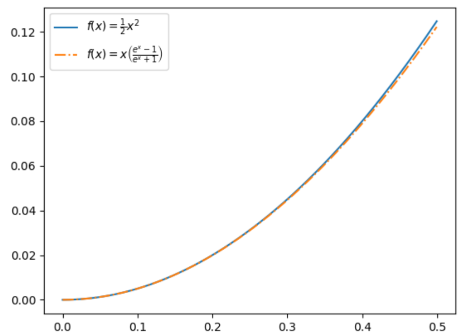
To prove our results, we leverage time-uniform concentration results for martingales (Howard et al., 2020, 2021). The bounds in these papers directly improve over related self-normalized concentration results (de la Pena et al., 2004; Chen et al., 2014). These latter bounds were leveraged in Rogers et al. (2016) to construct filters and odometers.

1.2. Summary of contributions

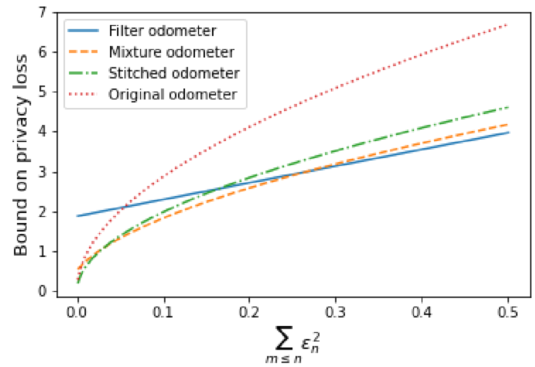
In this work, we provide two primary contributions:

- **Privacy filters:** In Theorem 1 of Section 4, we construct *privacy filters* that match the rate of advanced composition. Our filters improve over those of Rogers et al. (2016) and make fewer assumptions than Feldman and Zrnic (2021). In the supplement, we also address the case where the algorithms being composed satisfy a combination of differential privacy and zCDP (Bun and Steinke, 2016). We state our result in Informal Theorem 1 below.
- **Privacy odometers:** In Theorem 2 of Section 5, we construct improved *privacy odometers* — that is, sequences of upper bounds on privacy loss which are all simultaneously valid with high probability. Our three families of odometers theoretically, empirically, and asymptotically outperform those of Rogers et al. (2016). See Figure 1b for a comparison.

Informal Theorem 1 (Improved Privacy Filter) *Fix target privacy parameters $\epsilon > 0$ and $\delta > 0$, and suppose $(A_n)_{n\geq 1}$ is an adaptively selected sequence of algorithms. Assume that A_n is (ϵ_n, δ_n)-DP or $\frac{1}{2}\epsilon_n^2$-zCDP conditioned on the outputs of the first $n-1$ algorithms, where ϵ_n and δ_n may depend on outputs of $A_1, \ldots, A_{n-1}$. Roughly, if a data analyst stops interacting with the data before $\sqrt{2\log\left(\frac{1}{\delta}\right)\sum_{m\leq n+1}\epsilon_m^2} + \frac{1}{2}\sum_{m\leq n+1}\epsilon_m^2 > \epsilon$, then the entire interaction is (ϵ, δ)-DP.*



(a) Comparing lower order terms



(b) Comparing privacy odometers

Figure 1: Figure 1a compares the lower order terms of advanced composition and our privacy filter. Figure 1b compares the original odometer of Rogers et al. (2016) with our odometers (filter, mixture, and stitched).

Informal Theorem 1 almost recovers advanced composition when all parameters ϵ_n and δ_n are fixed prior to interacting with the dataset. The only difference is a slight gap in the lower order term, as $\epsilon\left(\frac{e^\epsilon-1}{e^\epsilon+1}\right) \le \frac{1}{2}\epsilon^2$. Figure 1a demonstrates that this gap is negligible for small values of ϵ, which is the natural setting for differential privacy.

Our second major contribution is the construction of several families of privacy odometers. These odometers give a running bound on privacy loss in settings where a target level of overall privacy is not known. Our constructed odometers are significantly tighter than the originals (Rogers et al., 2016), as can be seen in Figure 1b.

Our key insight is to view adaptive privacy composition as depending not on the number of algorithms being composed, but rather on the sums of squares of privacy parameters, $\sum_{m\le n}\epsilon_m^2$. This shift to looking at "intrinsic time" allows us to apply recent advances in time-uniform concentration (Howard et al., 2020, 2021) to privacy loss martingales. Overall, our results demonstrate that one has to pay essentially no privacy cost to perform fully adaptive private data analysis.

2. Background on Differential Privacy

Throughout, we assume all algorithms map from a space of datasets $\mathcal{X}$ to outputs in a measurable space, typically either denoted $(\mathcal{Y}, \mathcal{G})$ or $(\mathcal{Z}, \mathcal{H})$. For a sequence of algorithms $(A_n)_{n\ge1}$, we often consider the composed algorithm $A_{1:n} := (A_1, \dots, A_n)$. For more background on measure-theoretic matters, as well as on the notion of neighboring datasets, see Appendix A.

We start by formalizing a generalization of differential privacy in which the privacy parameters of an algorithm A_n can be functions of the outputs of $A_1, \dots, A_{n-1}$. In particular, we replace the probabilities in Equation (1) with conditional probabilities given relevant random variables.

Definition 1 (Conditional Differential Privacy) *Suppose A and B are algorithms mapping from a space $\mathcal{X}$ to measurable spaces $(\mathcal{Y}, \mathcal{G})$ and $(\mathcal{Z}, \mathcal{H})$ respectively. Suppose $\epsilon, \delta : \mathcal{Z} \to \mathbb{R}_{\ge 0}$ are measurable functions. We say the algorithm A is (ϵ, δ)-differentially private conditioned on B if, for any neighbors $x, x' \in \mathcal{X}$ and for all measurable sets $G \in \mathcal{G}$, we have*

$$\mathbb{P}\left(A(x) \in G \mid B(x)\right) \le e^{\epsilon(B(x))}\mathbb{P}\left(A(x') \in G \mid B(x)\right) + \delta(B(x)).$$

For conciseness, we will write either ϵ or $\epsilon(x)$ for $\epsilon(B(x))$ and likewise δ or $\delta(x)$ for $\delta(B(x))$.

In the nth round of adaptive composition, we will set $A := A_n$ and $B := A_{1:n-1}$. In this setting, the analyst has functions $\epsilon_n, \delta_n : \mathcal{Y}^{n-1} \to \mathbb{R}_{\geq 0}$ and takes the nth round privacy parameters to be $\epsilon_n(A_{1:n-1}(x))$ and $\delta_n(A_{1:n-1}(x))$. In other words, the analyst uses the outcome of the first $n-1$ algorithms to decide the level of privacy for the nth algorithm, ensuring that A_n is (ϵ_n, δ_n)-differentially private conditioned on $A_{1:n-1}$.

A key notion in differential privacy is *privacy loss*, which measures how much information is revealed about the underlying input dataset. For neighbors $x, x' \in \mathcal{X}$, let p^x and $p^{x'}$ be the densities of $A(x)$ and $A(x')$ respectively. The privacy loss between $A(x)$ and $A(x')$ is defined as

$$\mathcal{L}(x, x') := \log\left(\frac{p^x(A(x))}{p^{x'}(A(x))}\right). \tag{3}$$

By Equation (3), a negative privacy loss suggests that the input is more likely to be x', and likewise a positive privacy loss suggests that the input is more likely to be x. The next definition generalizes privacy loss to a setting in which an analyst may have additional information.

Definition 2 (Conditional Privacy Loss) *Suppose A and B are as in Definition 1. Suppose $x, x' \in \mathcal{X}$ are neighbors. Let $p^x(\cdot|\cdot), p^{x'}(\cdot|\cdot) : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}_{\geq 0}$ be conditional densities for $A(x)$ and $A(x')$ respectively given $B(x)$.*[1] *The privacy loss between $A(x)$ and $A(x')$ conditioned on B is given by*

$$\mathcal{L}_B(x, x') := \log\left(\frac{p^x(A(x)|B(x))}{p^{x'}(A(x)|B(x))}\right).$$

Suppose A_n is the nth algorithm being run and we have already observed $A_{1:n-1}(x)$ for some unknown input $x \in \mathcal{X}$. If we are trying to guess whether x or a neighbor x' produced the data, we would consider the privacy loss between $A_n(x)$ and $A_n(x')$ conditioned on $A_{1:n-1}(x)$.

It is straightforward to characterize the privacy loss of a composed algorithm $A_{1:n}$ in terms of the privacy loss of each constituent algorithm $A_1, \cdots, A_n$. Namely, from Bayes rule,

$$\mathcal{L}_{1:n}(x, x') = \sum_{m \leq n} \mathcal{L}_m(x, x'), \tag{4}$$

where $\mathcal{L}_m(x, x')$ is shorthand for the conditional privacy loss between $A_m(x)$ and $A_m(x')$ given $A_{1:m-1}(x)$, per Definition 2. Equation (4) also holds at arbitrary random times $N(x)$ that only depend on the dataset $x \in \mathcal{X}$ through observed algorithm outputs.

The simple decomposition of privacy loss noted above motivates the study of an "alternative", probabilistic definition of differential privacy. Intuitively, an algorithm should be differentially private if, with high probability, the privacy loss is small. More formally, an algorithm $A : \mathcal{X} \to \mathcal{Y}$ is said to be (ϵ, δ)-*probabilistically differentially private*, or (ϵ, δ)-pDP for short, if, for all neighboring inputs $x, x' \in \mathcal{X}$, we have $\mathbb{P}\left(|\mathcal{L}(x, x')| > \epsilon\right) \leq \delta$.

Unfortunately, as noted by Kasiviswanathan and Smith (2014) (in which pDP is called *pointwise indistinguishability*), pDP is a strictly stronger notion than differential privacy. In particular, if an algorithm is (ϵ, δ)-pDP, it is also (ϵ, δ)-DP. The converse in general requires a costly conversion.

1. To ensure the existence of conditional densities, it suffices to assume that $\mathcal{Y}$ and $\mathcal{Z}$ are *Polish spaces* under some metrics $d_{\mathcal{Y}}$ and $d_{\mathcal{Z}}$, and that $\mathcal{G}$ and $\mathcal{H}$ are the corresponding Borel σ-algebras associated with $d_{\mathcal{Y}}$ and $d_{\mathcal{Z}}$ (Durrett, 2019). These measurability assumptions are not restrictive, as Euclidean spaces, countable spaces, and Cartesian products of the two satisfy these assumption.

Lemma 1 (Conversions between DP and pDP (Kasiviswanathan and Smith, 2014)) *If A is (ϵ, δ)-probabilistically differentially private, then A is also (ϵ, δ)-differentially private. Conversely, if A is (ϵ, δ)-differentially private, then A is $(2\epsilon, \frac{2\delta}{\epsilon e^{\epsilon}})$-probabilistically differentially private.*

In this work, we prove most results on the level of probabilistic differential privacy, and hence only ever use the cost-free forward conversion between the notions of privacy. When we do handle the case where the algorithms being composed are conditionally differentially private, we leverage an extension of the fact that differentially private algorithms can be viewed as post-processings of the randomized response mechanism (Kairouz et al., 2015) to avoid a conversion cost.

Probabilistic differential privacy (pDP) admits a natural conditional counterpart.

Definition 3 (Conditional Probabilistic Differential Privacy) *Suppose $A : \mathcal{X} \to \mathcal{Y}$ and $B : \mathcal{X} \to \mathcal{Z}$ are algorithms, and $\epsilon, \delta : \mathcal{Z} \to \mathbb{R}_{\geq 0}$ are measurable. Then, A is said to be (ϵ, δ)-probabilistically differentially private conditioned on B if, for any neighbors $x, x' \in \mathcal{X}$, we have*

$$\mathbb{P}\left(|\mathcal{L}_B(x, x')| > \epsilon(B(x))|B(x)\right) \leq \delta(B(x)).$$

In sequential composition, we would assume the nth algorithm A_n is (ϵ_n, δ_n)-pDP conditioned on $A_{1:n-1}$. The privacy parameters would be given as functions of $A_{1:n-1}(x)$.

3. Filters, Odometers, and Privacy Loss Martingales

Privacy filters and odometers. To reiterate, our objectives are (i) to match the tightness of advanced composition while allowing privacy parameters to be chosen adaptively and (ii) to provide strong, time-uniform bounds on privacy loss for the case when target privacy levels may not be known.

To address these goals, we adopt the notions of privacy filters and odometers as introduced in Rogers et al. (2016).[2] A *privacy filter* is a stopping rule N that guarantees, if we stop by time N, the composed interaction will be differentially private with preset target parameters. Likewise, a *privacy odometer* is a sequence of time-uniform upper bounds on the accumulated privacy loss. To formalize these objects, we need filtrations and stopping times, as defined in Appendix A.

Definition 4 requires the algorithms being composed to satisfy conditional differential privacy. Previous filters (Rogers et al., 2016; Feldman and Zrnic, 2021) assume either pDP or RDP, and thus incur conversion costs to handle DP. Definition 5, which provides time-uniform privacy control, requires the assumption of probabilistic differential privacy (pDP).

Definition 4 (Privacy Filter (Rogers et al., 2016)) *Let $(A_n)_{n\geq 1}$ be an adaptive sequence of algorithms such that, for all $n \geq 1$, A_n is (ϵ_n, δ_n)-DP conditioned on $A_{1:n-1}$. Let $\epsilon > 0$ and $\delta > 0$ be target privacy parameters. Then, a function $N : \mathbb{R}^{\infty}_{\geq 0} \times \mathbb{R}^{\infty}_{\geq 0} \to \mathbb{N}$ is an (ϵ, δ)-privacy filter if*

1. *for all $x \in \mathcal{X}$, $N(x) := N\left((\epsilon_n(x))_{n\geq 1}, (\delta_n(x))_{n\geq 1}\right)$ is a stopping time with respect to the natural filtration generated by $(A_n(x))_{n\geq 1}$, and*
2. *the algorithm $A_{1:N(\cdot)}(\cdot)$ is (ϵ, δ)-differentially private.*

2. In Rogers et al. (2016), the authors define privacy filters and odometers in terms of an adversary selecting two sequences $(x_n^{(0)})_{n\geq 1}$ and $(x_n^{(1)})_{n\geq 1}$ of datasets such that $x_n^{(0)} \sim x_n^{(1)}$ for all n. This approach is no more general, as the tail bounds leveraged on privacy loss in each round remain the same.

Definition 5 (Privacy Odometer (Rogers et al., 2016)) *Let $(A_n)_{n\geq 1}$ be an adaptive sequence of algorithms such that, for all $n \geq 1$, A_n is (ϵ_n, δ_n)-pDP conditioned on $A_{1:n-1}$. Let $(u_n)_{n\geq 1}$ be a sequence of functions where $u_n : \mathbb{R}_{\geq 0}^{n-1} \times \mathbb{R}_{\geq 0}^{n-1} \to \mathbb{R}_{\geq 0}$. Let $\delta \in (0, 1)$ be a target confidence parameter. For $x \in \mathcal{X}, n \geq 1$, define $U_n(x) := u_n(\epsilon_{1:n-1}(x), \delta_{1:n-1}(x))$. Then, $(u_n)_{n\geq 1}$ is called a δ-privacy odometer if, for all $x, x' \in \mathcal{X}$ neighbors, we have*

$$\mathbb{P}\left(\exists n \geq 1 : \mathcal{L}_{1:n}(x, x') > U_n(x)\right) \leq \delta.$$

Privacy Loss martingales. A process $(X_n)_{n\in\mathbb{N}}$ is said to be a martingale with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if, for all $n \in \mathbb{N}$, (a) X_n is $\mathcal{F}_n$-measurable, (b) $\mathbb{E}|X_n| < \infty$, and (c) $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) = X_{n-1}$. Correspondingly, $(X_n)_{n\in\mathbb{N}}$ is a supermartingale if $\mathbb{E}(X_n \mid \mathcal{F}_{n-1}) \leq X_{n-1}$.

Martingales offer strong, time-uniform concentration of measure properties. We present a full exposition on time-uniform concentration in Appendix B (Ville, 1939; Howard et al., 2020, 2021). For a martingale $(M_n)_{n\in\mathbb{N}}$ and confidence level $\delta > 0$, time-uniform concentration inequalities provide sequences of bounds $(U_n)_{n\in\mathbb{N}}$ satisfying $\mathbb{P}(\exists n \in \mathbb{N} : M_n > U_n) \leq \delta$. If N is a stopping time, we additionally obtain the guarantee that $\mathbb{P}(M_N > U_N) \leq \delta$. This indicates that, if we can create a martingale from privacy loss, we can leverage time-uniform concentration to construct both filters and odometers. As such, let $(\mathcal{F}_n(x))_{n\in\mathbb{N}}$ be the natural filtration generated by $(A_n(x))_{n\geq 1}$. If, for all n, A_n is $(\epsilon_n, 0)$-pDP (we consider $\delta_n > 0$ later), we see that $(M_n)_{n\in\mathbb{N}}$ given by

$$M_n := M_n(x, x') := \mathcal{L}_{1:n}(x, x') - \sum_{m\leq n} \mathbb{E}\left(\mathcal{L}_m(x, x') | \mathcal{F}_{n-1}(x)\right), \qquad M_0 := 0 \tag{5}$$

is a martingale, aptly titled a *privacy loss martingale* (Dwork and Roth, 2014; Bun and Steinke, 2016; Dwork et al., 2010).

To construct their filters and odometers, Rogers et al. (2016) use self-normalized concentration inequalities (de la Pena et al., 2004; Chen et al., 2014). We instead use advances in time-uniform martingale concentration (Howard et al., 2020, 2021), which yields tighter results.

4. A Filter for Fully Adaptive Composition

We now provide a privacy filter that matches the rate of advanced composition. Our filter improves on the rate of the original filter presented in Rogers et al. (2016). In addition, we improve over the filters of Feldman and Zrnic (2021) in that we assume the algorithms being composed satisfy *conditional differential privacy*, whereas they assume *conditional probabilistic differential privacy* when approximating advanced composition, otherwise needing to pay the potentially hefty conversion price outlined in Lemma 1.[3]

Theorem 1 *Suppose $(A_n)_{n\geq 1}$ is a sequence of algorithms such that, for any $n \geq 1$, A_n is (ϵ_n, δ_n)-differentially private conditioned on $A_{1:n-1}$. Let $\epsilon > 0$ and $\delta = \delta' + \delta''$ be target privacy parameters such that $\delta' > 0, \delta'' \geq 0$. Consider the function $N : \mathbb{R}_{\geq 0}^{\infty} \times \mathbb{R}_{\geq 0}^{\infty} \to \mathbb{N}$ given by*

$$N((\epsilon_n)_{n\geq 1}, (\delta_n)_{n\geq 1}) := \inf\left\{n : \epsilon < \sqrt{2\log\left(\frac{1}{\delta'}\right)\sum_{m\leq n+1}\epsilon_m^2} + \frac{1}{2}\sum_{m\leq n+1}\epsilon_m^2 \quad \textit{or} \quad \delta'' < \sum_{m\leq n+1}\delta_m\right\}.$$

3. In Section 4.3 of their work, Feldman and Zrnic (2021) apply their Rényi filters to algorithms which satisfy conditional pDP. In general, a conversion from (ϵ, δ)-DP to (ϵ, δ)-pDP may be required to apply their filter.

Then, the algorithm $A_{1:N(\cdot)}(\cdot) : \mathcal{X} \to \mathcal{Y}^\infty$ is (ϵ, δ)-DP, where $N(x) := N((\epsilon_n(x))_{n\geq 1}, (\delta_n(x))_{n\geq 1})$. In other words, N is an (ϵ, δ)-privacy filter.

In Appendix E, we extend Theorem 1 to the setting where some of the algorithms satisfy *conditional* zCDP. Observe first that if all privacy parameters are fixed in advance, i.e. ϵ_n and δ_n are constants, then, for any $\delta' > 0$, taking $\epsilon := \sqrt{2\log\left(\frac{1}{\delta'}\right)\sum_{m\leq n}\epsilon_m^2} + \frac{1}{2}\sum_{m\leq n}\epsilon_m^2$ and $\delta'' := \sum_{m\leq n}\delta_m$ recovers advanced composition up to low order terms (See Figure 1a). Second, determining whether or not to stop at time n only depends on the privacy parameters known at time n. Lastly, being fully aware of the stopping rule, a user can easily choose ϵ_N and δ_N to exactly meet the condition defining N with equality, ensuring that the privacy budget of (ϵ, δ) is fully utilized.

As a first step to proving Theorem 1, it is easier to consider the case where each algorithm A_n satisfies conditional (ϵ_n, δ_n)-pDP, as this condition provides a high-probability bound on the privacy loss. This allows us to use the martingale machinery in Section 3 and Appendix B to prove tight composition results. We prove the following lemma in Appendix C, but sketch the proof below.

Lemma 2 *Theorem 1 holds under the stronger assumption that, for any $n \geq 1$, A_n is (ϵ_n, δ_n)-pDP conditioned on $A_{1:n-1}$.*

Proof [*sketch*] Assume $\delta_n = 0$ for all n. For neighbors $x, x' \in \mathcal{X}$, let $(M_n)_{n\in\mathbb{N}}$ be the privacy loss martingale, as constructed in Equation (5). For $n \geq 1$, letting $\Delta M_n := M_n - M_{n-1}$, it is clear ΔM_n is ϵ_n^2-subGaussian given $A_{1:n-1}(x)$. As such, for any $a, b > 0$, Theorem 3 yields $\mathbb{P}\left(\exists n \in \mathbb{N} : M_n > \frac{b}{2} + \frac{b}{2a}\sum_{m\leq n}\epsilon_m^2\right) \leq \exp\left(-\frac{b^2}{2a}\right)$. Carefully optimizing a and b, bounding privacy loss, and decomposing M_n yields that, with probability $\geq 1 - \delta'$, for all $n \leq N$,

$$\mathcal{L}_{1:n}(x, x') \leq \sqrt{2\sum_{m\leq n}\epsilon_m^2\log\left(\frac{1}{\delta'}\right)} + \frac{1}{2}\sum_{m\leq n}\epsilon_m^2 \leq \epsilon.$$

Lastly, a careful union bound argument generalizes the above to the case of general δ_n. ■

Our key insight above is to view filters as functions of the "intrinsic time" determined by privacy parameters, $\sum_{m\leq n}\epsilon_m^2$. Lemma 2 can also be obtained leveraging the analysis for Rényi filters (Feldman and Zrnic, 2021). However, our approach to proving Theorem 1 has the advantage that it does not require reductions between different modes of privacy. While Lemma 6, which bounds expected privacy loss, does require some complicated analysis, we only ever need to apply Lemma 2 to instances of randomized response, in which case computing the privacy loss bound is trivial.

We now use Lemma 2 to prove Theorem 1. Recall that Lemma 1 shows that algorithms that satisfy pDP also satisfy DP, but the converse is not true and may require a conversion cost. To avoid this cost, we define following generalization of randomized response.

Definition 6 (Conditional Randomized Response) *Let $\mathcal{R} := \{0, 1, \top, \bot\}$ and $2^{\mathcal{R}}$ be the corresponding power set of $\mathcal{R}$. Then, R taking inputs in $\{0, 1\}$ to outputs in the measurable space $(\mathcal{R}, 2^{\mathcal{R}})$ is an instance of (ϵ, δ)-randomized response if, for $b \in \{0, 1\}$, $R(b)$ outputs the following:*

$$R(b) = \begin{cases} b & \textit{with probability } (1-\delta)\frac{e^\epsilon}{1+e^\epsilon} \\ 1-b & \textit{with probability } (1-\delta)\frac{1}{1+e^\epsilon} \\ \top & \textit{with probability } \delta \textit{ if } b = 1 \\ \bot & \textit{with probability } \delta \textit{ if } b = 0. \end{cases}$$

More generally, suppose $B : \{0,1\} \to \mathcal{Z}$ is a randomized algorithm. For functions $\epsilon, \delta : \mathcal{Z} \to \mathbb{R}_{\geq 0}$, we say R is an instance of (ϵ, δ)-randomized response conditioned on B if, for any true input $b' \in \{0,1\}$ and hypothesized alternative $b \in \{0,1\}$, the conditional probability $\mathbb{P}(R(b) \in \cdot | B(b') = z)$ is the same as the law of $(\epsilon(z), \delta(z))$-randomized response with input bit b.

Conditional (ϵ, δ)-randomized response satisfies both conditional (ϵ, δ)-DP and conditional (ϵ, δ)-pDP. We will leverage the fact that it satisfies both privacy definitions with the same parameters. A surprising result in the nonadaptive setting is that *any* (ϵ, δ)-DP algorithm can be viewed as a randomized post-processing of (ϵ, δ)-randomized response (Kairouz et al., 2015). We generalize this result to the adaptive conditional setting below, providing a proof in Appendix D. In the language of Blackwell's comparison of experiments (Blackwell, 1953), instances of randomized response are "sufficient" for instances of arbitrary DP algorithms, and we prove that the same is true for conditional randomized response and conditionally DP algorithms. In what follows, by a transition kernel ν, we mean that for any $b \in \mathcal{Z}$ and $r \in \mathcal{R}$, $\nu(\cdot, r \mid b)$ is a probability measure on $(\mathcal{Y}, \mathcal{G})$.

Lemma 3 (Reduction to Conditional Randomized Response) *Let A and B map from $\mathcal{X}$ to measurable spaces $(\mathcal{Y}, \mathcal{G})$ and $(\mathcal{Z}, \mathcal{H})$, respectively. Suppose A is (ϵ, δ)-differentially private conditioned on B. Fix neighbors $x_0, x_1 \in \mathcal{X}$, and let R be an instance of (ϵ, δ)-randomized response conditioned on B', where $B' : \{0,1\} \to \mathcal{Z}$ is the restricted algorithm satisfying $B'(b) = B(x_b)$. Then, there is a transition kernel $\nu : \mathcal{G} \times \mathcal{R} \times \mathcal{Z} \to [0,1]$ such that, for all $b, b' \in \{0,1\}$, $\mathbb{P}\left(A(x_b) \in \cdot \mid B'(b')\right) = \nu_{b,b'}$, where $\nu_{b,b'} = \mathbb{E}\left(\nu(\cdot, R(b) \mid B'(b')) \mid B'(b')\right)$.*[4]

Lemma 3 tells us that the conditional distribution obtained by averaging the kernel $\nu(\cdot, R(b) \mid B'(b'))$ over the randomness in $R(b)$ matches the conditional distribution of $A(x_b)$. We now sketch the proof of Theorem 1 using the above lemma; the full proof is in Appendix D.

Proof [*Proof sketch of Theorem 1*] Fix neighbors $x_0, x_1 \in \mathcal{X}$, and, for any $n \geq 1$ consider the restricted algorithm $A'_n : \{0,1\} \to \mathcal{Y}$ given by $A'_n(b) := A_n(x_b)$. Let $(R_n)_{n \geq 1}$ be a sequence of algorithms such that R_n is an instance of (ϵ_n, δ_n)-randomized response given $A'_{1:n-1}$. For each n, R_n is (ϵ_n, δ_n)-pDP conditioned on $A'_{1:n-1}$, and hence by Lemma 2 $R_{1:N}$ is (ϵ, δ)-DP. Lastly, Lemma 3 allows us to view $A'_{1:N}$ as a randomized post-processing of $R_{1:N}$. Since differential privacy is closed under randomized post-processing, this proves the desired result. ■

5. Time Uniform Privacy Loss Bounds via Odometers

Previously, we constructed privacy filters that matched the rate of advanced composition (Equation (2)) while allowing *both* algorithms and privacy parameters to be chosen adaptively. While privacy filters require the total level of privacy (determined by ϵ, δ) to be fixed in advance, it is desirable to track the privacy loss at all steps without a pre-fixed privacy budget (Ligett et al., 2017).

Privacy odometers, as outlined in Definition 5, are sequences of upper bounds on accumulated privacy loss that are valid at all points in time simultaneously with high probability. Consequently,

4. By $\nu_{b,b'}(\cdot) := \mathbb{E}\left(\nu(\cdot, R(b) \mid B'(b')) \mid B'(b')\right)$, we mean that $\nu_{b,b'}$ is the (random) averaged probability measure:

$$\begin{aligned}\nu_{b,b'}(\cdot) &= \mathbb{P}(R(b) = 1 \mid B'(b'))\nu(\cdot, 1 \mid B'(b')) + \mathbb{P}(R(b) = 0 \mid B'(b'))\nu(\cdot, 0 \mid B'(b')) \\ &\quad + \mathbb{P}(R(b) = \bot \mid B'(b'))\nu(\cdot, \bot \mid B'(b')) + \mathbb{P}(R(b) = \top \mid B'(b'))\nu(\cdot, \top \mid B'(b')).\end{aligned}$$

odometers provide bounds on privacy loss under arbitrary stopping conditions (e.g. conditions based on model accuracy). Again, our insight is to view odometers not as functions of the number of algorithms being composed, but rather as functions of the intrinsic time $\sum_{m\leq n}\epsilon_m^2$. This reframing allows us to leverage the various time-uniform concentration inequalities discussed in Appendix B. We provide our novel privacy odometers in the following theorem. The proof of the following theorem is similar in spirit the proof of Lemma 2, and can also be found in Appendix C.

While in Theorem 1 we assumed that the algorithms being composed were *conditionally differentially private*, here, we need to assume *conditional probabilistic privacy*. This is because our goal is not differential privacy, but rather tight control over privacy loss. We conjecture that a version of Theorem 2 that replaces pDP by DP and leaves all else identical does not hold.

Theorem 2 *Suppose $(A_n)_{n\geq 1}$ is a sequence of algorithms such that, for any $n \geq 1$, A_n is (ϵ_n, δ_n)-pDP conditioned on $A_{1:n-1}$. Let $\delta = \delta' + \delta''$ be a target approximation parameter such that $\delta' > 0, \delta'' \geq 0$. Define $N := N((\delta_n)_{n\geq 1}) := \inf\left\{n \in \mathbb{N} : \delta'' < \sum_{m\leq n+1}\delta_m\right\}$ and $V(\epsilon_{1:n}) := \sum_{m\leq n}\epsilon_m^2$. Consider the following sequences of functions.*

1. ***Filter odometer.*** *For any $\epsilon > 0$, let $y^* := \left(-\sqrt{2\log\left(\frac{1}{\delta'}\right)} + \sqrt{2\log\left(\frac{1}{\delta'}\right)+\epsilon}\right)^2$. Define the functions $(u_n^F)_{n\geq 1}$ by*

$$u_n^F(\epsilon_{1:n}, \delta_{1:n}) := \begin{cases} \infty & n > N \\ \frac{\sqrt{2y^*\log\left(\frac{1}{\delta'}\right)}}{2} + \frac{\sqrt{2\log\left(\frac{1}{\delta'}\right)}}{2\sqrt{y^*}}V(\epsilon_{1:n}) + \frac{1}{2}V(\epsilon_{1:n}) & \textit{otherwise}. \end{cases}$$

2. ***Mixture odometer.*** *For any $\rho > 0$, define the sequence of functions $(u_n^M)_{n\geq 1}$ by*

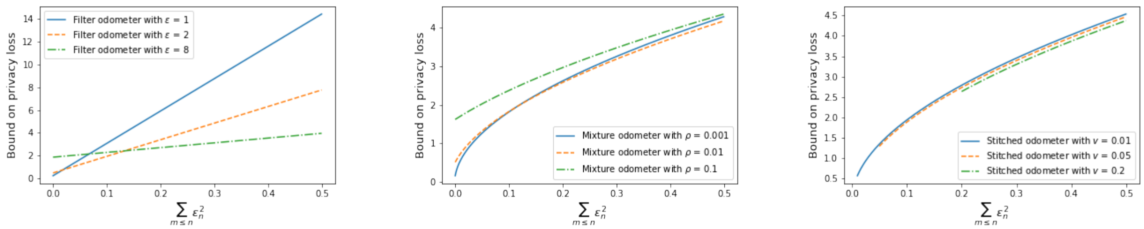
$$u_n^M(\epsilon_{1:n}, \delta_{1:n}) := \begin{cases} \infty & n > N \\ \sqrt{2\log\left(\frac{1}{\delta'}\sqrt{\frac{V(\epsilon_{1:n})+\rho}{\rho}}\right)(\rho + V(\epsilon_{1:n}))} + \frac{1}{2}V(\epsilon_{1:n}) & \textit{otherwise}. \end{cases}$$

3. ***Stitched odometer.*** *For any $v_0 > 0$, define the sequence of functions $(u_n^S)_{n\geq 1}$ by*

$$u_n^S(\epsilon_{1:n}, \delta_{1:n}) := \begin{cases} \infty & n > N \textit{ or } V(\epsilon_{1:n}) < v_0 \\ 1.7\sqrt{V(\epsilon_{1:n})\left(\log\log\left(\frac{2V(\epsilon_{1:n})}{v_0}\right) + 0.72\log\left(\frac{5.2}{\delta'}\right)\right)} + \frac{1}{2}V(\epsilon_{1:n}) & \textit{otherwise}. \end{cases}$$

Then, any of the sequences $(u_n^F)_{n\geq 1}$, $(u_n^M)_{n\geq 1}$, or $(u_n^S)_{n\geq 1}$ is a δ-privacy odometer.

We now provide intuition for our odometers, which are plotted in Figure 3. The filter odometer, dubbed as such because we derive it from Lemma 2, is the tightest odometer when $\sum_{m\leq n}\epsilon_m^2 \approx y^*$. This tightness drops off precipitously when $\sum_{m\leq n}\epsilon_m^2$ is far from y^*. The mixture odometer, which is named after the *the method of mixtures* (Robbins, 1970; de la Peña et al., 2007; Kaufmann and Koolen, 2021; Howard et al., 2021), sacrifices tightness at any fixed point in time to obtain overall tighter bounds on privacy loss. This odometer can be numerically optimized, in terms of ρ, for tightness at a predetermined value $\sum_{m\leq n}\epsilon_m^2$. The stitched odometer, whose name derives from



(a) Comparing filter odometers **(b)** Comparing mixture odometers **(c)** Comparing stitched odometers

Figure 2: Comparison of filter, mixture, and stitched odometers plotted as functions of $\sum_{m \leq n} \epsilon_m^2$. We set $\delta' = 10^{-6}$ and assume all algorithms being composed are purely differentially private for simplicity.

Theorem 5, is similarly tight across time. This odometer requires that $\sum_{m \leq n} \epsilon_m^2$ exceed some preselected "variance" v_0 before becoming nontrivial (i.e. finite). Larger values of v_0 will yield tighter odometers, albeit at the cost of losing bound validity when accumulated variance is small. With this intuition, we can compare our odometers to the original presented in Rogers et al. (2016).

Lemma 4 (Theorem 6.5 in Rogers et al. (2016)) *Assume the same setup as Theorem 2, and fix $\delta = \delta' + \delta''$, where $\frac{1}{e} \geq \delta' > 0$ and $\delta'' \geq 0$. Define the sequence of functions $(u_n^R)_{n \geq 1}$ by*

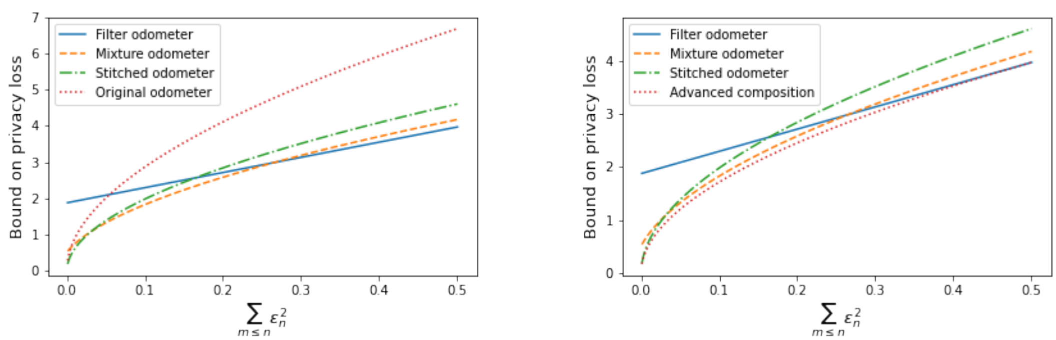
$$u_n^R(\epsilon_{1:n}, \delta_{1:n}) := \begin{cases} \infty & n > N \\ \sqrt{2V(\epsilon_{1:n})\left(\log(110e) + 2\log\left(\frac{\log(|x|)}{\delta'}\right)\right)} + \frac{1}{2}V(\epsilon_{1:n}) & n \leq N, V(\epsilon_{1:n}) \in \left[\frac{1}{|x|^2}, 1\right] \\ \sqrt{2\left(\frac{1}{|x|^2} + V(\epsilon_{1:n})\right)\left(1 + \frac{1}{2}\log\left(1 + |x|^2 V(\epsilon_{1:n})\right)\log\log\left(\frac{4}{\delta'}\log_2(|x|)\right)\right.} + \frac{1}{2}V(\epsilon_{1:n}) & \textit{otherwise} \end{cases},$$

where $|x|$ denotes the number of elements in dataset x. Then, $(u_n^R)_{n \geq 1}$ is a δ-privacy odometer.

Our new odometers improve over the one presented in Lemma 4. First, the above odometer has an explicit dependence on dataset size. In learning settings, datasets are large, degrading the quality of the odometer. Secondly, the tightness of the odometer drops off outside of the interval $\left[\frac{1}{|x|^2}, 1\right]$. If *any* privacy parameter of an algorithm being composed exceeds 1, the bound becomes significantly looser. Lastly, and perhaps most simply, the form of the odometer is complicated. Our odometers all have relatively straightforward dependence on the intrinsic time $\sum_{m \leq n} \epsilon_m^2$.

We now examine the rates of all odometers. For simplicity, let $v := \sum_{m \leq n} \epsilon_m^2$. The stitched odometer has a rate of $O(\sqrt{v \log\log(v)})$ in its leading term, asymptotically matching the law of the iterated logarithm (Robbins, 1970) up to constants. Both the original privacy odometer and the mixture odometer have a rate of $O(\sqrt{v \log(v)})$, demonstrating worse asymptotic performance. The filter odometer has the worst asymptotic performance, growing linearly as $O(v)$. This doesn't mean the stitched odometer is the best odometer, since target levels of privacy are often kept small.

To empirically compare odometers, it suffices to consider the setting of *pure* differential privacy, as the odometers identically depend on $(\delta_n)_{n \geq 1}$. Each presented odometer can be viewed as a function of v, allowing us to compare odometers by plotting their values for a continuum of v. Figure 3a shows that there is no clearly tightest odometer. All odometers, barring the original, dominate for some window of values of v. While the stitched odometer is asymptotically best, the mixture odometer is tighter for small values of v. Likewise, if one knows an approximate target privacy level, the filter odometer is tightest. This behavior is expected from our understanding of martingale concentration (Howard et al., 2020, 2021): there is no uniformly tightest boundary containing (with probability $1-\delta$) the entire path of a martingale; boundaries that are tight early must be looser later, and vice versa. In fact, we conjecture that our bounds are essentially unimprovable in general — this conjecture stems from the fact that the time-uniform martingale boundaries employed have error probability *essentially* equal to δ, which in turn stems from the deep fact that for continuous-path (and thus continuous-time) martingales, Ville's inequality (Fact 5)—that underlies the derivation of these boundaries—holds with exact equality. Since we operate in discrete-time, the only looseness in Ville's inequality stems from lower-order terms that reflect the possibility that at the stopping time, the value of the stopped martingale may not be *exactly* the value at the boundary.



(a) New odometers vs. original **(b)** New odometers vs. pointwise advanced composition

Figure 3: Figure 3a compares our odometers to the original. Figure 3b compares them with advanced composition optimized point-wise. The curve plotted for advanced composition is valid at any fixed time, but not uniformly over time. Our odometers nevertheless provide a close approximation.

In Figure 3b, we compare our new odometers with advanced composition optimized in a pointwise sense for all values of v simultaneously. Such a boundary *does not provide a valid odometer*, as advanced composition only holds at a single, prespecified point in intrinsic time v. Our novel odometers are almost tight with advanced composition for the values of v plotted. Our filter odometer lies tangent to the advanced composition curve, as expected from Section 5.2 of Howard et al. (2020). Overall, Figure 3 shows one can obtain almost tight, time-uniform control over privacy loss.

6. Summary of Results and Future Directions

In this work, we showed that one can match the rate of advanced composition while adaptively selecting both privacy parameters and algorithms. We constructed privacy filters that improve on both the original fully-adaptive composition results presented in Rogers et al. (2016) and the more recent results presented in Feldman and Zrnic (2021). We also constructed families of odometers that greatly outperform those presented in Rogers et al. (2016). Our insight is to view composition in terms of the "intrinsic time" $\sum_{m \leq n} \epsilon_m^2$. This allows us to leverage breakthroughs in self-normalized, time-uniform concentration (Howard et al., 2020, 2021). Our results show that full adaptivity is in fact a feature of differential privacy. Informally, advanced composition can simply be seen as a consequence of time-uniform martingale concentration.

There are many open problems related to fully-adaptive composition. While fully adaptive composition has been investigated for RDP and zCDP (Feldman and Zrnic, 2021), it has not been studied for f-DP (Dong et al., 2021). It also has not been investigated whether adaptivity in privacy parameter selection improves the performance of iterative algorithms such as private SGD. Intuitively, it should be beneficial to let the iterates of an algorithm guide future choices of privacy parameters. Optimal composition results (Kairouz et al., 2015; Murtagh and Vadhan, 2016; Zhu et al., 2021) have yet to be considered in a setting where privacy parameters are adaptively selected.

The most important direction for fully adaptive privacy composition pertains to applications to subsampled mechanisms (Wang et al., 2019). Rényi filters (Feldman and Zrnic, 2021) can be applied to subsampled methods when a target level of privacy in known. However, subsampled mechanisms are often used in settings where criteria such as accuracy define stopping conditions (Abadi et al., 2016). Existing Rényi odometers (Feldman and Zrnic, 2021; Lécuyer, 2021) do not provide high probability, time-uniform bounds on privacy loss. We believe the machinery introduced in this paper can be used in this setting to construct tight RDP odometers (in the sense of Rogers et al. (2016)).

7. Acknowledgements

AR acknowledges support from NSF DMS 1916320 and an ARL IoBT CRA grant. Research reported in this paper was sponsored in part by the DEVCOM Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196 (ARL IoBT CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

ZSW and JW were supported in part by the NSF CNS2120667, a CyLab 2021 grant, a Google Faculty Research Award, and a Mozilla Research Grant.

JW acknowledges support from NSF GRFP grants DGE1745016 and DGE2140739.

References

Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318, 2016.

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

Mark Bun and Thomas Steinke. Concentrated differential privacy: simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

Shanshan Chen, Zhenping Wang, Wenfei Xu, and Yu Miao. Exponential inequalities for self-normalized martingales. *Journal of Inequalities and Applications*, 2014(289):1–12, 2014.

Victor H de la Pena, Michael J Klass, and Tze Leung Lai. Self-normalized processes: exponential inequalities, moment bounds, and iterated logarithm laws. *Annals of Probability*, pages 1902–1933, 2004.

Victor H. de la Peña, Michael J. Klass, and Tze Leung Lai. Pseudo-maximization and self-normalized processes. *Probability Surveys*, 4:172 – 192, 2007. doi: 10.1214/07-PS119.

Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. In *Journal of the Royal Statistical Society: Series B*, pages 1–35, 2021.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, pages 371–380, 2009.

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

Cynthia Dwork and Guy N. Rothblum. Concentrated differential privacy. *CoRR*, abs/1603.01887, 2016.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006a.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006b. Springer Berlin Heidelberg.

Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 51–60. IEEE, 2010.

Vitaly Feldman and Tijana Zrnic. Individual privacy accounting via a Rényi filter. *Advances in Neural Information Processing Systems*, 2021.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257 – 317, 2020. doi: 10.1214/18-PS321.

Steven R. Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055 – 1080, 2021. doi: 10.1214/20-AOS1991.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *International Conference on Machine Learning*, pages 1376–1385. PMLR, 2015.

Shiva P Kasiviswanathan and Adam Smith. On the semantics of differential privacy: A Bayesian formulation. *Journal of Privacy and Confidentiality*, 6(1), 2014.

Emilie Kaufmann and Wouter M Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44, 2021.

Mathias Lécuyer. Practical privacy filters and odometers with Rényi differential privacy and applications to differentially private deep learning. *arXiv Preprint arXiv:2103.01379*, 2021.

Katrina Ligett, Seth Neel, Aaron Roth, Bo Waggoner, and Z Steven Wu. Accuracy first: Selecting a differential privacy level for accuracy-constrained ERM. *arXiv Preprint arXiv:1705.10829*, 2017.

Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275. IEEE, 2017.

Jack Murtagh and Salil Vadhan. The complexity of computing the optimal composition of differential privacy. In *Theory of Cryptography Conference*, pages 157–175. Springer, 2016.

Nicolas Papernot and Thomas Steinke. Hyperparameter tuning with renyi differential privacy. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=-70L8lpp9DF.

Herbert Robbins. Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41(5):1397–1409, 1970.

Ryan M Rogers, Aaron Roth, Jonathan Ullman, and Salil Vadhan. Privacy odometers and filters: pay-as-you-go composition. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Jean Ville. Etude critique de la notion de collectif. *Bull. Amer. Math. Soc*, 45(11):824, 1939.

Yu-Xiang Wang, Borja Balle, and Shiva Prasad Kasiviswanathan. Subsampled Rényi differential privacy and analytical moments accountant. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1226–1235. PMLR, 2019.

Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. Optimal accounting of differential privacy via characteristic function. *arXiv Preprint arXiv:2106.08567*, 2021.

Appendix A. Notation and Measure-Theoretic Formalism

In this section, we provide some measure-theoretic formalisms and details regarding datasets and neighboring relations.

Neighboring Datasets: Roughly speaking, an algorithm is differentially private if it difficult to distinguish between output distributions when the algorithm is run on similar inputs. In general, this notion of similarity amongst inputs is defined as a *neighboring relation* $\sim$ between elements on the input space $\mathcal{X}$. In particular, if two inputs (also referred to as datasets or databases) $x, x' \in \mathcal{X}$ satisfy the neighboring relation $x \sim x'$, the we say x and x' are *neighbors*.

There are several canonical examples of neighboring relations on the space of inputs $\mathcal{X}$. One example is where $\mathcal{X} = \mathbb{X}^n$ for some data domain $\mathbb{X}$. The data domain can be viewed as the set of all possible individual entries for a dataset, and the space $\mathbb{X}^n$ correspondingly contains all possible n element datasets. In this setting, databases $x, x' \in \mathcal{X}$ may be considered neighbors if x and x' differ in exactly one entry. Another slightly more general setting is when $\mathcal{X} = \mathbb{X}^*$, i.e., all possible datasets of finite size. In this situation, the earlier notion of neighboring still makes sense. However, in addition, we may say input datasets x and x' are neighbors if x can be obtained from x' by either adding or deleting an element. This is a very natural notion of neighboring, as under such a relation an algorithm would be differentially private if it were difficult to determine the presence or absence of an individual. Our work is agnostic to the precise choice of neighboring relation. As such, we choose to leave the notion as general as possible.

Algorithms and Random Variables: We will consider algorithms as randomized mappings $A : \mathcal{X} \to \mathcal{Y}$ taking inputs from $\mathcal{X}$ to some output space $\mathcal{Y}$. To be fully formal, we consider the output space $\mathcal{Y}$ as a *measurable space* $(\mathcal{Y}, \mathcal{G})$, where $\mathcal{G}$ is some σ-algebra denoting possible events. Recall that a σ-algebra $\mathcal{S}$ for a set S is simply a subset of 2^S containing S and $\emptyset$ that is closed under countable union, intersection, and complements. When we say A is an algorithm having inputs in some space $\mathcal{X}$, we really mean $A(x)$ is a $\mathcal{Y}$-valued random variable for any $x \in \mathcal{X}$. The space $\mathcal{X}$ need not have an associated σ-algebra, as algorithm inputs are essentially just indexing devices. Given a sequence of algorithms $(A_n)_{n\geq 1}$, $(A_n(x))_{n\geq 1}$ is a sequence of $\mathcal{Y}$-valued random variables, for any $x \in \mathcal{X}$.[5]

Since we are dealing with the composition of algorithms, we write $A_{1:n}(x)$ as shorthand for the random vector of the first n algorithm outputs, i.e. $A_{1:n}(x) = (A_1(x), \dots, A_n(x))$. Formally, the random vector $A_{1:n}(x)$ takes output values in the product measurable space $(\mathcal{Y}^n, \mathcal{G}^{\otimes n})$ where $\mathcal{G}^{\otimes n}$ denotes the n-fold product σ-algebra of $\mathcal{G}$ with itself. Likewise, since the number of algorithm outputs one views in fully-adaptive composition may be random, if N is a random time (i.e. a $\mathbb{N}$-valued random variable), we will often consider the random vector $A_{1:N}(x) = (A_1(x), \dots, A_N(x))$. We can view $A_{1:N}(x)$ as a random vector in the infinite product output space $(\mathcal{Y}^\infty, \mathcal{G}^{\otimes\infty})$. We do this by simply inserting a special terminating element $\Box$ into $\mathcal{Y}$ and repeating it ad infinitum after the last element of the random vector, i.e. $A_{1:N}(x) = (A_1(x), \dots, A_N(x), \Box, \Box, \dots)$. The reader should not concern themselves with the detail here — we just include this information to be fully formal and so it is explicit which sets of outcomes are considered measurable.

Filtrations and Stopping Times: Since privacy composition involves sequences of random outputs, we will use the measure-theoretic notion of a *filtration*. If we have fixed an input $x \in \mathcal{X}$, we can assume the random sequence $(A_n(x))_{n\geq 1}$ is defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Given such a probability space, a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$ is a sequence of σ-algebras satisfying: (i) $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \in \mathbb{N}$, and (ii) $\mathcal{F}_n \subset \mathcal{F}$ for all $n \in \mathbb{N}$. Given an arbitrary $\mathcal{Y}$-valued discrete-time stochastic process $(X_n)_{n\geq 1}$, it is often useful to consider the *natural filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ given by $\mathcal{F}_n := \sigma(X_m : m \leq n)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Intuitively, a filtration formalizes the notion of accumulating information over time. In particular, in the context of the natural filtration generated by a stochastic process, the nth σ-algebra in the filtration $\mathcal{F}_n$ essentially represents the entirety of information contained in the first n random variables. In other words, if one is given $\mathcal{F}_n$, they would know all possible events/outcomes that could have occurred up to and including timestep n.

Lastly, we briefly mention the notion of a *stopping time*, as this measure-theoretic object is necessary to define privacy filters. Given a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, a random time N is said to be a stopping time with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if, for any n, the event $\{N \leq n\} \in \mathcal{F}_n$. In words, a random time N is a stopping time if given the information in $\mathcal{F}_n$ we can determine whether or not we should have stopped by time n. Stopping times are essential to the study of fully-adaptive composition, as a practitioner of privacy will need to use the adaptively selected privacy parameters to determine whether or not to stop interacting with the underlying sensitive database.

5. Even if algorithms have different types of outputs (maybe some algorithms have categorical outputs while others output real-valued vectors), $\mathcal{Y}$ can still be made appropriately large to contain all possible outcomes.

Appendix B. Martingale Concentration Inequalities

In this appendix, we provide a thorough exposition into the concentration inequalities leveraged in this paper. First, at the heart of supermartingale concentration is Ville's inequality (Ville, 1939), which can be viewed as a time-uniform version of Markov's inequality.

Lemma 5 (Ville's Inequality (Ville, 1939)) *Let $(X_n)_{n\in\mathbb{N}}$ be a nonnegative supermartingale with respect to some filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Then, for any confidence parameter $\delta \in (0,1)$, we have $\mathbb{P}\left(\exists n \in \mathbb{N} : X_n \geq \frac{\mathbb{E}X_0}{\delta}\right) \leq \delta$.*

We do not directly leverage Ville's inequality in this work, but all inequalities we use can be directly proven from Lemma 5 (Howard et al., 2020, 2021). In short, each inequality in this supplement is proved by carefully massaging a martingale of interest into a non-negative supermartingale.

To obtain the same privacy composition rates as advanced composition, we leverage the following special case of a recent advance in time-uniform martingale concentration (Howard et al., 2020). We include the proof of this special case as its structure gives insight into how we can generalize our main results to also allow algorithms to satisfy other modes of privacy such as zero-concentrated differential privacy (Bun and Steinke, 2016). We emphasize that Theorem 3 is just a special case of the main result in Howard et al. (2020) — we are not reproving the more general, main result of that paper here.

Theorem 3 *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale with respect to some filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that $M_0 = 0$ almost surely. Moreover, let $(\sigma_n)_{n\geq 1}$ be a $(\mathcal{F}_n)_{n\in\mathbb{N}}$-predictable sequence of random variables such that, conditioned on $\mathcal{F}_{n-1}$, $\Delta M_n := M_n - M_{n-1}$ is σ_n^2-subGaussian. Define $V_n := \sum_{m\leq n}\sigma_m^2$. Then, we have, for all $a, b > 0$,*

$$\mathbb{P}\left(\exists n \in \mathbb{N} : M_n \geq \frac{b}{2} + \frac{b}{2a}V_n\right) \leq \exp\left(\frac{-b^2}{2a}\right).$$

Proof [**Theorem 3**] Let $(M_n)_{n\in\mathbb{N}}$ be the martingale listed in the theorem statement. Observe that, for any $a, b > 0$, the process $(X_n)_{n\in\mathbb{N}}$ given by

$$X_n := \exp\left(\frac{b}{a}M_n - \frac{b^2}{2a^2}\sum_{m\leq n}\sigma_n^2\right)$$

is a non-negative supermartingale. As such, applying Ville's inequality (Lemma 5) yields

$$\mathbb{P}\left(\exists n \in \mathbb{N} : X_n > \exp\left(\frac{b^2}{2a}\right)\right) \leq \exp\left(-\frac{b^2}{2a}\right).$$

Now, on such event, taking logs and rearranging yields

$$\frac{b}{a}M_n \leq \frac{b^2}{2a} + \frac{b^2}{2a^2}\sum_{m\leq n}\sigma_m^2.$$

Multiplying both sides by $\frac{a}{b}$ finishes the proof.

■

The predictable process $(V_n)_{n\in\mathbb{N}}$ is a proxy for the accumulated variance of $(M_n)_{n\in\mathbb{N}}$ up to any fixed point in time. In particular, the process $(V_n)_{n\in\mathbb{N}}$ can be thought of as yielding the "intrinsic time" of the process. The free parameters a and b thus allow us to optimize the tightness of the boundary for some intrinsic moment in time. This is ideal for us, as, for the sake of composition, the target privacy parameter ϵ can guide us in finding a point in intrinsic time (that is, in terms of the process $(V_n)_{n\in\mathbb{N}}$) to optimize for. We discuss how to apply this inequality to prove privacy composition results both in this supplement and in Section 4.

We also leverage the following martingale inequalities from Howard et al. (2021) in Section 5, where we construct various families of time-uniform bounds on privacy loss in fully-adaptive composition. These inequalities take on a more complicated form than Theorem 3, but we explain the intuition behind them in the sequel. The first bound we present relies on the method of mixtures for martingale concentration, which stems back to Robbins' work in the 1970s (Robbins, 1970). There are many good resources providing an introduction to the method of mixtures (de la Peña et al., 2007; Kaufmann and Koolen, 2021; Howard et al., 2021).

Theorem 4 *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale with respect to some filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that $M_0 = 0$ almost surely. Moreover, let $(\sigma_n)_{n\geq 1}$ be a $(\mathcal{F}_n)_{n\in\mathbb{N}}$-predictable sequence of random variables such that, conditioned on $\mathcal{F}_{n-1}$, both $\Delta M_n := M_n - M_{n-1}$ and $-\Delta M_n$ are σ_n^2-subGaussian. Define $V_n := \sum_{m\leq n}\sigma_m^2$ and choose a tuning parameter $\rho > 0$. Then, for any $\delta > 0$, we have*

$$\mathbb{P}\left(\exists n \in \mathbb{N} : M_n \geq \sqrt{2(V_n+\rho)\log\left(\frac{1}{\delta}\sqrt{\frac{V_n+\rho}{\rho}}\right)}\right) \leq \delta.$$

The next inequality relies on the recent technique of boundary stitching, first presented in Howard et al. (2021). Intuitively, the technique works by breaking intrinsic time — that is, time according to the accumulated variance process $(V_n)_{n\in\mathbb{N}}$ — into roughly geometrically spaced pieces. Then, one optimizes a tight-boundary in each region and takes a union bound. The actual details are more technical, but are not needed in this work.

Theorem 5 *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that $M_0 = 0$ almost surely. Moreover, let $(\sigma_n)_{n\geq 1}$ be a $(\mathcal{F}_n)_{n\in\mathbb{N}}$-predictable sequence of random variables such that, conditioned on $\mathcal{F}_{n-1}$, both $\Delta M_n := M_n - M_{n-1}$ and $-\Delta M_n$ are σ_n^2-subGaussian. Define $V_n := \sum_{m\leq n}\sigma_m^2$ and choose a starting intrinsic time $v_0 > 0$. Then, for any $\delta \in (0,1)$, we have*

$$\mathbb{P}\left(\exists n \in \mathbb{N} : M_n \geq 1.7\sqrt{V_n\left(\log\log\left(\frac{2V_n}{v_0}\right) + 0.72\log\left(\frac{5.2}{\delta}\right)\right)},\ V_n \geq v_0\right) \leq \delta.$$

Note that the original version of Theorem 5 as found in Howard et al. (2021) has more free parameters to optimize over, but we have already simplified the expression to make the result more readable. The free parameter $v_0 > 0$ in the above boundary gives the intrinsic time at which the boundary becomes non-trivial (i.e., the tightest available upper bound before $V_n \geq v_0$ is ∞).

We qualitatively compare these bounds in Section 5, wherein we construct various time-uniform bounds on privacy loss processes. For now, Theorem 3 can be thought of as providing a tight upper bound on a martingale at a single point in intrinsic time, providing loose guarantees elsewhere. On the other hand, Theorems 4 and 5 provide decently tight control over a martingale at all points in intrinsic time simultaneously, although at the cost of sacrificing tightness at any given fixed point.

Appendix C. Bounding Privacy Loss in Adaptive Composition

Before we can prove Lemma 2, we need to following bound on the conditional expectation of privacy loss, which can be immediately obtained from the bound on expected privacy loss presented in Bun and Steinke (2016).

Lemma 6 (Proposition 3.3 in Bun and Steinke (2016)) *Suppose A and B are algorithms such that A is ϵ-differentially private conditioned on B. Then, for any input dataset $x \in \mathcal{X}$ and neighboring dataset $x' \sim x$, we have that*

$$\mathbb{E}\left(\mathcal{L}(x, x')|B(x)\right) \leq \frac{1}{2}\left(\epsilon(B(x))\right)^2.$$

Now, we prove Lemma 2.

Proof [Lemma 2] To begin, we assume that the algorithms $(A_n)_{n\geq 1}$ satisfy $(\epsilon_n, 0)$-pDP conditioned on $A_{1:n-1}$. We will show how to alleviate this assumption on the approximation parameter in the second half of the proof. Fix an input database $x \in \mathcal{X}$. For convenience, we denote by $(\mathcal{F}_n(x))_{n\in\mathbb{N}}$ the natural filtration generated by $(A_n(x))_{n\geq 1}$. Since we have fixed $x \in \mathcal{X}$, for notational simplicity, we write ϵ_n for the random variable $\epsilon_n(A_{1:n-1}(x))$ and define δ_n similarly. Additionally, by N we mean the stopping time $N((\epsilon_n)_{n\in\mathbb{N}}, (\delta_n)_{n\in\mathbb{N}})$. Recall that we have already argued that, for any neighboring dataset $x' \sim x$, the process

$$M_n := M_n(x, x') = \mathcal{L}_{1:n}(x, x') - \sum_{m\leq n}\mathbb{E}\left(\mathcal{L}_m(x, x')|\mathcal{F}_{m-1}(x)\right)$$

is a martingale with respect to $(\mathcal{F}_n(x))_{n\in\mathbb{N}}$. Further observe that its increments $\Delta M_n := \mathcal{L}_n(x, x') - \mathbb{E}\left(\mathcal{L}_n(x, x')|\mathcal{F}_{n-1}(x)\right)$ are ϵ_n^2-subGaussian conditioned on $\mathcal{F}_{n-1}(x)$.

Thus, by Theorem 3, we know that, for any $b, a > 0$, we have

$$\mathbb{P}\left(\exists n \in \mathbb{N} : M_n \geq \frac{b}{2} + \frac{b}{2a}V_n\right) \leq \exp\left(\frac{-b^2}{2a}\right),$$

where the process $(V_n)_{n\in\mathbb{N}}$ given by $V_n := \sum_{m\leq n}\epsilon_m^2$ is the accumulated variance up to and including time n. Thus, it suffices to optimize the free parameters a and b to prove the result.

To do this, consider the following function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ given by

$$f(y) = \sqrt{2\log\left(\frac{1}{\delta'}\right)y} + \frac{1}{2}y.$$

Clearly, f is a quadratic polynomial in $\sqrt{y}$ which is strictly increasing. In particular, one can readily check that

$$y^* := \left(-\sqrt{2\log\left(\frac{1}{\delta'}\right)} + \sqrt{2\log\left(\frac{1}{\delta'}\right) + \epsilon}\right)^2 \tag{6}$$

solves the equation $f(y) = \epsilon$, where $\epsilon > 0$ is the target privacy parameter.

As such, setting $a := y^*$ and $b := \sqrt{2\log\left(\frac{1}{\delta'}\right)y^*}$ yields

$$\exp\left(\frac{-b^2}{a}\right) = \exp\left(\frac{-2y^*\log\left(\frac{1}{\delta'}\right)}{y^*}\right) = \delta'.$$

Furthermore, expanding the definition of $(M_n)_{n\in\mathbb{N}}$, we see that for the selected parameters the parameters yield, with probability at least $1-\delta'$, for all $n \le N$ we have:

$$\begin{aligned}
\mathcal{L}_{1:n}(x,x') &\le \frac{b}{2} + \frac{b}{2a}V_n + \sum_{m\le n}\mathbb{E}\left(\mathcal{L}_m(x,x') \mid \mathcal{F}_{m-1}\right) \le \frac{b}{2} + \frac{b}{2a}\sum_{m\le n}\epsilon_m^2 + \frac{1}{2}\sum_{m\le n}\epsilon_m^2 \\
&= \frac{1}{2}\sqrt{2\log\left(\frac{1}{\delta'}\right)y^*} + \frac{1}{2}\frac{\sqrt{2\log\left(\frac{1}{\delta'}\right)y^*}}{y^*}\sum_{m\le n}\epsilon_m^2 + \frac{1}{2}\sum_{m\le n}\epsilon_m^2 \\
&\le \frac{1}{2}\sqrt{2\log\left(\frac{1}{\delta'}\right)y^*} + \frac{1}{2}\sqrt{2\log\left(\frac{1}{\delta'}\right)y^*} + \frac{1}{2}\sum_{m\le n}\epsilon_m^2 \\
&= \sqrt{2\log\left(\frac{1}{\delta'}\right)y^*} + \frac{1}{2}\sum_{m\le n}\epsilon_m^2 \le \sqrt{2\log\left(\frac{1}{\delta'}\right)y^*} + \frac{1}{2}y^* = \epsilon.
\end{aligned}$$

Thus, we have proven the desired result in the case where all algorithms have $\delta_n = 0$.

Now, we show how to generalize our result to the case where the approximation parameters δ_n are not identically zero. Define the events

$$A := \left\{\exists n \le N : \mathcal{L}_{1:n}(x,x') > \epsilon\right\} \qquad \text{and} \qquad B := \left\{\exists n \le N : \mathcal{L}_n(x,x') > \epsilon_n\right\}.$$

Our goal is to show that, with N defined as in the statement of Theorem 1, that $\mathbb{P}(A) \le \delta$. Simply using Bayes rule, we have that

$$\mathbb{P}(A) = \mathbb{P}(A\cap B^c) + \mathbb{P}(A\cap B) \le \mathbb{P}(A|B^c) + \mathbb{P}(B) \le \delta' + \mathbb{P}(B),$$

where the second inequality follows from our already-completed analysis in the case that $\delta_n = 0$. Now, we show that $\mathbb{P}(B) \le \delta''$, which suffices to prove the result as we have, by assumption, $\delta = \delta' + \delta''$.

Define the modified privacy loss random variables $(\widetilde{\mathcal{L}}_n(x,x'))_{n\in\mathbb{N}}$ by

$$\widetilde{\mathcal{L}}_n(x,x') := \begin{cases} \mathcal{L}_n(x,x') & n \le N \\ 0 & \text{otherwise} \end{cases}.$$

Likewise, define the modified privacy parameter random variables $\widetilde{\epsilon}_n$ and $\widetilde{\delta}_n$ in an identical manner. Then, we can bound $\mathbb{P}(B)$ in the following manner:

$$\begin{aligned}
\mathbb{P}(\exists n \le N : \mathcal{L}_n(x,x') > \epsilon_n) &= \mathbb{P}\left(\exists n\in\mathbb{N} : \widetilde{\mathcal{L}}_n(x,x') > \widetilde{\epsilon}_n\right) \\
&\le \sum_{n=1}^{\infty}\mathbb{P}\left(\widetilde{\mathcal{L}}_n(x,x') > \widetilde{\epsilon}_n\right) = \sum_{n=1}^{\infty}\mathbb{E}\mathbb{P}\left(\widetilde{\mathcal{L}}_n(x,x') > \widetilde{\epsilon}_n | \mathcal{F}_{n-1}\right) \\
&\le \sum_{n=1}^{\infty}\mathbb{E}\widetilde{\delta}_n = \mathbb{E}\left[\sum_{n=1}^{\infty}\widetilde{\delta}_n\right] = \mathbb{E}\left[\sum_{n\le N}\delta_n\right] \le \delta''.
\end{aligned}$$

Thus, we have have proven the desired result in the general case, and are done.

■

Using the same general strategy as above, we can prove the validity of our privacy odometers presented in Theorem 2 in Section 5.

Proof [**Theorem 2**] As in the proof of Lemma 2, we first consider the case where $\delta_n = 0$ for all $n \geq 1$. In this case, fix an input dataset $x \in \mathcal{X}$ and a neighboring dataset $x' \in \mathcal{X}$. Let $(M_n)_{n\in\mathbb{N}}$ be the corresponding privacy loss martingale as outlined in Equation (5), where we implicitly hide the dependence on x, x', which are fixed. Let $(u_n)_{n\geq 1}$ be one of the sequences outlined in the theorem statement, and define $U_n := u_n(\epsilon_{1:n}, \delta_{1:n})$ for all $n \geq 1$, where once again we write ϵ_n and δ_n for $\epsilon_n(A_{1:n-1}(x))$ and $\delta_n(A_{1:n-1}(x))$ respectively. It follows from Theorems 3, 4, and 5 that

$$\mathbb{P}\left(\exists n \in \mathbb{N} : M_n > B_n\right) \leq \delta,$$

for $B_n = U_n - \frac{1}{2}\sum_{m\leq n} \epsilon_m^2$. Recalling that $M_n = \sum_{m\leq n}\{\mathcal{L}_m(x,x') - \mathbb{E}(\mathcal{L}_m(x,x')|\mathcal{F}_{n-1}(x))\}$ and that $\mathbb{E}(\mathcal{L}_n(x,x')|\mathcal{F}_{n-1}(x)) \leq \frac{1}{2}\epsilon_n^2$ for all $n \in \mathbb{N}$, it thus follows that

$$\mathbb{P}\Big(\exists n \in \mathbb{N} : \mathcal{L}_{1:n}(x,x') > U_n\Big) \leq \delta,$$

where $(\mathcal{F}_n(x))_{n\geq 1}$ is again the natural filtration generated by $(A_n(x))_{n\geq 1}$. Thus, since $x \sim x'$ were arbitrary, we have shown that $(u_n)_{n\geq 1}$ is a δ-privacy odometer in the case $\delta_n = 0$ for all $n \geq 1$.

To generalize to the case where δ_n may be nonzero, we can apply precisely the same argument used in the second part of the proof of Lemma 2, thus proving the general result. ■

Appendix D. Generalized Randomized Response

In this appendix, we ultimately aim to prove Theorem 1. First, recall the important fact that *any* differentially private algorithm can be viewed as a post-processing of randomized response (Kairouz et al., 2015), as stated in Lemma 7 below.

Lemma 7 (Reduction to Randomized Response (Kairouz et al., 2015)) *Let algorithm $A : \mathcal{X} \to \mathcal{Y}$ be (ϵ, δ)-DP. Let R be an instance of (ϵ, δ)-randomized response. Then, for any neighbors $x_0, x_1 \in \mathcal{X}$, there is a transition kernel $\nu : \mathcal{G} \times \mathcal{R} \to [0, 1]$ such that for $b \in \{0, 1\}$, we have $\mathbb{P}(A(x_b) \in \cdot) = \nu_b$, where*[6] *$\nu_b = \mathbb{E}\nu(\cdot, R(b))$.*

In Lemma 3 of Section 4, we generalized Lemma 7 to the case of conditional differential privacy. To do this, we introduced *conditional randomized response* in Definition 6. In conditional randomized response, on the event $\{B = z\}$, the conditional laws of $R(0)$ and $R(1)$ just become that of regular randomized response with some known privacy parameters $\epsilon(z)$ and $\delta(z)$. We now prove Lemma 3.

Proof [**Lemma 3**] Let $b, b' \in \{0, 1\}$ be arbitrary. For any outcome $\{B'(b') = z\}$, let $\mathbb{P}_z(A(x_b) \in \cdot)$ be the probability measure $\mathbb{P}(A(x_b) \in \cdot | B'(b') = z)$. In particular, this measure does not depend

6. By $\nu_b(\cdot) := \mathbb{E}\nu(\cdot, R(b))$, we mean ν_b is the averaged probability measure given by

$$\nu_b(\cdot) = \mathbb{P}(R(b) = 1)\nu(\cdot, 1) + \mathbb{P}(R(b) = 0)\nu(\cdot, 0) + \mathbb{P}(R(b) = \perp)\nu(\cdot, \perp) + \mathbb{P}(R(b) = \top)\nu(\cdot, \top).$$

on the input bit b'. By the assumptions of conditional differential privacy (Definition 1), it follows that under the probability measure $\mathbb{P}_z$, $A(x_b)$ is $(\epsilon(z), \delta(z))$-differentially private. Moreover, it also follows that R is an instance of $(\epsilon(z), \delta(z))$-randomized response under $\mathbb{P}_z$. Consequently, Lemma 7 yields the existence of a kernel ν_z such that $\mathbb{P}_z(A(x_b) \in \cdot) = \mathbb{E}_z \nu_z(\cdot, R(b))$, where the averaged measure is as defined in Footnote 6. Setting $\nu(\cdot, R(b)|z) := \nu_z(\cdot, R(b))$, we that

$$\mathbb{P}(A(x_b) \in \cdot \mid B'(b') = z) = \mathbb{E}\left(\nu(\cdot, R(b) \mid z) \mid B'(b') = z\right),$$

which thus yields

$$\mathbb{P}(A(x_b) \in \cdot \mid B'(b')) = \mathbb{E}\left(\nu(\cdot, R(b) \mid B'(b')) \mid B'(b')\right),$$

where the conditionally averaged measure is as described in Footnote 4 in the main body of the paper. This proves the desired result. ■

Lastly, before proving Theorem 1, we need the following lemma. This lemma essentially tells us that if A is (ϵ, δ)-pDP conditioned on B, and A' is a randomized post-processing algorithm, then releasing the vector (A, A') is also (ϵ, δ)-pDP conditioned on B. Note that this is *not* in contradiction with the converse direction of Lemma 1, as releasing the output of A' alone may not satisfy conditional (ϵ, δ)-pDP. But once we observe A, since A' is a post-processing, we can gleam no more information about the true underlying dataset.

Lemma 8 *Suppose A, B are algorithms with inputs in $\mathcal{X}$ and outputs in measurable spaces $(\mathcal{Y}, \mathcal{G})$ and $(\mathcal{Z}, \mathcal{H})$ respectively. Assume A is (ϵ, δ)-pDP conditioned on B. Let $(S, \mathcal{S})$ be a measurable space and suppose $\mu : \mathcal{S} \times \mathcal{Y} \times \mathcal{Z} \to [0, 1]$ is a conditional transition kernel. Suppose $A' : \mathcal{X} \to S$ is an algorithm satisfying*

$$\mathbb{P}\left(A'(x) \in \cdot | A(x') = y, B(x') = z\right) = \mu(\cdot, y \mid z), \tag{7}$$

for all $y \in \mathcal{Y}, z \in \mathcal{Z}$, and $x, x' \in \mathcal{X}$. Then, the joint algorithm $(A, A') : \mathcal{X} \to \mathcal{Y} \times S$ is also (ϵ, δ)-pDP conditioned on B.

Proof Let $x, x' \in \mathcal{X}$ be arbitrary neighboring datasets. Let $q_B^x, q_B^{x'}$ be the corresponding conditional joint densities of $(A(x), A'(x))$ and $(A(x'), A'(x'))$ given $B(x)$ respectively. Likewise, let $p_B^x, p_B^{x'}$ be the corresponding conditional densities of $A(x)$ and $A(x')$ respectively conditioned on $B(x)$, and $q_{B,A}^x, q_{B,A}^{x'}$ the conditional densities of $A'(x)$ and $A'(x')$ given $A(x)$ and $B(x)$. Let $\mathcal{L}_B^{(A,A')}(x, x')$ denote the joint privacy loss between $(A(x), A'(x))$ and $(A(x'), A'(x'))$ given $B(x)$, while $\mathcal{L}_B^A(x, x')$ denotes the privacy loss between $A(x)$ and $A(x')$ given $B(x)$. We have, using Bayes rule,

$$\begin{aligned}
\mathcal{L}_B^{(A,A')}(x, x') &= \log\left(\frac{q_B^x(A(x), A'(x) \mid B(x))}{q_B^{x'}(A(x), A'(x) \mid B(x))}\right) = \log\left(\frac{p_B^x(A(x) \mid B(x))}{p_B^{x'}(A(x) \mid B(x))} \cdot \frac{q_{B,A}^x(A'(x) \mid B(x), A(x))}{q_{B,A}^{x'}(A'(x) \mid B(x), A(x))}\right) \\
&= \log\left(\frac{p_B^x(A(x) \mid B(x))}{p_B^{x'}(A(x) \mid B(x))}\right) = \mathcal{L}_B^{(A)}(x, x'),
\end{aligned}$$

The first equality on the second line follows from the assumption outlined in Equation (7). More specifically, since we have

$$\mathbb{P}\left(A'(x) \in \cdot | A(x), B(x)\right) = \mu(\cdot, A(x) \mid B(x)) = \mathbb{P}\left(A'(x') \in \cdot | A(x), B(x)\right),$$

it follows that the conditional densities $q^x_{B,A}$ and $q^{x'}_{B,A}$ are equal almost surely. Since A is (ϵ, δ)-pDP conditioned on B, the result now follows. ■

We now prove the main result of Section 4.

Proof [**Theorem 1**] Fix arbitrary neighbors $x_0, x_1 \in \mathcal{X}$. Let $(R_n)_{n\geq 1}$ be a sequence of algorithms such that R_n is an instance of (ϵ_n, δ_n)-randomized response conditioned on $A'_{1:n-1} : \{0,1\} \to \mathcal{Y}^{n-1}$, where $A'_m : \{0,1\} \to \mathcal{Y}$ is the restricted algorithm given by $A'_m(b) := A_m(x_b)$, for all $m \geq 1$. Lemma 3 guarantees the existence of a sequence of transition kernels $(\nu_n)_{n\geq 1}$, $\nu_n : \mathcal{G} \times \mathcal{R} \times \mathcal{Y}^{n-1} \to [0,1]$ such that, for all $n \geq 1$ and $b, b' \in \{0,1\}$, we have $\mathbb{P}(A'_n(b) \in \cdot \mid A'_{1:n-1}(b')) = \nu^{(n)}_{b,b'}$ almost surely. Here, $\nu^{(n)}_{b,b'}$ is the averaged conditional probability, as defined in terms of ν_n in Lemma 3 and Footnote 4. This equality means we can find an underlying probability space (i.e. a coupling) such that the random post-processing draws from the kernel $\nu_n(\cdot, R_n(b) \mid A'_{1:n-1}(b'))$ equal $A'_n(b)$ almost surely, for all $n \geq 1$.

Now, for any $n \geq 1$, since R_n is an instance of (ϵ_n, δ_n)-randomized response conditioned on $A'_{1:n-1}$, it follows that R_n is in fact (ϵ_n, δ_n)-pDP conditioned on $A'_{1:n-1}$. Moreover, this also implies that R_n is (ϵ_n, δ_n)-pDP conditioned on $(A'_{1:n-1}, R_{1:n-1})$, since, by definition, ϵ_n and δ_n only depend on the realizations of $R_{1:n-1}$ through the outputs of $A'_{1:n-1}$. By Lemma 8, it follows that for all $n \geq 1$, the algorithm (R_n, A'_n) is (ϵ_n, δ_n)-pDP conditioned on $(R_{1:n-1}, A'_{1:n-1})$. Thus, by Lemma 2, it follows that the composed algorithm $(R_{1:N'(\cdot)}(\cdot), A'_{1:N'(\cdot)}(\cdot))$ is (ϵ, δ)-DP, where $N'(b) := N(x_b)$ and ϵ, δ and N, are as outlined in the statement of Theorem 1.

Lastly, since differential privacy is closed under arbitrary post-processing (Dwork and Roth, 2014), it follows that $A'_{1:N'(\cdot)}(\cdot)$ is (ϵ, δ)-differentially private. Since x_0 and x_1 were arbitrary neighboring inputs, the result follows, i.e. $A_{1:N(\cdot)}(\cdot) : \mathcal{X} \to \mathcal{Y}^\infty$ is (ϵ, δ)-differentially private. ■

Appendix E. Generalization to zCDP

In this section, we prove an extension to Theorem 1 that allows the algorithms being composed to either satisfy conditional differential privacy *or* conditional zero-concentrated privacy with probability one (this is made precise in the theorem statement). We provide this relatively straightforward extension because the privacy loss of many common algorithms (in particular, Gaussian mechanisms) can be more tightly characterized by zero-concentrated differential privacy.

Zero-concentrated differential privacy, constructed by Bun and Steinke (2016), is a form of privacy that involves characterizing the uncentered moment generating function of the privacy loss random variable. For our purposes, the definition of zero-concentrated differential privacy is as follows.

Definition 7 (Zero-concentrated differential privacy (Bun and Steinke, 2016))
An algorithm A is said to satisfy $\frac{1}{2}\epsilon^2$-zero-concentrated differential privacy (or $\frac{1}{2}\epsilon^2$-zCDP for short) if, for any $\lambda \geq 0$, we have

$$\mathbb{E}\left(e^{\lambda \mathcal{L}(x,x')}\right) \leq e^{\frac{1}{2}\lambda(\lambda+1)\epsilon_n^2}.$$

If we were considering zero-concentrated differentially private algorithms in isolation, the following would be a reasonable definition of conditional zCDP.

Definition 8 (Conditional zCDP) *Suppose A and B are algorithms having inputs in a space $\mathcal{X}$ and outputs in measurable spaces $(\mathcal{Y}, \mathcal{G})$ and $(\mathcal{Z}, \mathcal{H})$. Suppose $\epsilon : \mathcal{Z} \to \mathbb{R}_{\geq 0}$ is measurable. We say the algorithm A is $\frac{1}{2}\epsilon^2$-zCDP conditioned on B if, for any neighboring datasets $x \sim x'$ and all $\lambda \geq 0$,*

$$\mathbb{E}\left(e^{\lambda \mathcal{L}(x,x')} \mid B(x)\right) \leq e^{\frac{1}{2}\lambda(\lambda+1)\epsilon(x)^2}.$$

Note that the constant λ in Definition 8 can also be taken to be a $B(x)$-measurable random variable, adding increased generality.

While Definition 8 is a step in the right direction, what we actually want is a bit more subtle. In particular, in the context of adaptive data analysis, we want the data analyst to be able to select the type of privacy an algorithm satisfies on the fly. That is, given the outputs of the first $n-1$ algorithms, the data analyst can *then* decide whether to run an algorithm satisfying conditional DP or zCDP. The conditions of the following theorem establish precisely this desideratum.

Theorem 6 *Let $(A_n)_{n\geq 1}$ be an adaptive sequence of algorithms, and, for any x, let $(\mathcal{F}_n(x))_{n\in\mathbb{N}}$ be the natural filtration generated by $(A_n(x))_{n\in\mathbb{N}}$. For any $n \geq 1$, assume there are functions $\epsilon_n, \delta_n : \mathcal{Y}^{n-1} \times \mathcal{Z}^{n-1} \to \mathbb{R}_{\geq 0}$. For any $x \sim x'$ neighbors and any $n \geq 1$, assume that with probability 1 we have either*

1. *$\mathbb{P}(A_{1:n}(x) \in G \mid \mathcal{F}_{n-1}(x)) \leq e^{\epsilon_n(x)}\mathbb{P}(A_{1:n}(x') \in G \mid \mathcal{F}_{n-1}(x)) + \delta_n(x)$ for events G, or*
2. *$\delta_n(x) = 0$ and $\mathbb{E}\left(e^{\lambda\mathcal{L}_n(x,x')} \mid \mathcal{F}_{n-1}(x)\right) \leq e^{\frac{1}{2}\lambda(\lambda+1)\epsilon_n(x)^2}$ for all $\lambda \geq 0$.*

Then, for any $\epsilon > 0$ and $\delta = \delta' + \delta''$, letting N be as outlined in Theorem 1, we have that the algorithm $A_{1:N(\cdot)}(\cdot)$ is (ϵ, δ)-differentially private.

The assumptions on the algorithms in Theorem 6 essentially say that, with the information available up to time $n-1$, the analyst will then select the mode of privacy of the next algorithm, either being DP or zCDP. As was the case for Theorem 1, it is first easier to prove the case where conditional differential privacy is replaced by conditional probabilistic differential privacy. As such, we start by proving the following lemma.

Lemma 9 *Theorem 6 holds when the first of two conditions is replaced by the stronger assumption that, for all neighbors $x \sim x'$,*

$$\mathbb{P}\left(\mathcal{L}_n(x,x') > \epsilon_n(x) \mid \mathcal{F}_{n-1}(x)\right) \leq \delta_n(x).$$

Proof We proceed using the same general approach as in Theorem 2. That is, let $x \sim x'$ be fixed neighboring datasets, and assume that $\delta_n = 0$ for all $n \geq 1$. We let $(\mathcal{F}_n(x))_{n\geq 1}$ be the natural filtration associated with $(A_n(x))_{n\geq 1}$.

If the first condition holds (that is, if $\mathbb{P}(\mathcal{L}(x,x') > \epsilon_n(x) \mid \mathcal{F}_{n-1}(x)) = 0$), then applying Lemma 6 conditionally yields

$$\mathbb{E}\left(e^{\lambda\mathcal{L}_n(x,x')} \mid \mathcal{F}_{n-1}(x)\right) \leq e^{\frac{1}{2}\lambda(\lambda+1)\epsilon_n(x)^2}.$$

Thus, we see that the second condition actually holds for all algorithms being composed. Now, consider the process

$$M_n := \sum_{m\leq n}\left\{\mathcal{L}_n(x,x') - \frac{1}{2}\epsilon_n(x)^2\right\}.$$

While $(M_n)_{n\in\mathbb{N}}$ itself is not in general a martingale with respect to $(\mathcal{F}_n(x))_{n\in\mathbb{N}}$, from our above observation on the moment generating function of the privacy loss random variable, it follows that

$$X_n := \exp\left(\frac{b}{a}M_n - \frac{b^2}{2a^2}\sum_{m\leq n}\epsilon_m(x)^2\right)$$

is in fact a nonnegative supermartingale with respect to $(\mathcal{F}_n(x))_{n\in\mathbb{N}}$. In particular, the proof of Theorem 3 found in Appendix B still goes through, yielding, for any $a, b > 0$, the concentration inequality

$$\mathbb{P}\left(\exists n \in \mathbb{N} : M_n > \frac{b}{2} + \frac{b}{2a}\sum_{m\leq n}\epsilon_m(x)^2\right) \leq \exp\left(-\frac{b^2}{2a}\right).$$

Optimizing over a and b as in the proof of Theorem 2 and then union bounding to handle the case of general δ_n yields the desired result. ■

Now we prove the main theorem of this appendix.

Proof [**Theorem 6**] This proof follows in a similar vein to the proof of Theorem 1. As such, let $x_0 \sim x_1$ be neighboring inputs. Define, for all $m \geq 1$, A'_m to be the restricted algorithm, as outlined in the proof of Theorem 1. Let $(R_n)_{n\geq 1}$ be a sequence of algorithms such that on the event $E := \{\delta_n(A'_{1:n-1}(b')) = 0\}$ we have that

$$\mathbb{P}\left(R(b) \in \cdot \mid \mathcal{F}_{n-1}(x_{b'})\right) = \mathbb{P}\left(A'(b) \in \cdot \mid \mathcal{F}_{n-1}(x_{b'})\right)$$

and on the complementary event $E^c := \{\delta_n(A'_{1:n-1}(b')) > 0\}$ we have that R_n is an instance of (ϵ_n, δ_n)-randomized response given $A'_{1:n-1}$ (or, equivalently in terms of filtrations, given $\mathcal{F}_{n-1}$). It is straightforward to check that that there exist (random) transition kernels $(\rho_n)_{n\geq 1}$ such that

$$\mathbb{P}\left(A'_n(b) \in \cdot \mid \mathcal{F}_n(x_{b'})\right) = \rho^{(n)}_{b,b'}\left(\cdot \mid A'_{1:n-1}(b')\right) \quad \text{almost surely,} \tag{8}$$

where $\rho^{(n)}_{b,b'} = \mathbb{E}\left(\rho_n(\cdot, R_n(b) \mid A'_{1:n-1}(b')) \mid A'_{1:n-1}(b')\right)$, the averaged conditional probability measure as defined in Footnote 4. On the event E^c we can take $\rho_n = \nu_n$, where ν_n is as outlined in the proof of Theorem 1 in Appendix D. On E we can just take $\rho_n\left(\cdot, R_n(b)|A'_{1:n-1}(b')\right) = \delta_{R_n(b)}$, where δ_y indicates the measure putting a point mass on $y \in \mathcal{Y}$. If follows from Equation 8 that we can find an underlying probability space such that, for all $n \geq 1$, $A'_n(b)$ is equal to the random draws from $\rho_n(\cdot, R_n(b) \mid A'_{1:n-1}(b'))$ almost surely. This, combined with the observation that the sequence of algorithms $(R_n)_{n\geq 1}$ satisfies the assumptions of Lemma 6, allow the remainder of the proof to follow verbatim from the proof of Theorem 1. ■

Now we consider privacy filter for (conditional) approximate zCDP. First, we will introduce the following definition of (approximate) Renyi divergence.

Definition 9 ((Approximate) Renyi Divergence (Papernot and Steinke, 2022)) *Let P and Q be probability distributions over the same space. Let $\lambda \in [1, \infty]$. Assume that P is absolutely continuous w.r.t. Q. Let $P(y)$ and $Q(y)$ denote the densities of P and Q respectively. The Renyi divergence* *from P to Q of order λ is defined as*

$$D_\lambda(P\|Q) := \frac{1}{\lambda-1}\log\left(\mathbb{E}_{Y\sim P}\left[\left(\frac{P(Y)}{Q(Y)}\right)^{\lambda-1}\right]\right)$$

Let $\delta\in(0,1)$. Then the δ-approximate Renyi divergence from P to Q of order λ is defined as

$$D_\lambda^\delta(P\|Q) = \inf\left\{D_\lambda(P'\|Q') : P = (1-\delta)P' + \delta P'', Q = (1-\delta)Q' + \delta Q''\right\},$$

where $P = (1-\delta)P' + \delta P''$ denotes the fact that P can be expressed as a convex combination of two distributions P' and P'' with weights $1-\delta$ and δ respectively, and D_λ denotes the Renyi divergence of order λ.

Definition 10 (Conditional approximate zCDP) *Supppose A and B are algorithms with inputs in space $\mathcal{X}$ and outputs in measurable spaces $(\mathcal{Y},\mathcal{G})$ and $(\mathcal{Z},\mathcal{H})$. Suppose $\delta,\rho : \mathcal{Z}\to\mathbb{R}_{\geq 0}$ are measurable. We say the algorithm A is δ-approximate ρ-zCDP conditioned on B if, for any neighboring datasets $x\sim x'$ and all $\lambda\geq 1$,*

$$D_\lambda^{\delta(B(x))}(A(x)|B(x)\|A(x')|B(x)) \leq \rho(B(x))\lambda.$$

For succinctness, we will write $\rho(x)$ for $\rho(B(x))$ and $\delta(x)$ for $\delta(B(x))$.

Theorem 7 *Let $(A_n)_{n\geq 1}$ be an adaptive sequence of algorithms, and, for any x, let $(\mathcal{F}_n(x))_{n\in\mathbb{N}}$ be the natural filtration generated by $(A_n(x))_{n\in\mathbb{N}}$. For any $n\geq 1$, assume there are functions $\rho_n,\delta_n : \mathcal{Y}^{n-1}\times\mathcal{Z}^{n-1}\to\mathbb{R}_{\geq 0}$. For any $n\geq 1$, assume that A_n is δ_n-approximate ρ_n-zCDP conditioned on $A_{1:(n-1)}$. Consider the function $N\colon\mathbb{R}_{\geq 0}^\infty\times\mathbb{R}_{\geq 0}^\infty\to\mathbb{N}$ given by*

$$N((\rho_n)_{n\geq 1},(\delta_n)_{n\geq 1}) = \inf\left\{n\colon : \rho < \sum_{m\leq n+1}\rho_m \quad\text{or}\quad \delta < \sum_{m\leq n+1}\delta_m\right\}$$

Then the algorithm $A_{1:N(\cdot)}(\cdot)\colon\mathcal{X}\to\mathcal{Y}$ is δ-approximate ρ-zCDP, where $N(x) := N((\rho_n(x))_{n\geq 1},(\delta_n(x))_{n\geq 1})$ is the privacy filter.

Proof Let x and x' be a pair of neighboring datasets. For any n, let P_n and Q_n denote the densities of $A_{1:n}(x)$ and $A_{1:n}(x')$ respectively. By our assumption of conditional approximate zCDP at each step n, we can write P_n and Q_n as the following convex combinations:

$$P_n(A_n(x)\mid A_{1:n-1}(x)) = (1-\delta_n(x))P_n'(A_n(x)\mid A_{1:n-1}(x)) + \delta_n(x)P_n''(A_n(x)\mid A_{1:n-1}(x))$$
$$Q_n(A_n(x)\mid A_{1:n-1}(x)) = (1-\delta_n(x))Q_n'(A_n(x)\mid A_{1:n-1}(x)) + \delta_n(x)Q_n''(A_n(x)\mid A_{1:n-1}(x))$$

where the distributions P_n' and Q_n' satisfy: for all $\lambda\geq 1$,

$$D_\lambda\left(P_n'(\cdot\mid A_{1:n-1}(x))\|Q_n'(\cdot\mid A_{1:n-1}(x))\right)\leq \rho_n(x)\lambda.$$

Now we consider the Renyi divergence between the product measures $\prod_n P_n'$ and $\prod_n Q_n'$:

$$D_\lambda\left(\prod_n P_n'\|\prod_n Q_n'\right) = \frac{1}{\lambda-1}\log\left(\mathbb{E}_{A_{1:N(x)}(x)}\left[\left(\frac{\prod_{n\leq N(x)}P_n'(A_n(x)\mid A_{1:n-1}(x))}{\prod_{n\leq N(x)}Q_n'(A_n(x)\mid A_{1:n-1}(x))}\right)^{\lambda-1}\right]\right) \tag{9}$$

Note that by the law of iterated expectation, we have

$$\log\left(\mathbb{E}_{A_{1:N(x)}(x)}\left[\left(\frac{\prod_{n\le N(x)} P'_n(A_x \mid A_{1:n-1}(x))}{\prod_{n\le N(x)} Q'_n(A_x \mid A_{1:n-1}(x))}\right)^{\lambda-1}\right]\right) \tag{10}$$

$$= \log \prod_{n\le N(x)} \mathbb{E}_{A_n(x)\mid A_{1:n-1}(x)}\left[\left(\frac{P'_n(A_n(x) \mid A_{1:n-1}(x))}{Q'_n(A_n(x) \mid A_{1:n-1}(x))}\right)^{\lambda-1}\right] \tag{11}$$

$$= \sum_{n\le N(x)} \log\left(\mathbb{E}_{A_n(x)\mid A_{1:n-1}(x)}\left[\left(\frac{P'_n(A_n(x) \mid A_{1:n-1}(x))}{Q'_n(A_n(x) \mid A_{1:n-1}(x))}\right)^{\lambda-1}\right]\right) \tag{12}$$

As a result, we have

$$D_\lambda\left(\prod_n P'_n \| \prod_n Q'_n\right) = \sum_{n\le N(x)} D_\lambda(P'_n(\cdot \mid A_{1:n-1}(x)) \| Q'_n(\cdot \mid A_{1:n-1}(x))) \tag{13}$$

$$\le \sum_{n\le N(x)} \rho_n(x)\lambda \le \rho\lambda. \tag{14}$$

Finally, let P and Q denote the densities of $A(x)$ and $A(x')$ respectively. It suffices to show that P can be decomposed as a weighted mixture containing $\prod_n P'_n$, and Q can also be written as a weighted mixture containing $\prod_n Q'_n$. Moreover, the weights on $\prod_n P'_n$ and $\prod_n Q'_n$ are $1-\delta$. To see this, we have

$$P(A(x)) = \prod_{n\le N(x)} P_n(A_n(x) \mid A_{1:n-1}(x)) \tag{15}$$

$$= \prod_{n\le N(x)} \left[(1-\delta_n(x))P'_n(A_n(x) \mid A_{1:n-1}(x)) + \delta_n(x)P''_n(A_n(x) \mid A_{1:n-1}(x))\right] \tag{16}$$

$$= \prod_{n\le N(x)} (1-\delta_n(x)) \prod_{n\le N(x)} P'_n(A_n(x) \mid A_{1:n-1}(x)) + \underbrace{R(A(x))}_{\text{remaining probability}} \tag{17}$$

Note that $\prod_{n\le N(x)}(1-\delta_n(x)) \ge 1 - \sum_{n\le N(x)} \delta_n(x) \ge 1-\delta$. This allows us to further rewrite the convex combination as:

$$P(A(x)) = (1-\delta) \prod_{n\le N(x)} P'_n(A_n(x) \mid A_{1:n-1}(x))$$
$$+ \left[\underbrace{\left(\prod_{n\le N(x)} (1-\delta_n(x)) - (1-\delta)\right)}_{\ge 0} \prod_{n\le N(x)} P'_n(A_n(x) \mid A_{1:n-1}(x)) + R(A(x))\right]$$

We can rewrite decompose Q in a similar fashion. This completes the proof. ■