

Rényi Differential Privacy for Heavy-Tailed SDEs via Fractional Poincaré Inequalities

Benjamin Dupuis

*Inria, CNRS, Ecole Normale Supérieure
PSL Research University, Paris, France*

BENJAMIN.DUPUIS@INRIA.FR

Mert Gürbüzbalaban

*Department of Management Science and Information Systems
Rutgers Business School
Piscataway, NJ 08854, United States of America*

MG1366@RUTGERS.EDU

Umut Simsekli

*Inria, CNRS, Ecole Normale Supérieure
PSL Research University, Paris, France*

UMUT.SIMSEKLI@INRIA.FR

Jian Wang

*School of Mathematics and Statistics
Key Laboratory of Analytical Mathematics and Applications (Ministry of Education)
Fujian Provincial Key Laboratory of Statistics and Artificial Intelligence
Fujian Normal University
350007 Fuzhou, People's Republic of China*

JIANWANG@FJNU.EDU.CN

Sinan Yıldırım

*Faculty of Engineering and Natural Sciences
Sabancı University, Istanbul, Turkey*

SINANYILDIRIM@SABANCIUNIV.EDU

Lingjiong Zhu

*Department of Mathematics
Florida State University
Tallahassee, FL 32306, United States of America*

ZHU@MATH.FSU.EDU

Abstract

Characterizing the differential privacy (DP) of learning algorithms has become a major challenge in recent years. In parallel, many studies suggested investigating the behavior of stochastic gradient descent (SGD) with heavy-tailed noise, both as a model for modern deep learning models and to improve their performance. However, most DP bounds focus on light-tailed noise, where satisfactory guarantees have been obtained but the proposed techniques do not directly extend to the heavy-tailed setting. Recently, the first DP guarantees for heavy-tailed SGD were obtained. These results provide $(0, \delta)$ -DP guarantees without requiring gradient clipping. Despite casting new light on the link between DP and heavy-tailed algorithms, these results have a strong dependence on the number of parameters and cannot be extended to other DP notions like the well-established *Rényi differential privacy* (RDP). In this work, we propose to address these limitations by deriving the first RDP guarantees for heavy-tailed SDEs, as well as their discretized counterparts. Our framework is based on new Rényi flow computations and the use of well-established fractional Poincaré inequalities. Under the assumption that such inequalities are satisfied, we obtain DP guarantees that have a much weaker dependence on the dimension compared to prior art.

Keywords: Lévy-driven SDEs, Differential Privacy, Fractional Poincaré Inequalities

1 Introduction

Setup. Let $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$ be a data space endowed with a σ -algebra \mathcal{F} and a probability distribution μ_z , where \mathcal{X} denotes the space of input data and \mathcal{Y} denotes the space of output data. In the context of supervised learning, stochastic optimization algorithms are used to solve the following *empirical risk minimization* (ERM) problem

$$\min \left\{ \widehat{\mathcal{R}}_S(w) := \frac{1}{n} \sum_{i=1}^n \ell(w, z_i) : w \in \mathbb{R}^d \right\}, \quad (1)$$

where $\ell : \mathbb{R}^d \times \mathcal{Z} \rightarrow \mathbb{R}_+$ is a loss function and $S := (z_1, \dots, z_n) \sim \mu_z^{\otimes n}$ is a dataset sampled from the data distribution. With these notations, stochastic learning algorithms for solving the ERM problem can be seen as randomized mappings $\mathcal{A} : S \mapsto W_S \in \mathbb{R}^d$.

Due to the rapid development of machine learning, it has become essential to certify the *differential privacy* (DP) of learning algorithms. Informally, this notion can be understood as the question: given two *neighbouring* datasets $S, S' \in \mathcal{Z}^n$ (*i.e.*, differing by only one data point), how hard is it to distinguish the probability distributions of W_S and $W_{S'}$, *i.e.*, can we identify whether one particular data point has been used during training. A related question then is the creation of algorithms with provable DP guarantees. Several formalizations of DP have been proposed (Dwork, 2006; Dwork and Roth, 2013; Dwork and Rothblum, 2016). In our paper, we are in particular interested in the *Rényi differential privacy* (RDP) (Mironov, 2017), presented as a natural relaxation of classical DP conditions. More precisely, a learning algorithm $\mathcal{A} : S \mapsto W_S$ is said to be (β, κ) -RDP when $R_\beta(\text{Law}(W_S), \text{Law}(W_{S'})) \leq \kappa$, for neighboring datasets S and S' , where $R_\beta(\cdot, \cdot)$ denotes the Rényi divergence for $\beta > 1$. We provide in Section 2.1 a detailed presentation of these notions.

In our study, we aim to prove RDP guarantees for heavy-tailed SDEs. More precisely, given $S \in \mathcal{Z}^n$ and a time horizon $T > 0$, the learning algorithm is given by $\mathcal{A} : S \mapsto W_T$, where $(W_t)_{t \geq 0}$ is a solution of the following stochastic differential equation (SDE) in \mathbb{R}^d

$$dW_t = -\nabla \widehat{\mathcal{R}}_S(W_t)dt + \sigma_\alpha dL_t^\alpha + \sigma_2 \sqrt{2} dB_t, \quad (2)$$

where $\sigma_\alpha, \sigma_2 \geq 0$ are fixed, $(L_t^\alpha)_{t \geq 0}$ is a rotationally invariant α -stable Lévy process with tail index $\alpha \in (0, 2)$ (see Section 2.2), and $(B_t)_{t \geq 0}$ is a standard Brownian motion. We also consider Euler-Maruyama discretizations of Equation (2), which corresponds to heavy-tailed stochastic gradient descent (SGD), as studied by Şimşekli et al. (2024) in the case $\sigma_2 = 0$.

These heavy-tailed algorithms have recently gained a lot of attention in learning theory. First, it was shown in several studies, both empirically and theoretically, that SGD can produce heavy-tailed distributions under certain choices of hyperparameters (Simsekli et al., 2019; Gürbüzbalaban et al., 2021). Following these findings, it was shown that the presence or injection of heavy-tailed noise can ensure good generalization properties (Raj et al., 2023a,b; Dupuis and Şimşekli, 2024; Lim et al., 2022), and improve the compressibility of the model (Barsbey et al., 2021; Wan et al., 2023). In addition to these properties, it has been shown that heavy-tailed noise can lead to satisfactory minimization of the empirical risk in convex and non-convex settings (Wang et al., 2021; Simsekli et al., 2019), and in the

Paper	Type of DP	Assumptions	Guarantee
Abadi et al. (2016, Thm. 1)	(ε, δ) -DP	S.	$\varepsilon \simeq \mathcal{O}\left(\frac{b\sqrt{T \log(1/\delta)}}{n\sigma}\right)$
Asoodeh and Diaz (2023, Thm. 2)	(ε, δ) -DP	S.	$\varepsilon \simeq \mathcal{O}\left(\frac{1}{\sigma^2}\right)$ for $\delta \simeq \mathcal{O}\left(\frac{b}{n}\right)$
Chourasia et al. (2021, Thm. 2)	(β, κ) -RDP	S., LSI.	$\kappa \simeq \mathcal{O}\left(\frac{\beta S_g^2}{n^2 \sigma^2}\right)$
Ryffel et al. (2022, Thm 3.1)	(β, κ) -RDP	L., LSI.	$\kappa \simeq \mathcal{O}\left(\frac{\beta L^2}{n^2 \sigma^2}\right)$
Altschuler and Talwar (2022, Thm. 1.3)	(β, κ) -RDP	C., L., ∇ -L.	$\kappa \simeq \mathcal{O}\left(\frac{\beta L^2}{n^2 \sigma^2} \min(T, n)\right)$
Ye and Shokri (2022, Thm. 3.3)	(β, κ) -RDP	SC., S., ∇ -L.	$\kappa \simeq \mathcal{O}\left(\frac{\beta S_g^2}{\sigma^2 b^2}\right)$
Chien et al. (2024, Thm. 3.3)	(β, κ) -RDP	L., ∇ -L., LSI.	$\kappa \simeq \mathcal{O}\left(\frac{\beta T}{\sigma^2 n^2}\right)$
Chien et al. (2024, Thm. 3.3)	(β, κ) -RDP	L., ∇ -L., SC.	$\kappa \simeq \mathcal{O}\left(\frac{\beta}{\sigma^2 n^2}\right)$

Table 1: Comparison of existing guarantees for DP-SGD with Gaussian noise. Simplified versions are presented; we refer to the papers for details. *Abbreviations:* S. (finite sensitivity / clipping), LSI. (log-Sobolev inequality), ∇ -L. (gradient Lipschitz), C. (convex), SC. (strongly-convex), L. (L -Lipschitz). *notations:* T (number of iterations / time), \mathcal{S}_g (gradient sensitivity), $b \leq n$ (batch size).

context of one-hidden-layer neural networks (Wan et al., 2023). These observations support the interest in obtaining DP guarantees for heavy-tailed dynamics such as Equation (2).

Related works. Many differentially private mechanisms have been studied with Gaussian (Wang et al., 2017; Feldman et al., 2018) or Laplace noise (Chaudhuri et al., 2011; Kuru et al., 2022). The main inspiration for our work is the well-established differentially private SGD (DP-SGD) mechanism (Bassily et al., 2014), which involves adding (Gaussian) noise in SGD iterations, along with potential projections and gradient clipping operations (Abadi et al., 2016; Yu et al., 2019; Chen et al., 2020). In our study, this model corresponds to a discretized version of Equation (2) with $\sigma_\alpha = 0$, with projection or clipping of the gradient on a ball of finite radius, which we could also include in our model.

In particular, Chourasia et al. (2021) studied the following noisy SGD recursion

$$W_{k+1} = \Pi_C \left(W_k - \eta \nabla \widehat{\mathcal{R}}_S(W_k) + \sigma_2 \sqrt{2\eta} \xi_k \right), \quad \xi_k \stackrel{i.i.d.}{\sim} \mathcal{N}(0, I_d), \quad (3)$$

where Π_C denotes the projection on a convex set. They obtained a (β, κ) -RDP guarantee with $\kappa = \mathcal{O}(\beta S_g^2/n^2)$, where S_g is the gradient sensitivity, formally defined in Section 3.1. Similar RDP and DP guarantees have been derived by other authors for convex and non-convex losses, using a wide range of tools (Asoodeh and Diaz, 2023; Altschuler and Talwar, 2022). A large part of these works make crucial use of the *logarithmic Sobolev inequality* (LSI) (Gross, 1975) to derive RDP guarantees by exploiting the mixing properties of the underlying processes (Ye and Shokri, 2022; Chien et al., 2024; Ganesh and Talwar, 2020; Ryffel et al., 2022). These works identify assumptions on the loss such that the LSI can be satisfied, which they exploit to make their bounds time-uniform. We summarize some of the aforementioned results in Table 1.

Unfortunately, these works cannot be directly extended to the heavy-tailed case (*i.e.*, when $\sigma_\alpha > 0$ in Equation (2)) for two main reasons. First, the proposed derivations make explicit use of the Gaussian structure of the noise. For example, Chourasia et al. (2021) exploit the Fokker-Planck equations associated with continuous-time interpolations of Equation (3). Second, it has been noted by Dupuis and Şimşekli (2024) that the LSI that is instrumental to the Gaussian case might not be available in the heavy-tailed setting. So far, these issues have prevented the obtainment of RDP guarantees for heavy-tailed SDEs.

Recent studies derived DP guarantees under heavy-tailed noise. Ito et al. (2021) investigated the DP properties of linear dynamical systems under α -stable noise, while Zawacki and Abed (2025) proposed a general α -stable DP mechanism, which is, however, not directly related to SGD and, therefore, independent of our study. In the case of SGD, Asi et al. (2024) studied the DP of stochastic convex optimization under the assumption of heavy-tailed gradients, in the sense of finite moments up to order k . However, this study does not directly contain the case of α -stable noise, which is the focus of our paper. More recently, Şimşekli et al. (2024) investigated the following heavy-tailed noisy SGD recursion in \mathbb{R}^d

$$W_{k+1} = W_k - \eta \nabla \widehat{\mathcal{R}}_S(W_k) + \sigma_\alpha \eta^{1/\alpha} \xi_k, \quad \xi_k \stackrel{i.i.d.}{\sim} \mathcal{S}\alpha\mathcal{S} := \text{Law}(L_1^\alpha), \quad (4)$$

which is a discretization of Equation (2) with $\sigma_2 = 0$. These authors obtained $(0, \delta)$ -DP guarantees (see Definition 1) with $\delta = \mathcal{O}(d^{(1+\alpha)/2}/n)$ and without requiring gradient clipping. Their study is based on a Markov chain perturbation analysis and, most importantly, can handle unbounded gradients, while many works rely on bounded gradients or finite sensitivity assumptions (Chourasia et al., 2021). Despite successfully obtaining the first DP guarantees for heavy-tailed SGD, this work has three main disadvantages: (i) $(0, \delta)$ -DP is weaker than RDP, as we recall in Lemma 4, (ii) the proposed bound has a strong dependence on the dimension, which might render it vacuous in practice, and (iii) it has an intricate and not explicit dependence on the noise scale σ_α , making its impact on the bound unclear.

1.1 Contributions

In our work, we propose to address the issues mentioned above by proving RDP guarantees for heavy-tailed SDEs and their discrete counterparts. Our framework is based on the extension of the entropy flow computations of Chourasia et al. (2021) in the heavy-tailed setting. In contrast with the relative entropy flow derivations obtained by Dupuis and Şimşekli (2024) in the context of generalization bounds, we obtain *Rényi divergence flows* (*i.e.*, the time derivative of the Rényi divergence) for SDEs driven by general Lévy processes. In order to circumvent the lack of LSI in this heavy-tailed case, we show that we can obtain time-uniform RDP guarantees based on the fractional versions of the celebrated Poincaré inequalities (Wang and Wang, 2015; Mouhot et al., 2011). Based on this method, we obtained the first RDP guarantees for heavy-tailed (*i.e.*, α -stable) SGD. Our detailed contributions are summarized below and in Table 2.

- We provide a general framework to compute the flow of Rényi divergences along Lévy-driven SDEs and study the associated functional inequalities. Our proof technique applies to a very general class of Lévy processes, which makes it more general than existing works on heavy-tailed SGD.

Paper	Type of DP	Assumptions	Guarantee
Şimşekli et al. (2024, Thm. 10)	$(0, \delta)$ -DP	Ps.- ∇ -L., D.	$\delta \simeq \mathcal{O}\left(\frac{d^{(1+\alpha)/2}}{n}\right)$
Ours (Theorem 15)	(β, κ) -RDP	S., $\sigma_2 > 0$	$\kappa \simeq \mathcal{O}\left(\frac{\beta S_g^2 T}{n^2 \sigma_2^2}\right)$
Ours (Theorem 15)	(β, κ) -RDP	S., FPI, $\sigma_2 > 0$	$\kappa \simeq \mathcal{O}\left(\frac{\beta^2 S_g^2}{n^2 \sigma_2^2}\right)$
Ours (by Lemma 4)	$(0, \delta)$ -DP	S., FPI, $\sigma_2 > 0$	$\delta \simeq \mathcal{O}\left(\frac{S_g}{n \sigma_2}\right)$
Ours (Theorems 16 and 18)	(β, κ) -RDP	S., $\sigma_2 = 0$	$\kappa \simeq \mathcal{O}\left(\frac{\beta d^{1-\alpha/2} S_g^2 T}{n^2 \sigma_\alpha^\alpha R^{2-\alpha}}\right)$
Ours (Theorems 16 and 18)	(β, κ) -RDP	S., FPI, $\sigma_2 = 0$	$\kappa \simeq \mathcal{O}\left(\frac{\beta^2 d^{1-\alpha/2} S_g^2}{n^2 \sigma_\alpha^\alpha R^{2-\alpha}}\right)$
Ours (by Lemma 4)	$(0, \delta)$ -DP	S., FPI, $\sigma_2 = 0$	$\delta \simeq \mathcal{O}\left(\frac{d^{(2-\alpha)/4} S_g}{n \sigma_\alpha^{\alpha/2} R^{1-\alpha/2}}\right)$

Table 2: Comparison of guarantees for DP-SGD with α -stable noise. Simplified versions are presented, we refer to the papers for details. *Abbreviations:* S. (finite sensitivity / clipping), FPI. (fractional Poincaré inequality), D. (dissipativity), Ps.- ∇ -L. (pseudo gradient Lipschitz). *notations:* S_g (gradient sensitivity), R (quantity appearing in Theorems 16 and 18, which might depend on (d, T, β)).

- In the multifractal case (*i.e.*, when both $\sigma_\alpha > 0$ and $\sigma_2 > 0$), we obtain dimension-independent (β, κ) -RDP guarantees with $\kappa = \mathcal{O}(\beta^2/(n^2 \sigma_2^2))$, under a fractional Poincaré inequality (FPI) assumption. We also obtain a guarantee in $\kappa = \mathcal{O}(\beta T/(n^2 \sigma_2^2))$ when this assumption is removed.
- In the pure-jump α -stable case (*i.e.*, when $\sigma_2 = 0$), we draw a new link with Bourgain-Brezis-Mironescu's type formulas (Bourgain et al., 2001) and show that, under slightly stronger assumptions, we can obtain (β, κ) -RDP guarantees with $\kappa = \mathcal{O}(\beta^2 d^{1-\alpha/2}/(n^2 \sigma_\alpha^\alpha))$ under FPI and $\kappa = \mathcal{O}(\beta d^{1-\alpha/2} T/(n^2 \sigma_\alpha^\alpha))$ without FPI. Our results imply $(0, \delta)$ -DP guarantees which have much weaker dependence on the dimension d than in existing works in both the multifractal and the pure-jump case.
- Finally, we extend our RDP to the discrete-time setting, which is used in practical application. Moreover, as a sanity check, we investigate the satisfiability of fractional Poincaré inequalities in this case, hence providing theoretical foundation for our main assumptions. As a by-product of our analysis, we derive new stability results for fractional Poincaré inequalities.

Organization of the paper. We present some technical background on differential privacy, Lévy processes, and fractional Poincaré inequalities in Section 2. Our main results are discussed in Section 3, where we detail our setup, assumptions, the Rényi flow computations, and our DP guarantees in the multifractal and the pure-jump cases. Finally, Section 4 is dedicated to the analysis of the discrete-time algorithm and the stability properties of fractional Poincaré inequalities. All omitted proofs are postponed to the appendix.

Notations. For all $x \in \mathbb{R}^d$, we denote by $\|x\|$ the Euclidean norm of x . For a matrix $A \in \mathbb{R}^{d \times d}$, the notation $\|A\|_2$ refers to the spectral norm. For a matrix mapping $A : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$,

the notation $\|A\|_\infty$ is also understood with respect to the spectral norm. The open ball centered at x with radius $r > 0$ is denoted by $B_r(x)$. For any random variable X , the law of X is written as $\text{Law}(X)$. If μ is a probability measure and T is a measurable map, we write $T_\# \mu$ for the pushforward measure. We abbreviate the partial derivatives $\partial/\partial x$ by ∂_x . Let $\mathcal{C}_b^2(\mathbb{R}^d)$ denote the set of twice continuously differentiable functions $u : \mathbb{R}^d \rightarrow \mathbb{R}$ which are bounded, along with their first- and second-order derivatives. For $f \in \mathcal{C}_b^2(\mathbb{R}^d)$, we denote $\|f\|_{\mathcal{C}^2} := \max_{|\gamma| \leq 2} \|\partial^\gamma f\|_\infty$, where $\partial^\gamma f := \partial_{x_1}^{\gamma_1} \dots \partial_{x_d}^{\gamma_d}$ and $|\gamma| := \gamma_1 + \dots + \gamma_d$ for any $\gamma \in \mathbb{N}^d$. The Laplacian operator in \mathbb{R}^d is denoted as $\Delta := \partial_{x_1}^2 + \dots + \partial_{x_d}^2$.

2 Preliminaries

In this section, we introduce some necessary technical background. We start by discussing the classical notions of differential privacy and Rényi differential privacy in Section 2.1, and then present Lévy processes in Section 2.2. Finally, we provide an introduction to fractional Poincaré inequalities in Section 2.3.

2.1 Background on differential privacy

Let \mathcal{Z} be a data space endowed with a σ -algebra \mathcal{F} . Given two datasets $S, S' \in \mathcal{Z}^n$ for some $n \geq 0$, we say that they are *neighbors*, and we denote $S \simeq S'$, if both datasets differ by only one element. In this paper, we define a learning algorithm as a mapping $\mathcal{A} : \bigcup_{n=0}^{\infty} \mathcal{Z}^n \rightarrow \mathcal{P}(\mathbb{R}^d)$, $S \mapsto \mathcal{A}(S)$, where $\mathcal{P}(\mathbb{R}^d)$ denotes the set of Borel probability measures on \mathbb{R}^d . We will refer to the distribution $\mathcal{A}(S)$, for $S \in \bigcup_{n=0}^{\infty} \mathcal{Z}^n$, as a *posterior distribution*.

The classical notion of *Differential Privacy* (DP), as introduced in Dwork and Roth (2013); Dwork et al. (2010), regards the possibility of distinguishing the posterior distributions induced by two neighboring datasets.

Definition 1 ((ε, δ)-DP) *Let $\varepsilon \geq 0$ and $\delta \in [0, 1]$. The learning algorithm \mathcal{A} is said to satisfy (ε, δ) -DP if, for any $S \simeq S'$ and any Borel set B in \mathbb{R}^d , $\mathcal{A}(S)(B) \leq e^\varepsilon \mathcal{A}(S')(B) + \delta$.*

In Mironov (2017), the author defined a notion of DP based on the Rényi divergence between neighboring posterior distributions.

Definition 2 ((β, κ)-RDP) *Let $\kappa > 0$ and $\beta > 1$. The learning algorithm \mathcal{A} is said to satisfy (β, κ) -Rényi differential privacy (RDP) if, for any $S \simeq S'$, $R_\beta(\mathcal{A}(S), \mathcal{A}(S')) \leq \kappa$, where, given two Borel probability measures $\mathbb{Q} \ll \mathbb{P}$, we define their Rényi divergence by*

$$R_\beta(\mathbb{Q}, \mathbb{P}) := \frac{1}{\beta - 1} \log E_\beta(\mathbb{Q}, \mathbb{P}) \quad \text{with } E_\beta(\mathbb{Q}, \mathbb{P}) := \int \left(\frac{d\mathbb{Q}}{d\mathbb{P}} \right)^\beta d\mathbb{P}. \quad (5)$$

By convention, we set $R_\beta(\mathbb{Q}, \mathbb{P}) := +\infty$ when $\mathbb{Q} \not\ll \mathbb{P}$. Note that we also define $R_1(\mathbb{Q}, \mathbb{P}) := \text{KL}(\mathbb{Q} || \mathbb{P}) := \int \log(d\mathbb{Q}/d\mathbb{P}) d\mathbb{Q}$, which is the Kullback-Leibler divergence between \mathbb{Q} and \mathbb{P} . It can be seen that $\beta \mapsto R_\beta(\cdot, \cdot)$ is an increasing and continuous map for $\beta \geq 1$ (van Erven and Harremoës, 2014). The particular interest of the above notion comes from the fact that RDP implies DP.

Lemma 3 For any $\kappa > 0$, $\beta > 1$, and $\delta \in (0, 1]$, (β, κ) -RDP implies (ε, δ) -DP, with $\varepsilon = \kappa + \frac{\log(1/\delta)}{\beta-1}$.

The proof of this result is elementary and can be found in Abadi et al. (2016); Mironov (2017). This result is also mentioned in Asoodeh et al. (2021), who also prove tighter conversion results. In the case where $\varepsilon = 0$, we also have the following conversion lemma.

Lemma 4 Let $\kappa > 0$ and $\beta > 1$. Then, (β, κ) -RDP implies $(0, \sqrt{\kappa/2})$ -DP.

Proof Let $S \simeq S'$ in $\bigcup_{n=0}^{\infty} \mathcal{Z}^n$. By (van Erven and Harremoës, 2014, Theorem 3), the Rényi divergence $R_\beta(Q, P)$ is non-decreasing in β , and thus,

$$KL(\mathcal{A}(S)||\mathcal{A}(S')) \leq R_\beta(\mathcal{A}(S), \mathcal{A}(S')) \leq \kappa.$$

Therefore, by Pinsker's inequality (van Erven and Harremoës, 2014, Theorem 31), we obtain that $TV(\mathcal{A}(S), \mathcal{A}(S')) \leq \sqrt{\kappa/2}$. \blacksquare

We also define the following Rényi information, which plays a crucial role in Chourasia et al. (2021), as well as in the proofs of our main results. We follow the convention of Chourasia et al. (2021) for the normalization of this quantity.

Definition 5 Let Q and P be two probability measures on Ω such that $Q \ll P$ and $\beta > 1$. Assume that the Radon-Nikodym derivative dQ/dP is differentiable. Then, the Rényi information of Q with respect to P is defined as

$$I_\beta(Q, P) := \int \left(\frac{dQ}{dP} \right)^{\beta-2} \left\| \nabla \frac{dQ}{dP} \right\|^2 dP.$$

2.2 Lévy processes and infinitely divisible distributions

An interesting aspect of our approach based on Rényi flows is that the proof techniques extend to a very general class of Lévy processes. This fact is highlighted in Section B.1 in the appendix. In order to present it clearly, we give below some technical background on Lévy processes.

Let (Ω, \mathcal{F}, P) be a fixed probability space. A stochastic process $(L_t)_{t \geq 0}$ is called a Lévy process, if $L_0 = 0$ almost surely, and it satisfies the following properties.

- **Stationary increments:** For all $s \leq t$, $\text{Law}(L_t - L_s) = \text{Law}(L_{t-s})$.
- **Independent increments:** $L_t - L_s$ is independent from the σ -algebra $\sigma(L_u, u \leq s)$.
- **Stochastic continuity:** For all $\varepsilon > 0$, we have $\lim_{s \rightarrow t} \mathbb{P}(\|L_t - L_s\| > \varepsilon) = 0$.

Under these conditions, the distribution of L_t is *infinitely divisible* (Schilling, 2016), which means that its distribution can be written as an m -fold convolution for all $m \in \mathbb{N}^* := \{1, 2, \dots\}$. There is a one-to-one correspondence between Lévy processes and infinitely divisible distributions. Classically, this implies that the characteristic function of the Lévy process $(L_t)_{t \geq 0}$ can be expressed for all $\xi \in \mathbb{R}^d$ as $\mathbb{E}[\exp(i\xi^\top L_t)] = \exp(-t\psi(\xi))$, where

$i := \sqrt{-1}$ and ψ is called the characteristic exponent, which is given by the celebrated Lévy-Khintchine formula (Böttcher et al., 2013)

$$\psi(\xi) = -ib^\top \xi + \frac{1}{2}\xi^\top \Sigma \xi + \int_{\mathbb{R}^d \setminus \{0\}} \left(1 - e^{i\xi^\top z} + i\xi^\top z \chi(\|z\|)\right) d\nu(z), \quad \xi \in \mathbb{R}^d, \quad (6)$$

where $b \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ is a symmetric positive semi-definite matrix, χ satisfies¹ $\chi(s) := (1+s^2)^{-1}$, and ν is a positive Radon measure on \mathbb{R}^d such that $\int_{\mathbb{R}^d \setminus \{0\}} \min(1, \|z\|^2) d\nu(z) < +\infty$. The measure ν is called the *Lévy measure* and (b, Σ, ν) is the *Lévy triplet* of $(L_t)_{t \geq 0}$. For instance, the standard Brownian motion $(B_t)_{t \geq 0}$ in \mathbb{R}^d is a Lévy process with triplet $(0, I_d, 0)$. In our paper, we are in particular interested in the rotationally invariant α -stable Lévy process $(L_t^\alpha)_{t \geq 0}$, whose characteristic exponent is given by $\psi(\xi) = \|\xi\|^\alpha$, with $\alpha \in (0, 2]$. When $\alpha < 2$, its Lévy triplet is given by $(0, 0, \nu)$ (Böttcher et al., 2013, Example 2.4.d), with

$$d\nu(z) := C_{\alpha,d} \frac{dz}{\|z\|^{\alpha+d}} \quad \text{and} \quad C_{\alpha,d} := \alpha 2^{\alpha-1} \pi^{-d/2} \frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(1-\frac{\alpha}{2})}. \quad (7)$$

Lévy processes are characterized by their infinitesimal generator. A complete understanding of this notion is not absolutely necessary to understand our paper; however, the reader may consult Schilling (2016); Böttcher et al. (2013) for additional details on this topic. For a Lévy process with triplet $(0, \Sigma, \nu)$, the infinitesimal generator is given by

$$Au(x) := \frac{1}{2} \nabla \cdot (\Sigma \nabla u(x)) + \int_{\mathbb{R}^d \setminus \{0\}} (u(x+z) - u(x) - \nabla u(x) \cdot z \chi(\|z\|)) d\nu(z) \quad (8)$$

for any $u \in \mathcal{C}_b^2(\mathbb{R}^d)$. In the case of the α -stable Lévy process, it is well-known that this definition amounts to the fractional Laplacian, for which we give a definition below. We refer to Daoud and Laamri (2022); Nezza et al. (2012) for the other equivalent definitions of the fractional Laplacian.

Definition 6 (Fractional Laplacian) For $\alpha \in (0, 2)$, the fractional Laplacian of $u \in \mathcal{C}_b^2(\mathbb{R}^d)$ is given by $-(-\Delta)^{\frac{\alpha}{2}} u(x) := C_{\alpha,d} \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^d \setminus B_\varepsilon(0)} \frac{u(x+z) - u(x)}{\|z\|^{d+\alpha}} dz$. We can show that this definition is equivalent to Equation (8) in the case where $\Sigma = 0$ and ν is given by Equation (7), where it corresponds to the infinitesimal generator of $(L_t^\alpha)_{t \geq 0}$.

Remark 7 The fractional Laplacian is sometimes defined without the constant $C_{\alpha,d}$ (Tristani, 2013). This corresponds to different normalizations and our choice is motivated by the fact that we want this operator to be the infinitesimal generator of $(L_t^\alpha)_{t \geq 0}$.

2.3 Fractional Poincaré inequalities

The analysis of differential privacy in the presence of Gaussian noise makes crucial use of the *logarithmic Sobolev inequalities* (LSIs) (Gross, 1975; Bakry et al., 2014), which are instrumental in making the proposed bounds uniform in time (Chourasia et al., 2021; Chien et al., 2024; Ryffel et al., 2022; Ye and Shokri, 2022). In such cases, assuming that an LSI

1. χ can be chosen arbitrary as soon as it satisfies certain properties, see (Böttcher et al., 2013).

is satisfied is motivated by the fact that it is satisfied by the posterior distributions of the learning algorithms under reasonable assumptions. Unfortunately, as was noted by Dupuis and Şimşekli (2024) in their study of the generalization error of heavy-tailed SDEs, such LSIs may not always be applicable in the presence of α -stable noise.

In our paper, we show that, in the context of heavy-tailed dynamics, LSIs can be replaced by *fractional Poincaré inequalities*. Such inequalities have been widely studied (Wang and Wang, 2015; Chafaï, 2004; Mouhot et al., 2011) and have been used in the context of machine learning for generalization bounds (Dupuis and Şimşekli, 2024) and sampling (He et al., 2024). In particular, we have the following result (Gentil and Imbert, 2009; Wu, 2000).

Theorem 8 (Fractional Poincaré inequalities) *Let μ be an infinitely divisible distribution on \mathbb{R}^d with Lévy triplet $(0, \Sigma, \nu)$ in the sense of Equation (6). Then, for any differentiable and μ -square-integrable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$,*

$$\int_{\mathbb{R}^d} f^2 d\mu - \left(\int_{\mathbb{R}^d} f d\mu \right)^2 \leq \int_{\mathbb{R}^d} \|\nabla f\|_\Sigma^2 d\mu + \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} (f(x) - f(x+z))^2 d\mu(x) d\nu(z),$$

where $\|u\|_\Sigma^2 := u^\top \Sigma u$ for any $u \in \mathbb{R}^d$. In particular, if μ is the law of L_1^α , then

$$\int_{\mathbb{R}^d} f^2 d\mu - \left(\int_{\mathbb{R}^d} f d\mu \right)^2 \leq C_{\alpha,d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} (f(x) - f(x+z))^2 d\mu(x) \frac{dz}{\|z\|^{d+\alpha}}.$$

In particular, when $\nu = 0$, the above result recovers the classical Poincaré inequalities for Gaussian distributions. By analogy with the generalizations of classical Poincaré inequalities from the Gaussian measures to more general ones, we propose the following definition.

Definition 9 (α -stable Poincaré inequalities) *Let μ be a probability measure and $\alpha \in (0, 2)$, we say that μ satisfies an α -stable Poincaré inequality with constants (a, b) , if for any differentiable and μ -square-integrable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$,*

$$\int_{\mathbb{R}^d} f^2 d\mu - \left(\int_{\mathbb{R}^d} f d\mu \right)^2 \leq a C_{\alpha,d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} \frac{(f(x) - f(x+z))^2}{\|z\|^{d+\alpha}} d\mu(x) dz + b \int_{\mathbb{R}^d} \|\nabla f\|^2 d\mu.$$

We use the unusual denomination α -stable Poincaré inequality instead of fractional Poincaré inequality because we allow the inequality to contain a Gaussian component, corresponding to the term $\int_{\mathbb{R}^d} \|\nabla f\|^2 d\mu$. This allows our analysis to apply to the case where the noise of the algorithm is a combination of Gaussian and α -stable noises. We prove in Section 4.2 several properties related to the stability of such inequalities under certain transformations, thus showing that these inequalities can be satisfied by a wide variety of probability distributions. For more general conditions regarding the validity of these inequalities, we refer to Wang and Wang (2015); Mouhot et al. (2011).

In order to simplify the notation, we will use the following notation for Dirichlet forms.

Definition 10 *For $\alpha \in (0, 2)$, define*

$$\mathcal{E}_{\alpha,\mu}(f, f) := \frac{1}{2} C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{(f(x) - f(y))^2}{\|x - y\|^{d+\alpha}} d\mu(x) dy, \quad f \in \mathcal{C}_b^2(\mathbb{R}^d);$$

and for $\alpha = 2$,

$$\mathcal{E}_{2,\mu}(f, f) := \int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x), \quad f \in \mathcal{C}_b^2(\mathbb{R}^d).$$

These quantities correspond, respectively, to the Dirichlet forms associated with $(L_t^\alpha)_{t \geq 0}$ and $(\sqrt{2}B_t)_{t \geq 0}$. While a complete knowledge of Dirichlet forms is not required for this paper, we invite the reader to consult Bakry et al. (2014); Böttcher et al. (2013) for more details.

3 Differential Privacy of Lévy-Driven SDEs

In this section, we present our main results, which include RDP guarantees for Lévy-driven SDEs. We first present our setup and key technical lemmas in Sections 3.1 and 3.2 respectively, before presenting two cases. First, we derive RDP guarantees under multifractal noise (*i.e.*, a combination of Gaussian and α -stable noise) in Section 3.3. We will then explain how our analysis can be extended to pure-jump α -stable noise in Section 3.4.

3.1 Setup and assumptions

Given a dataset $S \in \mathcal{Z}^n$ and a loss function $\ell : \mathbb{R}^d \times \mathcal{Z} \rightarrow \mathbb{R}_+$, we define the empirical risk as $\widehat{\mathcal{R}}_S(w) := n^{-1} \sum_{i=1}^n \ell(w, z_i)$. We also denote by μ_z the data distribution on $(\mathcal{Z}, \mathcal{F})$.

Let us consider two neighboring datasets $S \simeq S' \in \mathcal{Z}^n$. Given a fixed common initial distribution, we consider the two following stochastic differential equations (SDEs):

$$\begin{cases} dW_t = -\nabla \widehat{\mathcal{R}}_S(W_t)dt + \sigma_\alpha dL_t^\alpha + \sigma_2 \sqrt{2} dB_t, \\ dW'_t = -\nabla \widehat{\mathcal{R}}_{S'}(W'_t)dt + \sigma_\alpha dL_t^\alpha + \sigma_2 \sqrt{2} dB_t, \end{cases} \quad (9)$$

where $(B_t)_{t \geq 0}$ is a d -dimensional standard Brownian motion, $(L_t^\alpha)_{t \geq 0}$ is a d -dimensional rotationally invariant α -stable Lévy process with $\alpha \in (1, 2)$, and $(\sigma_2, \sigma_\alpha)$ are positive constants. Inspired by Chourasia et al. (2021), we define the gradient sensitivity as:

$$\mathcal{S}_g := \text{ess sup}_{(z, z') \sim \mu_z \otimes \mu_z} \sup_{w \in \mathbb{R}^d} \|\nabla \ell(w, z') - \nabla \ell(w, z)\|. \quad (10)$$

In some of our results, we make the following finite sensitivity assumption.

Assumption 1 (Finite sensitivity) *The gradient sensitivity is finite, *i.e.*, $\mathcal{S}_g < +\infty$.*

Remark 11 In some of our main results, we use Assumption 1 to uniformly control the difference between the two drifts in Equation (9). This assumption is typically satisfied when $\ell(w, z)$ is a regularized loss of the form $\ell(w, z) = \ell_0(w, z) + \lambda \|w\|^2$, where $\ell_0(\cdot, z)$ is Lipschitz-continuous, for all $z \in \mathcal{Z}$. Alternatively, one can impose this condition by clipping all the individual gradients $\nabla \ell(\cdot, z)$ outside a compact set. Similar conditions appear in the literature related to Gaussian noise (Chourasia et al., 2021; Chien et al., 2024; Ryffel et al., 2022; Ye and Shokri, 2022), where it is often associated with projections on a compact set. For Brownian motion–driven SDEs, reflection on convex sets is well-defined via the Skorokhod problem (Lions and Sznitman, 1984), which constrains the dynamics to a convex domain by means of a boundary regulator. For Lévy processes, however, jumps

can overshoot the boundary, so that the reflection is not canonical and typically induces non-local, state-dependent boundary conditions. Existing constructions of regulators are case specific and apply only in restricted settings (Menaldi, 1985; Costantini et al., 2005; Słomiński, 2010; Piera et al., 2008). For this reason, we consider projections only in the discrete-time case.

Remark 12 In our proofs, the quantities $\|\nabla\ell(w, z') - \nabla\ell(w, z)\|$ are integrated with respect to certain probability distributions (see Section B). This suggests that Assumption 1 could be replaced by weaker but more intricate conditions. We leave this discussion for future work and focus our main results on the finite sensitivity case.

We denote by p_t and p'_t the probability density functions of W_t and W'_t . Throughout the paper, we assume that both SDEs are initialized from the same distribution. It is known that, under mild regularity assumptions (Duan, 2015; Umarov et al., 2018), p_t and p'_t are solutions (at least in a weak sense) of the following fractional Fokker-Planck equations,

$$\begin{cases} \partial_t p_t = -\sigma_\alpha^\alpha (-\Delta)^{\frac{\alpha}{2}} p_t + \sigma_2^2 \Delta p_t + \nabla \cdot (p_t \nabla \hat{\mathcal{R}}_S), \\ \partial_t p'_t = -\sigma_\alpha^\alpha (-\Delta)^{\frac{\alpha}{2}} p'_t + \sigma_2^2 \Delta p'_t + \nabla \cdot (p'_t \nabla \hat{\mathcal{R}}_{S'}), \end{cases} \quad (11)$$

where $(-\Delta)^{\frac{\alpha}{2}}$ is the fractional Laplacian operator given in Definition 6.

Our analysis of the differential privacy of heavy-tailed SDEs relies on estimates of the so-called *Rényi flow*, i.e., the time-derivative of the Rényi divergence between p_t and p'_t . In order to make these derivations perfectly rigorous, we need to impose enough regularity on the probability density functions of W_t and W'_t , which is made precise as follows.

Assumption 2 (Regularity conditions) Let $v_t := p_t/p'_t$. We assume that p_t , p'_t and v_t are positive, differentiable in t , and belong to $\mathcal{C}_b^2(\mathbb{R}^d)$. Moreover, we make the following non-explosion assumption, i.e., the maps $t \mapsto \|v_t\|_\infty$ are locally bounded on \mathbb{R}_+ , and the functions $|\partial_t p_t|$ and $|\partial_t p'_t|$ are locally uniformly integrable on \mathbb{R}_+ .

We say that a family of functions $\{f(t, \cdot)\}_{t \geq 0}$ is locally uniformly integrable if for all $t_0 \geq 0$, there exists $\varepsilon > 0$ such that $\sup_{t_0 - \varepsilon < t < t_0 + \varepsilon} |f(t, \cdot)| \in L^1(dx)$.

We make Assumption 2 for two reasons. First, it ensures that we can rigorously compute the entropy flow by differentiating $R_\beta(p_t, p'_t)$ under the integrals. Secondly, the uniform integrability near $t = 0$ ensures that the map $t \mapsto R_\beta(p_t, p'_t)$ is continuous at $t = 0$, which simplifies the derivations. Intuitively, the non-explosion condition ensures that the time-derivatives of the density functions of W_t and W'_t do not explode too fast, so that the derivative of their relative Rényi divergence can be controlled. Intuitively, we observe that the above condition suggests that the drifts $\nabla \hat{\mathcal{R}}_S$ and $\nabla \hat{\mathcal{R}}_{S'}$ do not grow too fast at infinity.

These conditions are reasonable in the context of SDEs driven by rotationally invariant α -stable Lévy processes. Indeed, some studies show that, under mild conditions, the distributions of $(L_t^\alpha)_{t \geq 0}$ and those generated by such SDEs have polynomial tails, typically in $\|x\|^{-d-\alpha}$ (Samorodnitsky and Grigoriu, 2003; Blumenthal and Getoor, 1960; Chen et al., 2017). This can be readily verified in the case of the α -stable Ornstein-Uhlenbeck process. Therefore, it is reasonable to expect that $v_t \in \mathcal{C}_b^2(\mathbb{R}^d)$. We believe that this (reasonable)

condition could be relaxed and we use this assumptions mainly to simplify some technical arguments, as checking such tail estimates is beyond the scope of this paper.

Finally, it should be noted that Assumption 2 could be replaced by other conditions. Indeed, Dupuis and Şimşekli (2024) computed (different but) similar entropy flow along heavy-tailed SDEs, and their main theorems make use of intricate regularity and domination conditions of the functions $v_t(x)$. In our work, we simplify this approach by noting that Assumption 2 is enough for the Rényi flow computations to be rigorous.

3.2 Rényi flows computations

Our work is based on a new upper bound of the Rényi flow, *i.e.*, the time derivative of the Rényi divergence along the fractional Fokker-Planck equations. With a slight abuse of notation, in what follows, we denote by p'_t the probability distribution whose density function is also denoted by p'_t .

Theorem 13 (Rényi flow for heavy-tailed SDEs) *Assume that Assumption 2 holds. Then, for any $\beta \geq 2$ and $t > 0$,*

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq -\frac{2\sigma_\alpha^\alpha}{\beta-1} \frac{\mathcal{E}_{\alpha,p'_t}(v_t^{\beta/2}, v_t^{\beta/2})}{E_\beta(p_t, p'_t)} - \frac{4\sigma_2^2}{\beta} \frac{\mathcal{E}_{2,p'_t}(v_t^{\beta/2}, v_t^{\beta/2})}{E_\beta(p_t, p'_t)} + R_{\text{potential}},$$

where the following term accounts for the contribution of the potentials

$$R_{\text{potential}} := \beta \int_{\mathbb{R}^d} v_t^{\beta-1} \frac{\langle \nabla v_t, \nabla \hat{\mathcal{R}}_{S'} - \nabla \hat{\mathcal{R}}_S \rangle}{E_\beta(p_t, p'_t)} p'_t dx.$$

Proof See Section B.2. ■

Theorem 13 shows that the Rényi flow is controlled by three terms: (i) the Dirichlet form $\mathcal{E}_{\alpha,p'_t}$ associated with the pure-jump part of the noise, (ii) the Dirichlet form \mathcal{E}_{2,p'_t} corresponding to the diffusive part of the noise, and (iii) the quantity $R_{\text{potential}}$, which is the contribution from the drift difference of both SDEs.

In the absence of pure-jump noise (*i.e.*, $\sigma_\alpha = 0$), Theorem 13 extends existing computations obtained in the case of Gaussian noise (Chourasia et al., 2021). In the case $\beta = 2$, related computations can be found in Dupuis and Şimşekli (2024); He et al. (2024), but without being related to differential privacy (in He et al. (2024), the simpler case of fractional heat flows is considered). However, the extension to arbitrary $\beta \geq 2$ is a major obstacle, as the Dirichlet form $\mathcal{E}_{\alpha,p'_t}$ naturally appears in the derivations when $\beta = 2$. We address this issue in the proof of Theorem 13 and directly compare the Rényi flow to $\mathcal{E}_{\alpha,p'_t}$, for $\beta \geq 2$.

The following lemma allows us to apply heavy-tailed Poincaré inequalities to the entropy flow computed in Lemma 29. It is a key component of our analysis.

Lemma 14 *Assume that Assumption 2 holds. Assume that, for all $t > 0$, p'_t satisfies an α -stable Poincaré inequality with constants $(\gamma\sigma_\alpha^\alpha, \gamma\sigma_2^2)$ for some $\gamma > 0$. Then, for $\beta \geq 2$ and $t > 0$, as long as $E_\beta(p_t, p'_t) < \infty$,*

$$\frac{2\sigma_\alpha^\alpha}{\beta-1} \mathcal{E}_{\alpha,p'_t}(v_t^{\beta/2}, v_t^{\beta/2}) + \frac{2\sigma_2^2}{\beta} \mathcal{E}_{2,p'_t}(v_t^{\beta/2}, v_t^{\beta/2}) \geq \frac{1}{\gamma\beta} E_\beta(p_t, p'_t) \left(1 - e^{-R_\beta(p_t, p'_t)}\right).$$

■

Moreover, when $\sigma_2 = 0$, the constant $1/(\gamma\beta)$ above can be replaced by $1/(\gamma(\beta - 1))$.

Proof Deferred to Section B.2.

3.3 Differential privacy of multifractal SDEs

Based on the technical lemmas of the previous subsection, we can now derive our main results, which consists in RDP guarantees for Equation (11). We first present the case where the noise has a non-trivial Gaussian component, *i.e.*, when $\sigma_2 > 0$, which we refer to as the *multifractal* setting. This case has the advantage of being technically simpler, due to the regularizing effect of the Gaussian noise.

The following theorem is a differential privacy bound for multifractal SDEs.

Theorem 15 Let $\beta \geq 2$. Suppose Assumptions 1 and 2 hold and p'_t satisfies an α -stable Poincaré inequality with constants $(\gamma\sigma_\alpha^\alpha, \gamma\sigma_2^2)$ for some $\gamma > 0$ and for all $t > 0$. Then,

$$R_\beta(p_t, p'_t) \leq \frac{\beta S_g^2}{2\sigma_2^2 n^2} t =: K_n t, \quad t > 0. \quad (12)$$

If moreover $K_n < a := \frac{1}{\gamma\beta}$, then,

$$R_\beta(p_t, p'_t) \leq -\log \left(1 - \frac{\gamma S_g^2 \beta^2}{2\sigma_2^2 n^2} \right), \quad t > 0. \quad (13)$$

Proof By Theorem 13, for $t > 0$,

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq -\frac{2\sigma_\alpha^\alpha}{\beta - 1} \frac{\mathcal{E}_{\alpha, p'_t}(v_t^{\beta/2}, v_t^{\beta/2})}{E_\beta(p_t, p'_t)} - \frac{4\sigma_2^2}{\beta} \frac{\mathcal{E}_{2, p'_t}(v_t^{\beta/2}, v_t^{\beta/2})}{E_\beta(p_t, p'_t)} + R_{\text{potential}}.$$

By the Cauchy-Schwarz and Young inequalities, we have that for all $\lambda > 0$,

$$\begin{aligned} R_{\text{potential}} &= \beta \int_{\mathbb{R}^d} v_t^{\beta-1} \frac{\langle \nabla v_t, \nabla \widehat{\mathcal{R}}_{S'} - \nabla \widehat{\mathcal{R}}_S \rangle}{E_\beta(p_t, p'_t)} p'_t dx \\ &\leq \frac{\beta}{E_\beta(p_t, p'_t)} \left(\frac{\lambda}{2} \int_{\mathbb{R}^d} v_t^{\beta-2} \|\nabla v_t\|^2 p'_t dx + \frac{1}{2\lambda} \int_{\mathbb{R}^d} \|\nabla \widehat{\mathcal{R}}_S - \nabla \widehat{\mathcal{R}}_{S'}\|^2 v_t^\beta p'_t dx \right) \\ &\leq \frac{2\lambda}{\beta} \frac{\mathcal{E}_{2, p'_t}(v_t^{\beta/2}, v_t^{\beta/2})}{E_\beta(p_t, p'_t)} + \frac{\beta S_g^2}{2\lambda n^2}. \end{aligned}$$

In the equation above, we choose $\lambda := \sigma_2^2$, which gives

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq -\frac{2\sigma_\alpha^\alpha}{\beta - 1} \frac{\mathcal{E}_{\alpha, p'_t}(v_t^{\beta/2}, v_t^{\beta/2})}{E_\beta(p_t, p'_t)} - \frac{2\sigma_2^2}{\beta} \frac{\mathcal{E}_{2, p'_t}(v_t^{\beta/2}, v_t^{\beta/2})}{E_\beta(p_t, p'_t)} + \frac{\beta S_g^2}{2\sigma_2^2 n^2}.$$

Now we use Lemma 14 and obtain that

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq -\frac{1}{\gamma\beta} \left(1 - e^{-R_\beta(p_t, p'_t)} \right) + \frac{\beta S_g^2}{2\sigma_2^2 n^2}. \quad (14)$$

We solve this type of differential inequality in Lemma 26 in Appendix A. In order to apply this lemma, we observe that Assumption 2 also implies the continuity of $t \mapsto R_\beta(p_t, p'_t)$ on $[0, \infty)$, by the dominated convergence theorem. The claim immediately then follows by noting that both SDEs are initialized with the same probability distributions. ■

The differential inequality (14) is a consequence of the fact that our analysis relies on the fractional Poincaré inequality, as the logarithmic Sobolev inequality is not available in our case (see Section 2.3). A similar differential inequality appears in the derivations of Cao et al. (2019) in their study of exponential decay of Rényi entropy along (Gaussian) Fokker-Planck equation. In the particular case where $\mathcal{S}_g = 0$, if we set p'_t to be the invariant distribution of Equation (11), then Equation (14) ensures the exponential decay of Rényi entropies along fractional Fokker-Planck equations, which is of independent interest.

We distinguish two regimes: (i) it always holds that $R_\beta(p_t, p'_t)$ grows at most linearly with time, corresponding to a regime where the Poincaré constant of the SDEs is too large, and (ii) if additionally $K_n < a$, then $R_\beta(p_t, p'_t)$ is bounded by a constant for all $t > 0$. Note that, as long as $\gamma < \infty$, we have $K_n < a$ as long as the sample size n is sufficiently large.

Theorem 15 has an interesting dependence on β (the order of the Rényi divergence). By Equation (12), we always have a RDP guarantee in $\rho\beta$, with $\rho = \mathcal{O}(t/n^2)$. This dependence in β corresponds to the zero-concentrated DP guarantee introduced by Bun and Steinke (2016) but comes at a cost of a potential linear dependence on time. On the other hand, we observe that when $K_n < a$, we can remove the time-dependence at the cost of a slightly worst dependence on β , yielding a bound in $\mathcal{O}(\beta^2/n^2)$. It can be seen from our proofs (in Lemma 14) that the presence of β^2 is inherently due to the non-local nature of the Dirichlet form $\mathcal{E}_{\alpha, p'_t}$, and cannot be avoided in our approach. We see this behavior as a *semi-concentrated* DP guarantee.

In the case $K_n < a$, we obtain (β, γ) -RDP with $\gamma = \mathcal{O}(\beta^2/n^2)$ as $n \rightarrow \infty$. By Lemma 4, this implies $(0, \delta)$ -DP with $\delta = \mathcal{O}(n^{-1})$, which is the same rate in n that was obtained by Şimşekli et al. (2024) for (S)GD under heavy-tailed noise. Moreover, our results provide a differential privacy guarantee that is independent of the dimension. This independence of the dimension is intrinsically related to the presence of Gaussian noise. In the next section, we will analyze in more detail the case of pure-jump noise and show that we still obtain a weaker dependence on the dimension compared to existing works.

3.4 Differential privacy of pure-jump SDEs

We now present the case of pure α -stable noise. That is, $\sigma_2 = 0$, and we consider the following SDEs, where we denote $\sigma := \sigma_\alpha$ for simplicity,

$$\begin{cases} dW_t = -\nabla \widehat{\mathcal{R}}_S(W_t)dt + \sigma dL_t^\alpha, \\ dW'_t = -\nabla \widehat{\mathcal{R}}_{S'}(W'_t)dt + \sigma dL_t^\alpha. \end{cases} \quad (15)$$

Before presenting the main result of this subsection, we first motivate the approach. In the proof of Theorem 15, the Young inequality is used, leading to a term proportional to $\mathcal{E}_{2, p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right)$, which is canceled out by the diffusive part of the Rényi flow. Unfortunately, this procedure cannot be applied when $\sigma_2 = 0$.

The intuition to address this issue comes from the celebrated Bourgain-Brezis-Mironescu's theorem (also known as BBM formula) (Bourgain et al., 2001), stating that for all u in the Sobolev space $H^1(\mathbb{R}^d) := \{u \in L^2(\mathbb{R}^d) : \nabla u \in L^2(\mathbb{R}^d)\}$,

$$\lim_{\alpha \rightarrow 2^-} \left(1 - \frac{\alpha}{2}\right) \iint_{\mathbb{R}^d \times \mathbb{R}^d} \frac{|u(x) - u(y)|^2}{\|x - y\|^{d+\alpha}} dx dy = \frac{\pi^{d/2}}{2\Gamma(1 + \frac{d}{2})} \int_{\mathbb{R}^d} \|\nabla u(x)\|^2 dx. \quad (16)$$

In our case, the Dirichlet form $\mathcal{E}_{\alpha,p'_t}(u, u)$ corresponds to a weighted version of the left-hand side of Equation (16). In fact, we are even able to prove a weighted version of the BBM formula; namely, for any $u \in \mathcal{C}_b^2(\mathbb{R}^d)$ and any Borel probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$\lim_{\alpha \rightarrow 2^-} \mathcal{E}_{\alpha,\mu}(u, u) = \mathcal{E}_{2,\mu}(u, u). \quad (17)$$

The proof of this result is deferred to Lemma 31 in Section B.3. Note that a similar observation can be found in (He et al., 2024). Equation (17) shows that it is reasonable to approximate $\mathcal{E}_{\alpha,p'_t}$ by \mathcal{E}_{2,p'_t} and, hence, makes the proof technique of Theorem 15 applicable to our case. In order to leverage this intuition more quantitatively and non-asymptotically, we require the following additional regularity assumption.

Assumption 3 *We assume that, with $v_t := p_t/p'_t$, for some $T > 0$ and all $\beta \geq 2$,*

$$\sup_{S \simeq S'} \sup_{t \leq T} \left(\left\| \nabla v_t^{\beta/2} \right\|_\infty + \left\| \nabla^2 v_t^{\beta/2} \right\|_\infty \right) < \infty \quad \text{and} \quad \inf_{S \simeq S'} \inf_{t \leq T} \left\| \nabla v_t^{\beta/2} \right\|_{L^2(p'_t)} > 0.$$

Equipped with this assumption, we can prove the following result, which is a RDP guarantee for SDEs driven by pure-jump α -stable noise.

Theorem 16 *Assume that Assumptions 1 to 3 hold, and that, for all $t \in (0, T]$, p'_t satisfies an α -stable Poincaré inequality with constants $(\gamma\sigma^\alpha, 0)$ for some $T, \gamma > 0$. Then, for $\beta \geq 2$, there exists $R > 0$ (which can depend on d, β, α, T) such that with the following constants*

$$a := \frac{1}{2\gamma(\beta - 1)}, \quad K_{\alpha,d} := \frac{4(2 - \alpha)d\Gamma(\frac{d}{2})\Gamma(1 - \frac{\alpha}{2})}{\alpha 2^\alpha R^{2-\alpha}\Gamma(\frac{d+\alpha}{2})}, \quad K_n := \frac{K_{\alpha,d}(\beta - 1)\mathcal{S}_g^2}{\sigma^\alpha n^2},$$

we have

$$R_\beta(p_t, p'_t) \leq \frac{K_{\alpha,d}(\beta - 1)\mathcal{S}_g^2 t}{\sigma^\alpha n^2} \quad \text{for all } t \in [0, T].$$

If moreover $K_n < a$, then

$$R_\beta(p_t, p'_t) \leq -\log \left(1 - \frac{2\gamma(\beta - 1)^2 K_{\alpha,d} \mathcal{S}_g^2}{\sigma^\alpha n^2} \right) \quad \text{for all } t \in [0, T].$$

Proof Deferred to Section B.3. ■

The main technical ingredient in the proof of Theorem 16 is to obtain a more quantitative version of the weighted BBM formula obtained above.

In Theorem 15, we observe the same two regimes as in the multifractal case, discussed in Section 3.3. Therefore, the same remarks regarding these two regimes apply. However, this result closely exploits the structure of the α -stable noise, which translates into an interesting dependence on α and d , which we will analyze later.

We provide in Lemma 28 in Appendix A an asymptotic analysis of the constant $K_{\alpha,d}$, showing that $K_{\alpha,d} = \mathcal{O}(1)$ as $\alpha \rightarrow 2^-$, as long as R does not depend on α or does not explode when $\alpha \rightarrow 2^-$. In particular, this shows that the effect of the unknown constant R vanishes as $\alpha \rightarrow 2^-$, as expected from the weighted BBM formula. On the other hand, if we assume that R grows slowly with d , in the high dimension limit ($d \rightarrow \infty$), Lemma 28 provides

$$K_{\alpha,d} = \mathcal{O}_{d \rightarrow \infty} \left(\frac{d^{1-\alpha/2}}{n^2 \sigma^\alpha} \right).$$

As $\alpha < 2$ (and in practice α is close to 2, see (Barsbey et al., 2021)) the dimension always has an exponent smaller than 1, which is therefore easily compensated for large sample sizes by the term n^{-2} . Moreover, we observe that the dimension dependence vanishes as $\alpha \rightarrow 2^-$, hence recovering the known behavior in the presence of Gaussian noise.

Moreover, if R grows slowly with d , Lemma 4, together with Theorem 16 implies a $(0, \delta)$ -DP guarantee with

$$\delta = \mathcal{O}_{d \rightarrow \infty} \left(\frac{d^{\frac{2-\alpha}{4}}}{n \sigma^{\alpha/2}} \right) = \mathcal{O}_{d \rightarrow \infty} \left(\frac{\sqrt{d}}{n} \left(\sigma \sqrt{d} \right)^{-\alpha/2} \right). \quad (18)$$

Again, we recover the dependence on n of Şimşekli et al. (2024), obtained in their study of heavy-tailed (S)GD. However, our result has a better dependence on d , as these authors obtained an estimate in $\mathcal{O}(d^{(1+\alpha)/2})$. However, it should be noted that the results of Şimşekli et al. (2024) do not require finite sensitivity, which we need in our study. This reveals a new tradeoff between the dimension dependence and the finite sensitivity assumption for heavy-tailed algorithms. Finally, while Şimşekli et al. (2024) suggested that heavier tails systematically lead to better guarantees, our bound reveals a more complex structure, indicating that heavier tails might be beneficial (resp. harmful) when $\sigma \sqrt{d} < 1$ (resp. $\sigma \sqrt{d} > 1$). A similar phenomenon has been reported by Dupuis and Şimşekli (2024) for generalization bounds.

Remark 17 The constant R appearing in the above theorem is of a different nature from that appearing in the entropy flow computations of Dupuis and Şimşekli (2024). Indeed, these authors relied on an intricate assumption (Assumption 4.3 in their paper) that directly imposes a lower bound on a Dirichlet form-like quantity. In our setting, we improve this analysis by directly obtaining a lower bound on the Dirichlet form under Assumption 3.

4 Extension to the Discrete-Time Setting

In this section, we discuss the extension of our results to the discrete-time setting, *i.e.*, (stochastic) gradient descent with heavy-tailed noise. We first present our discrete-time RDP guarantees in Section 4.1. Then, Section 4.2 focuses on the analysis of the associated Poincaré inequalities, thus providing theoretical foundations for our main technical assumptions.

4.1 Differential privacy of heavy-tailed (stochastic) gradient descent

In this subsection, we analyze the RDP guarantees of gradient descent under α -stable noise. These results can be easily extended to the case of stochastic gradient descent (*i.e.*, with mini-batch noise) and to the case of multifractal noise, in which case the constants obtained should be replaced by those appearing in Theorem 15.

Setup. Let $\mathcal{C} \subset \mathbb{R}^d$ be a closed convex set, and $S \simeq S'$ be fixed neighboring datasets in \mathcal{Z}^n for some $n \geq 1$, denoted $S = (z_1, \dots, z_n)$ and $S' = (z'_1, \dots, z'_n)$. Let $b \in \{1, \dots, n\}$ be a fixed batch size. Given a random mini-batch of indices $\Omega_k \subset \{1, \dots, n\}$ with $|\Omega_k| = b$, we denote the stochastic gradients by

$$\widehat{g}_S(x, \Omega_k) := \frac{1}{b} \sum_{i \in \Omega_k} \nabla \ell(w, z_i), \quad \widehat{g}_{S'}(x, \Omega_k) := \frac{1}{b} \sum_{i \in \Omega_k} \nabla \ell(w, z'_i).$$

In our study, we consider that Ω_k is a uniformly sampled random subset of $\{1, \dots, n\}$ of size b , which is the same sampling scheme as in prior works (Şimşekli et al., 2024). Moreover, we assume that $(\Omega_k)_{k \in \mathbb{N}}$ are *i.i.d.* and that they are independent of the other random variables (e.g., the Lévy processes and Brownian motions).

We consider the following recursions:

$$\begin{cases} X_{k+1} = \Pi_{\mathcal{C}} (X_k - \eta \widehat{g}_S(X_k, \Omega_k) + \sigma \eta^{1/\alpha} \xi_k), \\ X'_{k+1} = \Pi_{\mathcal{C}} (X'_k - \eta \widehat{g}_{S'}(X'_k, \Omega_k) + \sigma \eta^{1/\alpha} \xi_k), \end{cases} \quad (19)$$

where $\eta > 0$, $\xi_k \sim \mathcal{S}\alpha\mathcal{S} := \text{Law}(L_1^\alpha)$, $k \geq 0$, are *i.i.d.* α -stable random vectors in \mathbb{R}^d with $\alpha \in (1, 2)$, and $\Pi_{\mathcal{C}}$ denotes the orthogonal projection onto \mathcal{C} . Let $t_k := k\eta$ for all $k \geq 0$. In order to define the interpolation scheme, we define the following random mappings,

$$F_1^{(k)}(x) := \frac{1}{2} (\widehat{g}_S(x, \Omega_k) + \widehat{g}_{S'}(x, \Omega_k)), \quad F_2^{(k)}(x) := \frac{1}{2} (\widehat{g}_S(x, \Omega_k) - \widehat{g}_{S'}(x, \Omega_k)).$$

Inspired by Chourasia et al. (2021), we construct a time-continuous stochastic process Θ_t (resp. Θ'_t) by the following interpolation mechanism:

$$\begin{cases} \Theta_t = \Theta_{t_k} - \eta F_1^{(k)}(\Theta_{t_k}) - (t - t_k) F_2^{(k)}(\Theta_{t_k}) + \sigma(t - t_k)^{1/\alpha} \xi_k, & t_k \leq t < t_{k+1}, \\ \Theta'_t = \Theta'_{t_k} - \eta F_1^{(k)}(\Theta'_{t_k}) + (t - t_k) F_2^{(k)}(\Theta'_{t_k}) + \sigma(t - t_k)^{1/\alpha} \xi_k, & t_k \leq t < t_{k+1}, \\ \Theta_{t_{k+1}}^- := \Pi_{\mathcal{C}} \left(\lim_{t \rightarrow \eta^-} \Theta_{t_k+t} \right), \quad \Theta'_{t_{k+1}}^- := \Pi_{\mathcal{C}} \left(\lim_{t \rightarrow \eta^-} \Theta'_{t_k+t} \right). \end{cases}$$

We see that for all $k \in \mathbb{N}$, $\Theta_{t_k}^- \sim \text{Law}(X_k)$, and similarly for $\Theta'_{t_k}^-$.

Let p_t and p'_t denote the probability density function of Θ_t and Θ'_t , respectively. Under mild regularity conditions, we can show that they satisfy the following Fokker-Planck equations (Ryffel et al., 2022; Chourasia et al., 2021) for $t_k < t < t_{k+1}$,

$$\begin{cases} \partial_t p_t(\theta) = -\sigma^\alpha (-\Delta)^{\frac{\alpha}{2}} p_t(\theta) + \nabla \cdot (p_t(\theta) V_k(t, \theta)) \\ \partial_t p'_t(\theta) = -\sigma^\alpha (-\Delta)^{\frac{\alpha}{2}} p'_t(\theta) + \nabla \cdot (p'_t(\theta) V'_k(t, \theta)) \end{cases} \quad \text{with} \quad \begin{cases} V_k(t, \theta) = \mathbb{E} [F_2^{(k)}(\Theta_{t_k}) | \Theta_t = \theta], \\ V'_k(t, \theta) = -\mathbb{E} [F_2^{(k)}(\Theta'_{t_k}) | \Theta'_t = \theta]. \end{cases}$$

Because of the presence of the projection $\Pi_{\mathcal{C}}$, we adapt Assumption 1 as follows, hence weakening this condition.

Assumption 4 We assume that the sensitivity is finite on \mathcal{C} , i.e.,

$$\mathcal{S}_{g,\mathcal{C}} := \operatorname{ess\,sup}_{z,z' \sim \mu_z \otimes \mu_z} \sup_{w \in \mathcal{C}} \|\nabla \ell(w, z) - \nabla \ell(w, z')\| < +\infty.$$

Theorem 18 Assume that Assumptions 2 to 4 hold and that, for all $t > 0$, p'_t satisfies an α -stable Poincaré inequality with constant $(\gamma\sigma^\alpha, 0)$ for some $\gamma > 0$. Finally assume that $t \mapsto R_\beta(p_t, p'_t)$ is right-continuous. Then, there exists a constant R (which can depend on d , β , α and η) such that, with the following constants

$$a := \frac{1}{2\gamma(\beta-1)}, \quad K_{\alpha,d} := \frac{4(2-\alpha)d\Gamma\left(\frac{d}{2}\right)\Gamma\left(1-\frac{\alpha}{2}\right)}{\alpha 2^\alpha R^{2-\alpha}\Gamma\left(\frac{d+\alpha}{2}\right)}, \quad K_n := \frac{K_{\alpha,d}(\beta-1)\mathcal{S}_{g,\mathcal{C}}^2}{\sigma^\alpha n^2},$$

we have

$$R_\beta(\operatorname{Law}(X_k), \operatorname{Law}(X'_k)) \leq \frac{K_{\alpha,d}(\beta-1)\mathcal{S}_{g,\mathcal{C}}^2}{\sigma^\alpha n^2} k\eta, \quad k \in \mathbb{N}.$$

If moreover $K_n < a$, then

$$R_\beta(\operatorname{Law}(X_k), \operatorname{Law}(X'_k)) \leq -\log\left(1 - \frac{2\gamma(\beta-1)^2 K_{\alpha,d} \mathcal{S}_g^2}{\sigma^\alpha n^2}\right), \quad k \in \mathbb{N}$$

Proof Deferred to Section B.4. ■

This theorem shows that, under our assumptions, we obtain similar guarantees for heavy-tailed gradient descent as in the continuous case. Thus, as in the discussion of Theorem 16, we see that Theorem 18 implies a $(0, \delta)$ -DP guarantee that has the same dependence in n as in previous works, but with a much better dependence on the dimension d .

This improvement comes at the cost of assuming that the probability distributions generated by the learning algorithm satisfy a fractional Poincaré inequality. In the next subsection, we investigate in more detail the behavior of such inequalities in the context of heavy-tailed gradient descent.

4.2 Stability properties of the fractional Poincaré inequalities

In this section, we study the stability of fractional Poincaré inequalities through the noisy (S)GD iterations. This section is a sanity check. It ensures that the α -stable Poincaré inequalities can be satisfied in practice. In the case of pure Gaussian noise, similar stability properties have already been used in the context of differential privacy (Chourasia et al., 2021; Chien et al., 2024; Ryffel et al., 2022). These existing results are based on two properties of the classical Poincaré inequalities: (1) stability by Gaussian convolution and (2) stability under pushforward by Lipschitz mapping. It is also well known that Poincaré inequalities are stable by bounded perturbations (Holley and Stroock, 1987; Bakry et al., 2014). In our paper, we show that fractional Poincaré inequalities also satisfy comparable properties, which might be of independent interest beyond this work.

The stability of fractional Poincaré inequalities under convolution with an α -stable distribution was already observed by He et al. (2024). We extend this result in the following lemma to the α -stable Poincaré inequalities defined in Definition 9. The proof is similar, but we present it for the sake of completeness.

Lemma 19 (Stability under convolution with stable distributions) *Let μ and μ' be two probability measures on \mathbb{R}^d with densities $\mu(x)$ and $\mu'(x)$, respectively. Assume that μ and μ' satisfy the α -stable Poincaré inequality with constants (γ_1, γ_2) and (γ'_1, γ'_2) , respectively. Let us denote $m := \mu * \mu'$. Then, m satisfies an α -stable Poincaré's inequality with constants $(\gamma_1 + \gamma'_1, \gamma_2 + \gamma'_2)$.*

Proof See Section B.5. ■

Remark 20 This result recovers as a particular case the stability by convolution properties of the classical Poincaré inequalities (*i.e.*, when $\gamma_1 = \gamma'_1 = 0$) (Bakry et al., 2014).

Regarding the stability under Lipschitz mappings, the situation is more contrasted. This phenomenon is due to the non-local nature of the Dirichlet forms $\mathcal{E}_{\alpha, \mu}$ when $\alpha < 2$. Indeed, we find that bi-Lipschitz continuity is needed in the case of fractional Poincaré inequalities. This leads to the following stability property of α -stable Poincaré inequalities under bi-Lipschitz diffeomorphisms. This result is new to the best of our knowledge.

Lemma 21 (Stability under bi-Lipschitz diffeomorphisms) *Assume that μ has a density $\mu(x)$ with respect to the Lebesgue measure and satisfies an α -stable Poincaré inequality with constants (γ_1, γ_2) . Let $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a C^1 bi-Lipschitz diffeomorphism, *i.e.*, T is continuously differentiable and there exist $L_1, L_2 > 0$ such that for all $x, y \in \mathbb{R}^d$*

$$L_1 \|x - y\| \leq \|T(x) - T(y)\| \leq L_2 \|x - y\|.$$

Then the measure $T_{\#}\mu$ satisfies an α -stable Poincaré inequality with constants $\left(\gamma_1 \frac{L_2^{\alpha+d}}{L_1^d}, \gamma_2 L_2^2\right)$.

Proof See Section B.5. ■

Finally, we can prove a perturbation lemma, similar to the standard Holley-Stroock lemma (Holley and Stroock, 1987). Even though this result is not immediately used in the proofs of our main results, we present it to give a complete picture of the stability properties of the fractional Poincaré inequalities.

Lemma 22 (Bounded perturbation) *Assume that μ has a density $\mu(x)$ with respect to the Lebesgue measure and satisfies an α -stable Poincaré inequality with constants (γ_1, γ_2) . Let $\mu' \leq \mu$ be another Borel probability measure such that, almost surely, $e^{-b} \leq d\mu'/d\mu \leq e^b$ for some $b \geq 0$. Then μ' satisfies an α -stable Poincaré inequality with constants $(e^{2b}\gamma_1, e^{2b}\gamma_2)$.*

Proof See Section B.5. ■

Estimation of fractional Poincaré constants. Based on the stability properties derived above, we provide in the next proposition an upper bound of the fractional Poincaré constant of heavy-tailed (S)GD for smooth and strongly convex losses.

Proposition 23 Assume that $\ell(w, z)$ is λ -strongly convex and M -smooth in w . We consider Equation (19) with $\mathcal{C} = \mathbb{R}^d$. We assume that the initial distribution of X_0 satisfies an α -stable Poincaré inequality with constants $(\gamma, 0)$. We further assume that

$$\frac{\lambda}{M} \left(1 + \frac{\alpha}{d}\right) > 1. \quad (20)$$

Let us define $c_0 := \frac{\eta\sigma^\alpha}{1-F(\eta_0)}$ with $F(\eta) := \frac{(1-\eta\lambda)^{\alpha+d}}{(1-\eta M)^\alpha}$. Then $F(\eta_0) < 1$ and if $\gamma \leq c_0$, then the distribution of X_k satisfies an α -stable Poincaré inequality with constants $(c_0, 0)$ for all $k \in \mathbb{N}$.

Proof See Section B.5. ■

In practice, this result imposes a relatively strong condition on the condition number M/λ , so that it would apply only for well-conditioned losses. This behavior is due to Lemma 21, which introduces dimension-dependent quantities in the estimation of the α -stable Poincaré inequality constants. To better understand this aspect, we observe that our Lemma 21 does not recover the Gaussian limit as $\alpha \rightarrow 2^-$, which could however be expected by the weighted BBM formula, Equation (17). This shows that Lemma 21 might be largely improvable, leading to much better estimates of the fractional Poincaré constants. Improving such stability lemmas for fractional Poincaré inequalities is beyond the scope of this paper, and we leave it for future works.

5 Conclusion

In this work, we provided the first RDP guarantees for heavy-tailed SDEs with α -stable noise. We explored both the multifractal (*i.e.*, with a non-trivial Gaussian component) and the pure-jump case. In both cases, we obtain a semi-concentrated RDP guarantee with two regimes: (i) a time-uniform bound when under the assumption of an α -stable Poincaré inequality and (ii) a regime where the bound can grow at most linearly with time but has the same dependence on the order of the Rényi divergence than in concentrated DP. In both cases, we can convert our results to $(0, \delta)$ -DP, in which case we obtain a dependence on the dimension comparable to existing works. Finally, we extended our results in the discrete time setting and proved stability lemmas for α -stable Poincaré inequalities, hence, providing theoretical foundations to satisfy these inequalities in practice.

Limitations & future works. As already discussed above, some of our results rely on the existence of fractional Poincaré inequalities for the finite-time distributions generated by the heavy-tailed SDEs we consider. We provided several stability lemmas for these inequalities, allowing to verify them in certain cases. However, we believe that the estimate of the fractional Poincaré constants can be improved, and is an important (yet, difficult) direction for future works. We observed in Section 3 that our main results have two regimes, one that can grow at most linearly with time but with a concentrated DP guarantee (*i.e.*, $\mathcal{O}(\beta)$), and one that is uniform in time but grows as β^2 , under the assumption of a fractional Poincaré inequality. While this dependence on β in the second case seems unavoidable with our approach, we believe that it is a relevant direction for future work, as it might require to strengthen certain functional inequalities existing in the literature. Finally, relaxing or

verifying the regularity conditions made in this work is another promising research direction, which might be related to recent fractional heat kernel estimates.

Acknowledgments and Disclosure of Funding

Jian Wang is supported by the NSF of China the National Key R&D Program of China (2022YFA1006003) and the National Natural Science Foundation of China (Nos. 12225104 and 12531007). Lingjiong Zhu is partially supported by the grants NSF DMS-2053454 and NSF DMS-2208303. Umut Şimşekli and Benjamin Dupuis' research is supported by European Research Council Starting Grant DYNASTY-101039676.

Organization of the appendix. The appendix is organized as follows.

- Section A presents three technical lemmas that are used in several proofs.
- Section B is dedicated to the detailed omitted proofs of our main results.

Appendix A. Technical Lemmas

In this section, we first prove two technical lemmas that are essential to obtain our main results. First, Lemma 25 is a lower bound on certain Bregman divergences. The second result, Lemma 26, presents a solution to the differential inequality that naturally appears in our Rényi flow computations in Section 3.2.

Definition 24 (Bregman divergence) Let $\Phi : (0, \infty) \rightarrow \mathbb{R}$ be a convex function. Then we define the Bregman divergence, for $a, b \in (0, \infty)$, as

$$D_\Phi(a, b) := \Phi(a) - \Phi(b) - \Phi'(b)(a - b).$$

By convexity, it is clear that the above quantity is always non-negative. When $\Phi(x) = \Phi_\beta(x) := x^\beta$ for $\beta > 1$, we will shorten the notation and denote

$$D_\beta(a, b) := D_{\Phi_\beta}(a, b) = a^\beta + (\beta - 1)b^\beta - \beta ab^{\beta-1}. \quad (21)$$

Note in particular that $D_2(a, b) := (a - b)^2$.

Lemma 25 Let $\beta \geq 2$. Then, for all $a, b \geq 0$:

$$D_2(a^{\beta/2}, b^{\beta/2}) \leq D_\beta(a, b).$$

Proof If $\beta = 2$, the result is obvious; if $\beta > 2$, we first note that when $a = 0$,

$$D_2(a^{\beta/2}, b^{\beta/2}) - D_\beta(a, b) = (a^{\beta/2} - b^{\beta/2})^2 - (a^\beta + (\beta - 1)b^\beta - \beta ab^{\beta-1}) = (2 - \beta)b^\beta \leq 0.$$

Let us now assume $a > 0$. Then, we have

$$\begin{aligned} D_2(a^{\beta/2}, b^{\beta/2}) - D_\beta(a, b) &= (a^{\beta/2} - b^{\beta/2})^2 - (a^\beta + (\beta - 1)b^\beta - \beta ab^{\beta-1}) \\ &= -(\beta - 2)b^\beta - 2(ab)^{\beta/2} + \beta ab^{\beta-1} \\ &= -(ab)^{\beta/2} \left((\beta - 2) \left(\frac{b}{a} \right)^{\beta/2} - \beta \left(\frac{b}{a} \right)^{\beta/2-1} + 2 \right). \end{aligned}$$

Therefore, it makes sense to study the polynomial function

$$P(x) = (\beta - 2)x^{\frac{\beta}{2}} - \beta x^{\frac{\beta}{2}-1} + 2, \quad x \geq 0.$$

Clearly $P(0) = 2 > 0$. For any $x > 0$,

$$P'(x) = \frac{\beta(\beta - 2)}{2}x^{\frac{\beta}{2}-2}(x - 1).$$

Thus, the minimum of P is $P(1) = 0$. The result follows. ■

In the next lemma, we solve a class of differential inequalities appearing in our theory.

Lemma 26 Consider a non-negative function $f : (0, \infty) \rightarrow \mathbb{R}_+$ that is continuous and differentiable, admits a right limit as $t \rightarrow 0^+$, and satisfies

$$f'(t) \leq K - a \left(1 - e^{-f(t)}\right) \quad \text{for any } t > 0, \quad (22)$$

where $a, K > 0$ are two constants. Then, we have the following results:

- For any $t > 0$, $f(t) \leq f(0^+) + Kt$.
- If $K < a$ and $f(0^+) \leq \log\left(\frac{a}{a-K}\right)$, then for any $t > 0$, $f(t) \leq \log\left(\frac{a}{a-K}\right)$.
- If $K < a$ and $f(0^+) > \log\left(\frac{a}{a-K}\right)$, then for any $t > 0$,

$$f(t) \leq \log\left(\frac{a}{a-K}\right) + \log\left(1 + e^{-(a-K)t} \left(e^{f(0^+)} \frac{a-K}{a} - 1\right)\right).$$

Proof Case 1: We note that for any $t > 0$, $f'(t) \leq K$. Therefore, for any $\varepsilon > 0$ and any $t > 0$, $f(t) \leq f(\varepsilon) + K(t - \varepsilon)$. By taking the limit as $\varepsilon \rightarrow 0^+$, we conclude that $f(t) \leq Kt + f(0^+)$ for all $t \geq 0$.

In the following, we denote $f_0 := \log\left(\frac{a}{a-K}\right)$.

Case 2: Assume that $K < a$ and $f(0^+) \leq f_0$. Suppose that there exists $\tau > 0$ such that $f(\tau) > f_0$. By Equation (22), this implies that $f'(\tau) < 0$. By continuity of f , there exists $t_0 \in (0, \tau]$ such that $f(t_0) = \sup_{s \in (0, \tau]} f(s)$. Necessarily, $t_0 > 0$, and thus $f'(t_0) = 0$, as $f'(\tau) < 0$, which implies $f(t_0) \leq f_0$ by Equation (22); but that is absurd. The claim follows.

Case 3: Let $g(t) := f(t) - f_0$. Let $\tau := \inf\{t > 0 : g(t) \leq 0\}$. By assumption, we know that $\tau > 0$. For $t \in (0, \tau)$, a simple computation shows that

$$\frac{d}{dt} \left(e^{g(t)} - 1 \right) = g'(t) e^{g(t)} \leq -(a - K) e^{g(t)} \left(1 - e^{-g(t)}\right) = -(a - K) \left(e^{g(t)} - 1 \right).$$

By Grönwall's inequality, we deduce that, for $t \in (0, \tau)$:

$$e^{g(t)} \leq 1 + e^{-(a-K)t} \left(e^{g(0^+)} - 1 \right) = 1 + e^{-(a-K)t} \left(e^{f(0^+)} \frac{a-K}{a} - 1 \right).$$

Therefore,

$$f(t) \leq \log\left(\frac{a}{a-K}\right) + \log\left(1 + e^{-(a-K)t} \left(e^{f(0^+)} \frac{a-K}{a} - 1 \right)\right).$$

For $t \geq \tau$, we can apply **Case 2** and observe that the above inequality is still true; hence, it is proven for all $t > 0$. \blacksquare

Remark 27 (Local version of Lemma 26) We observe that the proof Lemma 26 still holds if we only have Equation (22) for $t \in (0, T]$ for some $T > 0$. In that case, all the claims are true for $t \in (0, T]$ instead of $t > 0$.

The next lemma provides asymptotics for the constant $C_{\alpha,d}$ appearing in the infinitesimal generator of α -stable Lévy processes (Equation (8)). Similar computations can be found in (Dupuis and Şimşekli, 2024); we only sketch the proof for the sake of completeness.

Lemma 28 *We have the following results:*

$$C_{\alpha,d} \underset{\alpha \rightarrow 2^-}{\sim} (2 - \alpha) \pi^{-d/2} d\Gamma\left(\frac{d}{2}\right), \quad C_{\alpha,d} \underset{d \rightarrow \infty}{\longrightarrow} \frac{\alpha 2^{\alpha/2-1} \pi^{-d/2}}{\Gamma(1 - \frac{\alpha}{2})} \Gamma\left(\frac{d}{2}\right) d^{\alpha/2}.$$

Proof Recall that $C_{\alpha,d} := \alpha 2^{\alpha-1} \pi^{-d/2} \frac{\Gamma(\frac{\alpha+d}{2})}{\Gamma(1 - \frac{\alpha}{2})}$. By the continuity of the Γ -function, we have $\Gamma(\frac{\alpha+d}{2}) \rightarrow \Gamma(1 + \frac{d}{2}) = \frac{d}{2} \Gamma(\frac{d}{2})$ as $\alpha \rightarrow 2^-$. On the other hand, by Euler's reflection formula, $\frac{1}{\Gamma(1 - \frac{\alpha}{2})} = \frac{\sin(\frac{\pi\alpha}{2}) \Gamma(\frac{\alpha}{2})}{\pi} \sim 1 - \frac{\alpha}{2}$ as $\alpha \rightarrow 2^-$. Similarly, we can apply Gautschi's formula (Gautschi, 1959) to obtain that $C_{\alpha,d} \sim \frac{\alpha 2^{\alpha-1} \pi^{-d/2}}{\Gamma(1 - \frac{\alpha}{2})} \Gamma(\frac{d}{2}) d^{\alpha/2} 2^{-\alpha/2}$ as $d \rightarrow \infty$. The results follow immediately. ■

Appendix B. Omitted Proofs of the Main Results

B.1 Renyi flow along general Lévy processes

We first prove a more general version of Theorem 13 we allow the driving noise in Equation (9) to be an arbitrary Lévy process (with a symmetric Lévy measure). This highlights one of the main advantages of our approach compared to the existing literature on differential privacy for heavy-tailed algorithms: our proof techniques are not restricted to α -stable noise.

In this section, we therefore study general Lévy processes, as they have been introduced in Section 2.2. Formally, it corresponds to replacing Equation (2) with the following SDE,

$$dW_t = \nabla \widehat{\mathcal{R}}_S(W_t) dt + dL_t, \quad (23)$$

where $(L_t)_{t \geq 0}$ is a Lévy process with triplet given by $(0, \sqrt{2}I_d, \nu)$, with ν a *symmetric* Lévy measure (*i.e.*, $\nu(dz) = \nu(-dz)$).

Moreover, in this section, we also allow the drift terms in the SDEs to be time-dependent, which will be particularly helpful to extend our theory to the discrete-time setting in Section 4.

More precisely, we consider two probability densities following fractional Fokker-Planck equations (FPEs) (Duan, 2015; Umarov et al., 2018) driven by different force fields. We define two time-dependent vector fields

$$F_t, F'_t : \mathbb{R}^d \longrightarrow \mathbb{R}^d,$$

and we consider p_t and p'_t two probability densities that are smooth solutions of the following fractional FPEs, for some tail-index $\alpha \in (0, 2)$:

$$\begin{cases} \partial_t p_t = I[p_t] + \sigma_2^2 \Delta p_t + \nabla \cdot (p_t F_t), \\ \partial_t p'_t = I[p'_t] + \sigma_2^2 \Delta p'_t + \nabla \cdot (p'_t F'_t), \end{cases} \quad (24)$$

where the pseudo-differential operator $A := \sigma_2 \Delta + I$ corresponds to the infinitesimal generator of the Lévy process $(L_t)_{t \geq 0}$, according to Equation (8). Therefore, with the notations of Equation (8), the non-local operator I is given for $u \in \mathcal{C}_b^2(\mathbb{R}^d)$ by

$$I[u](x) := \int_{\mathbb{R}^d \setminus \{0\}} (u(x+z) - u(x) - \nabla u(x) \cdot z \chi(\|z\|)) d\nu(z).$$

Note that the operator A is self-adjoint with respect to the Lebesgue measure (Gentil and Imbert, 2009) (for smooth enough functions). We will use this fact repeatedly in our proofs.

This leads us to the following lemma, which is a Rényi flow computation along SDEs driven by general Lévy processes. This lemma fundamentally differs from the derivations of Dupuis and Şimşekli (2024) for two main reasons: (1) we focus on the Rényi flow instead of the entropy flow and (2) in this lemma we allow both dynamics to be time-varying, instead of considering stationary distributions.

Lemma 29 *Assume that Assumption 2 holds. Then, for any $\beta > 1$ and $t > 0$,*

$$\frac{d}{dt} R_\beta(p_t, p'_t) = -\frac{\sigma_\alpha^\alpha}{\beta-1} \frac{B_\beta^\alpha(p_t, p'_t)}{E_\beta(p_t, p'_t)} - \beta \sigma_2^2 \frac{I_\beta(p_t, p'_t)}{E_\beta(p_t, p'_t)} + \beta \int_{\mathbb{R}^d} v_t^{\beta-1} \frac{\langle \nabla v_t, F_t - F'_t \rangle}{E_\beta(p_t, p'_t)} p'_t dx,$$

where the term $B_\beta^\alpha(p_t, p'_t)$ is called the Bregman integral and is defined by

$$B_\beta^\alpha(p_t, p'_t) := C_{\alpha,d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} D_\beta \left(\frac{p_t}{p'_t}(x), \frac{p_t}{p'_t}(x+z) \right) p'_t(x) dx d\nu(z).$$

Proof We consider the Fokker-Planck equations given by Equation (24), with the notations above. To simplify the notations, we use as above the operator

$$A[u] := I[u] + \sigma_2 \Delta u.$$

Let us denote $v_t := p_t/p'_t$ and $\Phi(x) := x^\beta$. Assumption 2 ensures that we can differentiate under the integral, and write

$$\begin{aligned} (\beta-1) E_\beta(p_t, p'_t) \frac{d}{dt} R_\beta(p_t, p'_t) &= \frac{d}{dt} \int_{\mathbb{R}^d} \Phi(v_t) p'_t dx \\ &= \int_{\mathbb{R}^d} \Phi'(v_t) (\partial_t p_t - v_t \partial_t p'_t) dx + \int_{\mathbb{R}^d} \Phi(v_t) \partial_t p'_t dx \\ &= \int_{\mathbb{R}^d} \Phi'(v_t) (A[p_t] + \nabla \cdot (p_t F_t) - v_t A[p'_t] - v_t \nabla \cdot (p'_t F'_t)) dx \\ &\quad + \int_{\mathbb{R}^d} \Phi(v_t) (A[p'_t] + \nabla \cdot (p'_t F'_t)) dx. \end{aligned}$$

Using the facts that I is self-adjoint in $L^2(\mathbb{R}^d; dx)$ (Gentil and Imbert, 2009) and that p_t and p'_t vanish at infinity (so that the boundary terms are equal to 0) by Assumption 2, we can integrate by parts to obtain that

$$(\beta-1) E_\beta(p_t, p'_t) \frac{d}{dt} R_\beta(p_t, p'_t) = C_{\text{diffusion}} + C_{\text{fractional}} + C_{\text{potential}},$$

where

$$\begin{aligned}
C_{\text{diffusion}} &:= \sigma_2^2 \int_{\mathbb{R}^d} p'_t (v_t \Delta \Phi'(v_t) - \Delta(v_t \Phi'(v_t)) + \Delta \Phi(v_t)) dx \\
&= \sigma_2^2 \int_{\mathbb{R}^d} p'_t \left(v_t \nabla \cdot (\Phi''(v_t) \nabla v_t) \right. \\
&\quad \left. - \nabla \cdot (\Phi'(v_t) \nabla v_t + v_t \Phi''(v_t) \nabla v_t) + \nabla \cdot (\Phi'(v_t) \nabla v_t) \right) dx \\
&= \sigma_2^2 \int_{\mathbb{R}^d} p'_t (v_t \nabla \cdot (\Phi''(v_t) \nabla v_t) - \nabla \cdot (v_t \Phi''(v_t) \nabla v_t)) dx \\
&= -\sigma_2^2 \int_{\mathbb{R}^d} \Phi''(v_t) \|\nabla v_t\|^2 p'_t dx = -\sigma_2^2 \beta(\beta-1) I_\beta(p_t, p'_t), \\
C_{\text{potential}} &:= - \int_{\mathbb{R}^d} p'_t (v_t \Phi''(v_t) \langle \nabla v_t, F_t \rangle + \langle \nabla(v_t \Phi'(v_t)), F'_t \rangle - \Phi'(v_t) \langle \nabla v_t, F'_t \rangle) dx \\
&= - \int_{\mathbb{R}^d} p'_t (v_t \Phi''(v_t) \langle \nabla v_t, F_t \rangle + v_t \Phi''(v_t) \langle \nabla v_t, F'_t \rangle) dx \\
&= \int_{\mathbb{R}^d} v_t \Phi''(v_t) \langle \nabla v_t, F'_t - F_t \rangle p'_t dx = \beta(\beta-1) \int_{\mathbb{R}^d} v_t^{\beta-1} \langle \nabla v_t, F'_t - F_t \rangle p'_t dx
\end{aligned}$$

and

$$C_{\text{fractional}} := \int_{\mathbb{R}^d} (I[\Phi'(v_t)] v_t - I[v_t \Phi'(v_t)] + I[\Phi(v_t)]) p'_t dx.$$

Now we use the expression of the non-local operator as

$$I[\phi](x) = \int_{\mathbb{R}^d \setminus \{0\}} (\phi(x+z) - \phi(x) - \langle \nabla \phi(x), z \rangle \chi(\|z\|)) d\nu(z).$$

The terms containing $\chi(\|z\|)$ cancel each other out, so that we obtain

$$\begin{aligned}
C_{\text{fractional}} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} \left\{ v_t(x) [\Phi'(v_t(x+z)) - \Phi'(v_t(x))] - v_t(x+z) \Phi'(v_t(x+z)) \right. \\
&\quad \left. + v_t(x) \Phi'(v_t(x)) + \Phi(v_t(x+z)) - \Phi(v_t(x)) \right\} p'_t(x) d\nu(z) dx.
\end{aligned}$$

After rearranging the terms, we obtain the following expression

$$C_{\text{fractional}} = -C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} D_\Phi(v_t(x), v_t(x+z)) p'_t(x) d\nu(z) dx.$$

The proof is complete. ■

Remark 30 We observe that the proof does not make use of the particular form of the Lévy measure of the rotationally invariant α -stable Lévy process. Indeed, the proof is valid for any symmetric Lévy process as the noise driving the two SDEs.

B.2 Omitted proofs of Section 3.2

We are now ready to present the proof of Theorem 13, which is a Rényi flow computation along heavy-tailed (multifractal) SDEs.

Proof of Theorem 13. As before, we denote $v_t := p_t/p'_t$. We apply Lemma 29 when $F_t = \nabla \widehat{\mathcal{R}}_S$ and $F'_t = \nabla \widehat{\mathcal{R}}_{S'}$. Moreover, we chose the Lévy process $(L_t)_{t \geq 0}$ in Equation (11) to be the α -stable Lévy process, as described in Section 2.2.

With these notations, we have that

$$\begin{aligned} B_\beta^\alpha(p_t, p'_t) &= C_{\alpha,d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} D_\beta(v_t(x), v_t(x+z)) p'_t(x) \frac{dx dz}{\|z\|^{d+\alpha}} \\ &\geq C_{\alpha,d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} D_2(v_t^{\beta/2}(x), v_t^{\beta/2}(x+z)) p'_t(x) \frac{dx dz}{\|z\|^{d+\alpha}} \quad (\text{Lemma 25}) \\ &= 2\mathcal{E}_{\alpha,p'_t}(v_t^{\beta/2}, v_t^{\beta/2}). \end{aligned}$$

We also have

$$I_\beta(p_t, p'_t) = \int_{\mathbb{R}^d} v_t^{\beta-2} \|\nabla v_t\|^2 dx = \frac{4}{\beta^2} \int_{\mathbb{R}^d} \left\| \nabla v_t^{\beta/2} \right\|^2 dx = \frac{4}{\beta^2} \mathcal{E}_{2,p'_t}(v_t^{\beta/2}, v_t^{\beta/2}).$$

The proof then follows from Lemma 29. ■

We conclude this section by presenting the proof of Lemma 14.

Proof of Lemma 14. Let us introduce the notation:

$$\mathcal{J} := \frac{2\sigma_\alpha^\alpha}{\beta-1} \mathcal{E}_{\alpha,p'_t}(v_t^{\beta/2}, v_t^{\beta/2}) + \frac{2\sigma_2^2}{\beta} \mathcal{E}_{2,p'_t}(v_t^{\beta/2}, v_t^{\beta/2}),$$

with $v_t := p_t/p'_t$. Let us denote $u_t := v_t^{\beta/2}$. By the α -stable Poincaré inequality, we have

$$\begin{aligned} \mathcal{J} &= \frac{\sigma_\alpha^\alpha C_{\alpha,d}}{\beta-1} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} (u_t(x) - u_t(x+z))^2 p'_t(x) \frac{dx dz}{\|z\|^{d+\alpha}} + \frac{2\sigma_2^2}{\beta} \int_{\mathbb{R}^d} \|\nabla u_t\|^2 dx \\ &\geq \frac{\sigma_\alpha^\alpha C_{\alpha,d}}{\beta} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} (u_t(x) - u_t(x+z))^2 p'_t(x) \frac{dx dz}{\|z\|^{d+\alpha}} + \frac{\sigma_2^2}{\beta} \int_{\mathbb{R}^d} \|\nabla u_t\|^2 dx \\ &\geq \frac{1}{\gamma\beta} \left\{ \int_{\mathbb{R}^d} \left(\frac{p_t}{p'_t} \right)^\beta p'_t dx - \left(\int_{\mathbb{R}^d} \left(\frac{p_t}{p'_t} \right)^{\beta/2} p'_t dx \right)^2 \right\} \\ &= \frac{1}{\gamma\beta} \left\{ e^{(\beta-1)R_\beta(p_t, p'_t)} - e^{(\beta-2)R_{\beta/2}(p_t, p'_t)} \right\} \\ &\geq \frac{1}{\gamma\beta} \left\{ e^{(\beta-1)R_\beta(p_t, p'_t)} - e^{(\beta-2)R_\beta(p_t, p'_t)} \right\} \quad (\beta \mapsto R_\beta \text{ is non-decreasing}) \\ &= \frac{1}{\gamma\beta} E_\beta(p_t, p'_t) \left(1 - e^{-R_\beta(p_t, p'_t)} \right). \end{aligned}$$

This completes the proof of the first assertion. For the second one, one can follow the same arguments as above. ■

B.3 Omitted proofs of Section 3.4

The following lemma is a formal proof of Equation (17) and serves as a motivation for our approach in Section 3.4. It can be seen as a weighted version of the celebrated Bourgain-Brezis-Mironescu theorem (BBM formula), under stronger assumptions. Note that a similar observation can be found in (He et al., 2024, Proposition B.1).

Lemma 31 (The weighted BBM formula) *Let $f \in \mathcal{C}_b^2(\mathbb{R}^d)$ and μ any Borel probability measure on \mathbb{R}^d . We have*

$$\mathcal{E}_{\alpha,\mu}(f, f) \xrightarrow[\alpha \rightarrow 2^-]{} \mathcal{E}_{2,\mu}(f, f).$$

Proof We write $\mathcal{E}_{\alpha,\mu}(f, f) = \frac{C_{d,\alpha}}{2} (I_1 + I_2)$, where

$$I_1 := \int_{\mathbb{R}^d} \int_{\|z\| \leq 1} \frac{(f(x+z) - f(x))^2}{\|z\|^{d+\alpha}} d\mu(x) dz,$$

and

$$I_2 := \int_{\mathbb{R}^d} \int_{\|z\| > 1} \frac{(f(x+z) - f(x))^2}{\|z\|^{d+\alpha}} d\mu(x) dz \leq 2 \|f\|_{\mathcal{C}^2}^2 \int_{\|z\| > 1} \frac{dz}{\|z\|^{d+\alpha}} = \frac{C}{\alpha}.$$

Here, $C < \infty$ is a constant independent of α . On the other hand, by the Taylor-Lagrange formula, for every $x, z \in \mathbb{R}^d$, $f(x+z) - f(x) = \langle z, \nabla f(x) \rangle + \langle z, \nabla^2 f(\xi_{x,z}) z \rangle / 2$, with $\xi_{x,z} \in \{x + az, a \in [0, 1]\}$. Therefore, we can further write $I_1 = I_3 + I_4$ as

$$\begin{aligned} I_3 &:= \int_{\mathbb{R}^d} \int_{\|z\| \leq 1} \frac{\frac{1}{4} \langle z, \nabla^2 f(\xi_{x,z}) z \rangle^2 + \langle z, \nabla^2 f(\xi_{x,z}) z \rangle \langle z, \nabla f(x) \rangle}{\|z\|^{d+\alpha}} d\mu(x) dz \\ &\lesssim \|f\|_{\mathcal{C}^2}^2 \int_{\|z\| \leq 1} \frac{dz}{\|z\|^{d+\alpha-3}} = \frac{C'}{3-\alpha} \end{aligned}$$

with $C' < \infty$ being another constant independent of α , and

$$I_4 := \int_{\mathbb{R}^d} \int_{\|z\| \leq 1} \frac{\langle z, \nabla f(x) \rangle^2}{\|z\|^{d+\alpha}} d\mu(x) dz = \frac{\text{Vol}(\mathbb{S}^{d-1})}{d} \int_{\mathbb{R}^d} \int_0^1 \|\nabla f(x)\|^2 d\mu(x) \frac{dr}{r^{\alpha-1}},$$

thanks to a spherical change of variable and the invariance by rotation.

Therefore, by Lemma 28, we have

$$\mathcal{E}_{\alpha,\mu}(f, f) \xrightarrow[\alpha \rightarrow 2^-]{} \frac{1}{2} (2-\alpha) \pi^{-d/2} d\Gamma\left(\frac{d}{2}\right) \frac{2\pi^{d/2}}{d\Gamma(\frac{d}{2})} \frac{1}{2-\alpha} \int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x) = \mathcal{E}_{2,\mu}(f, f).$$

This completes the proof. ■

We can now discuss the proof of Theorem 16, which is based on two technical lemmas. The next lemma is an integral representation of the Dirichlet form $\mathcal{E}_{\alpha,\mu}$ appearing in our Rényi flow computations.

Lemma 32 (Spherical parameterization of the Dirichlet form $\mathcal{E}_{\alpha,\mu}(f, f)$) Consider $u : \mathbb{R}^d \rightarrow \mathbb{R}$ a C^1 function with bounded gradient. Let $\mu \in \mathcal{P}(\mathbb{R}^d)$. Then,

$$\mathcal{E}_{\alpha,\mu}(u, u) = \frac{C_{\alpha,d}\sigma_{d-1}}{2d} \int_0^{+\infty} \mathcal{J}(r, u, \mu) \frac{dr}{r^{\alpha-1}},$$

where $\sigma_{d-1} := \text{Vol}(\mathbb{S}^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})}$ and $\mathcal{J}(\cdot, u, \mu) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a continuous function satisfying

$$\mathcal{J}(0, u, \mu) = \mathcal{E}_{2,\mu}(u, u) = \|\nabla u\|_{L^2(\mu)}^2.$$

Proof We apply a spherical change of coordinates to get

$$\begin{aligned} 2\mathcal{E}_{\alpha,\mu}(u, u) &= C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (u(x+z) - u(x))^2 \frac{d\mu(x)dz}{\|z\|^{d+\alpha}} \\ &= C_{\alpha,d} \int_0^{+\infty} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} \left(\int_0^r \langle \theta, \nabla u(x+s\theta) \rangle ds \right)^2 d\mu(x)d\sigma(\theta) \frac{dr}{r^{1+\alpha}}, \end{aligned}$$

where σ is the Hausdorff measure on the sphere \mathbb{S}^{d-1} . By Tonelli's theorem, we can rewrite the above integral as

$$2\mathcal{E}_{\alpha,\mu}(u, u) = K_{\alpha,d} \int_0^{+\infty} \mathcal{J}(r, u, \mu) \frac{dr}{r^{\alpha-1}},$$

where the function $\mathcal{J}(\cdot, u, \mu)$ defined for $r > 0$ by

$$\mathcal{J}(r, u, \mu) := \frac{d}{\sigma_{d-1}} \int_{\mathbb{R}^d \times \mathbb{S}^{d-1}} \left(\frac{1}{r} \int_0^r \langle \theta, \nabla u(x+s\theta) \rangle ds \right)^2 d(\mu \otimes \sigma)(x, \theta) \quad (25)$$

with $K_{\alpha,d} := \frac{C_{\alpha,d}\sigma_{d-1}}{d}$. We note that the integral above is clearly finite under our assumptions on u . Moreover, we note that for all $(x, \theta) \in \mathbb{R}^d \times \mathbb{S}^{d-1}$, we have by continuity of ∇u that

$$\frac{1}{r} \int_0^r \langle \theta, \nabla u(x+s\theta) \rangle ds \xrightarrow[r \rightarrow 0^+]{} \langle \theta, \nabla u(x) \rangle.$$

Moreover, by the Cauchy-Schwarz inequality,

$$\left(\frac{1}{r} \int_0^r \langle \theta, \nabla u(x+s\theta) \rangle ds \right)^2 \leq \|\nabla u\|_\infty^2 < +\infty.$$

Therefore, by the dominated convergence theorem with respect to $\mu \otimes \sigma$,

$$\mathcal{J}(r, u, \mu) \xrightarrow[r \rightarrow 0^+]{d} \frac{d}{\sigma_{d-1}} \int_{\mathbb{R}^d \times \mathbb{S}^{d-1}} \langle \theta, \nabla u(x) \rangle^2 d(\mu \otimes \sigma)(x, \theta).$$

By invariance by rotation and Tonelli's theorem, we deduce that

$$\mathcal{J}(r, u, \mu) \xrightarrow[r \rightarrow 0^+]{d} \frac{d}{\sigma_{d-1}} \int_{\mathbb{R}^d} \|\nabla u(x)\|^2 d\mu(x) \int_{\mathbb{S}^{d-1}} \langle \theta, e_1 \rangle^2 d\sigma(\theta) = \|\nabla u\|_{L^2(\mu)}^2,$$

where $e_1 \in \mathbb{S}^{d-1}$ is arbitrary. Therefore, we can continuously extend the function $\mathcal{J}(\cdot, u, \mu)$ on \mathbb{R}_+ by setting $\mathcal{J}(0, u, \mu) = \|\nabla u\|_{L^2(\mu)}^2 = \mathcal{E}_{2,\mu}(u, u)$. By the dominated convergence, we also obtain that $\mathcal{J}(\cdot, u, \mu)$ is continuous on \mathbb{R}_+ . This concludes the proof. \blacksquare

Lemma 33 (Lower bound on the Dirichlet form) *Assume that $\beta \geq 2$, and that v_t is twice continuously differentiable for all $t \in (0, T]$ so that*

$$\sup_{S \simeq S'} \sup_{t \leq T} \left(\|\nabla v_t^{\beta/2}\|_{\infty} + \|\nabla^2 v_t^{\beta/2}\|_{\infty} \right) < +\infty, \quad \inf_{S \simeq S'} \inf_{t \leq T} \|\nabla v_t^{\beta/2}\|_{L^2(p'_t)} > 0, \quad (26)$$

where $S \simeq S'$ ranges over all neighboring datasets in $\bigcup_{n \geq 1} \mathcal{Z}^n$. Then there exists a constant $R > 0$ (that may depend only on d and β as well as T) such that for all $t \in [0, T]$

$$2\mathcal{E}_{\alpha, p'_t}(v_t^{\beta/2}, v_t^{\beta/2}) \geq \frac{C_{\alpha, d\sigma_{d-1}}}{2d(2-\alpha)} R^{2-\alpha} \mathcal{E}_{2, p'_t}(v_t^{\beta/2}, v_t^{\beta/2}).$$

Proof fix $\beta > 0$. Let us introduce $u_t := v_t^{\beta/2}$ for all $t \in [0, T]$. By our assumptions, Lemma 32 gives us the spherical representation of the Dirichlet form $\mathcal{E}_{\alpha, p'_t}$:

$$\mathcal{E}_{\alpha, p'_t}(u_t, u_t) = \frac{C_{\alpha, d\sigma_{d-1}}}{2d} \int_0^{+\infty} \mathcal{J}(r, u_t, p'_t) \frac{dr}{r^{\alpha-1}}.$$

Let us denote the integrand in $\mathcal{J}(\cdot, u_t, p'_t)$ (see (25)) by

$$F(\theta, x, r) := \left(\frac{1}{r} \int_0^r \langle \theta, \nabla u(x + s\theta) \rangle ds \right)^2.$$

We have

$$\begin{aligned} \left| \frac{\partial F}{\partial r} \right| &= 2 \left| \frac{1}{r} \int_0^r \langle \theta, \nabla u(x + s\theta) \rangle ds \left(\frac{-1}{r^2} \int_0^r \langle \theta, \nabla u(x + s\theta) \rangle ds + \frac{1}{r^2} \int_0^r \langle \theta, \nabla u(x + r\theta) \rangle ds \right) \right| \\ &\leq 2 \|\nabla u\|_{\infty} \left| \frac{1}{r^2} \int_0^r \int_s^r \langle \theta, \nabla^2 u(x + l\theta) \theta \rangle dl ds \right| \\ &\leq \|\nabla u\|_{\infty} \|\nabla^2 u\|_{\infty} < +\infty. \end{aligned}$$

Therefore, we can differentiate under the sum and use the expression of $\mathcal{J}(\cdot, u_t, p'_t)$ (given in Lemma 32) to obtain that

$$\left| \frac{\partial \mathcal{J}}{\partial r}(r, u_t, p'_t) \right| \leq C_1 := d \|\nabla u\|_{\infty} \|\nabla^2 u\|_{\infty} < \infty.$$

In particular, for all $r > 0$ and $t \in [0, T]$,

$$|\mathcal{J}(r, u_t, p'_t) - \mathcal{J}(0, u_t, p'_t)| \leq C_1 r.$$

On the other hand, by our assumptions, we have $C_2 := \inf_{S \simeq S'} \inf_{t \leq T} \|\nabla u_t\|_{L^2(p'_t)}^2 > 0$. Thus, for all $t \in [0, T]$,

$$\mathcal{J}(0, u_t, p'_t) \geq C_2.$$

Therefore, there exists $R > 0$ (potentially depending only on β and d as well as T) such that for all $r \leq R$, $\mathcal{J}(r, u_t, p'_t) \geq (1/2)\mathcal{J}(0, u_t, p'_t)$. Thus, by Lemma 32, for all $r \leq R$, we have

$$2\mathcal{E}_{\alpha, p'_t}(u_t, u_t) \geq \frac{C_{\alpha, d\sigma_{d-1}}}{d} \int_0^R \mathcal{J}(r, u_t, p'_t) \frac{dr}{r^{\alpha-1}}$$

$$= \frac{C_{\alpha,d}\sigma_{d-1}}{2d} \mathcal{E}_{2,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right) \int_0^R \frac{dr}{r^{\alpha-1}} = \frac{C_{\alpha,d}\sigma_{d-1}}{2d(2-\alpha)} R^{2-\alpha} \mathcal{E}_{2,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right).$$

This completes the proof. \blacksquare

We now present the proof of our RDP guarantees for SDEs driven by pure-jump Lévy noise.

Proof of Theorem 16. We start the proof similarly to the proof of Theorem 15, with the difference that now $\sigma_2 = 0$ (recall that we denote $\sigma := \sigma_\alpha$ to simplify the notation). By Theorem 13, for $t > 0$,

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq -\frac{2\sigma^\alpha}{\beta-1} \frac{\mathcal{E}_{\alpha,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right)}{E_\beta(p_t, p'_t)} + R_{\text{potential}}.$$

We split the first term on the right-hand side into two halves and use Lemma 14 to obtain

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq -\frac{1}{2\gamma(\beta-1)} \left(1 - e^{-R_\beta(p_t, p'_t)} \right) - \frac{\sigma^\alpha}{(\beta-1)} \frac{\mathcal{E}_{\alpha,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right)}{E_\beta(p_t, p'_t)} + R_{\text{potential}}.$$

We can now apply Lemma 33, which gives that

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq -\frac{1 - e^{-R_\beta(p_t, p'_t)}}{2\gamma(\beta-1)} - \frac{\sigma^\alpha C_{\alpha,d}\sigma_{d-1} R^{2-\alpha}}{4d(\beta-1)(2-\alpha)} \frac{\mathcal{E}_{2,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right)}{E_\beta(p_t, p'_t)} + R_{\text{potential}}.$$

By the Cauchy-Schwarz and Young inequalities, we have for all $\lambda > 0$ that

$$\begin{aligned} R_{\text{potential}} &= \beta \int_{\mathbb{R}^d} v_t^{\beta-1} \frac{\langle \nabla v_t, \nabla \hat{\mathcal{R}}_{S'} - \nabla \hat{\mathcal{R}}_S \rangle}{E_\beta(p_t, p'_t)} p'_t dx \\ &\leq \frac{\beta}{E_\beta(p_t, p'_t)} \left(\frac{\lambda}{2} \int_{\mathbb{R}^d} v_t^{\beta-2} \|\nabla v_t\|^2 p'_t dx + \frac{1}{2\lambda} \int_{\mathbb{R}^d} \|\nabla \hat{\mathcal{R}}_S - \nabla \hat{\mathcal{R}}_{S'}\|^2 v_t^\beta p'_t dx \right) \\ &\leq \frac{2\lambda}{\beta} \frac{\mathcal{E}_{2,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right)}{E_\beta(p_t, p'_t)} + \frac{\beta \mathcal{S}_g^2}{2\lambda n^2}. \end{aligned}$$

Therefore, we make the following choice for the parameter $\lambda := \frac{\sigma^\alpha C_{\alpha,d}\sigma_{d-1}}{8d(\beta-1)(2-\alpha)} R^{2-\alpha}$. This implies that

$$\begin{aligned} \frac{d}{dt} R_\beta(p_t, p'_t) &\leq -\frac{1}{2\gamma(\beta-1)} \left(1 - e^{-R_\beta(p_t, p'_t)} \right) + \frac{4\mathcal{S}_g^2 d(2-\alpha)(\beta-1)}{n^2 \sigma^\alpha C_{\alpha,d}\sigma_{d-1}} \\ &= -\frac{1}{2\gamma(\beta-1)} \left(1 - e^{-R_\beta(p_t, p'_t)} \right) + \frac{4\mathcal{S}_g^2 (2-\alpha)d\Gamma\left(\frac{d}{2}\right)\Gamma\left(1-\frac{\alpha}{2}\right)(\beta-1)}{\alpha 2^\alpha \sigma^\alpha R^{2-\alpha} n^2 \Gamma\left(\frac{d+\alpha}{2}\right)}. \end{aligned}$$

We introduce $a := \frac{1}{2\gamma(\beta-1)}$ and $K_n := \frac{K_{\alpha,d}(\beta-1)\mathcal{S}_g^2}{\sigma^\alpha n^2}$, with $K_{\alpha,d} := \frac{4(2-\alpha)d\Gamma\left(\frac{d}{2}\right)\Gamma\left(1-\frac{\alpha}{2}\right)}{\alpha 2^\alpha R^{2-\alpha} \Gamma\left(\frac{d+\alpha}{2}\right)}$. Thus,

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq K_n - a \left(1 - e^{-R_\beta(p_t, p'_t)} \right). \quad (27)$$

We observe that Assumption 2 implies the continuity of $t \mapsto R_\beta(p_t, p'_t)$ at $t = 0$, by the dominated convergence theorem. The claim follows immediately from the local version of Lemma 26 (see Remark 27) by noting that both SDEs are initialized with the same probability distributions. \blacksquare

B.4 Omitted proofs of Section 4.1

We give below the proof of Theorem 18 that provides RDP guarantees for discretization of pure-jump SDEs.

Proof of Theorem 18. For any $t > 0$ and $\theta \in \mathbb{R}^d$, denote $\mathfrak{D}_t(\theta) = V_k(\theta, t) - V'_k(\theta, t)$. Let us fix $k \in \mathbb{N}$ and consider $t_k < t < t_{k+1}$. By utilizing the orthogonal projection and Assumption 4, we observe that, for all $\theta \in \mathbb{R}^d$,

$$\|\mathfrak{D}_t(\theta)\|^2 = \frac{1}{n^2} \left\| \mathbb{E} \left[F_2^{(k)}(\Theta_{t_k}) \mid \Theta_t = \theta \right] + \mathbb{E} \left[F_2^{(k)}(\Theta'_{t_k}) \mid \Theta'_t = \theta \right] \right\|^2.$$

Recall that the neighboring datasets $S \simeq S' \in \mathcal{Z}^n$ are fixed. Let $i_0 \in \{1, \dots, n\}$ be the (fixed) only index on which both datasets differ (*i.e.*, $z_{i_0} \neq z'_{i_0}$). As defined in Section 4, the random batches Ω_k are independent of the Lévy processes appearing in Equation (3). Therefore, we can condition on the event that $i_0 \in \Omega_k$, and write that

$$\left\| \mathbb{E} \left[F_2^{(k)}(\Theta_{t_k}) \mid \Theta_t = \theta \right] \right\| = \mathbb{P}(i_0 \in \Omega_k) \left\| \mathbb{E} \left[F_2^{(k)}(\Theta_{t_k}) \mid \Theta_t = \theta, i_0 \in \Omega_k \right] \right\| \leq \frac{b}{n} \cdot \frac{\mathcal{S}_{g,C}}{2b} = \frac{\mathcal{S}_{g,C}}{2n},$$

and similarly for $\mathbb{E} \left[F_2^{(k)}(\Theta'_{t_k}) \mid \Theta'_t = \theta \right]$. This gives us that

$$\|\mathfrak{D}_t(\theta)\|^2 \leq \frac{\mathcal{S}_{g,C}^2}{n^2}.$$

Therefore, we can replicate the lines of the proof of Theorem 16 for $t_k < t < t_{k+1}$, and the only difference is that we now have, for all $\lambda > 0$,

$$R_{\text{potential}} \leq \frac{2\lambda}{\beta} \frac{\mathcal{E}_{2,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right)}{\mathbb{E}_\beta(p_t, p'_t)} + \frac{\beta}{2\lambda} \sup_{\theta \in \mathbb{R}^d} \|\mathfrak{D}_t(\theta)\|^2 \leq \frac{2\lambda}{\beta} \frac{\mathcal{E}_{2,p'_t} \left(v_t^{\beta/2}, v_t^{\beta/2} \right)}{\mathbb{E}_\beta(p_t, p'_t)} + \frac{\beta \mathcal{S}_{g,C}^2}{\lambda n^2}.$$

Therefore, we obtain by the same lines as the proof of Theorem 16 an equation similar to Equation (27) (formally, $\mathcal{S}_{g,C}$ plays the same role as \mathcal{S}_g). Thus, we have

$$\frac{d}{dt} R_\beta(p_t, p'_t) \leq K_n - a \left(1 - e^{-R_\beta(p_t, p'_t)} \right), \quad t_k < t < t_{k+1}$$

with the quantities a and K_n defined in Theorem 18.

Case 1: By the local version of Lemma 26 (see Remark 27), we obtain

$$R_\beta(p_t, p'_t) \leq \lim_{s \rightarrow t_k^+} R_\beta(p_s, p'_s) + K_n(t - t_k), \quad t_k < t < t_{k+1}.$$

By taking the limit as $t \rightarrow t_{k+1}^-$ and applying the (assumed) right-continuity of the Rényi divergences, we have

$$\lim_{s \rightarrow t_{k+1}^-} R_\beta(p_s, p'_s) \leq R_\beta(p_{t_k^+}, p'_{t_k^+}) + K_n \eta,$$

where $p_{t_k^-}$ denotes the distribution of $\Theta_{t_k}^-$. Now we note that as $t \rightarrow t_k^+$, both Θ_t and Θ'_t are subjected to the same mapping; hence, by the data processing inequality and the lower semicontinuity of the Rényi divergence, we have

$$\lim_{s \rightarrow t_{k+1}^-} R_\beta(p_s, p'_s) \leq \lim_{s \rightarrow t_k^-} R_\beta(p_s, p'_s) + K_n \eta.$$

By recursion, we deduce that for all $N \in \mathbb{N}$,

$$R_\beta(p_{N\eta}, p'_{N\eta}) \leq \frac{K_{\alpha,d}(\beta-1)\mathcal{S}_{g,C}^2}{\sigma^\alpha n^2} N\eta.$$

Case 2: $K_n < a$. We show by recursion on $k \in \mathbb{N}$ that

$$\lim_{s \rightarrow t_k^-} R_\beta(p_s, p'_s) \leq \log\left(\frac{a}{a - K_n}\right). \quad (28)$$

For $k = 0$, we have $R_\beta(p_s, p'_s) = 0 < \log\left(\frac{a}{a - K}\right)$. Therefore, we apply again the local version of Lemma 26 (see Remark 27) to obtain,

$$\lim_{s \rightarrow t_1^-} R_\beta(p_s, p'_s) \leq \log\left(\frac{a}{a - K_n}\right).$$

Now let us assume that Equation (28) is true for some $k \in \mathbb{N}$. By the data processing inequality and the lower semicontinuity of the Rényi divergence, we have that

$$R_\beta(p_{t_k^-}, p'_{t_k^-}) \leq \lim_{s \rightarrow t_k^-} R_\beta(p_s, p'_s) \leq \log\left(\frac{a}{a - K_n}\right).$$

Therefore, we can apply Lemma 26 to obtain that

$$\lim_{s \rightarrow t_{k+1}^-} R_\beta(p_s, p'_s) \leq \log\left(\frac{a}{a - K_n}\right).$$

We conclude the proof by lower semicontinuity of the Rényi divergence. \blacksquare

We end this section by giving the multifractal version of the above theorem. This is an immediate application of the previous proof and the results of Section 3.3.

Theorem 34 Consider the same setting as in Section 4 but replace Equation (19) by

$$\begin{cases} X_{k+1} = \Pi_C(X_k - \eta \hat{g}_S(X_k, \Omega_k) + \sigma_\alpha \eta^{1/\alpha} \xi_k + \sigma_2 \sqrt{2\eta} \zeta_k), \\ X'_{k+1} = \Pi_C(X'_k - \eta \hat{g}_{S'}(X'_k, \Omega_k) + \sigma_\alpha \eta^{1/\alpha} \xi_k + \sigma_2 \sqrt{2\eta} \zeta_k) \end{cases} \quad (29)$$

with the same notations and $(\zeta_k)_{k \in \mathbb{N}} \sim \mathcal{N}(0, I_d)^{\otimes \mathbb{N}}$, independent of $(\xi_k)_{k \in \mathbb{N}}$. Assume that Assumptions 2 to 4 hold and that, for all $t > 0$, p'_t satisfies an α -stable Poincaré inequality with constant $(\gamma\sigma^\alpha, \gamma\sigma_2^2)$ for some $\gamma > 0$. Finally assume that $t \mapsto R_\beta(p_t, p'_t)$ is right-continuous. Then,

$$R_\beta(\text{Law}(X_k), \text{Law}((X'_k))) \leq \frac{\beta \mathcal{S}_{g,C}^2}{2\sigma_2^2 n^2} k\eta := K_n k\eta, \quad k \in \mathbb{N}.$$

If moreover $K_n < (\gamma\beta)^{-1}$, then,

$$R_\beta(\text{Law}(X_k), \text{Law}((X'_k))) \leq -\log\left(1 - \frac{\gamma \mathcal{S}_{g,C}^2 \beta^2}{2\sigma_2^2 n^2}\right), \quad k \in \mathbb{N}.$$

B.5 Omitted proofs of Section 4.2

In this section, we use the following classical notation for the variance with respect to a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, for any ν -square-integrable function f , $\text{Var}_\mu(f) := \int f^2 d\mu - (\int f d\mu)^2$. Equipped with this notation, we can prove the results of Section 4.2, starting with the stability properties of the fractional Poincaré inequalities.

B.5.1 PROOF OF THE STABILITY RESULTS FOR α -STABLE POINCARÉ INEQUALITIES

Proof of Lemma 19. Let f be a \mathcal{C}^1 function vanishing at infinity. Then, we have

$$\begin{aligned} \text{Var}_m(f) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x+y)^2 d\mu(x) d\mu'(y) - \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x+y) d\mu(x) d\mu'(y) \right)^2 \\ &= \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} f(x+y)^2 d\mu(x) - \left(\int_{\mathbb{R}^d} f(x+y) d\mu(x) \right)^2 \right\} d\mu'(y) \\ &\quad + \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x+y) d\mu(x) \right)^2 d\mu'(y) - \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x+y) d\mu(x) d\mu'(y) \right)^2 \\ &=: A_1 + A_2. \end{aligned}$$

We can apply the Poincaré's inequality for μ and Tonelli's theorem to obtain that

$$\begin{aligned} A_1 &\leq \gamma_1 C_{\alpha,d} \int_{\mathbb{R}^d} \left\{ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} (f(x+y) - f(x+y+z))^2 \frac{dz}{\|z\|^{d+\alpha}} d\mu(x) \right\} d\mu'(y) \\ &\quad + \gamma_2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \|\nabla f(x+y)\|^2 d\mu(x) d\mu'(y) \\ &= \gamma_1 C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} (f(x) - f(x+z))^2 \frac{dz}{\|z\|^{d+\alpha}} dm(x) + \gamma_2 \int_{\mathbb{R}^d} \|\nabla f(x)\|^2 dm(x). \end{aligned}$$

Now, we apply the Poincaré inequality for μ' on the function $y \mapsto \int_{\mathbb{R}^d} f(x+y) d\mu(x)$. This provides the following bound,

$$\begin{aligned} A_2 &\leq \gamma'_1 C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} \left\{ \int_{\mathbb{R}^d} f(x+y) d\mu(x) - \int_{\mathbb{R}^d} f(x+z+y) d\mu(x) \right\}^2 \frac{dz}{\|z\|^{d+\alpha}} d\mu'(y) \\ &\quad + \gamma'_2 \int_{\mathbb{R}^d} \left\{ \left\| \nabla_y \int_{\mathbb{R}^d} f(x+y) d\mu(x) \right\|^2 \right\} d\mu'(y). \end{aligned}$$

Therefore, by Jensen's inequality,

$$\begin{aligned} A_2 &\leq \gamma'_1 C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} (f(x+y) - f(x+z+y))^2 d\mu(x) \frac{dz}{\|z\|^{d+\alpha}} d\mu'(y) \\ &\quad + \gamma'_2 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \|\nabla f(x+y)\|^2 d\mu(x) d\mu'(y) \\ &= \gamma'_1 C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} (f(x) - f(x+z))^2 \frac{dz}{\|z\|^{d+\alpha}} dm(x) + \gamma'_2 \int_{\mathbb{R}^d} \|\nabla f(x)\|^2 dm(x). \end{aligned}$$

This gives the desired result. ■

Proof of Lemma 21. Let f be a \mathcal{C}^1 function vanishing at infinity. Using the change of variable formula, the Poincaré inequality and Tonelli's theorem, we have

$$\begin{aligned}\text{Var}_{T\#\mu}(f) &= \int_{\mathbb{R}^d} f^2 d(T\#\mu) - \left(\int_{\mathbb{R}^d} f d(T\#\mu) \right)^2 = \int_{\mathbb{R}^d} (f \circ T)^2 d\mu - \left(\int_{\mathbb{R}^d} f \circ T d\mu \right)^2 \\ &\leq \gamma_1 C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} (f(T(x)) - f(T(x+z)))^2 \frac{dz}{\|z\|^{d+\alpha}} d\mu(x) \\ &\quad + \gamma_2 \int_{\mathbb{R}^d} \|\nabla(f \circ T)\|^2(x) d\mu(x) \\ &= \gamma_1 C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{x\}} (f(T(x)) - f(T(y)))^2 \frac{dy}{\|x-y\|^{d+\alpha}} d\mu(x) \\ &\quad + \gamma_2 \int_{\mathbb{R}^d} \|\text{Jac}(T)(x)\|_2^2 \|\nabla f(T(x))\|^2 d\mu(x).\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Var}_{T\#\mu}(f) &\leq \gamma_1 C_{\alpha,d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{v\}} (f(u) - f(v))^2 \\ &\quad \cdot \frac{|\det \text{Jac}(T^{-1})(u)| |\det \text{Jac}(T^{-1})(v)|}{\|T^{-1}(u) - T^{-1}(v)\|^{d+\alpha}} \mu(T^{-1}(u)) du dv \\ &\quad + \gamma_2 L_2^2 \int_{\mathbb{R}^d} \|\nabla f(T(x))\|^2 d\mu(x) \\ &\leq \gamma_1 C_{\alpha,d} L_2^{\alpha+d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{v\}} (f(u) - f(v))^2 \\ &\quad \cdot \frac{|\det \text{Jac}(T^{-1})(u)| |\det \text{Jac}(T^{-1})(v)|}{\|u-v\|^{d+\alpha}} \mu(T^{-1}(u)) du dv \\ &\quad + \gamma_2 L_2^2 \int_{\mathbb{R}^d} \|\nabla f(T(x))\|^2 d\mu(x) \\ &= \gamma_1 C_{\alpha,d} L_2^{\alpha+d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{T(x)\}} (f(T(x)) - f(v))^2 \frac{|\det \text{Jac}(T^{-1})(v)|}{\|T(x) - v\|^{d+\alpha}} \mu(x) dv dx \\ &\quad + \gamma_2 L_2^2 \int_{\mathbb{R}^d} \|\nabla f(T(x))\|^2 d\mu(x).\end{aligned}$$

Now we note that T^{-1} is $(1/L_1)$ -Lipschitz continuous, so that, by Hadamard's inequality (Holland, 2007), we have $|\det \text{Jac}(T^{-1})(v)| \leq \prod_{i=1}^d \|\nabla(T^{-1})_i(v)\| \leq \frac{1}{L_1^d}$. Therefore,

$$\begin{aligned}\text{Var}_{T\#\mu}(f) &\leq \gamma_1 C_{\alpha,d} \frac{L_2^{\alpha+d}}{L_1^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{T(x)\}} (f(T(x)) - f(v))^2 \frac{dv}{\|T(x) - v\|^{d+\alpha}} \mu(x) dx \\ &\quad + \gamma_2 L_2^2 \int_{\mathbb{R}^d} \|\nabla f(T(x))\|^2 d\mu(x) \\ &\leq \gamma_1 C_{\alpha,d} \frac{L_2^{\alpha+d}}{L_1^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \setminus \{0\}} (f(T(x)) - f(T(x)+z))^2 \frac{dz}{\|z\|^{d+\alpha}} \mu(x) dx \\ &\quad + \gamma_2 L_2^2 \int_{\mathbb{R}^d} \|\nabla f(T(x))\|^2 d\mu(x),\end{aligned}$$

which yields the desired result. \blacksquare

Proof of Lemma 22. Letting $\Phi(x) := x^2$, we have the classical variational formula (Bakry et al., 2014):

$$\text{Var}_\nu(f) = \inf_{a \in \mathbb{R}} \int_{\mathbb{R}^d} D_\Phi(f(x), a) d\nu(x).$$

Therefore:

$$\begin{aligned} & \text{Var}_{\mu'}(f) \\ & \leq e^b \inf_{a \in \mathbb{R}} \int_{\mathbb{R}^d} D_\Phi(f(x), a) d\mu(x) = e^b \text{Var}_\mu(f) \\ & \leq e^b \left(\gamma_1 C_{\alpha,d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} \frac{(f(x) - f(x+z))^2}{\|z\|^{d+\alpha}} d\mu(x) dz + \gamma_2 \int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu(x) \right) \\ & \leq e^{2b} \left(\gamma_1 C_{\alpha,d} \int_{\mathbb{R}^d \setminus \{0\}} \int_{\mathbb{R}^d} \frac{(f(x) - f(x+z))^2}{\|z\|^{d+\alpha}} d\mu'(x) dz + \gamma_2 \int_{\mathbb{R}^d} \|\nabla f(x)\|^2 d\mu'(x) \right), \end{aligned}$$

which concludes the proof. \blacksquare

Proof of Proposition 23. Let us denote $T(x) := x - \eta \nabla \widehat{\mathcal{R}}_S(x)$. Using the strong convexity and smoothness assumptions, we classically prove that $T : \mathbb{R}^d \rightarrow \mathbb{R}$ is bi-Lipschitz with constants given by

$$(1 - \eta M) \|x - y\| \leq \|T(x) - T(y)\| \leq (1 - \eta \lambda) \|x - y\|, \quad x, y \in \mathbb{R}^d.$$

Now let X be a random variable whose probability distribution satisfies an α -stable Poincaré inequality with constant $(c, 0)$ and let $s \in (0, \eta)$. Then, by Lemmas 19 and 21, the distribution of $T(X) + s^{1/\alpha} \sigma L_1^\alpha$ satisfies an α -stable Poincaré inequality with constant $(c', 0)$, with

$$c' := c \frac{(1 - \eta \lambda)^{\alpha+d}}{(1 - \eta M)^d} + s \sigma^\alpha =: F(\eta)c + s \sigma^\alpha.$$

We easily see that $F : (0, M^{-1}) \rightarrow \mathbb{R}_+$ satisfies $F(0^+) = 1$ and $F(\eta) \rightarrow +\infty$ as $\eta \rightarrow M^{-1}$. Moreover, $F'(\eta_0) = 0$ if and only if $\eta_0 = \frac{(\alpha+d)\lambda - dM}{\alpha\lambda M}$. As $\lambda \leq M$, we have $\eta_0 < 1/M$, and, on the other hand, Equation (20) ensures that $\eta_0 > 0$. Therefore, we necessarily have $0 < F(\eta_0) < 1$. Now we observe that $c \leq \frac{\eta \sigma^\alpha}{1 - F(\eta_0)}$ implies $F(\eta)c + \eta \sigma^\alpha \leq \frac{\eta \sigma^\alpha}{1 - F(\eta_0)}$. Therefore, we have

$$c' \leq F(\eta)c + \eta \sigma^\alpha \leq \frac{\eta \sigma^\alpha}{1 - F(\eta_0)}.$$

The result follows by an immediate recursion. \blacksquare

References

Martín 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, Vienna, Austria, October 2016.

Jason Altschuler and Kunal Talwar. Privacy of Noisy Stochastic Gradient Descent: More Iterations without More Privacy Loss. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 3788–3800. Curran Associates, Inc., 2022.

Hilal Asi, Daogao Liu, and Kevin Tian. Private Stochastic Convex Optimization with Heavy Tails: Near-Optimality from Simple Reductions. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 59174–59215. Curran Associates, Inc., 2024.

Shahab Asoodeh and Mario Diaz. Privacy Loss of Noisy Stochastic Gradient Descent Might Converge Even for Non-Convex Losses. *arXiv:2305.09903*, 2023.

Shahab Asoodeh, Jiachun Liao, Flavio P. Calmon, Oliver Kosut, and Lalitha Sankar. Three Variants of Differential Privacy: Lossless conversion and applications. *IEEE Journal on Selected Areas in Information Theory*, 2(1):208–222, 2021.

Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer, 2014.

Melih Barsbey, Milad Sefidgaran, Murat A. Erdogdu, Gaël Richard, and Umut Şimşekli. Heavy Tails in SGD and Compressibility of Overparametrized Neural Networks. In *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, volume 34, pages 29364–29378. Curran Associates, Inc., June 2021.

Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 464–473, Philadelphia, PA, USA, 2014.

Robert M. Blumenthal and Ronald K. Getoor. Some Theorems on Stable Processes. *Transactions of the American Mathematical Society*, 95(2):263–273, 1960.

Björn Böttcher, René Schilling, and Jian Wang. *Lévy Matters III: Lévy-Type Processes: Construction, Approximation and Sample Path Properties*, volume 2099 of *Lecture Notes in Mathematics*. Springer International Publishing, Cham, 2013.

Jean Bourgain, Haïm Brezis, and Petru Mironescu. Another Look at Sobolev spaces. In J. L. Menaldi, E. Rofman, and A. Sulem, editors, *Optimal Control and Partial Differential Equations*, pages 439–455. IOS Press, 2001.

Mark Bun and Thomas Steinke. Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. In *Proceedings, Part I, of the 14th International Conference on Theory of Cryptography - Volume 9985*, pages 635–658, Berlin, Heidelberg, 2016. Springer-Verlag.

Yu Cao, Jianfeng Lu, and Yulong Lu. Exponential Decay of Rényi Divergence under Fokker-Planck Equations. *Journal of Statistical Physics*, 176(5):1172–1184, September 2019.

- Djalil Chafaï. Entropies, Convexity, and Functional Inequalities, On Φ -Entropies and Φ -Sobolev Inequalities. *Kyoto Journal of Mathematics*, 44(2):325–363, January 2004.
- Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially Private Empirical Risk Minimization. *Journal of Machine Learning Research*, 12:1069–1109, February 2011.
- Xiangyi Chen, Steven Z. Wu, and Mingyi Hong. Understanding Gradient Clipping in Private SGD: A Geometric Perspective. In *Advances in Neural Information Processing Systems*, volume 33, pages 13773–13782. Curran Associates, Inc., 2020.
- Zhen-Qing Chen, Eryan Hu, Longjie Xie, and Xicheng Zhang. Heat Kernels for Non-Symmetric Diffusion Operators with Jumps. *Journal of Differential Equations*, 263(10):6576–6634, 2017.
- Eli Chien, Haoyu Wang, Ziang Chen, and Pan Li. Langevin Unlearning: A New Perspective of Noisy Gradient Descent for Machine Unlearning. In *Advances in Neural Information Processing Systems*, volume 37, pages 79666–79703. Curran Associates, Inc., 2024.
- Rishav Chourasia, Jiayuan Ye, and Reza Shokri. Differential Privacy Dynamics of Langevin Diffusion and Noisy Gradient Descent. In *Advances in Neural Information Processing Systems*, volume 34, pages 14771–14781. Curran Associates, Inc., 2021.
- Cristina Costantini, Emmanuel Gobet, and Antonin Zadourian. Reflected Jump-Diffusions on the Half-Line. *Annals of Applied Probability*, 15(4):2701–2721, 2005.
- Maha Daoud and El Haj Laamri. Fractional Laplacians: A Short Survey. *Discrete and Continuous Dynamical Systems*, 15(1):95–116, 2022.
- Jinqiao Duan. *An Introduction to Stochastic Dynamics*. Cambridge Texts in Applied Mathematics, 2015.
- Benjamin Dupuis and Umut Şimşekli. Generalization Bounds for Heavy-Tailed SDEs through the Fractional Fokker-Planck Equation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 12087–12137. PMLR, July 2024.
- Cynthia Dwork. Differential Privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming*, pages 1–12, Berlin, Heidelberg, 2006. Springer.
- Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2013.
- Cynthia Dwork and Guy N. Rothblum. Concentrated Differential Privacy. *arXiv:1603.01887*, March 2016.
- 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, Las Vegas, NV, USA, 2010.

Vitaly Feldman, Ilya Mironov, Kunal Talwar, and Abhradeep Thakurta. Privacy Amplification by Iteration. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 521–532, Paris, France, October 2018.

Arun Ganesh and Kunal Talwar. Faster Differentially Private Samplers via Rényi Divergence Analysis of Discretized Langevin MCMC. In *Advances in Neural Information Processing Systems*, volume 33, pages 7222–7233. Curran Associates, Inc., 2020.

Walter Gautschi. Some Elementary Inequalities Relating to the Gamma and Incomplete Gamma Function. *Journal of Mathematical Physics*, 38(1):77–81, 1959.

Ivan Gentil and Cyril Imbert. Logarithmic Sobolev Inequalities: Regularizing Effect of Lévy Operators and symptotic Convergence in the Lévy-Fokker-Planck Equation. *Stochastics*, 81(3-4):401–414, 2009.

Leonard Gross. Logarithmic Sobolev Inequalities. *American Journal of Mathematics*, 97(4): 1061–1083, 1975.

Mert Gürbüzbalaban, Umut Şimşekli, and Lingjiong Zhu. The Heavy-Tail Phenomenon in SGD. In *International Conference on Machine Learning (ICML 2021)*, volume 139, pages 3964–3975. PMLR, 2021.

Ye He, Alireza Mousavi-Hosseini, Krishnakumar Balasubramanian, and Murat A. Erdogdu. A Separation in Heavy-Tailed Sampling: Gaussian vs. Stable Oracles for Proximal Samplers. In *Advances in Neural Information Processing Systems*, volume 37, pages 65260–65296. Curran Associates, Inc., December 2024.

Finbarr Holland. Another Proof of Hadamard’s Determinantal Inequality. *Irish Mathematical Society Bulletin*, 59:61–64, 2007.

Richard Holley and Daniel Stroock. Logarithmic Sobolev Inequalities and Stochastic Ising Models. *Journal of Statistical Physics*, 46:1159–1194, 1987.

Kaito Ito, Yu Kawano, and Kenji Kashima. Privacy Protection with Heavy-Tailed Noise for Linear Dynamical Systems. *Automatica*, 131:109732, 2021.

Nurdan Kuru, Ş İlker Birbil, Mert Gurbuzbalaban, and Sinan Yildirim. Differentially Private Accelerated Optimization Algorithms. *SIAM Journal on Optimization*, 32(2):795–821, June 2022.

Soon Hoe Lim, Yijun Wan, and Umut Şimşekli. Chaotic Regularization and Heavy-Tailed Limits for Deterministic Gradient Descent. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 26590–26602. Curran Associates, Inc., 2022.

Pierre-Louis Lions and Alain-Sol Sznitman. Stochastic Differential Equations with Reflecting Boundary Conditions. *Communications on Pure and Applied Mathematics*, 37(4):511–537, 1984.

- Jose-Luis Menaldi. Reflected Diffusion Processes with Jumps. In *Stochastic Differential Systems*, pages 170–186. Springer, 1985.
- Ilya Mironov. Renyi Differential Privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275, Santa Barbara, CA, USA, August 2017.
- Clément Mouhot, Emmanuel Russ, and Yannick Sire. Fractional Poincaré Inequalities for General Measures. *Journal de Mathématiques Pures et Appliquées*, 95(1):72–84, 2011.
- Eleonora Di Nezza, Giampiero Palatucci, and Enrico Valdinoci. Hitchhiker’s Guide to the Fractional Sobolev Spaces. *Bulletin des Sciences Mathématiques*, 136(5):521–573, 2012.
- Francisco J. Piera, Ravi R. Mazumdar, and Fabrice M. Guillemin. Boundary Behavior and Product-Form Stationary Distributions of Jump Diffusions in the Orthant with State-Dependent Reflections. *Advances in Applied Probability*, 40(4):913–956, 2008.
- Anant Raj, Melih Barsbey, Mert Gürbüzbalaban, Lingjiong Zhu, and Umut Şimşekli. Algorithmic Stability of Heavy-Tailed Stochastic Gradient Descent on Least Squares. In *Proceedings of The 34th International Conference on Algorithmic Learning Theory*, volume 201, pages 1292–1342. PMLR, 2023a.
- Anant Raj, Lingjiong Zhu, Mert Gürbüzbalaban, and Umut Şimşekli. Algorithmic Stability of Heavy-Tailed SGD with General Loss Functions. In *International Conference on Machine Learning (ICML 2023)*, volume 202, pages 28578–28597. PMLR, 2023b.
- Théo Ryffel, Francis Bach, and David Pointcheval. Differential Privacy Guarantees for Stochastic Gradient Langevin Dynamics. *arXiv:2201.11980*, 2022.
- Gennady Samorodnitsky and Mircea Grigoriu. Tails of Solutions of Certain Nonlinear Stochastic Differential Equations Driven by Heavy Tailed Lévy Motions. *Stochastic Processes and their Applications*, 105(1):69–97, May 2003.
- René L. Schilling. An Introduction to Lévy and Feller Processes. Advanced Courses in Mathematics - CRM Barcelona 2014. *arXiv:1603.00251*, 2016.
- Umut Simsekli, Levent Sagun, and Mert Gürbüzbalaban. A Tail-Index Analysis of Stochastic Gradient Noise in Deep Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, volume 97, pages 5827–5837. PMLR, 2019.
- Umut Şimşekli, Mert Gürbüzbalaban, Sinan Yıldırım, and Lingjiong Zhu. Privacy of SGD under Gaussian or Heavy-Tailed Noise: Guarantees without Gradient Clipping. *arXiv:2403.02051*, 2024.
- Leszek Ślomiński. Stochastic Differential Equations with Jump Reflection at Time-Dependent Barriers. *Stochastic Processes and their Applications*, 120(6):935–966, 2010.
- Isabelle Tristani. Fractional Fokker-Planck Equation. *Communications in Mathematical Sciences*, 13(5):1243–1260, 2013.

- Sabir Umarov, Marjorie Hahn, and Kei Kobayashi. *Beyond the Triangle: Brownian Motion, Ito Calculus and Fokker-Planck Equation - Fractional Generalizations*. World Scientific Publishing, 2018.
- Tim van Erven and Peter Harremoës. Rényi Divergence and Kullback-Leibler Divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, July 2014.
- Yijun Wan, Melih Barsbey, Abdellatif Zaidi, and Umut Simsekli. Implicit Compressibility of Overparametrized Neural Networks Trained with Heavy-Tailed SGD. <https://arxiv.org/abs/2306.08125v2>, June 2023.
- Di Wang, Minwei Ye, and Jinhui Xu. Differentially Private Empirical Risk Minimization Revisited: Faster and More General. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- Feng-Yu Wang and Jian Wang. Functional Inequalities for Stable-Like Dirichlet Forms. *Journal of Theoretical Probability*, 28(2):423–448, June 2015.
- Hongjian Wang, Mert Gürbüzbalaban, Lingjiong Zhu, Umut Şimşekli, and Murat A. Erdogdu. Convergence Rates of Stochastic Gradient Descent under Infinite Noise Variance. In *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, volume 34, pages 18866–18877. Curran Associates, Inc., 2021.
- Liming Wu. A New Modified Logarithmic Sobolev Inequality for Poisson Point Processes and Several Applications. *Probability Theory and Related Fields*, 118:427–438, 2000.
- Jiayuan Ye and Reza Shokri. Differentially Private Learning Needs Hidden State (Or Much Faster Convergence). In *Advances in Neural Information Processing Systems*, volume 35, pages 703–715. Curran Associates, Inc., 2022.
- Lei Yu, Ling Liu, Calton Pu, Mehmet Emre Gursoy, and Stacey Truex. Differentially Private Model Publishing for Deep Learning. In *2019 IEEE Symposium on Security and Privacy (SP)*, pages 332–349, San Francisco, CA, USA, May 2019.
- Christopher C. Zawacki and Eyad H. Abed. Heavy-Tailed Privacy: The Symmetric alpha-Stable Privacy Mechanism. *arXiv:2504.18411*, April 2025.