

DEEP RELU NETWORK EXPRESSION RATES FOR OPTION PRICES IN HIGH-DIMENSIONAL, EXPONENTIAL LÉVY MODELS

LUKAS GONON AND CHRISTOPH SCHWAB

ABSTRACT. We study the expression rates of deep neural networks (DNNs for short) for option prices written on baskets of d risky assets, whose log-returns are modelled by a multivariate Lévy process with general correlation structure of jumps. We establish sufficient conditions on the characteristic triplet of the Lévy process X that ensure ε error of DNN expressed option prices with DNNs of size that grows polynomially with respect to $\mathcal{O}(\varepsilon^{-1})$, and with constants implied in $\mathcal{O}(\cdot)$ which grow polynomially with respect d , thereby overcoming the curse of dimensionality and justifying the use of DNNs in financial modelling of large baskets in markets with jumps.

In addition, we exploit parabolic smoothing of Kolmogorov partial integrodifferential equations for certain multivariate Lévy processes to present alternative architectures of ReLU DNNs that provide ε expression error in DNN size $\mathcal{O}(|\log(\varepsilon)|^a)$ with exponent $a \sim d$, however, with constants implied in $\mathcal{O}(\cdot)$ growing exponentially with respect to d . Under stronger, dimension-uniform non-degeneracy conditions on the Lévy symbol, we obtain algebraic expression rates of option prices in exponential Lévy models which are free from the curse of dimensionality. In this case the ReLU DNN expression rates of prices depend on certain sparsity conditions on the characteristic Lévy triplet. We indicate several consequences and possible extensions of the present results.

CONTENTS

1. Introduction	2
2. Exponential Lévy models and PIDEs	4
2.1. Lévy processes	4
2.2. Exponential Lévy models	4
2.3. PIDEs for option prices	5
3. Deep neural networks (DNNs)	5
4. DNN approximations for univariate Lévy models	7
4.1. DNN expression rates: probabilistic argument	8
4.2. DNN expression of European calls	11
4.3. ReLU DNN exponential expressivity	12
5. DNN approximation rates for multivariate Lévy models	15
5.1. DNN expression rate bounds via probabilistic argument	15
5.2. Discussion of related results	20
5.3. Exponential ReLU DNN expression rates via PIDE	21
5.4. Remarks on the Curse of Dimensionality	24
6. Conclusion and Generalizations	29
References	30

1. INTRODUCTION

Recent years have seen a dynamic development in applications of deep neural networks (DNNs for short) in expressing high-dimensional input-output relations. This development was driven mainly by the need for quantitative modelling of input-output relationships subject to large sets of observation data. Rather naturally, therefore, DNNs have found a large number of applications in computational finance and in financial engineering. We refer to the survey [RW20] and to the references there. Without going into details, we only state that the majority of activity addresses techniques to employ DNNs in demanding tasks in computational finance. The often striking efficient computational performance of DNN based algorithms raises naturally the question for theoretical, in particular mathematical, underpinning of successful algorithms. Recent years have seen progress, in particular in the context of option pricing for Black–Scholes type models, for DNN based numerical approximation of diffusion models on possibly large baskets (see, e.g. [BGJ18, EGJS18] and [IRZ19, RZ19] for game-type options). These references prove that DNN based approximations of option prices on possibly large baskets of risky assets can overcome the so-called curse of dimensionality in the context of affine diffusion models for the dynamics of the (log-)prices of the underlying risky assets. These results could be viewed also as particular instances of DNN expression rates of certain PDEs on high-dimensional state spaces, and indeed corresponding DNN expressive power results have been shown for their solution sets in [GHJvW18], [GGJ⁺19] and the references there.

Since the turn of the century, models beyond the classical diffusion setting have been employed increasingly in financial engineering. In particular, Lévy processes and their nonstationary generalizations such as Feller–Lévy processes (see, e.g., [BSW13] and the references there) have received wide attention. This can in part be explained by their ability to account for heavy tails of financial data and by Lévy–based models constituting *hierarchies* of models, comprising in particular classical diffusion (“Black–Scholes”) models with constant volatility that are still widely used in computational finance as a benchmark. Therefore, all results for geometric Lévy processes in the present paper apply in particular to the Black–Scholes model.

The “Feynman Kac correspondence” which relates conditional expectations of sufficiently regular functionals over diffusions to (viscosity) solutions of corresponding Kolmogorov PDEs, extends to multivariate Lévy processes. We mention only [NS01, CT04, CV05, Gla16], [EK19, Chap. 5.4] and the references there. The Kolmogorov PDE (“Black–Scholes equation”) in the diffusion case is then replaced by a so-called *Partial Integro-differential Equation (PIDE)* where the fractional integrodifferential operator accounting for the jumps is related in a one-to-one fashion with the Lévy measure ν^d of the LP $X^d \subset \mathbb{R}^d$. In particular, Lévy type models for (log-)returns of risky assets result in *nonlocal partial integrodifferential equations* for the option price, which generalize the linear parabolic differential equations which arise in classical diffusion models. We refer to [Ber96, Sat99] for fundamentals on Lévy processes and to [BSW13] for extensions to certain nonstationary settings. For the use of Lévy processes in financial modelling we refer to [CT04, EK19] and to the references there. We refer to [CV05, CV06, MvPS04, HRSW13] for a presentation and for numerical methods for option pricing in Lévy models.

The results on DNNs in the context of option pricing mentioned above are exclusively concerned with models with continuous price processes. This naturally raises the question whether DNN based approximations are still capable of overcoming the curse of dimensionality in high-dimensional financial models with jumps, which have a much richer mathematical structure. This question is precisely the subject of this article. We study the expression rates of DNNs for prices of options (and the associated PIDEs) written on possibly large baskets of risky assets, whose log-returns are modelled by a multivariate Lévy process with general correlation structure of jumps. In particular,

we establish sufficient conditions on the characteristic triplet of the Lévy process X^d that ensure ε error of DNN expressed option prices with DNNs of size $\mathcal{O}(\varepsilon^{-2})$, and with constants implied in $\mathcal{O}(\cdot)$ which grow polynomially with respect d . This shows that DNNs are capable to overcome the curse of dimensionality also for general exponential Lévy models.

Scope of results. The DNN expression rate results proved here give a theoretical justification for neural network based non-parametric option pricing methods. These have become very popular recently, see for instance the recent survey [RW20]. Our results show that if option prices result from an exponential Lévy model, as described e.g. in [EK19, Chap.3.7], under mild conditions on the Lévy-triplets these prices can be expressed efficiently by (ReLU) neural networks, also for high dimensions. The result covers, in particular, rather general, multivariate correlation structure in the jump part of the Lévy process, for example parametrized by a so-called *Lévy copula*, see [KT06, FRS07] [EK19, Chap.8.1] and the references there. This extends, at least to some extent, the theoretical foundation to the widely used neural network based non-parametric option pricing methodologies to market models with jumps.

We prove two types of results on DNN expression rate bounds for European options in exponential Lévy models, with one probabilistic and one “deterministic” proof. The former one is based on concepts from statistical learning theory, and provides for relevant payoffs (baskets, call on max, ...) an expression error $\mathcal{O}(\varepsilon)$ with DNN sizes of $\mathcal{O}(\varepsilon^{-2})$, whereas the latter is based on parabolic smoothing of the Kolmogorov equation, and allows us to prove *exponential expressivity of prices for positive maturities*.

For the latter approach certain non-degeneracy is required on the symbol of the underlying Lévy process. The probabilistic proof of DNN approximation rate results, on the other hand, does not require any such assumptions. It only relies on the additive structure of the semigroup associated to the Lévy process and existence of moments. Thus, the results proved here are specifically tailored to the class of option pricing functions (or more generally expectations of exponential Lévy processes) under European style, plain vanilla payoffs.

The structure of this paper is as follows. In Section 2 we review terminology, basic results, and financial modelling with exponential Lévy processes. In particular, we also recapitulate the corresponding fractional, partial integrodifferential Kolmogorov equations which generalize the classical Black–Scholes equations to Lévy models. Section 3 recapitulates notation and basic terminology for deep neural networks to the extent required in the ensuing expression rate analysis. We focus mainly on so-called ReLU DNNs, but add that corresponding definitions and also results do hold for more general activation functions. In Section 4 we present a first set of DNN expression rate results, still in the univariate case. This is, on the one hand, for presentation purposes, as this setting allows for lighter notation, and to introduce mathematical concepts which will be used subsequently also for contracts on possibly large basket of Lévy-driven risky assets. We also present an application of the results to neural-network based call option pricing. Section 5 then has the main results of the present paper: expression rate bounds for ReLU DNNs for multivariate, exponential Lévy models. We identify sufficient conditions to obtain expression rates which are free from the curse of dimensionality via mathematical tools from statistical learning theory. We also develop a second argument based on parabolic Gevrey regularity with quantified derivative bounds, which even yield exponential expressivity of ReLU DNNs, albeit with constants that generally depend on the basket size in a possibly exponential way. Finally, we develop an argument based on quantified sparsity in polynomial chaos expansions and corresponding ReLU expression rates from [SZ19] to prove high algebraic expression rates for ReLU DNNs, with constants that are independent of the basket size. We also provide a brief discussion of recent, related results. We conclude in Section 6 and indicate several possible generalizations of the present results.

2. EXPONENTIAL LÉVY MODELS AND PIDEs

2.1. Lévy processes. Fix a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random elements are defined.

We start with the univariate case. We recall that an \mathbb{R} -valued continuous-time process $(X_t)_{t \geq 0}$ is called a Lévy process if it is stochastically continuous, it has almost surely RCLL sample paths, it satisfies $X_0 = 0$ almost surely, and it has stationary and independent increments. See, e.g. [Ber96, Sat99] for discussion and for detailed statements of definitions.

It is shown in these references that a Lévy process (LP for short) $X \subset \mathbb{R}$ is characterized by its so-called Lévy triplet (σ^2, γ, ν) , where $\sigma \geq 0$, $\gamma \in \mathbb{R}$ and where ν is a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $\nu(\{0\}) = 0$, the so-called *jump-measure*, or *Lévy-measure* of the LP X which satisfies $\int_{\mathbb{R}} (x^2 \wedge 1) \nu(dx) < \infty$. For more details on both, univariate LPs $X \subset \mathbb{R}$, and for the multivariate situation we refer to [Sat99].

As in the univariate case, multivariate LPs $X^d \subset \mathbb{R}^d$ are completely described by their characteristic triplet (A^d, γ^d, ν^d) where $\gamma^d \in \mathbb{R}^d$ is a drift vector, $A^d \in \mathbb{R}^{d \times d}$ is a symmetric, nonnegative definite matrix denoting the covariance matrix of the Brownian motion part of X^d , and ν^d is the Lévy measure describing the jump structure of X^d .

To characterize the dependence structure of a Lévy process the drift parameter γ^d does not play a role. The dependence structure of the diffusion part of X^d is characterized by A^d . Since the continuous part and the jump part of X^d are stochastically independent, the dependence structure of the jump part of X^d is characterized by the Lévy measure ν^d .

In [KT06], a characterization of admissible jump measures ν^d of the LP $X^d \subset \mathbb{R}^d$ has been obtained as *superposition* of marginal, univariate Lévy measures with a so-called *Lévy copula* function.

2.2. Exponential Lévy models. In a widely used class of financial models, log-returns of risky assets are modelled by LPs. This leads to price processes which are exponential Lévy models, generalizing geometric Brownian motion.

Let the stochastic process $(S_t)_{t \in [0, T]} \subset \mathbb{R}$ model the price of one risky financial asset. Here $T \in (0, \infty)$ is a fixed, finite time horizon. An *exponential Lévy model* assumes that $S_t = s_0 e^{rt + X_t}$, $t \in [0, T]$, where $r \in \mathbb{R}$ denotes the (constant) interest rate. The model could be specified either under a real-world measure or directly under a risk-neutral measure (constructed using the general change of measure result in [Sat99, Theorems 33.1 and 33.2] of which the Esscher transform [GS94] is a particular case, or by minimizing certain functionals over the family of equivalent martingale measures, see for instance [JKM07], [ES05] and the references therein). For the purposes of this section we assume the latter situation, that is, we require that $(S_t e^{-rt})_{t \in [0, T]}$ is a martingale. This is equivalent to the following condition on the Lévy triplet of X (e.g. [HRSW13, Lem. 10.1.5])

$$(2.1) \quad \gamma = -\frac{\sigma^2}{2} - \int_{\mathbb{R}} (e^y - 1 - y \mathbb{1}_{\{|y| \leq 1\}}) \nu(dy).$$

Here, and throughout, we assume the Lévy measure ν satisfies the *exponential integrability condition*: for some $p \geq 2$,

$$(2.2) \quad \int_{|y| > 1} e^{py} \nu(dy) < \infty.$$

For a d -dimensional Lévy process X^d , [Sat99, Theorems 25.17] shows that the multivariate geometric Lévy process $(e^{X_{t,1}^d}, \dots, e^{X_{t,d}^d})_{t \geq 0}$ is a martingale if and only if

$$(2.3) \quad \int_{\mathbb{R}^d \setminus \{\|y\| \leq 1\}} e^{y_i} \nu^d(dy) < \infty, \quad \text{for } i = 1, \dots, d,$$

$$\gamma_i^d = -\frac{A_{ii}^d}{2} - \int_{\mathbb{R}^d} (e^{y_i} - 1 - y_i \mathbb{1}_{\{\|y\| \leq 1\}}) \nu^d(dy), \quad \text{for } i = 1, \dots, d.$$

This condition ensures that the functions defined in (2.4) and (5.1) below represent option prices. However, the condition is not needed for the proof of the results later, so we do not impose (2.1) or (2.3) in any of the results proved in the article.

For more details on exponential Lévy models, with particular attention to their use in financial modelling, we refer to [CT04], [LM08], and [EK19] and the references there.

2.3. PIDEs for option prices. Let us first discuss the case of a univariate exponential Lévy model. For the multivariate case we refer to Section 5 (cf. (5.1) and (5.16) below).

Consider a European style option with payoff function $\varphi: (0, \infty) \rightarrow [0, \infty)$ and at most polynomial (p -th order) growth at infinity.

The value of the option (under the chosen risk-neutral measure) at time $t \in [0, T]$ is given as the conditional expectation $C_t = \mathbb{E}[e^{-r(T-t)} \varphi(S_T) | \mathcal{F}_t]$ with $\mathcal{F}_t = \sigma(S_v : v \in [0, t])$. By the Markov property $C_t = C(t, S_t)$ and so, switching to time-to-maturity $\tau = T - t$, $u(\tau, s) = C(T - \tau, s)$ we can rewrite the option price as follows:

$$(2.4) \quad u(\tau, s) = \mathbb{E}[e^{-r\tau} \varphi(S_T) | S_t = s] = \mathbb{E}[e^{-r\tau} \varphi(s \exp(r\tau + X_\tau))]$$

for $\tau \in [0, T]$, $s \in (0, \infty)$, where the second step uses that $X_T - X_t$ is independent of X_t and has the same distribution as X_{T-t} . If the payoff function φ is Lipschitz-continuous on \mathbb{R} and the Lévy process fulfils either $\sigma > 0$ or a certain non-degeneracy condition on ν , then u is continuous on $[0, T] \times (0, \infty)$, it is $C^{1,2}$ on $(0, T) \times (0, \infty)$ and it satisfies the *linear, parabolic partial integrodifferential equation* (PIDE for short)

$$(2.5) \quad \frac{\partial u}{\partial \tau}(\tau, s) - rs \frac{\partial u}{\partial s}(\tau, s) - \frac{\sigma^2 s^2}{2} \frac{\partial^2 u}{\partial s^2}(\tau, s) - ru(\tau, s) - \int_{\mathbb{R}} \left[u(\tau, se^y) - u(\tau, s) - s(e^y - 1) \frac{\partial u}{\partial s}(\tau, s) \right] \nu(dy) = 0$$

on $[0, T] \times (0, \infty)$ with initial condition $u(0, \cdot) = \varphi$, see for instance Proposition 2 in [CV05]. If the non-degeneracy condition on ν is dropped, one can still characterize u (transformed to log-price variables) as the unique viscosity solution to the PIDE above. This is established e.g. in [CV05] (see also Proposition 3.3 in [CV06]). For our purposes the representation (2.4) is more suitable. However, by using this characterization (also called Feynman-Kac representation for viscosity-solutions of PIDEs, see [BBP97]) the results formulated below also provide DNN approximations for PIDEs. Finally, note that the interest rate r may also be directly modelled as a part of X by modifying γ . To simplify the notation we set $r = 0$ in what follows. We also remark that all expression rate results hold verbatim for assets with a constant dividend payment (see, e.g., [LM08, Eqn. (3.1)] for the functional form of the exponential Lévy model in that case).

3. DEEP NEURAL NETWORKS (DNNs)

Roughly speaking, a deep neural network (DNN for short) is a function built by multiple concatenations of affine transformations with a (typically non-linear) activation function. This gives rise

to a parametrized family of nonlinear maps, see for instance [PV18] or [BGTW19, Section 4.1] and the references there.

Here we follow current practice and refer to the collection of parameters Φ as “*the neural network*” and denote by $R(\Phi)$ its realization, that is, the function defined by these parameters. More specifically, we use the following terminology (see for instance Section 2 in [OPS20]): firstly, we fix a function $\varrho: \mathbb{R} \rightarrow \mathbb{R}$ (referred to as the activation function) which is applied componentwise to vector-valued inputs.

Definition 3.1. Let $d, L \in \mathbb{N}$. A neural network (with L layers and d -dimensional input) is a collection

$$\Phi = ((A_1, b_1), \dots, (A_L, b_L)),$$

where $N_0 := d$, $N_i \in \mathbb{N}$, $A_i \in \mathbb{R}^{N_i \times N_{i-1}}$, $b_i \in \mathbb{R}^{N_i}$ for $i = 1, \dots, L$ and (A_i, b_i) are referred to as the weights of the i -th layer of the NN.

The associated realization of Φ is the mapping

$$R(\Phi): \mathbb{R}^d \rightarrow \mathbb{R}^{N_L}, \quad x \mapsto R(\Phi)(x) = A_L x_{L-1}(x) + b_L$$

where $x_{L-1}(x)$ is given as

$$x_0 := x, \quad x_l(x) = \varrho(A_l x_{l-1}(x) + b_l) \text{ for } l = 1, \dots, L-1.$$

We call $M_j(\Phi) = \|A_j\|_0 + \|b_j\|_0$ the number of (non-zero) weights in the j -th layer and $M(\Phi) = \sum_{j=1}^L M_j(\Phi)$ the number of weights of the neural network Φ . We also refer to $M(\Phi)$ as the *size* of the neural network, write $L(\Phi) = L$ for the number of layers of Φ and refer to $N_o(\Phi) = N_L$ as the output dimension.

We refer to Section 2 in [OPS20] for further details.

The following lemma shows that concatenating n affine transformations with distinct neural networks and taking their weighted average can itself be represented as a neural network. The number of non-zero weights in the resulting neural network can be controlled by the number of non-zero weights in the original neural networks. The proof of the lemma is based on a simple extension of the *full parallelization* operation for neural networks (see [OPS20, Proposition 2.5]) and refines [GHJvW18, Lemma 3.8].

Lemma 3.2. Let $d, L, n \in \mathbb{N}$ and let Φ^1, \dots, Φ^n be neural networks with L layers, d -dimensional input and equal output dimensions. Let D_1, \dots, D_n be $d \times d$ -matrices, $c_1, \dots, c_n \in \mathbb{R}^d$ and $w_1, \dots, w_n \in \mathbb{R}$.

Then there exists a neural network ψ such that

$$(3.1) \quad R(\psi)(x) = \sum_{i=1}^n w_i R(\Phi^i)(D_i x + c_i) \quad \text{for all } x \in \mathbb{R}^d$$

and $M_j(\psi) \leq \sum_{i=1}^n M_j(\Phi^i)$ for $j = 2, \dots, L$. If, in addition, D_1, \dots, D_n are diagonal matrices and $c_1 = \dots = c_n = 0$, then $M(\psi) \leq \sum_{i=1}^n M(\Phi^i)$.

Proof. Write for $i = 1, \dots, n$

$$\Phi^i = ((A_1^i, b_1^i), \dots, (A_L^i, b_L^i))$$

and define the block matrices

$$\begin{aligned} A_1^{n+1} &= \begin{pmatrix} A_1^1 D_1 \\ \vdots \\ A_1^n D_n \end{pmatrix}, & b_1^{n+1} &= \begin{pmatrix} A_1^1 c_1 + b_1^1 \\ \vdots \\ A_1^n c_n + b_1^n \end{pmatrix}, \\ A_j^{n+1} &= \begin{pmatrix} A_j^1 & & 0 \\ & \ddots & \\ 0 & & A_j^n \end{pmatrix}, & b_j^{n+1} &= \begin{pmatrix} b_j^1 \\ \vdots \\ b_j^n \end{pmatrix} \text{ for } j = 2, \dots, L-1, \\ A_L^{n+1} &= (w_1 A_L^1 \quad \dots \quad w_n A_L^n) \text{ and } b_L^{n+1} = w_1 b_L^1 + \dots + w_n b_L^n. \end{aligned}$$

Set $\psi = ((A_1^{n+1}, b_1^{n+1}), \dots, (A_L^{n+1}, b_L^{n+1}))$. Then, for $l = 1, \dots, L-1$ and $x \in \mathbb{R}^d$, it is straightforward to verify that $x_l(x)$ has a block structure

$$x_l(x) = \begin{pmatrix} x_l^1(x) \\ \vdots \\ x_l^n(x) \end{pmatrix},$$

with $x_1^i(x) = \varrho(A_1^i(D_1 x + c_1) + b_1^i)$, $x_l^i(x) = \varrho(A_l^i x_{l-1}^i + b_l^i)$ for $l = 2, \dots, L-1$ and

$$R(\psi)(x) = A_L^{n+1} x_{L-1}(x) + b_L^{n+1} = \sum_{i=1}^n w_i [A_L^i x_{L-1}^i(x) + b_L^i].$$

Hence, (3.1) is satisfied and

$$\begin{aligned} M_j(\psi) &= M_j(\Phi^1) + \dots + M_j(\Phi^n) \text{ for } j = 2, \dots, L-1, \\ M_L(\psi) &\leq M_L(\Phi^1) \mathbb{1}_{\{w_1 \neq 0\}} + \dots + M_L(\Phi^n) \mathbb{1}_{\{w_n \neq 0\}}. \end{aligned}$$

If in addition D_1, \dots, D_n are diagonal matrices and $c_1 = \dots = c_n = 0$, then $\|A_1^i D_i\|_0 = \|A_1^i\|_0$ and therefore $M_1(\psi) = M_1(\Phi^1) + \dots + M_1(\Phi^n)$. Thus, in this situation, $M(\psi) = \sum_{j=1}^L M_j(\psi) \leq \sum_{j=1}^L \sum_{i=1}^n M_j(\Phi^i) = \sum_{i=1}^n M(\Phi^i)$, as claimed. \square

4. DNN APPROXIMATIONS FOR UNIVARIATE LÉVY MODELS

We study DNN expression rates for option prices under (geometric) Lévy models for asset prices, initially here in one spatial dimension. We present two expression rate estimates for ReLU DNNs, which are based on distinct mathematical arguments: the first, probabilistic argument builds on ideas used in recent works [GGJ⁺19, BGJ20] and the references there. However, for the key step of the proof a different technique is used, which is based on the Ledoux-Talagrand contraction principle (Theorem 4.12 in [LT13]) and statistical learning. This new approach is not only technically less involved (in comparison to, e.g., the techniques used in [GGJ⁺19]), but it also allows for weaker assumptions on the activation function, see Proposition 4.1 below. Alternatively, under slightly different hypotheses on the activation function one can also rely on [GGJ⁺19, Lemma 2.16], see Proposition 4.3 below. The probabilistic arguments result in, essentially, ε -complexity of DNN expression of order ε^{-2} . The second argument draws on parabolic (analytic) regularity furnished by the corresponding Kolmogorov equations, and results in far stronger, exponential expression rates, i.e., with an ε -complexity of DNN expression scaling, essentially, polylogarithmic with respect to $0 < \varepsilon < 1$. As we shall see in the next section, however, the latter argument is in general subject to the curse of dimensionality.

4.1. DNN expression rates: probabilistic argument. We fix $0 < a < b < \infty$ and measure the approximation error in the uniform norm on $[a, b]$. Recall that $M(\Phi)$ denotes the number of (non-zero) weights of a neural network Φ and $R(\Phi)$ is the realization of Φ .

Proposition 4.1. *Suppose the moment condition (2.2) holds. Suppose the payoff φ can be approximated well by neural networks, that is, given a payoff function $s \mapsto \varphi(s)$ there exists constants $c > 0, q \geq 0$ such that for any $\varepsilon \in (0, 1]$ there exists a neural network ϕ_ε with*

$$(4.1) \quad |\varphi(s) - R(\phi_\varepsilon)(s)| \leq \varepsilon c(1 + |s|), \quad s \in (0, \infty),$$

$$(4.2) \quad M(\phi_\varepsilon) \leq c\varepsilon^{-q},$$

$$(4.3) \quad \text{Lip}(R(\phi_\varepsilon)) \leq c.$$

Then there exists $\kappa \in [c, \infty)$ (depending on the interval $[a, b]$) and neural networks ψ_ε , $\varepsilon \in (0, 1]$, such that for any target accuracy $\varepsilon \in (0, 1]$ the number of weights is bounded by $M(\psi_\varepsilon) \leq \kappa\varepsilon^{-2-q}$ and the approximation error between the neural network ψ_ε and the option price is at most ε , that is,

$$\sup_{s \in [a, b]} |u(T, s) - R(\psi_\varepsilon)(s)| \leq \varepsilon.$$

Remark 4.2. In relevant examples such as, e.g., plain vanilla European options, the initial condition can be represented *exactly* as a neural network ϕ . Then one can choose $\phi_\varepsilon = \phi$ for all $\varepsilon \in (0, 1]$ and so (4.1)–(4.3) is satisfied with $q = 0$, $c = \max(M(\phi), \text{Lip}(R(\phi)))$. Examples include call options, straddles, and butterfly payoff functions (when ϱ is the ReLU activation function $x \mapsto \max(x, 0)$).

Proof. Let $\varepsilon \in (0, 1]$ be the given target accuracy and fix $\bar{\varepsilon} \in (0, 1]$ (to be specified later). Denote $\phi = \phi_{\bar{\varepsilon}}$. Firstly, (4.1) and (4.3) show that for any $s \in (0, \infty)$

$$\begin{aligned} |\varphi(s)| &\leq |\varphi(s) - R(\phi)(s)| + |R(\phi)(s) - R(\phi)(0)| + |R(\phi)(0)| \\ &\leq \bar{\varepsilon}c(1 + |s|) + c|s| + |R(\phi)(0)|. \end{aligned}$$

Thus, φ is at most linearly growing at ∞ and so we obtain $\mathbb{E}[\varphi(se^{X_T})] < \infty$, since even the second exponential moment is finite,

$$(4.4) \quad \mathbb{E}[e^{2X_T}] < \infty,$$

due to the assumed integrability (2.2) of the Lévy measure and [Sat99, Theorem 25.17].

Now recall that

$$u(T, s) = \mathbb{E}[\varphi(se^{X_T})].$$

Combining this with assumption (4.1) yields for all $s \in [a, b]$

$$(4.5) \quad |u(T, s) - \mathbb{E}[R(\phi)(se^{X_T})]| \leq \mathbb{E}[|\varphi(se^{X_T}) - R(\phi)(se^{X_T})|] \leq \bar{\varepsilon}c(1 + |s|\mathbb{E}[e^{X_T}]) \leq \bar{\varepsilon}c_1$$

with the constant $c_1 = c(1 + b\mathbb{E}[e^{X_T}])$ being finite due to (4.4).

In the second step, let X^1, \dots, X^n denote n i.i.d. copies of X and introduce an independent collection of Rademacher random variables $\varepsilon_1, \dots, \varepsilon_n$. Write $f(s) = R(\phi)(s) - R(\phi)(0)$. Note that the mapping $\mathbb{R}^n \times \mathbb{R}^n \ni (x, y) \mapsto \sup_{s \in [a, b]} |\sum_{k=1}^n y_k f(se^{x_k})|$ is Borel-measurable, because the supremum over $s \in [a, b]$ equals the supremum over $s \in [a, b] \cap \mathbb{Q}$ due to continuity of f and the pointwise supremum of a countable collection of measurable functions is itself measurable. The same reasoning guarantees that the suprema over $s \in [a, b]$ in (4.6) below are indeed random variables, because they are equal to the respective suprema over $s \in [a, b] \cap \mathbb{Q}$.

Using independence and symmetrization (see for instance [BLM13, Lemma 11.4]) in the first step, elementary properties of conditional expectations in the second step, and Theorem 4.12 in [LT13] (with $T_{x_1, \dots, x_n} = \{t \in \mathbb{R}^n : t_1 = se^{x_1}, \dots, t_n = se^{x_n} \text{ for some } s \in [a, b]\}$) in the third step shows that

$$\begin{aligned}
 & \mathbb{E} \left[\sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k}) \right| \right] \\
 & \leq 2 \mathbb{E} \left[\sup_{s \in [a, b]} \left| \frac{1}{n} \sum_{k=1}^n \varepsilon_k f(se^{X_T^k}) \right| \right] \\
 & = \frac{2}{n} \mathbb{E} \left[\mathbb{E} \left[\sup_{t \in T_{x_1, \dots, x_n}} \left| \sum_{k=1}^n \varepsilon_k f(t_k) \right| \right] \Bigg|_{x_1 = X_T^1, \dots, x_n = X_T^n} \right] \\
 (4.6) \quad & \leq \frac{4}{n} \text{Lip}(\mathbf{R}(\phi)) \mathbb{E} \left[\mathbb{E} \left[\sup_{t \in T_{x_1, \dots, x_n}} \left| \sum_{k=1}^n \varepsilon_k t_k \right| \right] \Bigg|_{x_1 = X_T^1, \dots, x_n = X_T^n} \right] \\
 & = \frac{4}{n} \text{Lip}(\mathbf{R}(\phi)) \mathbb{E} \left[\sup_{s \in [a, b]} \left| \sum_{k=1}^n \varepsilon_k se^{X_T^k} \right| \right] \\
 & \leq \frac{4b}{n} \text{Lip}(\mathbf{R}(\phi)) \mathbb{E} \left[\left| \sum_{k=1}^n \varepsilon_k e^{X_T^k} \right| \right].
 \end{aligned}$$

On the other hand, one may apply Jensen's inequality, independence and $\mathbb{E}[\varepsilon_k \varepsilon_l] = \delta_{k,l}$ to estimate

$$\begin{aligned}
 \mathbb{E} \left[\left| \sum_{k=1}^n \varepsilon_k e^{X_T^k} \right| \right] & \leq \mathbb{E} \left[\left| \sum_{k=1}^n \varepsilon_k e^{X_T^k} \right|^2 \right]^{1/2} = \left(\sum_{k=1}^n \mathbb{E} \left[e^{2X_T^k} \right] \right)^{1/2} \\
 & = \sqrt{n} \mathbb{E} \left[e^{2X_T} \right]^{1/2}.
 \end{aligned}$$

Combining this with the previous estimate (4.6) and the hypothesis on the Lipschitz-constant of the neural network (4.3) we obtain that

$$(4.7) \quad \mathbb{E} \left[\sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k}) \right| \right] \leq \frac{c_2}{\sqrt{n}}$$

with $c_2 = 4bc \mathbb{E} \left[e^{2X_T} \right]^{1/2}$, which is finite again due to the existence of exponential moments (4.4). In a third step we can now apply Markov's inequality for the first estimate and then insert (4.7) to estimate

$$\begin{aligned}
 & \mathbb{P} \left(\sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k}) \right| \geq \frac{3c_2}{2\sqrt{n}} \right) \\
 (4.8) \quad & \leq \frac{2\sqrt{n}}{3c_2} \mathbb{E} \left[\sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k}) \right| \right] \\
 & \leq \frac{2}{3}.
 \end{aligned}$$

This proves in particular that

$$(4.9) \quad \mathbb{P} \left(\sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k}) \right| \leq \frac{2c_2}{\sqrt{n}} \right) > 0.$$

Therefore (as $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$ necessarily needs to satisfy $A \neq \emptyset$) there exists $\omega \in \Omega$ with

$$(4.10) \quad \sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k(\omega)}) \right| \leq \frac{2c_2}{\sqrt{n}}.$$

Lemma 3.2 proves that $s \mapsto \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k(\omega)})$ is itself the realization of a neural network $\tilde{\psi}$ with $M(\tilde{\psi}) \leq nM(\phi)$ and hence we have proved the existence of a neural network $\tilde{\psi}$ with

$$(4.11) \quad \sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \mathbf{R}(\tilde{\psi})(s) \right| \leq \frac{2c_2}{\sqrt{n}}.$$

The final step consists in selecting $\bar{\varepsilon} = \varepsilon(c_1 + 1)^{-1}$, choosing $n = \lceil (2c_2\bar{\varepsilon}^{-1})^2 \rceil$, setting $\psi_\varepsilon = \tilde{\psi}$, noting (with $\kappa = c(1 + 4c_2^2)(c_1 + 1)^{2+q}$)

$$M(\psi_\varepsilon) = M(\tilde{\psi}) \leq nM(\phi) \leq (1 + (2c_2\bar{\varepsilon}^{-1})^2)c\bar{\varepsilon}^{-q} \leq c(1 + 4c_2^2)\bar{\varepsilon}^{-2-q} = \kappa\varepsilon^{-2-q}$$

and combining (4.11) with (4.5) to estimate

$$\begin{aligned} & \sup_{s \in [a, b]} |u(T, s) - \mathbf{R}(\psi_\varepsilon)(s)| \\ & \leq \sup_{s \in [a, b]} |u(T, s) - \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})]| + \sup_{s \in [a, b]} |\mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \mathbf{R}(\tilde{\psi})(s)| \\ & \leq \bar{\varepsilon}(c_1 + 1) = \varepsilon. \end{aligned}$$

□

Proposition 4.3. *Consider the same situation as in Proposition 4.1, but instead of (4.3) assume that $\mathbf{R}(\phi_\varepsilon)$ is C^1 and there is a constant $c > 0$ such that for every $s \in (0, \infty)$ holds*

$$(4.12) \quad |\mathbf{R}(\phi_\varepsilon)(s)| \leq c(1 + |s|), \quad |\mathbf{R}(\phi_\varepsilon)'(s)| \leq c.$$

Then the assertion of Proposition 4.1 remains valid.

Proof. The same proof as for Proposition 4.1 applies, only the second step needs to be adapted. In other words, we prove the estimate (4.7) with a different constant c_2 by using a different technique. To do this, again we let X^1, \dots, X^n denote n i.i.d. copies of X . Applying Lemma 2.16 in [GGJ⁺19] (with random fields $\xi_k(s, \omega) = \mathbf{R}(\phi)(se^{X_T^k(\omega)})$, $k = 1, \dots, n$, which satisfy the hypotheses of Lemma 2.16 in [GGJ⁺19] thanks to (4.4) and (4.12)) in the first inequality and using (4.12) for the second inequality then proves that

$$(4.13) \quad \begin{aligned} & \mathbb{E} \left[\sup_{s \in [a, b]} \left| \mathbb{E}[\mathbf{R}(\phi)(se^{X_T})] - \frac{1}{n} \sum_{k=1}^n \mathbf{R}(\phi)(se^{X_T^k}) \right| \right] \\ & \leq \frac{32\sqrt{e}}{\sqrt{n}} \sup_{s \in [a, b]} \left[\mathbb{E}[|\mathbf{R}(\phi)(se^{X_T})|^2]^{1/2} + (b-a)\mathbb{E}[|\mathbf{R}(\phi)'(se^{X_T})e^{X_T}|^2]^{1/2} \right] \\ & \leq \frac{32c\sqrt{e}}{\sqrt{n}} \left[1 + b\mathbb{E}[e^{2X_T}]^{1/2} + (b-a)\mathbb{E}[e^{2X_T}]^{1/2} \right], \end{aligned}$$

which is a bound as in (4.7) with constant

$$c_2 = 32c\sqrt{e} \left[1 + b\mathbb{E}[e^{2X_T}]^{1/2} + (b-a)\mathbb{E}[e^{2X_T}]^{1/2} \right].$$

□

Remark 4.4. The architecture of the neural network approximations constructed using probabilistic arguments in Proposition 4.1, Proposition 4.3 and also Theorem 5.1 ahead differ from architectures obtained by analytic arguments, see Proposition 4.6 and Theorem 5.4 ahead. While the neural networks in the latter results are deep in any situation, the architecture of the neural networks in the former situation depends heavily on the architecture of the neural network ϕ_ε used to approximate the payoff function φ . Therefore, in certain simple situations, the approximating neural network ψ_ε may be a shallow neural network, that is, a neural network with only $L = 2$ layers. E.g., by (4.5) or (2.4) the function φ is specified in the variable $s > 0$, and not in log-return variable x . This implies, e.g., for a plain-vanilla European call that $\varphi(s) = (s - K)^+$ must be emulated by a ReLU NN, which can be done using a simple 2-layer neural network, see Proposition 4.5 below.

4.2. DNN expression of European calls. In this section we illustrate how the results of Proposition 4.1 can be used to bound DNN expression rates of call options on exponential Lévy models. Suppose we observe call option prices for a fixed maturity T and N different strikes $K_1, \dots, K_N > 0$. Denote these prices by $\hat{C}(T, K_1), \dots, \hat{C}(T, K_N)$. A task frequently encountered in practice is to extrapolate from these prices to prices corresponding to unobserved maturities or to learn a non-parametric option pricing function. A widely used approach is to solve

$$(4.14) \quad \min_{\phi \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^N \left[\frac{\hat{C}(T, K_i)}{K_i} - \phi(S_0/K_i) \right]^2.$$

Here \mathcal{H} is a suitable collection of (realizations of) neural networks, for instance all networks with an a-priori fixed architecture. In fact, many of the papers listed in the recent review [RW20] use this approach or a variation of it, where for instance an absolute value is inserted instead of a square or $\hat{C}(T, K_i)/K_i$ is replaced by $\hat{C}(T, K_i)$ and S/K_i by K_i .

In this section we assume that the observed call prices are generated from an (assumed unknown) exponential Lévy model and \mathcal{H} consists of ReLU networks. Then we show that the error in (4.14) can be controlled and we can give bounds on the number of non-zero parameters of the minimizing neural network. The following result is a direct consequence of Proposition 4.1. It shows that $\mathcal{O}(\varepsilon^{-1})$ weights suffice to achieve an error of at most ε in (4.14).

Proposition 4.5. *Assume that*

$$\hat{C}(T, K_i) = \mathbb{E}[(S_T - K_i)^+], \text{ for } i = 1, \dots, N,$$

with $S_T = S_0 \exp(X_T)$ and X an (unknown) Lévy process satisfying (2.2). For any $\kappa > 0$, $\varepsilon \in (0, 1]$ we let $\mathcal{H}_{\kappa, \varepsilon}$ denote the set of all (realizations of) neural networks with at most $\kappa \varepsilon^{-1}$ non-zero weights and choose $\varrho(x) = \max(x, 0)$ as activation function. Then there exists $\kappa \in [3, \infty)$ such that for all $\varepsilon \in (0, 1]$

$$\min_{\phi \in \mathcal{H}_{\kappa, \varepsilon}} \frac{1}{N} \sum_{i=1}^N \left[\frac{\hat{C}(T, K_i)}{K_i} - \phi(S_0/K_i) \right]^2 \leq \varepsilon.$$

Proof. Firstly, choose $a = \min\{S_0/K_1, \dots, S_0/K_N\}$ and $b = \max\{S_0/K_1, \dots, S_0/K_N\}$. We note that the function $\varphi(s) = (s-1)^+$ can be represented by the 2-layer neural network $\phi_0 = ((1, -1), (1, 0))$, that is, $R(\phi_0) = \varphi$. Thus, Proposition 4.1 can be applied (with $\phi_\varepsilon = \phi_0$ for all $\varepsilon \in (0, 1]$ and $q = 0$,

$c = 3$) and so there exists $\kappa \in [3, \infty)$ and neural networks ψ_δ , $\delta \in (0, 1]$, such that for any $\delta \in (0, 1]$ we have $M(\psi_\delta) \leq \kappa\delta^{-2}$ and

$$\sup_{s \in [a, b]} |u(T, s) - \mathbf{R}(\psi_\delta)(s)| \leq \delta$$

with $u(T, s) = \mathbb{E}[(se^{X_T} - 1)^+]$. Therefore,

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left[\frac{\hat{C}(T, K_i)}{K_i} - \mathbf{R}(\psi_\delta)(S_0/K_i) \right]^2 &= \frac{1}{N} \sum_{i=1}^N [u(T, S_0/K_i) - \mathbf{R}(\psi_\delta)(S_0/K_i)]^2 \\ &\leq \delta^2. \end{aligned}$$

Setting $\varepsilon = \delta^2$ and noting $\mathbf{R}(\psi_\delta) \in \mathcal{H}_{\kappa, \varepsilon}$ then finishes the proof. \square

4.3. ReLU DNN exponential expressivity. We now develop a second argument for bounding the expressivity of ReLU DNNs for the option price $u(\tau, s)$ solution of (2.5), subject to the initial condition $u(0, s) = \varphi(s)$. In particular, in this subsection we choose $\varrho(x) = \max(x, 0)$ as activation function.

As in the preceding, probabilistic argument, we consider the DNN expression error in a bounded interval $[a, b]$ with $0 < a < s < b < \infty$. The argument is based on *parabolic smoothing* of the linear, parabolic PIDE (2.5). This, in turn, ensures smoothness of $s \mapsto u(\tau, s)$ at positive times $\tau > 0$, i.e. smoothness in the ‘‘spatial’’ variables $s \in [a, b]$ resp. in the log-return variable $x = \log(s) \in [\log(a), \log(b)]$, even for nonsmooth payoff functions φ (so, in particular, binary options with discontinuous payoffs φ are admissible, albeit at the cost of non-uniformity of derivative bounds at $\tau \downarrow 0$). It is a classical result that this implies spectral, possibly exponential convergence of *polynomial approximations* of $u(\tau, \cdot)|_{[a, b]}$ in $L^\infty([a, b])$. As we observed in [OSZ19, Section 3.2], this exponential polynomial convergence rate implies also exponential expressivity of ReLU DNNs of $u(\tau, \cdot)|_{[a, b]}$ in $L^\infty([a, b])$ for any $\tau > 0$.

To ensure smoothing properties of the solution operator of the PIDE, we require additional assumptions (see (4.19) below) on the Lévy triplet. To formulate these, recall that the Lévy symbol ψ of the LP X is defined as

$$(4.15) \quad \psi(\xi) = \frac{\sigma^2}{2} \xi^2 - i\gamma\xi - \int_{\mathbb{R}} \left[e^{i\xi x} - 1 - i\xi x \mathbb{1}_{\{|x| \leq 1\}} \right] \nu(dx), \quad \xi \in \mathbb{R}.$$

Next, we provide a sufficient condition on the Lévy triplet which guarantees that (4.19) is satisfied.

Assumption 1. *Let X be a Lévy process with characteristic triplet (σ^2, γ, ν) and Lévy density $k(z)$ where $\nu(dz) = k(z)dz$ satisfies*

(1) *There are constants $\beta_- > 0$, $\beta_+ > 1$ and $C > 0$ such that*

$$(4.16) \quad k(z) \leq C \begin{cases} e^{-\beta_-|z|}, & z < -1, \\ e^{-\beta_+z}, & z > 1. \end{cases}$$

(2) *Furthermore, there exist constants $0 < \alpha < 2$ and $C_+ > 0$ such that*

$$(4.17) \quad k(z) \leq C_+ \frac{1}{|z|^{1+\alpha}}, \quad 0 < |z| < 1.$$

(3) *If $\sigma = 0$, we assume additionally that there is a $C_- > 0$ such that*

$$(4.18) \quad \frac{1}{2}(k(z) + k(-z)) \geq C_- \frac{1}{|z|^{1+\alpha}}, \quad 0 < |z| < 1.$$

Proposition 4.6. *Suppose that the symbol of the LP X is such that there exists $\rho \in (0, 1]$ and constants $C_i > 0$, $i = 1, 2, 3$ such that for all $\xi \in \mathbb{R}$ holds*

$$(4.19) \quad \Re\psi(\xi) \geq C_1|\xi|^{2\rho}, \quad |\psi(\xi)| \leq C_2|\xi|^{2\rho} + C_3.$$

Then, for every v_0 such that $v_0 = \varphi \circ \exp \in L^2(\mathbb{R})$, and for every $0 < \tau \leq T < \infty$, and for every interval $I = [a, b]$ with $0 < a < b < \infty$, and for every $0 < \varepsilon < 1/2$ exists neural networks ψ_ε^u which express the solution $u(\tau, \cdot)|_{[a, b]}$ to accuracy ε , i.e.,

$$\|u(\tau, \cdot) - \mathbf{R}(\psi_\varepsilon^u)\|_{L^\infty([a, b])} \leq \varepsilon.$$

Furthermore, there exists a constant $C' > 0$ such that with $\delta := 1/\min\{1, 2\rho\} \geq 1$ holds

$$M(\psi_\varepsilon^u) \leq C'|\log(\varepsilon)|^{2\delta}, \quad L(\psi_\varepsilon^u) \leq C'|\log(\varepsilon)|^\delta |\log(|\log(\varepsilon)|)|.$$

Remark 4.7. If the Lévy triplet (σ^2, γ, ν) is such that Assumption 1 holds, then (4.19) is satisfied (see [HRSW13, Lemma 10.4.2]). Here, $\rho := 1$ if $\sigma > 0$ and otherwise $\rho := \alpha/2$ with $\alpha \in (0, 2)$ in Assumption 1.

Proof. We change coordinates to $x = \log(s) \in (-\infty, \infty)$ so that $v(\tau, x) = u(\tau, \exp(x))$. Then, the PIDE (2.5) takes the form (e.g. [MvPS04, Section 3], [LM08, Section 3.1])

$$(4.20) \quad \frac{\partial v}{\partial \tau} - \frac{\sigma^2}{2} \frac{\partial^2 v}{\partial x^2} - (\gamma + r) \frac{\partial v}{\partial x} + A[v] + rv = 0 \quad \text{in } (0, T) \times \mathbb{R}$$

where A denotes the following integrodifferential operator

$$A[f](x) := - \int_{\mathbb{R}} \{f(x+y) - f(x) - yf'(x)\mathbb{1}_{\{|y| \leq 1\}}\} \nu(dy)$$

together with the initial condition

$$(4.21) \quad v|_{\tau=0} = \varphi(e^x) = (\varphi \circ \exp)(x).$$

Then $C(t, s) := v(T-t, \ln(s))$ satisfies

$$(4.22) \quad C(t, S_t) = \mathbb{E}[e^{r(t-T)}\varphi(S_T)|\mathcal{F}_t].$$

Conversely, if $C(t, s)$ in (4.22) is sufficiently regular, then $v(\tau, x) := C(T-\tau, e^x)$ is solution of (4.20), (4.21). Recall that we assume $r = 0$ for notational simplicity.

The Lévy-Khintchine formula describes the LP $X \subset \mathbb{R}$ by the log-characteristic function ψ of the RV X_1 . From the time-homogeneity of the LP X ,

$$(4.23) \quad \forall t > 0: \quad \mathbb{E}[e^{i\xi X_t}] = e^{-t\psi(\xi)}.$$

The Lévy exponent ψ of the LP X admits the explicit representation (4.15).

The Lévy exponent ψ is the symbol of the pseudo-differential operator $-\mathcal{L}$, where \mathcal{L} is the infinitesimal generator of the semi-group of the LP X . $\mathcal{A} := -\mathcal{L}$ is the spatial operator in (4.20) given by

$$(4.24) \quad \mathcal{A}[f](x) := -\frac{\sigma^2}{2} \frac{d^2 f}{dx^2}(x) - \gamma \frac{df}{dx}(x) + A[f](x).$$

For $f, g \in C_0^\infty(\mathbb{R})$ we associate with operator \mathcal{A} the bilinear form

$$(4.25) \quad a(f, g) := \int_{\mathbb{R}} \mathcal{A}[f](x)g(x)dx.$$

The translation invariance of the operator \mathcal{A} (implied by stationarity of the LP X) in (4.24) and Parseval's equality (see [HRSW13, Remark 10.4.1]) imply that ψ is the symbol of \mathcal{A} , i.e.

$$\forall f, g \in C_0^\infty(\mathbb{R}) : \quad a(f, g) = \int_{\mathbb{R}} \psi(\xi) \hat{f}(\xi) \overline{\hat{g}(\xi)} d\xi .$$

Fix $0 < \tau \leq T < \infty$, $x \in \mathbb{R}$ arbitrary. Then the variational solution v of (4.20) (whose existence and uniqueness is ensured by (4.19), see [EG14]) satisfies

$$v(\tau, x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(ix\xi) \hat{v}(\tau, \xi) d\xi = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(ix\xi) \exp(-\tau\psi(\xi)) \widehat{\varphi \circ \exp}(\xi) d\xi .$$

For every $k \in \mathbb{N}_0$, Parseval's equality implies with the lower bound in (4.19)

$$\begin{aligned} \int_{\mathbb{R}} |(D_x^k v)(\tau, x)|^2 dx &= \int_{\mathbb{R}} |\xi|^{2k} \exp(-2\tau\psi(\xi)) |\widehat{\varphi \circ \exp}(\xi)|^2 d\xi \\ &\leq \int_{\mathbb{R}} |\xi|^{2k} \exp(-2\tau C_1 |\xi|^{2\rho}) |\widehat{\varphi \circ \exp}(\xi)|^2 d\xi . \end{aligned}$$

An elementary calculation shows that for any $m, \kappa, \mu > 0$ holds

$$(4.26) \quad \max_{\eta > 0} \{\eta^m \exp(-\kappa\eta^\mu)\} = \left(\frac{m}{\kappa\mu e} \right)^{m/\mu} .$$

We employ (4.26) with $m = 2k$, $\kappa = 2\tau C_1$, $\mu = 2\rho$ and $\eta = |\xi|$ to obtain

$$(4.27) \quad \|(D_x^k v)(\tau, \cdot)\|_{L^2(\mathbb{R})}^2 \leq \left(\frac{k}{2\tau C_1 \rho e} \right)^{k/\rho} \|v_0\|_{L^2(\mathbb{R})}^2 .$$

Taking square roots and using the (rough) Stirling bound $k^k \leq k! e^k$ valid for all $k \in \mathbb{N}$, we obtain

$$(4.28) \quad \forall \tau > 0, \forall k \in \mathbb{N} : \quad \|(D_x^k v)(\tau, \cdot)\|_{L^2(\mathbb{R})} \leq \left[\left(\frac{1}{2\tau C_1 \rho} \right)^{1/(2\rho)} \right]^k (k!)^{1/(2\rho)} \|v_0\|_{L^2(\mathbb{R})} .$$

This implies, with the Sobolev embedding, that for any bounded interval $I = [x_-, x_+] \subset \mathbb{R}$, $-\infty < x_- < x_+ < \infty$, and for every fixed $\tau > 0$, there exist constants $C = C(x_+, x_-) > 0$ and $A(\tau, \rho) > 0$ such that

$$(4.29) \quad \forall k \in \mathbb{N} : \quad \|(D_x^k v)(\tau, \cdot)\|_{L^\infty(I)} \leq C(A(\tau, \rho))^k (k!)^{1/\min\{1, 2\rho\}} .$$

I.e., $v(\tau, \cdot)|_I$ is Gevrey- δ regular with $\delta = 1/\min\{1, 2\rho\}$.

To construct the DNNs ψ_ε^u in the claim, we proceed in several steps: we first use a (analytic, in the bounded interval $I = [x_-, x_+] \subset \mathbb{R}$) change of variables $s = \exp(x)$ and the fact that Gevrey regularity is preserved under analytic changes of variables to infer Gevrey regularity in $[a, b] \subset \mathbb{R}_{>0}$ of $s \mapsto u(\tau, s)$, for every fixed $\tau > 0$.

This, in turn, implies the existence of a sequence $\{u_p(s)\}_{p \geq 1}$ of polynomials of degree $p \in \mathbb{N}$ in $[a, b]$ converging in $W^{1,\infty}([a, b])$ to $u(\tau, \cdot)$ at rate $\exp(-b'p^{1/\delta})$ for some constant $b' > 0$ depending on a , b and on $\delta \geq 1$, but independent of p .

The asserted DNNs will then be obtained by approximately expressing the u_p through ReLU DNNs, again at exponential rates, using results from [OPS20].

The interval $s \in [a, b]$ in the assertion corresponds to $x \in [\log(a), \log(b)]$ under the analytic (in the bounded interval $[a, b]$) change of variables $x = \log(s)$. As Gevrey regularity is preserved under analytic changes of variables (e.g. [Rod93, Proposition 1.4.6]), also $u(\tau, s)|_{s \in [a, b]}$ is Gevrey- δ regular, with the same index $\delta = 1/\min\{1, 2\rho\} \geq 1$ and with constants in the derivative bounds

which depend on $0 < a < b < \infty$, $\rho \in (0, 1]$, $\tau > 0$. In particular, for $\rho \geq 1/2$, $u(\tau, s)|_{s \in [a, b]}$ is real analytic in $[a, b]$.

Expression rates of deep ReLU NNs for Gevrey-regular (and in particular, analytic functions) of one real variable have been obtained in [OPS20, Theorem 5.12]. It follows from the proof of that result (which allows for functions with an endpoint singularity and introduces continuous, *piecewise* polynomial functions on N pieces; we require only polynomials on *one* piece) that for every $N \in \mathbb{N}$, there exists a ReLU NN $\Phi^{u, N}$ of depth $\mathcal{O}(N^\delta \log(N))$ and of size (i.e., number of nonzero weights) $\mathcal{O}(N^{2\delta+1})$ such that the expression error in $H^1(I)$ (and, hence, also in $L^\infty(I)$) is bounded, as $N \rightarrow \infty$, by $\mathcal{O}(\exp(-bN))$ for some constant $b > 0$ independent of N (but possibly depending on $a, b, \delta \geq 1, \tau > 0$).

This implies that for every $0 < \varepsilon < 1/2$, a pointwise error of $\mathcal{O}(\varepsilon)$ in $[a, b]$ can be achieved by some ReLU NN ψ_ε^u of depth $\mathcal{O}(|\log(\varepsilon)|^\delta |\log(|\log(\varepsilon)|)|)$ and of size $\mathcal{O}(|\log(\varepsilon)|^{2\delta})$.

This completes the proof. \square

We remark that in the particular case when $\delta = 1$, the derivative bounds (4.28) imply analyticity of the map $s \mapsto u(\tau, s)$ for $s \in [a, b]$ which implies the assertion also with the exponential expression rate bound for analytic functions in [OSZ19].

We also remark that the smoothing of the solution operator in Proposition 4.6 accommodated payoff functions which belong merely to L^2 , i.e. in particular binary contracts. This is a consequence of the assumption (4.19) (which holds, for instance, under Assumption 1) which, on the other hand, excludes one-sided Lévy processes. These are covered by Proposition 4.1.

5. DNN APPROXIMATION RATES FOR MULTIVARIATE LÉVY MODELS

We now turn to DNN expression rates for multivariate geometric Lévy models. This is a typical situation when option prices on baskets of d risky assets are of interest, whose log-returns are modelled by multivariate Lévy processes. We admit rather general jump measures with, in particular, fully correlated jumps in the marginals, as provided, for example, by so-called Lévy copula constructions in [KT06].

As in the univariate case, we prove two results on ReLU DNN expression rates of option prices for European style contracts. The first argument is developed in Section 5.1 below and overcomes, in particular, the curse of dimensionality. Its proof is again based on probabilistic arguments from statistical learning theory. As exponential LPs X^d generalize geometric Brownian motions, Theorem 5.1 generalizes several results from the classical Black–Scholes setting and we comment on the relation of Theorem 5.1 to these recent results in Section 5.2. Owing to the method of proof, the DNN expression rate in Theorem 5.1 will deliver an ε -complexity of $\mathcal{O}(\varepsilon^{-2})$, achieved with potentially shallow DNNs, see Remark 4.4.

The second argument is based on parabolic regularity of the deterministic Kolmogorov PIDE associated to the LP X^d . We show in Theorem 5.4 that polylogarithmic in ε expression rate bounds can be achieved by allowing DNN depth to increase essentially as $\mathcal{O}(|\log \varepsilon|)$. The result in Theorem 5.4 is, however, prone to the curse of dimensionality: constants implied in the $\mathcal{O}(\cdot)$ bounds may (and, generally, will) depend exponentially on d . We also show that under a hypothesis on sufficiently large time $t > 0$, parabolic smoothing will allow to overcome the curse of dimension, with dimension-independent expression rate bounds which are possibly larger than the rates furnished by the probabilistic argument (which is, however, valid uniformly for all $t > 0$).

5.1. DNN expression rate bounds via probabilistic argument. We start by remarking that in this subsection, there is no need to assume ReLU activation.

The following result proves that neural networks are capable of approximating option prices in multivariate exponential Lévy models without the curse of dimensionality given that the corresponding Lévy triplets (A^d, γ^d, ν^d) are bounded uniformly with respect to the dimension d .

For any dimension $d \in \mathbb{N}$ we assume given a payoff $\varphi_d: \mathbb{R}^d \rightarrow \mathbb{R}$, a d -variate LP X^d and we denote the option price in time-to-maturity by

$$(5.1) \quad u_d(\tau, s) = \mathbb{E}[\varphi_d(s_1 \exp(X_{\tau,1}^d), \dots, s_d \exp(X_{\tau,d}^d))], \quad \tau \in [0, T], s \in (0, \infty)^d.$$

We refer to [Sat99] for more details on multivariate Lévy processes and to [CT04, EK19] for more details on multivariate geometric Lévy models in finance.

The next theorem is a main result of the present paper. It states that DNNs can efficiently express prices on possibly large baskets of risky assets whose dynamics are driven by multivariate Lévy processes with general jump correlation structure. The expression rate bounds are polynomial in the number d of assets and, therefore, not prone to the curse of dimensionality. This result generalizes earlier work on DNN expression rates for diffusion models in [EGJS18, GHJvW18].

Theorem 5.1. *Assume that for any $d \in \mathbb{N}$, the payoff $\varphi_d: \mathbb{R}^d \rightarrow \mathbb{R}$ can be approximated well by neural networks, that is, there exists constants $c > 0, p \geq 2, q \geq 0$ and, for all $\varepsilon \in (0, 1]$, there exists a neural network $\phi_{\varepsilon,d}$ with*

$$(5.2) \quad |\varphi_d(s) - \mathbb{R}(\phi_{\varepsilon,d})(s)| \leq \varepsilon c d^p (1 + \|s\|^p), \quad \text{for all } s \in (0, \infty)^d,$$

$$(5.3) \quad M(\phi_{\varepsilon,d}) \leq c d^p \varepsilon^{-q},$$

$$(5.4) \quad \text{Lip}(\mathbb{R}(\phi_{\varepsilon,d})) \leq c d^p.$$

In addition, assume that the Lévy triplets (A^d, γ^d, ν^d) of X^d are bounded in the dimension, that is, there exists a constant $B \geq 0$ such that for each $d \in \mathbb{N}$, $i, j = 1, \dots, d$,

$$(5.5) \quad A_{ij}^d \leq B, \quad \gamma_i^d \leq B, \quad \int_{\mathbb{R}^d \setminus \{\|y\| \leq 1\}} e^{py_i} \nu^d(dy) \leq B, \quad \int_{\{\|y\| \leq 1\}} y_i^2 \nu^d(dy) \leq B.$$

Then there exist constants $\kappa, \mathbf{p}, \mathbf{q} \in [0, \infty)$ and neural networks $\psi_{\varepsilon,d}$, $\varepsilon \in (0, 1]$, $d \in \mathbb{N}$ such that for any target accuracy $\varepsilon \in (0, 1]$ and for any $d \in \mathbb{N}$ the number of weights grows only polynomially $M(\psi_{\varepsilon,d}) \leq \kappa d^{\mathbf{p}} \varepsilon^{-\mathbf{q}}$ and the approximation error between the neural network $\psi_{\varepsilon,d}$ and the option price is at most ε , that is,

$$\sup_{s \in [a,b]^d} |u_d(T, s) - \mathbb{R}(\psi_{\varepsilon,d})(s)| \leq \varepsilon.$$

Remark 5.2. The statement of Theorem 5.1 is still valid, if we admit logarithmic growth of B with d in (5.5).

Remark 5.3. As in the univariate case (cf. Remark 4.2), in relevant examples of options written on $d > 1$ underlyings (such as basket options, call on max/min options, put on max/min options, ...) the payoff can be represented *exactly* as a ReLU DNN and so, we may choose $q = 0$ in (5.3) and obtain $\mathbf{q} = 2$ in Theorem 5.1 (cf. (5.14)).

Proof. Let $\varepsilon \in (0, 1]$ be the given target accuracy and consider $\bar{\varepsilon} \in (0, 1]$ (to be selected later). To simplify notation we write for $s \in [a, b]^d$

$$se^{X_T^d} = (s_1 \exp(X_{T,1}^d), \dots, s_d \exp(X_{T,d}^d)).$$

The proof consists in four steps:

- Step 1 bounds the error that arises when the payoff φ_d is replaced by the neural network approximation $\phi_{\bar{\varepsilon},d}$. As a part of Step 1 we also prove that the p -th exponential moments of the components $X_{T,i}^d$ of the Lévy process are bounded uniformly in the dimension d .

- Step 2 is a technical step that is required for Step 3; it bounds the error that arises when the Lévy process is capped at a threshold $D > 0$. If we assumed in addition that the output of the neural network $\phi_{\bar{\varepsilon},d}$ were bounded (this is for instance the case if the activation function ϱ is bounded), then Step 2 could be omitted.
- Step 3 is the key step in the proof. We introduce n i.i.d. copies of (the capped version of) X_T^d and use statistical learning techniques (symmetrization, Gaussian and Rademacher complexities) to estimate the expected maximum difference between the option price (with neural network payoff) and its sample average. This is then used to construct the approximating neural networks.
- Step 4 combines the estimates from Steps 1-3 and concludes the proof.

Step 1: Assumption (5.2) and Hölder's inequality yield for all $s \in [a, b]^d$

$$\begin{aligned}
 |u_d(T, s) - \mathbb{E}[\mathbf{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^d})]| &\leq \mathbb{E}[|\varphi_d(se^{X_T^d}) - \mathbf{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^d})|] \leq \bar{\varepsilon}cd^p(1 + \mathbb{E}[\|se^{X_T^d}\|^p]) \\
 &= \bar{\varepsilon}cd^p \left(1 + \mathbb{E} \left[\left(\sum_{i=1}^d s_i^2 e^{2X_{T,i}^d} \right)^{p/2} \right] \right) \\
 &\leq \bar{\varepsilon}cd^p \left(1 + b^p \mathbb{E} \left[d^{(p-1)/2} \left(\sum_{i=1}^d e^{2pX_{T,i}^d} \right)^{1/2} \right] \right) \\
 &\leq \bar{\varepsilon}c_1 d^{\frac{3}{2}p + \frac{1}{2}}
 \end{aligned} \tag{5.6}$$

with the constant $c_1 = c \max(1, b^p)(1 + \sup_{d,i} \mathbb{E}[e^{pX_{T,i}^d}])$ and we used $\|\cdot\| \leq \|\cdot\|_1$ in the last step. To see that c_1 is indeed finite, note that (5.5) and [Sat99, Theorem 25.17] (with the vector $w \in \mathbb{R}^d$ in that result being pe_i) imply that for any $d \in \mathbb{N}$, $i = 1, \dots, d$, the exponential moment can be bounded as

$$\begin{aligned}
 \mathbb{E}[e^{pX_{T,i}^d}] &= \exp \left(T \left[\frac{p^2}{2} A_{ii}^d + \int_{\mathbb{R}^d} (e^{py_i} - 1 - py_i \mathbb{1}_{\{\|y\| \leq 1\}}) \nu^d(dy) + p\gamma_i^d \right] \right) \\
 &\leq \exp \left(T \left[\frac{3p^2}{2} B + \int_{\{\|y\| \leq 1\}} (e^{py_i} - 1 - py_i) \nu^d(dy) + \int_{\mathbb{R}^d \setminus \{\|y\| \leq 1\}} (e^{py_i} - 1) \nu^d(dy) \right] \right) \\
 &\leq \exp \left(T \left[\frac{5p^2}{2} B + p^2 e^p \int_{\{\|y\| \leq 1\}} y_i^2 \nu^d(dy) \right] \right) \\
 &\leq \exp \left(T \left[\frac{5p^2}{2} B + p^2 e^p B \right] \right),
 \end{aligned} \tag{5.7}$$

where in the second inequality we used that $|e^z - 1 - z| \leq z^2 e^p$ for all $z \in [-p, p]$ which can be seen e.g. from the (mean value form of the) Taylor remainder formula.

Step 2: Before proceeding with the key step of the proof, we need to introduce a cut-off in order to ensure that the neural network output is bounded. Let $D > 0$ and consider the random variable $X_T^{d,D} = \min(X_T^d, D)$, where the minimum is understood componentwise. Then the Lipschitz property (5.4) implies

$$\begin{aligned}
(5.8) \quad |\mathbb{E}[\mathbf{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^d})] - \mathbb{E}[\mathbf{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}})]| &\leq cd^p \mathbb{E}[\|se^{X_T^d} - se^{X_T^{d,D}}\|] \\
&\leq bcd^p \mathbb{E} \left[\sum_{i=1}^d |e^{X_{T,i}^d} - e^{X_{T,i}^{d,D}}| \right] \\
&\leq bcd^p \sum_{i=1}^d \mathbb{E} \left[2e^{X_{T,i}^d} \mathbb{1}_{\{X_{T,i}^d > D\}} \right] \\
&\leq 2bcd^p \sum_{i=1}^d \mathbb{E}[e^{2X_{T,i}^d}]^{1/2} \mathbb{P}(X_{T,i}^d > D)^{1/2} \\
&\leq 2e^{-D} bcd^p \sum_{i=1}^d \mathbb{E}[e^{2X_{T,i}^d}] \\
&\leq \tilde{c}_1 e^{-D} d^{p+1},
\end{aligned}$$

where $\tilde{c}_1 = 2bc \exp(5TpB + 2Te^p pB)$ and we used $\|\cdot\| \leq \|\cdot\|_1$, Hölder's inequality, Chernoff's bound and finally again Hölder's inequality and (5.7).

Step 3: Let X_1, \dots, X_n denote n i.i.d. copies of the random vector $X_T^{d,D}$ and let Z_1, \dots, Z_n denote i.i.d. standard normal variables, independent of X_1, \dots, X_n . For any separable class of functions $\mathcal{H} \subset C(\mathbb{R}^d, \mathbb{R})$ define the random variable (the so-called empirical Gaussian complexity)

$$\hat{G}_n(\mathcal{H}) = \mathbb{E} \left[\sup_{f \in \mathcal{H}} \left| \frac{2}{n} \sum_{k=1}^n Z_k f(X_k) \right| \mid X_1, \dots, X_n \right].$$

Consider now for $i = 1, \dots, d$ the function classes

$$\mathcal{H}_i = \{(-\infty, D]^d \ni x \mapsto s \exp(x_i) : s \in [a, b]\}$$

and, with the notation $s \exp(x) = (s_1 \exp(x_1), \dots, s_d \exp(x_d))$, the class

$$\mathcal{H} = \{(-\infty, D]^d \ni x \mapsto \mathbf{R}(\phi_{\bar{\varepsilon},d})(s \exp(x)) - \mathbf{R}(\phi_{\bar{\varepsilon},d})(0) : s \in [a, b]^d\}.$$

Denoting by $\tilde{\mathcal{H}} \subset C((-\infty, D]^d, \mathbb{R}^d)$ the direct sum of $\mathcal{H}_1, \dots, \mathcal{H}_d$, we have that

$$\mathcal{H} = \phi(\tilde{\mathcal{H}})$$

where $\phi = \mathbf{R}(\phi_{\bar{\varepsilon},d})(\cdot) - \mathbf{R}(\phi_{\bar{\varepsilon},d})(0)$ is a Lipschitz-function with Lipschitz-constant cd^p (due to hypothesis on the Lipschitz-constant of the neural network (5.4)), $\phi(0) = 0$ and ϕ is bounded on the range of $\tilde{\mathcal{H}}$ (which is contained in $[0, b \exp(D)]^d$).

Consequently, Theorem 14 in [BM03] implies that

$$(5.9) \quad \hat{G}_n(\mathcal{H}) \leq 2cd^p \sum_{i=1}^d \hat{G}_n(\mathcal{H}_i).$$

Therefore, denoting by $\varepsilon_1, \dots, \varepsilon_n$ an independent collection of Rademacher random variables, we estimate

$$\begin{aligned}
 & \mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon}, d})(se^{X_T^{d, D}})] - \frac{1}{n} \sum_{k=1}^n \mathbb{R}(\phi_{\bar{\varepsilon}, d})(se^{X_k}) \right| \right] \\
 & \leq 2 \mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \frac{1}{n} \sum_{k=1}^n \varepsilon_k \phi(se^{X_k}) \right| \right] \\
 & \leq \tilde{c}_2 \mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \frac{2}{n} \sum_{k=1}^n Z_k \phi(se^{X_k}) \right| \right] \\
 (5.10) \quad & = \tilde{c}_2 \mathbb{E} \left[\sup_{f \in \mathcal{H}} \left| \frac{2}{n} \sum_{k=1}^n Z_k f(X_k) \right| \right] \\
 & \leq 2\tilde{c}_2 cd^p \sum_{i=1}^d \mathbb{E}[\hat{G}_n(\mathcal{H}_i)] \\
 & \leq \frac{4\tilde{c}_2 cd^p b}{n} \sum_{i=1}^d \mathbb{E} \left[\left| \sum_{k=1}^n Z_k e^{X_{k,i}} \right| \right],
 \end{aligned}$$

where the first inequality follows by symmetrization (see for instance [BLM13, Lemma 11.4]), the second inequality follows from the comparison results on Gaussian and Rademacher complexities (see for instance [BM03, Lemma 4]) with some absolute constant \tilde{c}_2 and the third inequality uses (5.9).

In fact, it is possible to prove that \tilde{c}_2 may be chosen as $\tilde{c}_2 = 1/\mathbb{E}[|Z_1|] = \sqrt{\pi/2}$. Indeed, setting $\mathcal{G} = \sigma(\varepsilon_1, \dots, \varepsilon_n, X_1, \dots, X_n)$ and using independence yields

$$\begin{aligned}
 \mathbb{E}[|Z_1|] \mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \frac{1}{n} \sum_{k=1}^n \varepsilon_k \phi(se^{X_k}) \right| \right] &= \mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \frac{1}{n} \sum_{k=1}^n \mathbb{E}[|Z_k| | \mathcal{G}] \varepsilon_k \phi(se^{X_k}) \right| \right] \\
 &= \mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n |Z_k| \varepsilon_k \phi(se^{X_k}) \mid \mathcal{G} \right] \right| \right] \\
 &\leq \mathbb{E} \left[\mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \frac{1}{n} \sum_{k=1}^n |Z_k| \varepsilon_k \phi(se^{X_k}) \mid \mathcal{G} \right| \right] \right] \\
 &= \mathbb{E} \left[\sup_{s \in [a, b]^d} \left| \frac{1}{n} \sum_{k=1}^n Z_k \phi(se^{X_k}) \right| \right].
 \end{aligned}$$

To further simplify (5.10) we now apply Jensen's inequality and use independence and $\mathbb{E}[Z_k Z_l] = \delta_{k,l}$ to derive for $i = 1, \dots, d$

$$\begin{aligned}
 \mathbb{E} \left[\left| \sum_{k=1}^n Z_k e^{X_{k,i}} \right| \right] &\leq \mathbb{E} \left[\left| \sum_{k=1}^n Z_k e^{X_{k,i}} \right|^2 \right]^{1/2} = \left(\sum_{k=1}^n \mathbb{E}[e^{2X_{k,i}}] \right)^{1/2} \leq \sqrt{n} \mathbb{E}[e^{2X_{T,i}^d}]^{1/2} \\
 &\leq \sqrt{n} \mathbb{E}[e^{pX_{T,i}^d}]^{1/p}.
 \end{aligned}$$

Combining this with the previous estimate (5.10) and with the exponential moment estimate (5.7) we obtain that

$$(5.11) \quad \mathbb{E} \left[\sup_{s \in [a,b]^d} \left| \mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}})] - \frac{1}{n} \sum_{k=1}^n \mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_k}) \right| \right] \leq \frac{c_2 d^{p+1}}{\sqrt{n}}$$

with $c_2 = 4\sqrt{\pi/2}cb \exp(5BTp/2 + BTpe^p)$. By applying Markov's inequality (see (4.8)-(4.10)) this proves that there exists $\omega \in \Omega$ with

$$(5.12) \quad \sup_{s \in [a,b]^d} \left| \mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}})] - \frac{1}{n} \sum_{k=1}^n \mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_k(\omega)}) \right| \leq \frac{2c_2 d^{p+1}}{\sqrt{n}}.$$

Using that $s \mapsto \frac{1}{n} \sum_{k=1}^n \mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_k(\omega)})$ is the realization of a neural network $\tilde{\psi}_{\bar{\varepsilon},d}$ with $M(\tilde{\psi}_{\bar{\varepsilon},d}) \leq nM(\phi_{\bar{\varepsilon},d})$ (see Lemma 3.2) we have therefore proved that there exists a neural network $\tilde{\psi}_{\bar{\varepsilon},d}$ with

$$(5.13) \quad \sup_{s \in [a,b]^d} \left| \mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}})] - \mathbb{R}(\tilde{\psi}_{\bar{\varepsilon},d})(s) \right| \leq \frac{2c_2 d^{p+1}}{\sqrt{n}}.$$

Step 4: In the final step we now provide appropriate choices of the hyperparameters. We select $\bar{\varepsilon} = \varepsilon(c_1 d^{\frac{3}{2}p+\frac{1}{2}} + 2)^{-1}$, choose $n = \lceil (2c_2 d^{p+1} \bar{\varepsilon}^{-1})^2 \rceil$, $D = \log(\bar{\varepsilon}^{-1} d^{p+1} \tilde{c}_1)$ and set $\psi_{\varepsilon,d} = \tilde{\psi}_{\bar{\varepsilon},d}$. Then the total number of parameters of the approximating neural network can be estimated using assumption (5.3) as

$$(5.14) \quad \begin{aligned} M(\psi_{\varepsilon,d}) &= M(\tilde{\psi}_{\bar{\varepsilon},d}) \leq nM(\phi_{\bar{\varepsilon},d}) \\ &\leq (1 + (2c_2 d^{p+1} \bar{\varepsilon}^{-1})^2) c d^p \bar{\varepsilon}^{-q} \\ &\leq (1 + 4c_2^2) c d^{3p+2} \bar{\varepsilon}^{-2-q} \\ &\leq [(1 + 4c_2^2) c (c_1 + 2)^{2+q}] d^{(\frac{3}{2}p+\frac{1}{2})(2+q)+3p+2} \varepsilon^{-2-q}, \end{aligned}$$

which shows the number of weights to be bounded polynomially in d and ε^{-1} , as claimed.

Finally, we combine (5.6), (5.8) and (5.13) to estimate the approximation error as

$$\begin{aligned} &\sup_{s \in [a,b]^d} |u_d(T, s) - \mathbb{R}(\psi_{\varepsilon,d})(s)| \\ &\leq \sup_{s \in [a,b]^d} \left\{ |u_d(T, s) - \mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^d})]| + |\mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^d})] - \mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}})]| \right. \\ &\quad \left. + |\mathbb{E}[\mathbb{R}(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}})] - \mathbb{R}(\tilde{\psi}_{\bar{\varepsilon},d})(s)| \right\} \\ &\leq \bar{\varepsilon} c_1 d^{\frac{3}{2}p+\frac{1}{2}} + \tilde{c}_1 e^{-D} d^{p+1} + \frac{2c_2 d^{p+1}}{\sqrt{n}} \\ &\leq \bar{\varepsilon} (c_1 d^{\frac{3}{2}p+\frac{1}{2}} + 2) = \varepsilon, \end{aligned}$$

as claimed. \square

5.2. Discussion of related results. As recently there have been several results on DNN expression rates in high dimensional diffusion models, a discussion on the relation of the multivariate DNN expression rate result, Thm.5.1, to other recent mathematical results on DNN expression rate bounds is in order. Given that geometric diffusion models are particular cases of the presently considered models (corresponding to $\nu^d = 0$ in the Lévy triplet), it is of interest to consider to which extent the DNN expression error bound Thm.5.1 relates to these results.

Firstly, we note that with the exception of [GGJ⁺19] and [EGJS18], previous results in the literature which are concerned with DNN approximation rates for Kolmogorov equations for diffusion processes (see, e.g., [GGJ⁺19, GHJZ19, BGJ18, EGJS18, GHJvW18, RZ19] and the references therein) study approximation with respect to the L^p -norm ($p < \infty$), whereas in Thm.5.1 we study approximation with respect to the L^∞ -norm, which requires entirely different techniques. While the results in [EGJS18] rely on specific structure of the payoff, the proof of the expression rates in [GGJ⁺19] has some similarities with the proof of Thm.5.1. However, the novelty in the proof of Thm.5.1 is the use of statistical learning techniques (symmetrization, Gaussian and Rademacher complexities) which allow for weaker assumptions on the activation function than in [GGJ⁺19]. In addition, the class of PDEs considered in [GGJ⁺19] (heat equation and related) is different than the one considered in Thm.5.1 (Black–Scholes PDE and Lévy PIDE).

Secondly, Thm.5.1 is the first result on ReLU DNN expression rates for option prices in models with jumps or, equivalently, for *partial-integro-differential equations* in non-divergence form

$$\begin{aligned}
 \partial_t v_d(\tau, x) &= \frac{1}{2} \text{Trace}(A^d D_x^2 v_d(\tau, x)) + D_x v_d(\tau, x) \gamma^d \\
 (5.15) \quad &+ \int_{\mathbb{R}^d} [v_d(\tau, x + y) - v_d(\tau, x) - D_x v_d(\tau, x) y \mathbb{1}_{\{\|y\| \leq 1\}}] \nu^d(dy), \\
 v_d(0, x) &= (\varphi_d \circ \exp)(x)
 \end{aligned}$$

for $x \in \mathbb{R}^d, \tau > 0$ or, when transformed from log-price variables x_i to actual price variables s_i via $(s_1, \dots, s_d) = (\exp(x_1), \dots, \exp(x_d))$,

$$\begin{aligned}
 \partial_t u_d(\tau, s) &= \frac{1}{2} \sum_{i,j=1}^d A_{i,j}^d s_i s_j \partial_{s_i} \partial_{s_j} u_d(\tau, s) + \sum_{i=1}^d s_i \gamma_i^d \partial_{s_i} u_d(\tau, s) \\
 &+ \int_{\mathbb{R}^d} [u_d(\tau, s_1 e^{y_1}, \dots, s_d e^{y_d}) - u_d(\tau, s) - \sum_{i=1}^d s_i (e^{y_i} - 1) \partial_{s_i} u_d(\tau, s)] \nu^d(dy), \\
 (5.16) \quad u_d(0, s) &= \varphi_d(s)
 \end{aligned}$$

for $s \in (0, \infty)^d, \tau > 0$ (see for instance [HRSW09, Theorem 4.1]). As in our assumptions also $A^d = 0$ is admissible under suitable conditions on ν^d , the present ReLU DNN expression rates are not mere generalizations of the diffusion case, but cover indeed the case of pure jump models, of finite and of infinite activity Lévy processes.

In the case $\nu^d = 0$, however, the Lévy PIDE reduces to a Black–Scholes PDE. In this particular case, we can therefore compare the result in Thm.5.1 to the results that have been proved in [GHJvW18]. The results in the latter article are specialized to the Black–Scholes case in Section 4 [GHJvW18], where Setting 4.1 specifies the coefficients $(A^d)_{i,j}$ (in our notation) as $\beta_i^d \beta_j^d (B^d (B^d)^\top)_{i,j}$ for some $\beta^d \in \mathbb{R}^d, B^d \in \mathbb{R}^{d \times d}$ satisfying $(B^d (B^d)^\top)_{k,k} = 1$ for all $d \in \mathbb{N}, i, j, k = 1, \dots, d$ and $\sup_{d,i} |\beta_i^d| < \infty$. The coefficient γ^d is chosen as α^d satisfying $\sup_{d,i} |\alpha_i^d| < \infty$. Using that $\Sigma = (B^d (B^d)^\top)$ is symmetric, positive definite we obtain $\Sigma_{i,j} \leq \sqrt{\Sigma_{i,i} \Sigma_{j,j}} = 1$ and hence these assumptions imply that (5.5) is satisfied. Therefore, the DNN expression rate results from Section 4 in [GHJvW18] can also be deduced from Thm.5.1, here in the case when the probability measure used to quantify the L^p -error in [GHJvW18] is compactly supported, as in that case the L^∞ -bounds proved here imply the L^p -bounds proved in [GHJvW18].

5.3. Exponential ReLU DNN expression rates via PIDE. We now extend the univariate case discussed in Section 4.3, and prove an exponential expression rate bound similar to Proposition 4.6 for baskets of $d \geq 2$ Lévy-driven assets. In this subsection we assume ReLU activation function $\varrho(x) = \max(x, 0)$. As in Section 5.1, we admit general correlation structure of the marginal processes' jumps. To prove DNN expression rate bounds, we exploit once more the fact that the stationarity and homogeneity of the LP $X^d \subset \mathbb{R}^d$ imply that the Kolmogorov equation (5.15) has constant coefficients. Under the provision that in (5.15) holds $v_d(0, \cdot) \in L^2(\mathbb{R}^d)$, this allows to write

for every $\tau > 0$ the Fourier transform $F_{x \rightarrow \xi} v_d(\tau, \cdot) = \hat{v}_d(\tau, \xi)$ as

$$(5.17) \quad \hat{v}_d(\tau, \xi) = \exp(-\tau\psi(\xi))\hat{v}_d(0, \xi), \quad \xi \in \mathbb{R}^d.$$

Here, for $\xi \in \mathbb{R}^d$ the symbol $\psi(\xi) = \exp(-ix^\top \xi) \mathcal{A}(\partial_x) \exp(ix^\top \xi)$, with $\mathcal{A}(\partial_x)$ denoting the constant coefficient spatial integrodifferential operator in (5.15), and (4.23) becomes

$$(5.18) \quad \mathbb{E}[\exp(i\xi^\top X_\tau^d)] = \exp(-\tau\psi(\xi)), \quad \xi \in \mathbb{R}^d.$$

In fact, ψ can be expressed in terms of the characteristic triplet (A^d, γ^d, ν^d) of the LP X^d as

$$(5.19) \quad \psi(\xi) = \frac{1}{2}\xi^\top A^d \xi - i\xi^\top \gamma^d - \int_{\mathbb{R}^d} \left[e^{i\xi^\top y} - 1 - i\xi^\top y \mathbb{1}_{\{\|y\| \leq 1\}} \right] \nu^d(dy), \quad \xi \in \mathbb{R}^d.$$

We impose again the strong ellipticity assumption (4.19), however now with $|\xi|$ understood as $|\xi|^2 = \xi^\top \xi$ for $\xi \in \mathbb{R}^d$. Then reasoning exactly as in the proof of Proposition 4.6 we obtain with $C_1 > 0$ as in (4.19) for every $\tau > 0$ for the variational solution v_d of (5.15) the bound

$$(5.20) \quad \forall k \in \mathbb{N}_0 : \quad \|(D_x^k v_d)(\tau, \cdot)\|_{L^2(\mathbb{R}^d)}^2 \leq \left(\frac{k}{2\tau C_1 \rho e} \right)^{k/\rho} \|v_d(0, \cdot)\|_{L^2(\mathbb{R}^d)}^2.$$

Here, D_x^k denotes any weak derivative of total order $k \in \mathbb{N}_0$ with respect to $x \in \mathbb{R}^d$.

With the Sobolev embedding we again obtain for any bounded cube $I^d := [x_-, x_+]^d \subset \mathbb{R}^d$ with $-\infty < x_- < x_+ < \infty$, and for every fixed $\tau > 0$, that there exist constants $C(d) > 0$ and $A(\tau, \rho) > 0$ such that

$$(5.21) \quad \forall k \in \mathbb{N} : \quad \|(D_x^k v_d)(\tau, \cdot)\|_{L^\infty(I^d)} \leq C(d) (A(\tau, \rho))^k (k!)^{1/\min\{1, 2\rho\}}.$$

The constant $C(d)$ is independent of x_-, x_+ , but depends in general exponentially on the basket size (resp. the dimension) $d \geq 2$, and the constant $A(\tau, \rho) = (2\tau C_1 \rho)^{-1/(2\rho)}$ denotes the constant from (5.20) and Stirling's bound. If $\rho = 1$ (which corresponds to the case of non-degenerate diffusion) and if $\tau > 0$ is sufficiently large (so that $(2\tau C_1)^{1/(2\rho)} \geq 1$) then the constant is bounded uniformly w.r. to the dimension d .

More generally, $v_d(\tau, \cdot)|_{I^d}$ is Gevrey- δ regular with $\delta = 1/\min\{1, 2\rho\}$. In particular, for $\delta = 1$, i.e. when $\rho \geq 1/2$, for every fixed $\tau > 0$, $x \mapsto v_d(\tau, x)$ is real analytic in I^d , which is the case we consider first.

In this case, $x \mapsto v_d(\tau, x)$ admits, with respect to each co-ordinate $x_i \in [x_-, x_+]$ of x , a holomorphic extension to an open neighborhood of $[x_-, x_+]$ in \mathbb{C} . By Hartogs' theorem (see, e.g., [Hö66]), for every fixed $\tau > 0$, $x \mapsto v_d(\tau, x)$ admits a holomorphic extension to a polyellipse in \mathbb{C}^d with foci at x_-, x_+ or, in normalized coordinates

$$(5.22) \quad \hat{x}_i = (T^{-1}(x))_i := 2[x_i - (x_- + x_+)/2]/(x_+ - x_-), \quad i = 1, \dots, d,$$

the map $[-1, 1]^d \ni \hat{x} \mapsto v_d(\tau, T(\hat{x})) =: v_c(\tau, \hat{x})$ admits a holomorphic extension to a Bernstein polyellipse $\mathcal{E}_\varrho \subset \mathbb{C}^d$ with foci at $\hat{x}_i = \pm 1$, and semiaxis sums $1 < \varrho = \mathcal{O}(A(\tau, \rho)^{-1})$. As $\tau \mapsto A(\tau, \rho)^{-1}$ is increasing for every fixed value of ρ , for $\rho \geq 1/2$ parabolic smoothing increases the domain of holomorphy with τ .

In the general case $\delta = 1/\min\{1, 2\rho\}$ with $\rho > 0$ as in (4.19), ReLU DNN expression rates of multivariate holomorphic (if $\rho \geq 1/2$) and Gevrey-regular (if $0 < \rho < 1/2$) functions such as $\hat{x} \mapsto v_c(\tau, \hat{x})$ have been studied in [OSZ19].

The holomorphy or Gevrey-regularity [depending on δ] of the map $\hat{x} \mapsto v_c(\tau, \hat{x})$ implies, with [OSZ19, Theorem 3.6, Proposition 4.1] that there exist constants $\beta' = \beta'(\varrho, d) > 0$ and $C =$

$C(u_d, \varrho, d) > 0$, and for every $\mathcal{N} \in \mathbb{N}$ there exists a ReLU DNN $\tilde{u}_{\mathcal{N}} : [-1, 1]^d \rightarrow \mathbb{R}$ such that

$$(5.23) \quad M(\tilde{u}_{\mathcal{N}}) \leq \mathcal{N}, \quad L(\tilde{u}_{\mathcal{N}}) \leq C\mathcal{N}^{\min\{\frac{1}{2}, \frac{1}{d+1/\delta}\}} \log(\mathcal{N})$$

and such that the error bound

$$(5.24) \quad \|v_c(\tau, \cdot) - \tilde{u}_{\mathcal{N}}(\cdot)\|_{W^{1,\infty}([-1,1]^d)} \leq C \exp\left(-\beta' \mathcal{N}^{\min\{\frac{1}{2\delta}, \frac{1}{\delta d+1}\}}\right)$$

holds. Reverting the affine change of variables (5.22) in the input layer, we obtain the following result on the ε -complexity of the ReLU DNN expression error for $x \mapsto v_d(\tau, x)$ at fixed $0 < \tau \leq T$.

Theorem 5.4. *Assume that the symbol ψ of the LP $X^d \subset \mathbb{R}^d$ satisfies (4.19) with $|\xi|^2 = \xi^\top \xi$ and with some $\rho \in (0, 1]$.*

Then, for every φ_d with $v_d(0, \cdot) = \varphi_d \circ \exp \in L^2(\mathbb{R}^d)$, for every $\tau > 0$ and with $\delta = 1/\min\{1, 2\rho\} \geq 1$, on every closed, bounded cube $I^d = [x_-, x_+]^d \subset \mathbb{R}^d$ and, resp., $J^d = [s_-, s_+]^d \subset (0, \infty)^d$ with $s_{\pm} = \exp(x_{\pm})$ the variational solutions v_d of the Kolmogorov PIDE (5.15) at τ and $u_d(\tau, s) = v_d(\tau, \log(s))$ can be expressed on I^d, J^d by ReLU DNNs $\tilde{v}_{d,\varepsilon}, \tilde{u}_{d,\varepsilon}$ at exponential rate.

Specifically, there exists a constant $C = C(x_-, x_+, \delta, d, \tau) > 0$ such that for every $0 < \varepsilon \leq 1/2$ exist ReLU DNNs $\tilde{v}_{d,\varepsilon}, \tilde{u}_{d,\varepsilon}$ for which there holds

$$\|v_d(\tau, \cdot) - \mathbf{R}(\tilde{v}_{d,\varepsilon})(\cdot)\|_{L^\infty(I^d)}, \quad \|u_d(\tau, \cdot) - \mathbf{R}(\tilde{u}_{d,\varepsilon})(\cdot)\|_{L^\infty(J^d)} \leq \varepsilon,$$

and,

$$M(\tilde{v}_{d,\varepsilon}) + M(\tilde{u}_{d,\varepsilon}) \leq C|\log(\varepsilon)|^{\max\{2\delta, \delta d+1\}}, \quad L(\tilde{v}_{d,\varepsilon}) + L(\tilde{u}_{d,\varepsilon}) \leq C|\log(\varepsilon)|^\delta |\log(|\log(\varepsilon)|)|.$$

Here, the constants $C = C(\delta, d, \tau) > 0$ depend on I and J and, generally, exponentially on the basket size d .

Proof. The asserted bounds for $\mathbf{R}(\tilde{v}_{d,\varepsilon})$ follow by elementary manipulations from insisting that the expression error bound (5.24) equal $\varepsilon \in (0, 1/2]$ and subsequently inserting the resulting expression $\mathcal{N} \simeq |\log(\varepsilon)|^{\max\{2\delta, \delta d+1\}}$ into the bounds (5.23) for the DNN size and depth.

The bounds for $\mathbf{R}(\tilde{u}_{d,\varepsilon})$ are then deduced from those for $\mathbf{R}(\tilde{v}_{d,\varepsilon})$ and the fact that the transformation $\log(\cdot) : J^d \rightarrow I^d$ (understood component-wise) is real analytic. Hence, it admits a holomorphic extension to an open neighbourhood of J^d in \mathbb{C}^d . Then [OSZ19, Thm. 3.6], combined with the affine transformation $T : [-1, 1]^d \rightarrow J^d$, implies that there are constants $C, \beta' > 0$ such that for every $\mathcal{N} \in \mathbb{N}$ exists a ReLU DNN $\widetilde{\log}_{\mathcal{N}}$ such that

$$(5.25) \quad M(\widetilde{\log}_{\mathcal{N}}(\cdot)) \leq \mathcal{N}, \quad L(\widetilde{\log}_{\mathcal{N}}(\cdot)) \leq C\mathcal{N}^{\frac{1}{d+1}} \log_2(\mathcal{N})$$

and the error bound

$$(5.26) \quad \left\| (\log \circ T)(\cdot) - \mathbf{R}(\widetilde{\log}_{\mathcal{N}}) \circ T(\cdot) \right\|_{W^{1,\infty}([-1,1]^d)} \leq C \exp\left(-\beta' \mathcal{N}^{\frac{1}{d+1}}\right).$$

Due to (5.26), for every $\mathcal{N} \in \mathbb{N}$ the set $\widetilde{I}^d := \mathbf{R}(\widetilde{\log}_{\mathcal{N}})(J^d) \cup \log(J^d) \subset (-\infty, \infty)^d$ is compact. For given $\varepsilon \in (0, 1/2]$, we choose $\mathcal{N} \in \mathbb{N}$ as before. Using that $\mathcal{N}^{\min\{\frac{1}{2\delta}, \frac{1}{\delta d+1}\}} \leq \mathcal{N}^{\frac{1}{d+1}}$, this choice guarantees in (5.26) that $C \exp\left(-\beta' \mathcal{N}^{\frac{1}{d+1}}\right) \leq \varepsilon$. Then we define $\tilde{u}_d(\tau, \cdot) = \mathbf{R}(\tilde{v}_{d,\varepsilon})(\cdot) \circ \mathbf{R}(\widetilde{\log}_{\mathcal{N}})(\cdot)$

and estimate

$$\begin{aligned}
\|u_d(\tau, \cdot) - \tilde{u}_d(\tau, \cdot)\|_{L^\infty(J^d)} &= \|v_d(\tau, \cdot) \circ \log(\cdot) - \tilde{u}_d(\tau, \cdot)\|_{L^\infty(J^d)} \\
&\leq \|v_d(\tau, \cdot) \circ \log(\cdot) - v_d(\tau, \cdot) \circ \mathbf{R}(\log_{\mathcal{N}})(\cdot)\|_{L^\infty(J^d)} \\
&\quad + \|v_d(\tau, \cdot) \circ \mathbf{R}(\log_{\mathcal{N}})(\cdot) - \mathbf{R}(\tilde{v}_{d,\varepsilon})(\cdot) \circ \mathbf{R}(\log_{\mathcal{N}})(\cdot)\|_{L^\infty(J^d)} \\
&\leq \|v_d(\tau, \cdot)\|_{W^{1,\infty}(\tilde{I}^d)} \|\log(\cdot) - \mathbf{R}(\log_{\mathcal{N}})(\cdot)\|_{L^\infty(J^d)} \\
&\quad + \|v_d(\tau, \cdot) - \mathbf{R}(\tilde{v}_{d,\varepsilon})(\cdot)\|_{L^\infty(\tilde{I}^d)} \\
&\leq C\varepsilon.
\end{aligned}$$

Since the DNN size and DNN depth are additive under composition of ReLU DNNs, the assertion for $\tilde{u}_{d,\varepsilon}$ follows (possibly adjusting the value of the constant C). \square

5.4. Remarks on the Curse of Dimensionality. The result Theorem 5.1 demonstrated DNN expression rates that are polynomial in ε^{-1} for European style options in multivariate, exponential Lévy models, with constants that depend polynomially on the number d of assets. In particular, in Theorem 5.1 the curse of dimensionality was proved to be overcome for a market model with jumps: a DNN expression rate was shown that is algebraic in terms of the target accuracy $\varepsilon > 0$ with constants that depend polynomially on the dimension d .

Theorem 5.4, on the other hand, stated *exponential expressivity* of deep ReLU NNs, i.e. maximum expression error at time $\tau > 0$ with accuracy $\varepsilon > 0$ can be attained by a deep ReLU NN of size and depth which grow polylogarithmically with respect to $|\log(\varepsilon)|$. This exponential expression rate bound was, however, still prone to the curse of dimensionality. We show in the present section that *parabolic smoothing* of the Kolmogorov PDE allows, *for sufficiently large times* for polynomial with respect to d bounds on the constants. We start by discussing more precisely the dependence of the constants in the proof of Theorem 5.4 on the dimension d .

Remark 5.5. The constant $C(d)$ in the derivative bound (5.21) need not be exponential in d . To see it, we bound (5.21) by the inverse Fourier transform and the Cauchy-Schwarz inequality. For $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = \sum_{i=1}^d \alpha_i = k$, we find with the Cauchy-Schwarz inequality and with the lower bound (4.19)

$$\begin{aligned}
&\|(D_x^\alpha v_d)(\tau, \cdot)\|_{L^\infty(I^d)} \\
&= \sup_{x \in I^d} \left| \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} (i\xi)^\alpha \exp(ix^\top \xi) \exp(-\tau\psi(\xi)) \hat{v}_d(0, \xi) d\xi \right| \\
(5.27) \quad &\leq \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} |\xi|^k \exp(-\tau\psi(\xi)) |\hat{v}_d(0, \xi)| d\xi \\
&\leq \frac{1}{(2\pi)^{d/2}} \left(\int_{\mathbb{R}^d} \exp(-2\tau C_1 |\xi|^{2\rho}) d\xi \right)^{1/2} \left(\int_{\mathbb{R}^d} |\xi|^{2k} \exp(-2\tau C_1 |\xi|^{2\rho}) |\hat{v}_d(0, \xi)|^2 d\xi \right)^{1/2}.
\end{aligned}$$

The last factor can be bounded precisely by the square-root of the right hand side of (5.20) (by using (4.26)) and so, using $k^k \leq k!e^k$ we obtain the bound (5.21) as

$$(5.28) \quad \|(D_x^\alpha v_d)(\tau, \cdot)\|_{L^\infty(I^d)} \leq C(d, \tau) (A(\tau, \rho))^k (k!)^{1/\min\{1, 2\rho\}} \|v_d(0, \cdot)\|_{L^2(\mathbb{R}^d)}$$

with constant $A(\tau, \rho) = (2\tau C_1 \rho)^{-1/(2\rho)}$ and the explicit constant

$$\begin{aligned}
 (5.29) \quad C(d, \tau) &= \frac{1}{(2\pi)^{d/2}} \left(\int_{\mathbb{R}^d} \exp(-2\tau C_1 |\xi|^{2\rho}) d\xi \right)^{1/2} \\
 &= \frac{1}{(2\pi)^{d/2}} \left(2 \frac{\pi}{d} \omega_d \int_0^\infty r^{d-1} \exp(-2\tau C_1 r^{2\rho}) dr \right)^{1/2} \\
 &= \frac{1}{(2\pi)^{d/2}} \left(\frac{\pi}{\rho d} \frac{1}{(2\tau C_1)^{d/(2\rho)}} \omega_d \Gamma\left(\frac{d}{2\rho}\right) \right)^{1/2} \\
 &= \frac{1}{(2\pi)^{d/2}} \left(\frac{\pi}{\rho d} \frac{1}{(2\tau C_1)^{d/(2\rho)}} \frac{\pi^{d/2} \Gamma(\frac{d}{2\rho})}{\Gamma(\frac{d}{2} + 1)} \right)^{1/2},
 \end{aligned}$$

where ω_d denotes the volume of the unit ball in \mathbb{R}^d . Inspecting the constant $C(d, \tau)$ in (5.29), we observe that e.g. for $\rho = 1$ and $\tau_0 = \tau_0(C_1) = 1/(8\pi C_1)$, $\tau \geq \tau_0 > 0$ sufficiently large implies that the constant $C(d, \tau)$ is bounded independent of τ and d .

Remark 5.6. In certain cases, the parabolic smoothing implied by the ellipticity assumption (4.19) on the generator \mathcal{A} entails that the constant C in the regularity estimates (5.21) grows only polynomially with respect to d . This allows to derive an explicit and dimension-independent bound on the series of Taylor coefficients. This, in turn, allows to obtain bounds on the constant in (5.24) which scale polynomially with respect to d . Consider, for example, $\rho = 1$ (i.e. non-degenerate diffusion) and assume that $\tau > 0$ is sufficiently large: specifically, $(2\tau C_1)^{1/(2\rho)} \geq 1$ and $dA(\tau, \rho) < 1$, where $A(\tau, \rho) = (2\rho\tau C_1)^{-1/(2\rho)}$ denotes the constant in the $[\cdot]$ parentheses of (4.28). This holds if

$$(5.30) \quad \tau > \frac{d^{2\rho}}{2\rho C_1}.$$

With (5.30) and using $\sum_{\alpha \in \mathbb{N}_0^d, |\alpha|=k} \binom{k}{\alpha} = d^k$, we may estimate with the multinomial theorem

$$\begin{aligned}
 \sum_{\alpha \in \mathbb{N}_0^d} \frac{\|(D_x^\alpha v_d)(\tau, \cdot)\|_{L^\infty(I^d)}}{\alpha!} &\leq C(d, \tau) \sum_{\alpha \in \mathbb{N}_0^d} \frac{A(\tau, \rho)^{|\alpha|} (|\alpha|)!}{\alpha!} \\
 &= C(d, \tau) \sum_{k=0}^{\infty} [dA(\tau, \rho)]^k = C(d, \tau) \frac{1}{1 - dA(\tau, \rho)}.
 \end{aligned}$$

By Remark 5.6, (5.30) implies that $C(d, \tau)$ in (5.29) is bounded uniformly with respect to d .

The constant $C > 0$ in the exponential expression rate bounds established in Theorem 5.4 depends in general exponentially on the basket size d , resp. on the dimension of the solution space of the PIDE (5.15), due to the reliance on the ReLU DNN expression rate analysis in [OSZ19]. Furthermore, the DNN size grows polylogarithmically with respect to the dimension d , in terms of $|\log(\varepsilon)|$. Considering exponential expression rate bounds, this degeneracy with respect to the dimension d seems, in general, not avoidable, as can be seen from [OSZ19, Theorem 3.5]. Nevertheless, in Remark 5.6 we already hinted at parabolic smoothing implying sufficient regularity for polynomial w.r. to d constants in DNN expression rate bounds.

In the following paragraphs, we settle for *algebraic DNN expression rates* and overcome exponential dependence on d in ReLU DNN expression error bounds under certain *sparsity assumptions on polynomial chaos expansions*, as shown in [SZ19], [CDS10] and the references there. We develop a variation of the results in [SZ19] in the present context.

We impose the following hypothesis, which takes the place of (4.19).

Assumption 2. *There exists a constant $C_1 > 0$ and $(\rho_j)_{j \in \mathbb{N}}$ with $\frac{1}{2} < \rho_j \leq 1$, such that for each $d \in \mathbb{N}$, the symbol ψ_{X^d} of the LP X^d satisfies that*

$$(5.31) \quad \forall \xi \in \mathbb{R}^d : \quad \Re \psi_{X^d}(\xi) \geq C_1 \sum_{j=1}^d |\xi_j|^{2\rho_j}.$$

Furthermore,

$$(5.32) \quad \rho := \inf_{j \in \mathbb{N}} \rho_j > \frac{1}{2}.$$

The payoff function φ_d in (5.15) is such that $v_d(0, \cdot) = \varphi_d \circ \exp \in L^2(\mathbb{R}^d)$.

Remark 5.7. Consider the pure diffusion case, i.e., when the characteristic triplet is $(A^d, 0, 0)$ with a symmetric, positive definite diffusion matrix A^d and Lévy-symbol $\psi_{X^d} : \mathbb{R}^d \rightarrow \mathbb{R} : \xi \mapsto \xi^\top A^d \xi$. A sufficient condition for assumption (5.31) to hold is that the eigenvalues $(\lambda_i^d)_{i=1, \dots, d}$ of A^d be lower bounded away from zero,

$$(5.33) \quad C_1 := \inf_{i,d} \lambda_i^d > 0.$$

To see this, write $Q^\top A^d Q = D$ for a diagonal matrix D containing the eigenvalues of A^d and an orthogonal matrix Q . Then we obtain for arbitrary $\xi \in \mathbb{R}^d$

$$\psi_{X^d}(\xi) = \xi^\top A^d \xi = (\xi^\top) Q D Q^\top \xi = \sum_{i=1}^d \lambda_i (Q^\top \xi)_i^2 \geq \left[\min_i \lambda_i \right] |Q^\top \xi|^2 = \left[\min_i \lambda_i \right] |\xi|^2.$$

Therefore condition (5.31) is satisfied with C_1 as in (5.33) and $\rho_j = 1$ for all $j \in \mathbb{N}$.

This condition imposes, in applications, that different assets (modelled by different components of the LP X^d) should not become asymptotically (perfectly) dependent as the dimension grows.

Remark 5.8. Consider characteristic triplets (A^d, γ^d, ν^d) and the more general case of non-degenerate diffusion, i.e. with A^d satisfying the condition (5.33) formulated in Remark 5.7. Then the real part of the Lévy symbol ψ_{X^d} of X^d satisfies for all $\xi \in \mathbb{R}^d$

$$\Re \psi_{X^d}(\xi) = \frac{1}{2} \xi^\top A^d \xi - \int_{\mathbb{R}^d} \left[\cos(\xi^\top y) - 1 \right] \nu^d(dy) \geq \frac{1}{2} \xi^\top A^d \xi \geq C_1 |\xi|^2$$

with C_1 as in (5.33). Hence, Assumption 2 is satisfied also in this more general situation. Further examples of LP satisfying Assumption 2 are based on stable-like processes and copula-based constructions as e.g. in [FRS07].

In comparison to the conditions in Section 5.2, the *anisotropic lower bound* in Assumption 2 is a different type of condition.

As we shall see below, Assumption 2 ensures good “separation” and “anisotropy” properties of the symbol (5.19) of the corresponding Lévy process X^d .

For $\tau > 0$ satisfying (5.30), we analyze the regularity of the map $x \mapsto v_d(\tau, x)$. From Assumption 2 we find that for every $\tau > 0$, $x \mapsto v_d(\tau, x) \in L^2(\mathbb{R}^d)$ and that its Fourier transform has the explicit form

$$(5.34) \quad \hat{v}_d(\tau, \xi) = F_{x \rightarrow \xi} v_d(\tau, \cdot) = \exp(-\tau \psi_{X^d}(\xi)) \hat{v}_d(0, \xi).$$

For a multi-index $\nu = (\nu_1, \dots, \nu_d) \in \mathbb{N}_0^d$, denote by ∂_x^ν the mixed partial derivative of total order $|\nu| = \nu_1 + \dots + \nu_d$ with respect to $x \in \mathbb{R}^d$. Formula (5.34) and Assumption 2 can be used to show that for every $\tau > 0$, $x \mapsto v_d(\tau, x)$ is analytic at any $x \in \mathbb{R}^d$. This is of course the well-known

smoothing property of the generator of certain non-degenerate Lévy processes. To address the curse of dimensionality, we quantify the smoothing effect in a d -explicit fashion.

To this end, with Assumption 2 we calculate for any $\boldsymbol{\nu} \in \mathbb{N}_0^d$ at $x = 0$ (by stationarity, the same bounds hold for the Taylor coefficients at any $x \in \mathbb{R}^d$)

$$(2\pi)^{d/2} |\partial_x^\nu v_d(\tau, 0)| = \left| \int_{\xi \in \mathbb{R}^d} (i\xi)^\nu \hat{v}_d(\tau, \xi) d\xi \right| \leq \int_{\xi \in \mathbb{R}^d} |\hat{v}_d(0, \xi)| \prod_{j=1}^d |\xi_j|^{\nu_j} \exp(-\tau C_1 |\xi_j|^{2\rho_j}) d\xi.$$

We use (4.26) with $m \leftarrow \nu_j$, $\kappa \leftarrow C_1 \tau$, $\mu = 2\rho_j$ to bound the product as

$$\prod_{j=1}^d |\xi_j|^{\nu_j} \exp(-\tau C_1 |\xi_j|^{2\rho_j}) \leq \prod_{j=1}^d \left(\frac{\nu_j}{2\rho_j \tau C_1 e} \right)^{\nu_j / (2\rho_j)}.$$

We arrive at the following bound for the Taylor coefficient of order $\boldsymbol{\nu} \in \mathbb{N}_0^d$ of $v_d(t, \cdot)$ at $x = 0$:

$$(5.35) \quad |t_\nu| = \left| \frac{1}{\boldsymbol{\nu}!} \partial_x^\nu v_d(\tau, x) \Big|_{x=0} \right| \leq \frac{1}{(2\pi)^{d/2}} \|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)} \prod_{j=1}^d \frac{1}{\nu_j!} \left(\frac{\nu_j}{2\rho_j \tau C_1 e} \right)^{\nu_j / (2\rho_j)}.$$

Stirling's inequality

$$(5.36) \quad \forall n \in \mathbb{N} : \quad n! \geq n^n e^{-n} \sqrt{2\pi n} \geq n^n e^{-n}$$

implies in (5.35) the bound

$$(5.37) \quad \forall \boldsymbol{\nu} \in \mathbb{N}_0^d : \quad |t_\nu| \leq \frac{1}{(2\pi)^{d/2}} \|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)} ((\boldsymbol{\nu}!)^{-1} \mathbf{b}^\nu)^{\rho'}.$$

Here, $\rho' = 1 - \frac{1}{2\rho} > 0$ and the positive weight sequence $\mathbf{b} = (b_j)_{j \geq 1}$ is given by $b_j := (2\rho_j \tau C_1)^{-1/(2\rho_j \rho')}$, $j = 1, 2, \dots$ and multi-index notation is employed: $\boldsymbol{\nu}^{-\nu} := (\nu_1^{\nu_1} \nu_2^{\nu_2} \dots)^{-1}$, $\mathbf{b}^\nu = b_1^{\nu_1} b_2^{\nu_2} \dots$ and $\boldsymbol{\nu}! = \nu_1! \nu_2! \dots$, with the convention $0! := 1$ and $0^0 := 1$.

We raise (5.37) to a power $q > 0$, with $q < 1/\rho'$ and sum the resulting inequality over all $\boldsymbol{\nu} \in \mathbb{N}_0^d$ to estimate (generously)

$$\sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} |t_\nu|^q \leq \frac{\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}^q}{(2\pi)^{dq/2}} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} \left(\frac{1}{\boldsymbol{\nu}!} \mathbf{b}^\nu \right)^{q\rho'} \leq \frac{\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}^q}{(2\pi)^{dq/2}} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} \left(\frac{|\boldsymbol{\nu}!}{\boldsymbol{\nu}!} \mathbf{b}^\nu \right)^{q\rho'}.$$

To obtain the estimate (5.37), one could also use the L^2 -bound with explicit constant derived in (5.28), (5.29).

Under hypothesis (5.31) and for $\tau > 0$ satisfying (5.30), we can now prove q -summability of the Taylor coefficients as follows.

$$\begin{aligned} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} |t_\nu|^q &\leq \frac{\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}^q}{(2\pi)^{dq/2}} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} \left(\frac{|\boldsymbol{\nu}!}{\boldsymbol{\nu}!} \mathbf{b}^\nu \right)^{q\rho'} \\ &\leq \frac{\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}^q}{(2\pi)^{dq/2}} \sum_{k=0}^{\infty} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d: |\boldsymbol{\nu}|=k} \left(\frac{|\boldsymbol{\nu}!}{\boldsymbol{\nu}!} (2\rho\tau C_1)^{-k/(2\rho\rho')} \right)^{q\rho'} \\ &\leq \frac{\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}^q}{(2\pi)^{dq/2}} \sum_{k=0}^{\infty} (2\rho\tau C_1)^{-qk/(2\rho)} \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d: |\boldsymbol{\nu}|=k} \left(\frac{|\boldsymbol{\nu}!}{\boldsymbol{\nu}!} \right)^{q\rho'}. \end{aligned}$$

Using that $|\boldsymbol{\nu}'| \geq \boldsymbol{\nu}'$ and that $1 \geq q\rho' > 0$ we obtain with the multinomial theorem

$$\sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d: |\boldsymbol{\nu}|=k} \left(\frac{|\boldsymbol{\nu}'|}{\boldsymbol{\nu}'} \right)^{q\rho'} \leq \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d: |\boldsymbol{\nu}|=k} \frac{|\boldsymbol{\nu}'|}{\boldsymbol{\nu}'} = d^k$$

and so, provided that

$$(5.38) \quad \tau > \tau_0(d) \text{ with } \tau_0(d) := \frac{d^{2\rho/q}}{2\rho C_1},$$

it follows that

$$(5.39) \quad \|\{t_{\boldsymbol{\nu}}\}\|_{\ell^q(\mathbb{N}_0^d)}^q = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} |t_{\boldsymbol{\nu}}|^q \leq \frac{\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}^q}{(2\pi)^{dq/2}} \frac{1}{1 - d(2\rho\tau C_1)^{-q/(2\rho)}}.$$

Therefore, we have proved q -summability of the Taylor coefficients of $x \mapsto v_d(\tau, x)$ at $x = 0$ for any $\tau > \tau_0(d)$. The q -norm $\|\{t_{\boldsymbol{\nu}}\}\|_{\ell^q(\mathbb{N}_0^d)}$ is bounded independently of d , provided that $\tau > \tau_0(d)$ and $\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}(2\pi)^{-d/2}$ is bounded independently of d .

The q -summability (5.39) of the Taylor coefficients of $x \mapsto v_d(\tau, x)$ at $x = 0$ with $q = 1$ implies for $\tau > \tau_0(d)$ unconditional, pointwise convergence in the cube $[-1, 1]^d$ of

$$(5.40) \quad v_d(\tau, x) = \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d} t_{\boldsymbol{\nu}} x^{\boldsymbol{\nu}}, \quad x^{\boldsymbol{\nu}} := x_1^{\nu_1} x_2^{\nu_2} \dots$$

Furthermore, as was shown in [SZ19, Lemma 2.8], the fact that the sequence $\{t_{\boldsymbol{\nu}}\}$ is q -summable for some $0 < q < 1$ and the coefficient bound (5.37) imply that for $\tau > \tau_0(d)$ exists a sequence $\{\Lambda_n\}_{n \geq 1} \subset \mathbb{N}_0^d$ of *nested, downward closed multi-index sets*¹ $\Lambda_n \subset \mathbb{N}_0^d$ with $\#(\Lambda_n) \leq n$ such that general polynomial chaos (gpc for short) approximations given by the partial sums

$$(5.41) \quad v_d^{\Lambda_n}(\tau, x) := \sum_{\boldsymbol{\nu} \in \Lambda_n} t_{\boldsymbol{\nu}} x^{\boldsymbol{\nu}}$$

converge at dimension-independent rate $r = 1/q - 1$ (see, e.g., [CDS10, Lemma 5.5])

$$(5.42) \quad \|v_d(\tau, \cdot) - v_d^{\Lambda_n}(\tau, \cdot)\|_{L^\infty([-1, 1]^d)} \leq \sum_{\boldsymbol{\nu} \in \mathbb{N}_0^d \setminus \Lambda_n} |t_{\boldsymbol{\nu}}| \leq n^{-(1/q-1)} \|\{t_{\boldsymbol{\nu}}\}\|_{\ell^q(\mathbb{N}_0^d)}.$$

The summability (5.39) of the coefficients in the Taylor gpc expansion (5.40) also implies quantitative bounds on the expression rates of ReLU DNNs. With [SZ19, Thm.2.7, (ii)], we find that there exists a constant $C > 0$ independent of d such that

$$\sup_{\boldsymbol{\nu} \in \Lambda_n} |\boldsymbol{\nu}|_1 \leq C(1 + \log(n)).$$

We now refer to [SZ19, Thm. 3.9] (with q in place of p in the statement of that result) and, observing that in the proof of that theorem, only the p -summability of the Taylor coefficient sequence $\{t_{\boldsymbol{\nu}}\}$ was used, we conclude that for $\tau > 0$ satisfying (5.38) there exists a constant $C > 0$ that is independent of d and, for every $n \in \mathbb{N}$ exists a ReLU DNN \tilde{v}_d^n with input dimension d , such that

$$(5.43) \quad M(\tilde{v}_d^n) \leq C(1 + n \log(n) \log(\log(n))), \quad L(\tilde{v}_d^n) \leq C(1 + \log(n) \log(\log(n))),$$

$$\|v_d(\tau, \cdot) - \mathbf{R}(\tilde{v}_d^n)\|_{L^\infty([-1, 1]^d)} \leq Cn^{-(1/q-1)}.$$

¹I.e., if $\mathbf{e}_j \in \Lambda_n$ then $\mathbf{e}_i \in \Lambda_n$ for all $0 \leq i \leq j$.

6. CONCLUSION AND GENERALIZATIONS

We proved that prices of European style derivative contracts on baskets of $d \geq 1$ asset in exponential Lévy models can be expressed by ReLU DNNs to accuracy $\varepsilon > 0$ with DNN size polynomially growing in ε^{-1} and d , thereby overcoming the curse of dimensionality. The technique of proof was based on probabilistic arguments and provides expression rate bounds that scale algebraically in terms of the DNN size. We then also provided an alternative, analytic argument, that allows to prove *exponential expressivity of ReLU DNNs* of the option price, i.e. of the map $s \mapsto u(t, s)$ at any fixed time $0 < t < T$, with DNN size growing polynomially w.r. to $\log(\varepsilon)$ to achieve accuracy $\varepsilon > 0$. For sufficiently large $t > 0$, based on analytic arguments involving parabolic smoothing and sparsity of generalized polynomial chaos expansions, we established in (5.43) a second, algebraic expression rate bound for ReLU DNNs that is free from the curse of dimensionality. The main result, Thm. 5.1, could be extended in various directions.

First, the expression rates are, almost certainly, not optimal in general; for high-dimensional diffusions, which are a particular case with $A^d = I$ and $\nu^d = 0$, in [EGJS18] we established for particular payoff functions a spectral expression rate in terms of the DNN size, free from the curse of dimensionality.

Solving Hamilton-Jacobi partial integrodifferential equations (HJPIDEs for short) by DNNs: it is classical that the Kolmogorov equation for the exponential LP X^d in Section 2.2 is, in fact, a HJPIDE (e.g. [BBP97, BI08]). The expression rate bounds obtained in Section 5 therefore imply corresponding expression rate bounds for ReLU DNNs which are free from the curse of dimensionality for viscosity solutions of the HJPIDEs of the LP X^d and for its exponential counterparts. The present results constitute to our knowledge the first ReLU DNN expression rate bounds which scale polynomially w.r. to d for HJPIDEs.

Barriers: We considered payoff functions corresponding to European style contracts. Here, the stationarity of the LP X^d and exponential Lévy modelling allowed to reduce our analysis to Cauchy problems of the Kolmogorov equations of X^d in \mathbb{R}^d . In Lévy models in the presence of barriers, option prices generally exhibit singularities at the barriers. More involved versions of the Fourier transform based representations are available (involving a so-called Wiener-Hopf factorization of the Fourier symbol, see, e.g., [BL02]). For LPs X^d with bounded exponential moments, the present regularity analysis may be localized to compact subsets, well separated from the barriers, subject to an exponentially small localization error term; see [HRSW13, Chap. 10.5]. Here, the semiheavy tails of the LPs X^d enter crucially in the analysis. We therefore expect the present DNN expression rate bounds to remain valid also for barrier contracts, at least far from the barriers, for the LPs X^d considered here.

Dividends: We assumed throughout that contracts do not pay dividends; however, including a dividend stream (with constant over $(0, T]$ rate) on the underlying does not change the mathematical arguments; we refer to [LM08, Section 3.1] for a complete statement of exponential Lévy models with constant dividend payment rate $\delta > 0$, and for the corresponding pricing of European and American style contracts for such models.

American style contracts: Deep learning based algorithms for the numerical solution of optimal stopping problems for Markovian models have been recently proposed in [BCJ19]. For the particular case of American style contracts in exponential Lévy models, [LM08] provide an analysis in the univariate case, and establish qualitative properties of the exercise boundary $\{(b(t), t) : 0 < t < T\}$. Here, for geometric Lévy models, in certain situations ($d = 1$, i.e. single risky asset, monotonic, piecewise analytic payoff function) the option price, as a function of $x \in \mathbb{R}$ at fixed $0 < t < T$, is shown in [LM08] to be a piecewise analytic function which is, globally, Hölder continuous with a possibly algebraic singularity at the exercise boundary $b(t)$. This holds, likewise, for the price

expressed in the logarithmic coordinate $x = \log(s)$. The ReLU DNN expression rate of such functions has been analyzed in [OPS20, Section 5.4]. In higher dimensions $d > 1$, recently also higher Hölder regularity of the price in symmetric, stable Lévy models has been obtained for smooth payoffs in [BFRO18].

REFERENCES

- [BBP97] Guy Barles, Rainer Buckdahn, and Etienne Pardoux, *Backward stochastic differential equations and integral-partial differential equations*, Stochastics Stochastics Rep. **60** (1997), no. 1-2, 57–83. MR 1436432
- [BCJ19] Sebastian Becker, Patrick Cheridito, and Arnulf Jentzen, *Deep optimal stopping*, Journal of Machine Learning Research **20** (2019), no. 74, 1–25.
- [Ber96] Jean Bertoin, *Lévy processes*, Cambridge Tracts in Mathematics, vol. 121, Cambridge University Press, Cambridge, 1996. MR 1406564
- [BFRO18] Begoña Barrios, Alessio Figalli, and Xavier Ros-Oton, *Free boundary regularity in the parabolic fractional obstacle problem*, Comm. Pure Appl. Math. **71** (2018), no. 10, 2129–2159. MR 3861075
- [BGJ18] Julius Berner, Philipp Grohs, and Arnulf Jentzen, *Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations*, Revision requested from SIAM Journal on Mathematics of Data Science; arXiv:1809.03062 (2018), 35 pages.
- [BGJ20] Christian Beck, Lukas Gonon, and Arnulf Jentzen, *Overcoming the curse of dimensionality in the numerical approximation of high-dimensional semilinear elliptic partial differential equations*, Tech. Report 2020-16, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2020.
- [BGTW19] Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood, *Deep hedging*, Quantitative Finance **19** (2019), no. 8, 1271–1291.
- [BI08] Guy Barles and Cyril Imbert, *Second-order elliptic integro-differential equations: viscosity solutions’ theory revisited*, Ann. Inst. H. Poincaré Anal. Non Linéaire **25** (2008), no. 3, 567–585. MR 2422079
- [BL02] Svetlana Boyarchenko and Sergei Levendorskiĭ, *Barrier options and touch-and-out options under regular Lévy processes of exponential type*, Ann. Appl. Probab. **12** (2002), no. 4, 1261–1298. MR 1936593
- [BLM13] Stéphane Boucheron, Gabor Lugosi, and Pascal Massart, *Concentration inequalities: A nonasymptotic theory of independence*, OUP Oxford, 2013.
- [BM03] Peter L. Bartlett and Shahar Mendelson, *Rademacher and Gaussian complexities: Risk bounds and structural results*, Journal of Machine Learning Research **3** (2003), no. 3, 463–482.
- [BSW13] Björn Böttcher, René Schilling, and Jian Wang, *Lévy matters. III*, Lecture Notes in Mathematics, vol. 2099, Springer, Cham, 2013, Lévy-type processes: construction, approximation and sample path properties, With a short biography of Paul Lévy by Jean Jacod, Lévy Matters. MR 3156646
- [CDS10] Albert Cohen, Ronald DeVore, and Christoph Schwab, *Convergence rates of best N -term Galerkin approximations for a class of elliptic sPDEs*, Found. Comput. Math. **10** (2010), no. 6, 615–646. MR 2728424
- [CT04] Rama Cont and Peter Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, 2004.
- [CV05] Rama Cont and Ekaterina Voltchkova, *Integro-differential equations for option prices in exponential lévy models*, Finance and Stochastics **9** (2005), 299–325.
- [CV06] ———, *A Finite Difference Scheme for Option Pricing in Jump Diffusion and Exponential Lévy Models*, SIAM Journal on Numerical Analysis **43** (2006), no. 4, 1596–1626.
- [EG14] Ernst Eberlein and Kathrin Glau, *Variational solutions of the pricing PIDEs for European options in Lévy models*, Appl. Math. Finance **21** (2014), no. 5, 417–450. MR 3253890
- [EGJS18] Dennis Elbrächter, Philipp Grohs, Arnulf Jentzen, and Christoph Schwab, *Dnn expression rate analysis of high-dimensional pdes: Application to option pricing*, Tech. Report 2018-33, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2018.
- [EK19] Ernst Eberlein and Jan Kallsen, *Mathematical finance*, Springer Finance, Springer, Cham, [2019] ©2019. MR 3971545
- [ES05] Felix Esche and Martin Schweizer, *Minimal entropy preserves the Lévy property: how and why*, Stochastic Processes and their Applications **115** (2005), no. 2, 299 – 327.
- [FRS07] Walter Farkas, Nils Reich, and Christoph Schwab, *Anisotropic stable Lévy copula processes—analytical and numerical aspects*, Math. Models Methods Appl. Sci. **17** (2007), no. 9, 1405–1443. MR 2353148
- [GGJ⁺19] Lukas Gonon, Philipp Grohs, Arnulf Jentzen, David Kofler, and David Šiška, *Uniform error estimates for artificial neural network approximations for heat equations*, no. 2019-61.

- [GHJvW18] Philipp Grohs, Fabian Hornung, Arnulf Jentzen, and Philippe von Wurstemberger, *A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations*, To appear in *Memoirs of the American Mathematical Society*; arXiv:1809.02362 (2018), 124 pages.
- [GHJZ19] Philipp Grohs, Fabian Hornung, Arnulf Jentzen, and Philipp Zimmermann, *Space-time error estimates for deep neural network approximations for differential equations*, arXiv:1908.03833 (2019), 86 pages.
- [Gla16] Kathrin Glau, *A Feynman-Kac-type formula for Lévy processes with discontinuous killing rates*, *Finance Stoch.* **20** (2016), no. 4, 1021–1059. MR 3551859
- [GS94] Hans U. Gerber and Elias S. W. Shiu, *Option pricing by Esscher transforms*, *Transactions of the Society of Actuaries* **46** (1994), 99–191.
- [Hö66] Lars Hörmander, *An introduction to complex analysis in several variables*, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1966. MR 0203075
- [HRSW09] Norbert Hilber, Nils Reich, Christoph Schwab, and Christoph Winter, *Numerical methods for Lévy processes*, *Finance and Stochastics* **13** (2009), 471–500.
- [HRSW13] Norbert Hilber, Oleg Reichmann, Christoph Schwab, and Christoph Winter, *Computational Methods for Quantitative Finance*, Springer Finance, Springer, Heidelberg, 2013, Finite element methods for derivative pricing. MR 3026658
- [IRZ19] Kazufumi Ito, Christoph Reisinger, and Yufei Zhang, *A neural network based policy iteration algorithm with global H^2 -superlinear convergence for stochastic games on domains*, 2019.
- [JKM07] Monique Jeanblanc, Susanne Klöppel, and Yoshio Miyahara, *Minimal $f q$ -martingale measures for exponential Lévy processes*, *Ann. Appl. Probab.* **17** (2007), no. 5-6, 1615–1638.
- [KT06] Jan Kallsen and Peter Tankov, *Characterization of dependence of multidimensional Lévy processes using Lévy copulas*, *J. Multivariate Anal.* **97** (2006), no. 7, 1551–1572. MR 2275419
- [LM08] Damien Lamberton and Mohammed Mikou, *The critical price for the American put in an exponential Lévy model*, *Finance Stoch.* **12** (2008), no. 4, 561–581. MR 2447412
- [LT13] Michel Ledoux and Michel Talagrand, *Probability in Banach Spaces*, Springer Berlin Heidelberg, 2013.
- [MvPS04] Ana-Maria Matache, Tobias von Petersdorff, and Christoph Schwab, *Fast deterministic pricing of options on Lévy driven assets*, *M2AN Math. Mod. and Num. Anal.* **38** (2004), 37–71.
- [NS01] David Nualart and Wim Schoutens, *Backward stochastic differential equations and Feynman-Kac formula for Lévy processes, with applications in finance*, *Bernoulli* **7** (2001), no. 5, 761–776. MR 1867081
- [OPS20] Joost A. A. Opschoor, Philipp C. Petersen, and Christoph Schwab, *Deep ReLU networks and high-order finite element methods*, *Analysis and Applications* (2020), 1–56.
- [OSZ19] Joost A. A. Opschoor, Christoph Schwab, and Jakob Zech, *Exponential ReLU DNN expression of holomorphic maps in high dimension*, Tech. Report 2019-35, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2019, (revised version).
- [PV18] Philipp Petersen and Felix Voigtlaender, *Optimal approximation of piecewise smooth functions using deep ReLU neural networks*, *Neural Netw.* **108** (2018), 296 – 330.
- [Rod93] Luigi Rodino, *Linear partial differential operators in Gevrey spaces*, World Scientific Publishing Co., Inc., River Edge, NJ, 1993. MR 1249275
- [RW20] Johannes Ruf and Weiguan Wang, *Neural networks for option pricing and hedging: a literature review*, Preprint (2020).
- [RZ19] Christoph Reisinger and Yufei Zhang, *Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems*, arXiv:1903.06652 (2019), 34 pages.
- [Sat99] Ken-Iti Sato, *Lévy processes and infinitely divisible distributions*, Cambridge University Press, 1999.
- [SZ19] Christoph Schwab and Jakob Zech, *Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ* , *Analysis and Applications*, Singapore **17** (2019), no. 1, 19–55.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MUNICH, THERESIENSTRASSE 39, 80333 MUNICH, GERMANY
Email address: gonon@math.lmu.de

SEMINAR FOR APPLIED MATHEMATICS, ETH ZÜRICH, RÄMISTRASSE 101, CH-8092 ZÜRICH, SWITZERLAND
Email address: christoph.schwab@sam.math.ethz.ch