Deep ReLU network expression rates for option prices in high-dimensional, exponential Lévy models

Deep ReLU network expression rates for option prices in high-dimensional, exponential Lévy models

Lukas Gonon[1] · Christoph Schwab[2]

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 in d, thereby overcoming the curse of dimensionality (CoD) 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 integro-differential equations for certain multivariate Lévy processes to present alternative architectures of ReLU ("rectified linear unit") DNNs that provide ε expression error in DNN size $\mathcal{O}(|\log(\varepsilon)|^a)$ with exponent a proportional to d, but 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 presented results.

✉ L. Gonon
gonon@math.lmu.de

C. Schwab
christoph.schwab@sam.math.ethz.ch

[1] Department of Mathematics, University of Munich, Theresienstrasse 39, 80333 Munich, Germany

[2] Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland

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 financial engineering. We refer to the survey by Ruf and Wang [47] and 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. Berner et al. [9], Elbrächter et al. [22] and Ito et al. [34], Reisinger and Zhang [45] 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 Grohs et al. [29], Gonon et al. [27] 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 non-stationary generalisations such as Feller–Lévy processes (see e.g. Böttcher et al. [11, Chap. 2] 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 Nualart and Schoutens [41], Cont and Tankov [16, Sect. 12.2], Cont and Voltchkova [18], Glau [26], Eberlein and Kallsen [21, 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 integro-differential operator accounting for the jumps is related in a one-to-one fashion with the Lévy measure ν^d of the $\mathbb{R}^d$-valued Lévy process X^d. In particular, Lévy-type models for (log-)returns of risky assets result in *nonlocal partial integro-differential equations* for the option price which generalise the linear parabolic differential equations which arise in classical diffusion models. We refer to Bertoin [10, Chap. 1], Sato [48, Chaps. 1–5] for fundamentals on Lévy processes and to Böttcher et al.

[11, Chap. 2] for extensions to certain non-stationary settings. For the use of Lévy processes in financial modelling, we refer to Cont and Tankov [16, Chap. 11], Eberlein and Kallsen [21, Sect. 8.1] and the references there. We refer to Cont and Voltchkova [18, 17], Matache et al. [40], Hilber et al. [32, Chap. 14] 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 to d. This shows that DNNs are capable of overcoming the curse of dimensionality also for general exponential Lévy models.

Let us outline the scope of our 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 by Ruf and Wang [47]. Our results show that if option prices result from an exponential Lévy model, as described e.g. in Eberlein and Kallsen [21, Chap. 3.7], these prices can under mild conditions on the Lévy triplets 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 parametrised by a so-called *Lévy copula*; see Kallsen and Tankov [36], Farkas et al. [24], Eberlein and Kallsen [21, 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 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 $\mathcal{O}(\varepsilon^{-2})$ and with constants implied in $\mathcal{O}(\cdot)$ which grow polynomially in d, thereby overcoming the curse of dimensionality. The latter bound is based on parabolic smoothing of the Kolmogorov equation and allows us to prove *exponential expressivity of prices for positive maturities*, i.e., an expression error $\mathcal{O}(\varepsilon)$ with DNN sizes of $\mathcal{O}(|\log \varepsilon|^a)$ for some $a > 0$, albeit with constants implied in $\mathcal{O}(\cdot)$ possibly growing exponentially in d.

For the latter approach, a certain non-degeneracy is required for the symbol of the underlying Lévy process. The probabilistic proof of the 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 on existence of moments. Thus the results proved here are specifically tailored to the class of op-

tion 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 Sect. 2, we review terminology, basic results and financial modelling with exponential Lévy processes. In particular, we also recapitulate the corresponding fractional, partial integro-differential Kolmogorov equations which generalise 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 hold for more general activation functions. In Sect. 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 lighter notation, and for introducing 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 Schwab and Zech [49] 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 Sect. 6 and indicate several possible generalisations 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, has almost surely RCLL sample paths, satisfies $X_0 = 0$ almost surely and has stationary and independent increments. See e.g. Bertoin [10, Sect. I.1], Sato [48, Definition 1.6] for discussion and for detailed statements of definitions.

It is shown in these references that a Lévy process (LP for short) X is characterised by its so-called Lévy triplet (σ^2, γ, ν), where $\sigma \geq 0$, $\gamma \in \mathbb{R}$ and 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 and the multivariate situation, we refer to [48, Chap. 2].

As in the univariate case, multivariate ($\mathbb{R}^d$-valued) LPs X^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 characterise 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 characterised 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 characterised by the Lévy measure ν^d.

In Kallsen and Tankov [36], a characterisation of admissible jump measures ν^d for $\mathbb{R}^d$-valued LPs X^d has been obtained as *superposition* of marginal, univariate Lévy measures with a so-called *Lévy copula* function.

2.2 Exponential Lévy models

In this article, we are interested in estimating expression rates of deep neural networks for approximating the function $s \mapsto \mathbb{E}[\varphi(s S_T)]$, where S is an exponential of a d-dimensional Lévy process and $\varphi \colon \mathbb{R}^d \to \mathbb{R}$ an appropriate function. The key motivation for studying such expectations comes from the context of option valuation. Thus, we now outline this relation and always use the language of option pricing, i.e., we refer to these expectations as option prices and to φ as the payoff. This interpretation is justified if S is a martingale, and we state below conditions on the Lévy process that guarantee this.

Let the $\mathbb{R}$-valued stochastic process $(S_t)_{t \in [0,T]}$ 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 Sato [48, Theorems 33.1 and 33.2] of which the Esscher transform in Gerber and Shiu [25] is a particular case, or by minimising certain functionals over the family of equivalent martingale measures; see for instance Jeanblanc et al. [35], Esche and Schweizer [23] and the references therein). The latter situation means that $(S_t e^{-rt})_{t \in [0,T]}$ is a martingale, which is equivalent to the condition on the Lévy triplet of X (see e.g. Hilber et al. [32, Lemma 10.1.5]) that

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

For a d-dimensional Lévy process X^d, [48, Theorem 25.17] shows that the multivariate geometric Lévy process $(e^{X_{t,1}^d}, \ldots, e^{X_{t,d}^d})_{t \geq 0}$ is a martingale if and only if

$$\int_{\{|y| > 1\}} e^{y_i} \nu^d(dy) < \infty \qquad \text{for } i = 1, \ldots, d,$$

$$\gamma_i^d = -\frac{A_{ii}^d}{2} - \int_{\mathbb{R}^d} (e^{y_i} - 1 - y_i \mathbb{1}_{\{|y| \leq 1\}}) \nu^d(dy) \qquad \text{for } i = 1, \ldots, d. \qquad (2.2)$$

This condition ensures that the functions defined in (2.3) and (5.1) below represent option prices. However, the condition is not needed for the proof of the results later,

and so we do not need to impose (2.1) or (2.2) in any of the results proved in the article. We shall, however, impose certain moment or regularity conditions.

For more details on exponential Lévy models, with particular attention to their use in financial modelling, we refer to Cont and Tankov [16, Chap. 11], Lamberton and Mikou [38] and Eberlein and Kallsen [21, Sect. 8.1] 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 Sect. 5 (cf. (5.1) and (5.14) below).

Consider a European-style option with payoff function $\varphi \colon (0, \infty) \to [0, \infty)$ and at most polynomial (pth order) growth at infinity. Assume for this subsection that (2.1) is satisfied.

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 the σ-field $\mathcal{F}_t = \sigma(S_v : v \in [0, t])$. 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

$$u(\tau, s) = \mathbb{E}[e^{-r\tau}\varphi(S_T)|S_t = s] = \mathbb{E}\big[e^{-r\tau}\varphi\big(s \exp(r\tau + X_\tau)\big)\big] \qquad (2.3)$$

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 $[0, T) \times (0, \infty)$, $C^{1,2}$ on $(0, T) \times (0, \infty)$ and satisfies the *linear, parabolic partial integro-differential equation* (PIDE for short)

$$\frac{\partial u}{\partial \tau}(\tau, s) - rs\frac{\partial u}{\partial s}(\tau, s) - \frac{\sigma^2 s^2}{2}\frac{\partial^2 u}{\partial s^2}(\tau, s) - ru(\tau, s)$$

$$- \int_{\mathbb{R}}\left(u(\tau, se^y) - u(\tau, s) - s(e^y - 1)\frac{\partial u}{\partial s}(\tau, s)\right)\nu(dy) = 0 \qquad (2.4)$$

on $[0, T) \times (0, \infty)$ with initial condition $u(0, \cdot) = \varphi$; see for instance Cont and Voltchkova [18, Proposition 2]. If the non-degeneracy condition on ν is dropped, one can still characterise u (transformed to log-price variables) as the unique viscosity solution to the PIDE above. This is established e.g. in [18] (see also Cont and Voltchkova [17, Proposition 3.3]). For our purposes, the representation (2.3) is more suitable. However, by using this characterisation (also called Feynman–Kac representation for viscosity solutions of PIDEs; see Barles et al. [2]), 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. Lamberton and Mikou [38, Eq. (3.1)] for the functional form of the exponential Lévy model in that case).

3 Deep neural networks (DNNs)

This article is concerned with establishing expression rate bounds of deep neural networks (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. The term "expression rate" denotes the rate of convergence to 0 of the error between the option price and its DNN approximation. This rate can be directly translated to quantify the DNN size required to achieve a given approximation accuracy. For instance, in Theorem 5.1 below, an expression rate of q^{-1} is established and one may even choose $q = 2$ in many relevant cases. We now give a brief introduction to DNNs.

Roughly speaking, a deep neural network (DNN for short) is a function built by multiple concatenations of affine transformations with a (typically nonlinear) activation function. This gives rise to a parametrised family of nonlinear maps; see for example Petersen and Voigtlaender [44] or Buehler et al. [14, Sect. 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 realisation, that is, the function defined by these parameters. More specifically, we use the following terminology (see for example Opschoor et al. [42, Sect. 2]). We first fix a function $\varrho \colon \mathbb{R} \to \mathbb{R}$ (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

$$\Phi = \big((A_1, b_1), \ldots, (A_L, b_L)\big),$$

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 $i = 1, \ldots, L$ and (A_i, b_i) are referred to as the *weights* of the ith layer of the NN. The associated *realisation* of Φ is the mapping

$$R(\Phi) \colon \mathbb{R}^d \to \mathbb{R}^{N_L}, \qquad x \mapsto R(\Phi)(x) = A_L x_{L-1} + b_L,$$

where x_{L-1} is given as

$$x_0 = x, \qquad x_\ell = \varrho(A_\ell x_{\ell-1} + b_\ell) \quad \text{for } \ell = 1, \ldots, L - 1.$$

We call $M_j(\Phi) = \|A_j\|_0 + \|b_j\|_0$, with $\|\cdot\|_0$ being the number of non-zero entries in $\cdot$, the number of (non-zero) weights in the jth 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 Opschoor et al. [42, Sect. 2] 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 parallelisation* operation for neural networks (see [42, Proposition 2.5]) and refines Grohs et al. [29, Lemma 3.8].

Lemma 3.2 *Let $d, L, n \in \mathbb{N}$ and let $\Phi^1, \ldots, \Phi^n$ be neural networks with L layers, d-dimensional input and equal output dimensions. Let $D_1, \ldots, D_n$ be $d \times d$-matrices, $c_1, \ldots, c_n \in \mathbb{R}^d$ and $w_1, \ldots, w_n \in \mathbb{R}$. Then there exists a neural network ψ such that*

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

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

Proof Write for $i = 1, \ldots, n$

$$\Phi^i = \left((A_1^i, b_1^i), \ldots, (A_L^i, b_L^i)\right)$$

and define the block matrices

$$A_1^{n+1} = \begin{pmatrix} A_1^1 D_1 \\ \vdots \\ A_1^n D_n \end{pmatrix}, \qquad b_1^{n+1} = \begin{pmatrix} A_1^1 c_1 + b_1^1 \\ \vdots \\ A_1^n c_n + b_1^n \end{pmatrix},$$

$$A_j^{n+1} = \begin{pmatrix} A_j^1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & A_j^n \end{pmatrix}, \qquad b_j^{n+1} = \begin{pmatrix} b_j^1 \\ \vdots \\ b_j^n \end{pmatrix} \qquad \text{for } j = 2, \ldots, L-1,$$

$$A_L^{n+1} = \begin{pmatrix} w_1 A_L^1 & \cdots & w_n A_L^n \end{pmatrix}, \qquad b_L^{n+1} = w_1 b_L^1 + \cdots + w_n b_L^n.$$

Set $\psi = ((A_1^{n+1}, b_1^{n+1}), \ldots, (A_L^{n+1}, b_L^{n+1}))$. Then for $\ell \in \{1, \ldots, L-1\}$ and $x \in \mathbb{R}^d$, it is straightforward to verify that x_ℓ has a block structure (with subscripts indicating the layers and superscripts indicating the blocks)

$$x_\ell = \begin{pmatrix} x_\ell^1 \\ \vdots \\ x_\ell^n \end{pmatrix},$$

with $x_1^i = \varrho(A_1^i (D_i x + c_i) + b_1^i)$, $x_\ell^i = \varrho(A_\ell^i x_{\ell-1}^i + b_\ell^i)$ for $\ell = 2, \ldots, L-1$ and

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

Hence (3.1) is satisfied and

$$M_j(\psi) = M_j(\Phi^1) + \cdots + M_j(\Phi^n) \qquad \text{for } j = 2, \ldots, L-1,$$

$$M_L(\psi) \leq M_L(\Phi^1) \mathbb{1}_{\{w_1 \neq 0\}} + \cdots + M_L(\Phi^n) \mathbb{1}_{\{w_n \neq 0\}}.$$

If in addition $D_1, \ldots, D_n$ are diagonal matrices and $c_1 = \cdots = c_n = 0$, then $\|A_1^i D_i\|_0 = \|A_1^i\|_0$ and therefore $M_1(\psi) = M_1(\Phi^1) + \cdots + M_1(\Phi^n)$. 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. $\qquad \square$

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 by Gonon et al. [27], Beck et al. [7] 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 (see Ledoux and Talagrand [39, Theorem 4.12]) and statistical learning. This new approach is not only technically less involved (in comparison to e.g. the techniques used in [27]), but also allows for weaker assumptions on the activation function; see Proposition 4.1 below. Alternatively, under stronger hypotheses on the activation function, one can also rely on [27, Lemma 2.16]; see Proposition 4.4 below. The probabilistic arguments result in, essentially, an expression error $\mathcal{O}(\varepsilon)$ with DNN sizes of $\mathcal{O}(\varepsilon^{-2})$. 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 expression error $\mathcal{O}(\varepsilon)$ with DNN sizes which are 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 realisation of Φ. Consider the following exponential integrability condition on the Lévy measure ν: for some $p \geq 2$,

$$\int_{\{|y|>1\}} e^{py} \nu(dy) < \infty. \tag{4.1}$$

Furthermore, for any function g, we denote by $\mathrm{Lip}(g)$ the best Lipschitz constant for g.

Proposition 4.1 *Suppose the moment condition* (4.1) *holds. Suppose further the payoff φ can be approximated by neural networks, i.e., given a payoff function $s \mapsto \varphi(s)$,*

there exist constants $c > 0$, $q \geq 0$ such that for any $\varepsilon \in (0, 1]$, there exists a neural network ϕ_ε with

$$|\varphi(s) - \mathrm{R}(\phi_\varepsilon)(s)| \leq \varepsilon c(1 + |s|), \qquad s \in (0, \infty), \qquad (4.2)$$

$$M(\phi_\varepsilon) \leq c\varepsilon^{-q}, \qquad (4.3)$$

$$\mathrm{Lip}\big(\mathrm{R}(\phi_\varepsilon)\big) \leq c. \qquad (4.4)$$

Then there exist $\kappa \in [c, \infty)$ (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 u from (2.3) is at most ε, that is,

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

Remark 4.2 In relevant examples such as e.g. plain vanilla European options, the payoff can be represented *exactly* as a neural network ϕ. Then one can choose $\phi_\varepsilon = \phi$ for all $\varepsilon \in (0, 1]$ and so (4.2)–(4.4) are satisfied with $q = 0$, $c = \max\{M(\phi), \mathrm{Lip}(\mathrm{R}(\phi))\}$. Examples include call options, straddles and butterfly payoff functions (when ϱ is the ReLU activation function given by $x \mapsto \max\{x, 0\}$).

Remark 4.3 In Proposition 4.1, the time horizon $T > 0$ is finite and fixed. As is evident from the proof, the constant κ depends on T.

Proof of Proposition 4.1 Let $\varepsilon \in (0, 1]$ be the given target accuracy and fix $\bar{\varepsilon} \in (0, 1]$ (to be specified later). Denote $\phi = \phi_{\bar{\varepsilon}}$. First, (4.2) and (4.4) show for any $s \in (0, \infty)$ that

$$|\varphi(s)| \leq |\varphi(s) - \mathrm{R}(\phi)(s)| + |\mathrm{R}(\phi)(s) - \mathrm{R}(\phi)(0)| + |\mathrm{R}(\phi)(0)|$$

$$\leq \bar{\varepsilon}c(1 + |s|) + c|s| + |\mathrm{R}(\phi)(0)|.$$

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

$$\mathbb{E}[e^{2X_T}] < \infty, \qquad (4.5)$$

due to the assumed integrability (4.1) of ν and Sato [48, Theorem 25.17].

Now recall that

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

Combining this with assumption (4.2) yields for all $s \in [a, b]$ that

$$|u(T, s) - \mathbb{E}[\mathrm{R}(\phi)(se^{X_T})]| \leq \mathbb{E}[|\varphi(se^{X_T}) - \mathrm{R}(\phi)(se^{X_T})|]$$

$$\leq \bar{\varepsilon}c(1 + |s|\mathbb{E}[e^{X_T}]) \leq \bar{\varepsilon}c_1 \qquad (4.6)$$

with the constant $c_1 = c(1 + b\mathbb{E}[e^{X_T}])$ being finite due to (4.5).

In the second step, let $X^1, \ldots, X^n$ denote n i.i.d. copies of X and introduce an independent collection of Rademacher random variables $\varepsilon_1, \ldots, \varepsilon_n$. This means that $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d., take the values ± 1 with probabilities $\frac{1}{2}$ each and are independent of all other random variables introduced before. Write $f(s) = \mathrm{R}(\phi)(s) - \mathrm{R}(\phi)(0)$ for $s \in (0, \infty)$. 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. The same reasoning guarantees that the suprema over $s \in [a, b]$ in (4.7) and (4.8) below are indeed random variables.

Using independence and a standard symmetrisation argument (see for example Boucheron et al. [12, Lemma 11.4]), we obtain

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

$$
\leq 2\mathbb{E}\left[\sup_{s \in [a,b]} \left| \frac{1}{n} \sum_{k=1}^n \varepsilon_k f(se^{X_T^k}) \right| \right]. \tag{4.7}
$$

Elementary properties of conditional expectations in the first step and Ledoux and Talagrand [39, Theorem 4.12], with T in that result chosen as

$$
T_{x_1, \ldots, x_n} = \{t \in \mathbb{R}^n : t_1 = se^{x_1}, \ldots, t_n = se^{x_n} \text{ for some } s \in [a, b]\},
$$

in the second step show that

$$
2\mathbb{E}\left[\sup_{s \in [a,b]} \left| \frac{1}{n} \sum_{k=1}^n \varepsilon_k f(se^{X_T^k}) \right| \right]
$$

$$
= \frac{2}{n} \mathbb{E}\left[\mathbb{E}\left[\sup_{t \in T_{x_1, \ldots, x_n}} \left| \sum_{k=1}^n \varepsilon_k f(t_k) \right| \right] \Big|_{x_1 = X_T^1, \ldots, x_n = X_T^n} \right]
$$

$$
\leq \frac{4}{n} \mathrm{Lip}(\mathrm{R}(\phi)) \mathbb{E}\left[\mathbb{E}\left[\sup_{t \in T_{x_1, \ldots, x_n}} \left| \sum_{k=1}^n \varepsilon_k t_k \right| \right] \Big|_{x_1 = X_T^1, \ldots, x_n = X_T^n} \right]
$$

$$
= \frac{4}{n} \mathrm{Lip}(\mathrm{R}(\phi)) \mathbb{E}\left[\sup_{s \in [a,b]} \left| \sum_{k=1}^n \varepsilon_k se^{X_T^k} \right| \right]
$$

$$
\leq \frac{4b}{n} \mathrm{Lip}(\mathrm{R}(\phi)) \mathbb{E}\left[\left| \sum_{k=1}^n \varepsilon_k e^{X_T^k} \right| \right]. \tag{4.8}
$$

On the other hand, using Jensen's inequality, independence and $\mathbb{E}[\varepsilon_k \varepsilon_\ell] = \delta_{k,\ell}$ yields

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

Combining this with (4.7), (4.8) and (4.4), we obtain that

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

with $c_2=4bc\mathbb{E}[e^{2X_T}]^{1/2}$, which is finite again due to (4.5).

In a third step, we now apply Markov's inequality and then insert (4.9) to estimate

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

$$\leq\frac{2\sqrt{n}}{3c_2}\mathbb{E}\left[\sup_{s\in[a,b]}\left|\mathbb{E}[R(\phi)(se^{X_T})]-\frac{1}{n}\sum_{k=1}^{n}R(\phi)(se^{X_T^k})\right|\right]\leq\frac{2}{3}.\tag{4.10}$$

This proves in particular that

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

Therefore there exists $\omega\in\Omega$ with

$$\sup_{s\in[a,b]}\left|\mathbb{E}[R(\phi)(se^{X_T})]-\frac{1}{n}\sum_{k=1}^{n}R(\phi)(se^{X_T^k(\omega)})\right|\leq\frac{2c_2}{\sqrt{n}}.\tag{4.11}$$

Lemma 3.2 proves that $s\mapsto\frac{1}{n}\sum_{k=1}^{n}R(\phi)(se^{X_T^k(\omega)})$ is itself the realisation 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

$$\sup_{s\in[a,b]}|\mathbb{E}[R(\phi)(se^{X_T})]-R(\tilde{\psi})(s)|\leq\frac{2c_2}{\sqrt{n}}.\tag{4.12}$$

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 $\kappa=c(1+4c_2^2)(c_1+1)^{2+q}$) that

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

and combining (4.12) with (4.6) to estimate

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

$$\leq\sup_{s\in[a,b]}|u(T,s)-\mathbb{E}[R(\phi)(se^{X_T})]|+\sup_{s\in[a,b]}|\mathbb{E}[R(\phi)(se^{X_T})]-R(\tilde{\psi})(s)|$$

$$\leq\bar{\varepsilon}(c_1+1)=\varepsilon.\qquad\qquad\square$$

Proposition 4.4 *Consider the setting of Proposition* 4.1, *but instead of* (4.4) *assume that* $\mathrm{R}(\phi_\varepsilon)$ *is* C^1 *and there is a constant* $c > 0$ *such that for every* $s \in (0, \infty)$, *we have*

$$|\mathrm{R}(\phi_\varepsilon)'(s)| \leq c. \tag{4.13}$$

Then the assertion of Proposition 4.1 *remains valid.*

Proof We present two different approaches to prove Proposition 4.4. The first combines (4.13) with the mean value theorem to obtain that (4.4) is satisfied. Hence, Proposition 4.1 can be applied. The second approach is based on a different technique to obtain (4.9). First, let us verify that (4.13) and (4.2) yield a linear growth condition for $\mathrm{R}(\phi_\varepsilon)$. Indeed, we may use the triangle inequality to estimate for any $\varepsilon \in (0, 1]$, $s \in (0, \infty)$ that

$$|\mathrm{R}(\phi_\varepsilon)(s)| \leq |\mathrm{R}(\phi_\varepsilon)(s) - \mathrm{R}(\phi_\varepsilon)(0)| + |\mathrm{R}(\phi_\varepsilon)(0) - \varphi(0)| + |\varphi(0)|$$

$$\leq \max\{c, |\varphi(0)|\}(1 + |s|). \tag{4.14}$$

Now the same proof as for Proposition 4.1 applies; only the second step needs to be adapted. In other words, we prove the estimate (4.9) with a different constant c_2 by using a different technique.

To do this, we again let $X^1, \ldots, X^n$ denote n i.i.d. copies of X. Applying Gonon et al. [27, Lemma 2.16] (with random fields $\xi_k(s, \omega) = \mathrm{R}(\phi)(se^{X_T^k(\omega)})$, $k = 1, \ldots, n$, which satisfy the hypotheses of [27, Lemma 2.16] thanks to (4.5) and (4.13)) in the first inequality and using (4.13) and (4.14) for the second inequality then proves that

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

$$\leq \frac{32\sqrt{e}}{\sqrt{n}} \sup_{s \in [a,b]} \left(\mathbb{E}[|\mathrm{R}(\phi)(se^{X_T})|^2]^{1/2} + (b-a)\mathbb{E}[|\mathrm{R}(\phi)'(se^{X_T})e^{X_T}|^2]^{1/2}\right)$$

$$\leq \frac{32\max\{c, |\varphi(0)|\}\sqrt{e}}{\sqrt{n}}\left(1 + b\mathbb{E}[e^{2X_T}]^{1/2} + (b-a)\mathbb{E}[e^{2X_T}]^{1/2}\right),$$

which is a bound as in (4.9) with constant

$$c_2 = 32\max\{c, |\varphi(0)|\}\sqrt{e}\left(1 + b\mathbb{E}[e^{2X_T}]^{1/2} + (b-a)\mathbb{E}[e^{2X_T}]^{1/2}\right). \qquad \square$$

Remark 4.5 The architecture of the neural network approximations constructed using probabilistic arguments in Proposition 4.1, Proposition 4.4 and also Theorem 5.1 below differ from architectures obtained by analytic arguments; see Proposition 4.8 and Theorem 5.4 below. 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. For example, by (4.6) or (2.3), the function φ is specified in the variable $s > 0$ and

not in the 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 the simple 2-layer neural network $\phi_0 = ((1, -K), (1, 0))$, that is, $R(\phi_0) = \varphi$.

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 $K_1, \ldots, K_N > 0$. Denote these prices by $\hat{C}(T, K_1), \ldots, \hat{C}(T, K_N)$. 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

$$\min_{\phi \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{C}(T, K_i)}{K_i} - \phi(S_0/K_i) \right)^2. \tag{4.15}$$

Here $\mathcal{H}$ is a suitable collection of (realisations of) neural networks, for example all networks with an a-priori fixed architecture. In fact, many of the papers listed in the recent review by Ruf and Wang [47] use this approach or a variation of it, where for example 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 (4.15) can be controlled and that we can give bounds on the number of non-zero parameters of the minimising 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 (4.15).

Proposition 4.6 *Assume that*

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

with $S_T = S_0 \exp(X_T)$ and X an (unknown) Lévy process satisfying (4.1). For any $\kappa > 0$, $\varepsilon \in (0, 1]$, we let $\mathcal{H}_{\kappa, \varepsilon}$ denote the set of all (realisations 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 $\kappa \in (0, \infty)$ 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 First, choose the interval $[a, b]$ by setting $a = \min\{S_0/K_1, \ldots, S_0/K_N\}$ and $b = \max\{S_0/K_1, \ldots, S_0/K_N\}$. 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 there exist $\kappa \in [3, \infty)$ 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

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

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

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{C}(T, K_i)}{K_i} - \mathrm{R}(\psi_\delta)(S_0/K_i) \right)^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(u(T, S_0/K_i) - \mathrm{R}(\psi_\delta)(S_0/K_i) \right)^2 \leq \delta^2.$$

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

Remark 4.7 The proof shows that κ is independent of N. This can also be seen by observing that the result directly generalises to an infinite number of call options with strikes in a compact interval $\mathcal{K} = [\underline{K}, \overline{K}]$ with $\underline{K} > 0$, $\overline{K} < \infty$. Indeed, let μ be a probability measure on $(\mathcal{K}, \mathcal{B}(\mathcal{K}))$. Then choosing ψ_δ, $\delta = \varepsilon^2$ as in the proof of Proposition 4.6 and $a = S_0/\overline{K}$, $b = S_0/\underline{K}$ yields $\mathrm{R}(\psi_\delta) \in \mathcal{H}_{\kappa, \varepsilon}$ and

$$\int_{\mathcal{K}} \left(\frac{\hat{C}(T, K)}{K} - \mathrm{R}(\psi_\delta)(S_0/K) \right)^2 \mu(dK)$$

$$= \int_{\mathcal{K}} \left(u(T, S_0/K) - \mathrm{R}(\psi_\delta)(S_0/K) \right)^2 \mu(dK) \leq \varepsilon.$$

4.3 ReLU DNN exponential expressivity

We now develop a second argument for bounding the expressivity of ReLU DNNs for the option price $u(\tau, s)$ solving (2.4) with 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 first, probabilistic argument, we consider the DNN expression error in a bounded interval $[a, b]$ with $0 < a < s < b < \infty$. The second argument is based on *parabolic smoothing* of the linear parabolic PIDE (2.4). This in turn ensures smoothness of $s \mapsto u(\tau, s)$ at positive times $\tau > 0$, i.e., smoothness in the "spatial" variable $s \in [a, b]$ resp. in the log-return variable $x = \log s \in [\log a, \log b]$, even for non-smooth 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 observed in Opschoor et al. [43, Sect. 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 (4.17) below) on the Lévy triplet (σ^2, γ, ν). To formulate

these, we recall the Lévy symbol ψ of the $\mathbb{R}$-valued LP X as

$$\psi(\xi) = \frac{\sigma^2}{2}\xi^2 - i\gamma\xi - \int_{\mathbb{R}}(e^{i\xi x} - 1 - i\xi x \mathbb{1}_{\{|x|\leq 1\}})\nu(dx), \qquad \xi \in \mathbb{R}. \quad (4.16)$$

Proposition 4.8 *Suppose that the symbol ψ of the LP X is such that there exist $\rho \in (0, 1]$ and constants $C_i > 0$, $i = 1, 2, 3$, such that for all $\xi \in \mathbb{R}$, we have*

$$\mathrm{Re}\psi(\xi) \geq C_1|\xi|^{2\rho}, \qquad |\psi(\xi)| \leq C_2|\xi|^{2\rho} + C_3. \quad (4.17)$$

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

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

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

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

Remark 4.9 A sufficient condition on the Lévy triplet which ensures (4.17) is as follows. 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 the following:

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

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

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

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

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

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

Then (4.17) is satisfied (see Hilber et al. [32, Lemma 10.4.2]). Here, $\rho = 1$ if $\sigma > 0$ and otherwise $\rho = \alpha/2$.

Proof of Proposition 4.8 The proof proceeds in several steps. First, we apply the change of variables $x = \log s \in \mathbb{R}$ in order to leverage the stationarity of the LP X for obtaining a constant coefficient Kolmogorov PIDE. Assumption (4.17) then ensures well-posedness of the PIDE in a suitable variational framework. We then exploit that stationarity of the LP X facilitates the use of Fourier transformation; the

lower bound on ψ in (4.17) will allow deriving sharp, explicit bounds on high spatial derivatives of (variational) solutions of the PIDE which imply Gevrey-regularity of these solutions on bounded intervals $[a, b] \subseteq (0, \infty)$. Gevrey-regularity in turn implies exponential rates of convergence of polynomial and deep ReLU NN approximations of $s \mapsto u(\tau, s)$ for $\tau > 0$, whence we obtain the assertion of the theorem. We recall that for $\delta \geq 1$, a smooth function $x \mapsto f(x)$ is *Gevrey-δ-regular* in an open subset $D \subseteq \mathbb{R}^d$ if $f \in C^\infty(D)$ and for every compact set $\kappa \subseteq D$, there exists $C_\kappa > 0$ such that for all $\alpha \in \mathbb{N}_0^d$ and every $x \in \kappa$, we have $|D_x^\alpha f(x)| \leq C_\kappa^{|\alpha|+1}(\alpha!)^\delta$. Note that $\delta = 1$ implies that f is real analytic in κ. We refer to Rodino [46, Sect. 1.4] for details, examples and further references.

We change coordinates to $x = \log s \in (-\infty, \infty)$ so that $v(\tau, x) = u(\tau, e^x)$. Then the PIDE (2.4) takes the form (see e.g. Matache et al. [40, Sect. 3], Lamberton and Mikou [38, Sect. 3.1])

$$\frac{\partial v}{\partial \tau} - \frac{\sigma^2}{2}\frac{\partial^2 v}{\partial x^2} - (\gamma + r)\frac{\partial v}{\partial x} + A[v] + rv = 0 \qquad \text{in } (0, T) \times \mathbb{R}, \qquad (4.18)$$

where A denotes the integro-differential operator

$$A[v](\tau, x) = -\int_{\mathbb{R}}\left(v(\tau, x+y) - v(\tau, x) - y\frac{\partial v}{\partial x}(\tau, x)\mathbb{1}_{\{|y| \leq 1\}}\right)\nu(dy),$$

together with the initial condition

$$v(0, x) = \varphi(e^x) = (\varphi \circ \exp)(x). \qquad (4.19)$$

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

$$C(t, S_t) = \mathbb{E}[e^{r(t-T)}\varphi(S_T)|\mathcal{F}_t]. \qquad (4.20)$$

Conversely, if $C(t, s)$ in (4.20) is sufficiently regular, then $v(\tau, x) = C(T - \tau, e^x)$ is a solution of (4.18), (4.19) (recall that we assume $r = 0$ for notational simplicity).

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

$$\mathbb{E}[e^{i\xi X_t}] = e^{-t\psi(\xi)}, \qquad \forall t > 0. \qquad (4.21)$$

The Lévy exponent ψ of the LP X admits the explicit representation (4.16).

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. Here $\mathcal{A} = -\mathcal{L}$ is the spatial operator in (4.18) given by

$$\mathcal{A}[f](x) = -\frac{\sigma^2}{2}\frac{d^2 f}{dx^2}(x) - \gamma\frac{df}{dx}(x) + A[f](x). \qquad (4.22)$$

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

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

The translation invariance of $\mathcal{A}$ (implied by stationarity of the LP X) in (4.22) and Parseval's equality (see Hilber et al. [32, Remark 10.4.1]) imply that ψ is the symbol of $\mathcal{A}$, i.e.,

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

Here $\hat{f}(\xi) = (2\pi)^{-\frac{1}{2}} \int_{\mathbb{R}} e^{-ix\xi} f(x)dx$, $\xi \in \mathbb{R}$, denotes the Fourier transform of f. The assumption (4.17) on ψ implies continuity and coercivity of the bilinear form $a(\cdot, \cdot)$ on $H^{\rho/2}(\mathbb{R}) \times H^{\rho/2}(\mathbb{R})$ so that for $v_0 \in L^2(\mathbb{R})$, there exists a unique variational solution $v \in C([0, T]; L^2(\mathbb{R})) \cap L^2(0, T; H^{\rho/2}(\mathbb{R}))$ of the PIDE (4.18) with initial condition (4.19); see e.g. Eberlein and Glau [20].

Fix $0 < \tau \leq T < \infty$ and $x \in \mathbb{R}$. The variational solution v of (4.18), (4.19) satisfies

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

For every $k \in \mathbb{N}_0$, Parseval's equality implies with the lower bound in (4.17) that

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

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

$$\max_{\eta > 0} \left(\eta^m \exp(-\kappa\eta^{\mu})\right) = \left(\frac{m}{\kappa\mu e}\right)^{m/\mu}. \tag{4.23}$$

We use (4.23) with $m = 2k$, $\kappa = 2\tau C_1$, $\mu = 2\rho$ and $\eta = |\xi|$ to obtain

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

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

$$\|(D_x^k v)(\tau, \cdot)\|_{L^2(\mathbb{R})} \leq \left(\left(\frac{1}{2\tau C_1 \rho}\right)^{\frac{1}{2\rho}}\right)^k (k!)^{\frac{1}{2\rho}} \|v_0\|_{L^2(\mathbb{R})}, \qquad \forall \tau > 0, \forall k \in \mathbb{N}. \tag{4.24}$$

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

$$\sup_{x \in I} |(D_x^k v)(\tau, x)| \leq C\big(A(\tau, \rho)\big)^k (k!)^{1/\min\{1, 2\rho\}}, \qquad \forall k \in \mathbb{N}.$$

This means that $v(\tau, \cdot)|_I$ 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 $I = [x_-, x_+] \subseteq \mathbb{R}$) change of variables $s = \exp(x)$ and the fact that Gevrey-regularity is preserved under analytic changes of variables to infer Gevrey-δ-regularity in $[a, b] \subseteq \mathbb{R}_{++}$ 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 $u(\tau, \cdot)$ for $\tau > 0$ 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 are then obtained by approximately expressing the u_p through ReLU DNNs, again at exponential rates, via Opschoor et al. [43]. The details are as follows.

The interval $s \in [a, b]$ in the assertion of the proposition corresponds to the interval $x \in [\log a, \log b]$ under the analytic (in the bounded interval $[a, b]$) change of variables $x = \log s$. As Gevrey-regularity is known to be preserved under analytic changes of variables (see e.g. Rodino [46, 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 $[a, b]$.

With Gevrey-δ-regularity of $s \mapsto u(\tau, s)$ for $s \in [a, b]$ established, we may invoke expression rate bounds for deep ReLU NNs for such functions. In Opschoor et al. [43, Proposition 4.1], it was shown that for such functions in space dimension $d = 1$, there exist constants $C' > 0$, $\beta' > 0$ such that for every $\mathcal{N} \in \mathbb{N}$, there exists a deep ReLU NN $\tilde{u}_\mathcal{N}$ with

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

$$\|u - \mathrm{R}(\tilde{u}_\mathcal{N})\|_{W^{1,\infty}([-1,1]^d)} \leq C' \exp\big(-\beta' \mathcal{N}^{\min\{\frac{1}{2\delta}, \frac{1}{d\delta+1}\}}\big).$$

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. □

4.4 Summary and discussion

For prices of derivative contracts on one risky asset whose log-returns are modelled by an LP X, we have analysed the expression rate of deep ReLU NNs. We have provided two mathematically distinct approaches to the analysis of the expressive power of deep ReLU NNs. The first, probabilistic approach furnishes algebraic expression rates, i.e., pointwise accuracy $\varepsilon > 0$ on a bounded interval $[a, b]$ is furnished with DNNs of size $\mathcal{O}(\varepsilon^{-q})$ with suitable $q \geq 0$. The argument is based on approximating the option price by Monte Carlo sampling, estimating the uniform error on $[a, b]$ and then emulating the resulting average by a DNN. The second, "analytic" approach, leverages regularity of (variational) solutions of the corresponding Kolmogorov partial integro-differential equations and furnishes exponential DNN expression rates.

That is, an expression error $\varepsilon > 0$ is achieved with DNNs of size $\mathcal{O}(|\log \varepsilon|^a)$ for suitable $a > 0$. Key in the second approach were stronger conditions (4.17) on the characteristic exponent of the LP X, which imply, as we showed, Gevrey-δ-regularity of the map $s \mapsto u(\tau, s)$ for suitable $\tau > 0$. This regularity implies in turn exponential rates of polynomial approximation (in the uniform norm on $[a, b]$) of $s \mapsto u(\tau, s)$, which is a result of independent interest, and subsequently, by emulation of polynomials with deep ReLU NNs, the corresponding exponential rates.

We remark that in the particular case $\delta = 1$, the derivative bounds (4.24) 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 Opschoor et al. [43].

We also remark that the smoothing of the solution operator in Proposition 4.8 accommodates payoff functions which belong merely to L^2, as they arise e.g. in particular binary contracts. This is a consequence of the assumption (4.17), which on the other hand excludes Lévy processes with one-sided jumps. Such processes 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, in particular with fully correlated jumps in the marginals, as provided for example by so-called Lévy copula constructions in Kallsen and Tankov [36].

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 Sect. 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 generalise geometric Brownian motions, Theorem 5.1 generalises several results from the classical Black–Scholes setting, and we comment on the relation of Theorem 5.1 to these recent results in Sect. 5.2. Owing to the method of proof, the DNN expression rate in Theorem 5.1 delivers an ε-complexity of $\mathcal{O}(\varepsilon^{-2})$, achieved with potentially shallow DNNs; see Remark 4.5.

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: the constants implied in the $\mathcal{O}(\cdot)$ bounds may (and in general will) depend exponentially on d. We also show that under a hypothesis of sufficiently large time $t > 0$, parabolic smoothing allows overcoming the curse of dimensionality, with dimension-independent expression rates which are possibly larger than the rate furnished by the probabilistic argument (which is, however, valid uniformly for all $t > 0$).

5.1 DNN expression rate bounds via probabilistic arguments

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 if 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 $\varphi_d \colon \mathbb{R}^d \to \mathbb{R}$ and a d-variate LP X^d, and we denote the option price in time-to-maturity by

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

We refer to Sato [48, Chap. 2] for more details on multivariate Lévy processes and to Cont and Tankov [16, Chap. 11], Eberlein and Kallsen [21, Sect. 8.1] 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 partially generalises earlier work on DNN expression rates for diffusion models in Elbrächter et al. [22], Grohs et al. [29].

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

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

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

$$\mathrm{Lip}\big(\mathrm{R}(\phi_{\varepsilon,d})\big) \leq c d^{\tilde{q}}. \tag{5.4}$$

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 $B > 0$ such that for each $d \in \mathbb{N}$, $i, j = 1, \ldots, d$,

$$\max\left\{ A_{ij}^d, \gamma_i^d, \int_{\{|y|>1\}} e^{py_i} \nu^d(dy), \int_{\{|y|\leq 1\}} y_i^2 \nu^d(dy) \right\} \leq B. \tag{5.5}$$

Then there exist constants $\kappa, \mathfrak{p}, \mathfrak{q} \in [0, \infty)$ 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, i.e., $M(\psi_{\varepsilon,d}) \leq \kappa d^{\mathfrak{p}} \varepsilon^{-\mathfrak{q}}$, and the approximation error between the neural network $\psi_{\varepsilon,d}$ and the option price is at most ε, that is,

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

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. Thus we may choose $q = 0$ in (5.3) and obtain $\mathsf{q} = 2$ in Theorem 5.1 (cf. (5.12) below).

Proof of Theorem 5.1 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 $s \in [a, b]^d$

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

The proof consists of 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 pth 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 were to assume in addition that the output of the neural network $\phi_{\bar{\varepsilon},d}$ were bounded (this is for example 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 (symmetrisation, 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 $s \in [a, b]^d$ that

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

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 Sato [48, Theorem 25.17] (with the vector $w \in \mathbb{R}^d$ in that result being pe_i) imply that for any $d \in \mathbb{N}$, $i = 1, \ldots, d$, the exponential moment can be bounded as

$$\mathbb{E}[e^{pX_{T,i}^d}] = \exp\left(T\left(\frac{p^2}{2}A_{ii}^d + \int_{\mathbb{R}^d}(e^{py_i} - 1 - py_i\mathbb{1}_{\{|y|\leq 1\}})\nu^d(dy) + p\gamma_i^d\right)\right)$$

$$\leq \exp\left(T\left(\frac{3p^2}{2}B + \int_{\{|y|\leq 1\}}(e^{py_i} - 1 - py_i)\nu^d(dy)\right.\right.$$

$$\left.\left. + \int_{\{|y|>1\}}(e^{py_i} - 1)\nu^d(dy)\right)\right)$$

$$\leq \exp\left(T\left(\frac{5p^2}{2}B + p^2e^p\int_{\{|y|\leq 1\}}y_i^2\nu^d(dy)\right)\right)$$

$$\leq \exp\left(T\left(\frac{5p^2}{2}B + p^2e^pB\right)\right), \tag{5.7}$$

where the second inequality uses 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 that

$$|\mathbb{E}[R(\phi_{\bar{\varepsilon},d})(se^{X_T^d})] - \mathbb{E}[R(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}})]| \leq cd^{\tilde{q}}\mathbb{E}[|se^{X_T^d} - se^{X_T^{d,D}}|]$$

$$\leq bcd^{\tilde{q}}\mathbb{E}\left[\sum_{i=1}^d |e^{X_{T,i}^d} - e^{X_{T,i}^{d,D}}|\right]$$

$$\leq bcd^{\tilde{q}}\sum_{i=1}^d \mathbb{E}[2e^{X_{T,i}^d}\mathbb{1}_{\{X_{T,i}^d > D\}}]$$

$$\leq 2bcd^{\tilde{q}}\sum_{i=1}^d \mathbb{E}[e^{2X_{T,i}^d}]^{1/2}\mathbb{P}[X_{T,i}^d > D]^{1/2}$$

$$\leq 2e^{-D}bcd^{\tilde{q}}\sum_{i=1}^d \mathbb{E}[e^{2X_{T,i}^d}]$$

$$\leq \tilde{c}_1 e^{-D}d^{\tilde{q}+1}, \tag{5.8}$$

where $\tilde{c}_1 = 2bc\exp(5TpB + 2Te^ppB)$ 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 $X_1, \ldots, X_n$ denote n i.i.d. copies of the random vector $X_T^{d,D}$ and $Z_1, \ldots, Z_n$ i.i.d. standard normal variables, independent of $X_1, \ldots, X_n$. For any separable class of functions $\mathcal{H} \subseteq C(\mathbb{R}^d; \mathbb{R})$, define the random variable (the so-called

empirical Gaussian complexity)

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

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

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

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

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

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

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

where $\phi = \mathrm{R}(\phi_{\bar{\varepsilon},d})(\cdot) - \mathrm{R}(\phi_{\bar{\varepsilon},d})(0)$ is a Lipschitz function with Lipschitz constant $cd^{\tilde{q}}$ (due to the hypothesis on the Lipschitz constant of the neural network (5.4)), satisfies $\phi(0) = 0$ and is bounded on the range of $\tilde{\mathcal{H}}$ (which is contained in $[0, b\exp(D)]^d$). Consequently, Bartlett and Mendelson [6, Theorem 14] implies that

$$\hat{G}_n(\mathcal{H}) \le 2cd^{\tilde{q}}\sum_{i=1}^{d}\hat{G}_n(\mathcal{H}_i). \tag{5.9}$$

Let $\varepsilon_1, \ldots, \varepsilon_n$ be an independent collection of Rademacher random variables. We then estimate

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

$$\le 2\mathbb{E}\left[\sup_{s \in [a,b]^d}\left|\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k\phi(se^{X_k})\right|\right]$$

$$\le \tilde{c}_2\mathbb{E}\left[\sup_{s \in [a,b]^d}\left|\frac{2}{n}\sum_{k=1}^{n}Z_k\phi(se^{X_k})\right|\right]$$

$$= \tilde{c}_2\mathbb{E}\left[\sup_{f \in \mathcal{H}}\left|\frac{2}{n}\sum_{k=1}^{n}Z_k f(X_k)\right|\right]$$

$$\le 2\tilde{c}_2cd^{\tilde{q}}\sum_{i=1}^{d}\mathbb{E}[\hat{G}_n(\mathcal{H}_i)]$$

$$\le \frac{4\tilde{c}_2cd^{\tilde{q}}b}{n}\sum_{i=1}^{d}\mathbb{E}\left[\left|\sum_{k=1}^{n}Z_k e^{X_{k,i}}\right|\right]. \tag{5.10}$$

Here, the first inequality follows by symmetrisation (see for example Boucheron et al. [12, Lemma 11.4]), the second follows from the comparison results on Gaussian and Rademacher complexities (see for instance Bartlett and Mendelson [6, Lemma 4]) with some absolute constant $\tilde{c}_2$ and the third uses (5.9).

We note that the constant $\tilde{c}_2$ in (5.10) may be chosen as $\tilde{c}_2 = 1/\mathbb{E}[|Z_1|] = \sqrt{\pi/2}$. Indeed, setting $\mathcal{G} = \sigma(\varepsilon_1, \ldots, \varepsilon_n, X_1, \ldots, X_n)$ and using independence yields

$$\mathbb{E}[|Z_1|]\mathbb{E}\left[\sup_{s\in[a,b]^d}\left|\frac{1}{n}\sum_{k=1}^{n}\varepsilon_k\phi(se^{X_k})\right|\right]$$

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

$$= \mathbb{E}\left[\sup_{s\in[a,b]^d}\left|\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n}|Z_k|\varepsilon_k\phi(se^{X_k})\Big|\mathcal{G}\right]\right|\right]$$

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

$$= \mathbb{E}\left[\sup_{s\in[a,b]^d}\left|\frac{1}{n}\sum_{k=1}^{n}Z_k\phi(se^{X_k})\right|\right].$$

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

$$\mathbb{E}\left[\left|\sum_{k=1}^{n}Z_ke^{X_{k,i}}\right|\right] \leq \mathbb{E}\left[\left|\sum_{k=1}^{n}Z_ke^{X_{k,i}}\right|^2\right]^{1/2} = \left(\sum_{k=1}^{n}\mathbb{E}[e^{2X_{k,i}}]\right)^{1/2}$$

$$\leq \sqrt{n}\mathbb{E}[e^{2X_{T,i}^d}]^{1/2} \leq \sqrt{n}\mathbb{E}[e^{pX_{T,i}^d}]^{1/p}.$$

Combining this with (5.10) and (5.7), we obtain that

$$\mathbb{E}\left[\sup_{s\in[a,b]^d}\left|\mathbb{E}\left[R(\phi_{\tilde{\varepsilon},d})(se^{X_T^{d,D}})\right] - \frac{1}{n}\sum_{k=1}^{n}R(\phi_{\tilde{\varepsilon},d})(se^{X_k})\right|\right] \leq \frac{c_2d^{\tilde{q}+1}}{\sqrt{n}}$$

with $c_2 = 4\sqrt{\pi/2}\,cb\exp(5BTp/2 + BTpe^p)$. By applying Markov's inequality (see (4.10) and (4.11)), this proves that there exists $\omega \in \Omega$ with

$$\sup_{s\in[a,b]^d}\left|\mathbb{E}\left[R(\phi_{\tilde{\varepsilon},d})(se^{X_T^{d,D}})\right] - \frac{1}{n}\sum_{k=1}^{n}R(\phi_{\tilde{\varepsilon},d})(se^{X_k(\omega)})\right| \leq \frac{2c_2d^{\tilde{q}+1}}{\sqrt{n}}.$$

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

that for arbitrary $n \in \mathbb{N}$, there exists a neural network $\tilde{\psi}_{\bar{\varepsilon},d}$ with

$$\sup_{s \in [a,b]^d} \left| \mathbb{E}\big[R(\phi_{\bar{\varepsilon},d})(se^{X_T^{d,D}}) \big] - R(\tilde{\psi}_{\bar{\varepsilon},d})(s) \right| \leq \frac{2c_2 d^{\tilde{q}+1}}{\sqrt{n}}. \tag{5.11}$$

Step 4: In the final step, we now provide appropriate choices of the hyper-parameters. We select $\bar{\varepsilon} = \varepsilon(c_1 d^{\tilde{q}+\frac{1}{2}p+\frac{1}{2}} + 2)^{-1}$, choose $n = \lceil (2c_2 d^{\tilde{q}+1}\bar{\varepsilon}^{-1})^2 \rceil$, $D = \log(\bar{\varepsilon}^{-1} d^{\tilde{q}+1}\tilde{c}_1)$ 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

$$
\begin{aligned}
M(\psi_{\varepsilon,d}) &= M(\tilde{\psi}_{\bar{\varepsilon},d}) \\
&\leq n M(\phi_{\bar{\varepsilon},d}) \\
&\leq \big(1 + (2c_2 d^{\tilde{q}+1}\bar{\varepsilon}^{-1})^2\big) c d^{\tilde{q}} \bar{\varepsilon}^{-q} \\
&\leq (1 + 4c_2^2) c d^{3\tilde{q}+2} \bar{\varepsilon}^{-2-q} \\
&\leq \big((1 + 4c_2^2) c(c_1 + 2)^{2+q}\big) d^{(\tilde{q}+\frac{1}{2}p+\frac{1}{2})(2+q)+3\tilde{q}+2} \varepsilon^{-2-q}. \tag{5.12}
\end{aligned}
$$

Thus the number of weights is bounded polynomially in d and ε^{-1}, as claimed. Finally, we combine (5.6), (5.8) and (5.11) to estimate the approximation error as

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

as claimed.   □

The proof of Theorem 5.1 is very similar to the proof of Proposition 4.1. Steps 1 and 4 in the proof of Theorem 5.1 are essentially identical in both proofs. The key difference is in Step 3: in the d-dimensional case we cannot use the comparison theorem for Rademacher complexities in Ledoux and Talagrand [39, Theorem 4.12], but instead need to use a comparison result for Gaussian complexities from Bartlett and Mendelson [6, Theorem 14]. In the d-dimensional case, the truncation with D in Step 2 is needed to guarantee that the hypotheses of [6, Theorem 14] are satisfied; in the 1-dimensional case, this is not required for [39, Theorem 4.12].

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 in Theorem 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 in Theorem 5.1 relates to these results.

Firstly, we note that with the exception of Gonon et al. [27] and Elbrächter et al. [22], previous results in the literature which are concerned with DNN approximation rates for Kolmogorov equations for diffusion processes (see e.g. Grohs et al. [30], Berner et al. [9], Grohs et al. [29], Reisinger and Zhang [45] and the references therein) study approximation with respect to the L^p-norm ($p < \infty$), whereas in Theorem 5.1 we study approximation with respect to the L^∞-norm, which requires entirely different techniques. While the results in [22] rely on a specific structure of the payoff, the proof of the expression rates in [27] has some similarities with the proof of Theorem 5.1. However, the novelty in the proof of Theorem 5.1 is the use of statistical learning techniques (symmetrisation, Gaussian and Rademacher complexities) which allow weaker assumptions on the activation function than in [27]. In addition, the class of PDEs considered in [27] (heat equation and related) is different from the one considered in Theorem 5.1 (Black–Scholes PDE and Lévy PIDE).

Secondly, Theorem 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

$$
\begin{aligned}
\partial_t v_d(\tau, x) = &\frac{1}{2}\text{Trace}\big(A^d D_x^2 v_d(\tau, x)\big) + D_x v_d(\tau, x)\gamma^d \\
&+ \int_{\mathbb{R}^d} \big(v_d(\tau, x+y) - v_d(\tau, x) - D_x v_d(\tau, x) y \mathbb{1}_{\{|y|\leq 1\}}\big)\nu^d(dy),
\end{aligned}
$$
$$
v_d(0, x) = (\varphi_d \circ \exp)(x) \tag{5.13}
$$

for $x \in \mathbb{R}^d$, $\tau > 0$, or, when transformed from log-price variables x_i to actual price variables s_i via $(s_1, \ldots, s_d) = (\exp(x_1), \ldots, \exp(x_d))$ (and with the convention $se^y = (s_1 e^{y_1}, \ldots, s_d e^{y_d})$),

$$
\begin{aligned}
\partial_t u_d(\tau, s) = &\frac{1}{2}\sum_{i,j=1}^d A_{i,j}^d s_i s_j \partial_{s_i} \partial_{s_j} u_d(\tau, s) + \sum_{i=1}^d s_i \tilde{\gamma}_i^d \partial_{s_i} u_d(\tau, s) \\
&+ \int_{\mathbb{R}^d} \bigg(u_d(\tau, se^y) - u_d(\tau, s) - \sum_{i=1}^d s_i(e^{y_i} - 1)\partial_{s_i} u_d(\tau, s)\bigg)\nu^d(dy),
\end{aligned}
$$
$$
u_d(0, s) = \varphi_d(s) \tag{5.14}
$$

for $s \in (0, \infty)^d$, $\tau > 0$ and with $\tilde{\gamma}_i^d = \gamma_i^d + \frac{A_{i,i}^d}{2} + \int_{\mathbb{R}^d}(e^{y_i} - 1 - y_i \mathbb{1}_{\{|y|\leq 1\}})\nu^d(dy)$ (see for instance Hilber et al. [31, 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 generalisations of the diffusion case, but cover indeed the case of arbitrary pure jump models for both finite and infinite activity Lévy processes satisfying (5.5).

In the case of X being a diffusion with drift, i.e., for $\nu^d = 0$, the Lévy PIDE reduces to a Black–Scholes PDE. In this particular case, we may compare the result in Theorem 5.1 to the recent results e.g. in Grohs et al. [29]. The results in the latter article are specialised to the Black–Scholes case in [29, Sect. 4], where Setting 4.1 specifies the coefficients $A_{i,j}^d$ (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 $d \in \mathbb{N}$, $i, j, k = 1, \ldots, d$ 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 and 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 [29, Sect. 4] can also be deduced from our Theorem 5.1 in the case when the probability measure used to quantify the L^p-error in [29] is compactly supported, as in that case the L^∞-bounds proved here imply the L^p-bounds proved in [29].

5.3 Exponential ReLU DNN expression rates via PIDEs

We now extend the univariate case discussed in Sect. 4.3, and prove an exponential expression rate bound similar to Proposition 4.8 for baskets of $d \geq 2$ Lévy-driven assets. **In this subsection**, we assume the ReLU activation function $\varrho(x) = \max\{x, 0\}$. As in Sect. 5.1, we admit a general correlation structure for the marginal processes' jumps. To prove DNN expression rate bounds, we exploit once more the fact that the stationarity and homogeneity of the $\mathbb{R}^d$-valued LP X^d imply that the Kolmogorov equation (5.13) has constant coefficients. Under the provision that $v_d(0, \cdot) \in L^2(\mathbb{R}^d)$ holds in (5.13), this allows writing for every $\tau > 0$ the Fourier transform $F_{x \to \xi} v_d(\tau, \cdot) = \hat{v}_d(\tau, \xi)$ as

$$\hat{v}_d(\tau, \xi) = \exp\left(-\tau \psi(\xi)\right)\hat{v}_d(0, \xi), \qquad \xi \in \mathbb{R}^d. \tag{5.15}$$

Here, for $\xi \in \mathbb{R}^d$, the symbol is given by $\psi(\xi) = \exp(-ix^\top \xi)\mathcal{A}(\partial_x)\exp(ix^\top \xi)$ with $\mathcal{A}(\partial_x)$ denoting the constant coefficient spatial integro-differential operator in (5.13) by Courrège's second theorem (see e.g. Applebaum [1, Theorem 3.5.5]), and (4.21) becomes

$$\mathbb{E}[\exp(i\xi^\top X_\tau^d)] = \exp\left(-\tau \psi(\xi)\right), \qquad \xi \in \mathbb{R}^d. \tag{5.16}$$

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

$$\psi(\xi) = \frac{1}{2}\xi^\top A^d \xi - i\xi^\top \gamma^d - \int_{\mathbb{R}^d} (e^{i\xi^\top y} - 1 - i\xi^\top y \mathbb{1}_{\{|y| \leq 1\}})\nu^d(dy), \qquad \xi \in \mathbb{R}^d. \tag{5.17}$$

We impose again the strong ellipticity assumption (4.17), but 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 4.8,

we obtain with $C_1 > 0$ as in (4.17) for every $\tau > 0$ for the variational solution v_d of (5.13) the bound

$$\|(D_x^k v_d)(\tau, \cdot)\|_{L^2(\mathbb{R}^d)}^2 \leq \left(\frac{k}{2\tau C_1 \rho e}\right)^{k/\rho} \|v_d(0, \cdot)\|_{L^2(\mathbb{R}^d)}^2, \qquad \forall k \in \mathbb{N}_0. \quad (5.18)$$

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 theorem, we again obtain for any bounded cube $I^d = [x_-, x_+]^d \subseteq \mathbb{R}^d$ 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

$$\sup_{x \in I^d} |(D_x^k v_d)(\tau, x)| \leq C(d)\big(A(\tau, \rho)\big)^k (k!)^{1/\min\{1, 2\rho\}}, \qquad \forall k \in \mathbb{N}. \quad (5.19)$$

In (5.19), the constant $C(d)$ is independent of x_-, x_+, but depends in general exponentially on the basket size (respectively the dimension) $d \geq 2$, and the constant $A(\tau, \rho) = (2\tau C_1 \rho)^{-1/(2\rho)}$ denotes the constant from (5.18) and Stirling's bound. If $\rho = 1$ (which corresponds to the case of a non-degenerate diffusion part) and if $\tau > 0$ is sufficiently large (so that $(2\tau C_1)^{1/(2\rho)} \geq 1$), the constant is bounded uniformly with respect to the dimension d. The derivative bound (5.19) implies that $v_d(\tau, \cdot)|_{I^d}$ 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 $\tau > 0$, the function $x \mapsto v_d(\tau, x)$ is real analytic in I^d, which is the case we consider first. In this case, we perform an affine change of coordinates to transform $v_d(\tau, \cdot)$ to the real analytic function $[-1, 1]^d \ni \hat{x} \mapsto v_c(\tau, \hat{x})$. This function admits a holomorphic extension to some open set $O \subseteq \mathbb{C}^d$ containing $[-1, 1]^d$. By choosing $\bar{\varrho} > 1$ (the "semiaxis sums") sufficiently close to 1, we obtain that $\mathcal{E}_{\bar{\varrho}} \subseteq O$, i.e., $v_c(\tau, \cdot)$ admits a holomorphic extension to $\mathcal{E}_{\bar{\varrho}}$, where the Bernstein polyellipse $\mathcal{E}_{\bar{\varrho}} \subseteq \mathbb{C}^d$ is defined as the d-fold Cartesian product of the Bernstein ellipse $\{(z + z^{-1})/2 : z \in \mathbb{C}, 1 \leq |z| < \bar{\varrho}\}$. More precisely, $x \mapsto v_d(\tau, x)$ 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 $\mathbb{C}$ (see e.g. Krantz and Parks [37, Sect. 1.2]). By Hartogs' theorem (see e.g. Hörmander [33, Theorem 2.2.8]), for every fixed $\tau > 0$, the function $x \mapsto v_d(\tau, x)$ admits a holomorphic extension to a polyellipse in $\mathbb{C}^d$ with foci at x_-, x_+ or, in normalised coordinates

$$\hat{x}_i = \big(T^{-1}(x)\big)_i = 2[x_i - (x_- + x_+)/2]/(x_+ - x_-), \qquad i = 1, \ldots, d, \quad (5.20)$$

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 $\mathcal{E}_{\bar{\varrho}} \subseteq \mathbb{C}^d$ with foci at $\hat{x}_i = \pm 1$ and semiaxis sums $1 < \bar{\varrho} = \mathcal{O}(A(\tau, \rho)^{-1})$. As $\tau \mapsto A(\tau, \rho)^{-1}$ is increasing for every fixed value of ρ, parabolic smoothing increases for $\rho \geq 1/2$ the domain of holomorphy with τ.

In the general case $\delta = 1/\min\{1, 2\rho\}$ with $\rho > 0$ as in (4.17), 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 Opschoor et al. [43]. The holomorphy or Gevrey-δ-regularity of the map $\hat{x} \mapsto v_c(\tau, \hat{x})$ implies with [43, Theorem 3.6, Proposition 4.1] that there exist constants $\beta' = \beta'(\bar{\varrho}, d) > 0$

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

$$M(\tilde{u}_{\mathcal{N}}) \leq \mathcal{N}, \qquad L(\tilde{u}_{\mathcal{N}}) \leq C\mathcal{N}^{\min\{\frac{1}{2}, \frac{1}{d+1/\delta}\}} \log \mathcal{N} \tag{5.21}$$

and such that the error bound

$$\|v_c(\tau, \cdot) - \tilde{u}_{\mathcal{N}}(\cdot)\|_{W^{1,\infty}([-1,1]^d)} \leq C \exp(-\beta' \mathcal{N}^{\min\{\frac{1}{2\delta}, \frac{1}{\delta d+1}\}}) \tag{5.22}$$

holds. Reversing the affine change of variables (5.20) 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 $\mathbb{R}^d$-valued LP X^d satisfies (4.17) with $|\xi|^2 = \xi^\top \xi$ and some $\rho \in (0, 1]$. For every φ_d with $v_d(0, \cdot) = \varphi_d \circ \exp \in L^2(\mathbb{R}^d)$, for every $\tau > 0$, on every closed, bounded hypercube $I^d = [x_-, x_+]^d \subseteq \mathbb{R}^d$ and, respectively, $J^d = [s_-, s_+]^d \subseteq (0, \infty)^d$ with $s_\pm = \exp(x_\pm)$, the variational solutions v_d of the Kolmogorov PIDE (5.13) at τ and $u_d(\tau, s) = v_d(\tau, \log s)$ can then 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 with $\delta = 1/\min\{1, 2\rho\} \geq 1$, for every $0 < \varepsilon \leq 1/2$, there exist ReLU DNNs $\tilde{v}_{d,\varepsilon}$, $\tilde{u}_{d,\varepsilon}$ for which we have*

$$\sup_{x \in I^d} |v_d(\tau, x) - \mathrm{R}(\tilde{v}_{d,\varepsilon})(x)|, \qquad \sup_{s \in J^d} |u_d(\tau, s) - \mathrm{R}(\tilde{u}_{d,\varepsilon})(s)| \leq \varepsilon$$

and

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

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

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

Proof The asserted bounds for $\mathrm{R}(\tilde{v}_{d,\varepsilon})$ follow by elementary manipulations from insisting that the expression error bound (5.22) must 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 (5.21) for the DNN size and depth.

The bounds for $\mathrm{R}(\tilde{u}_{d,\varepsilon})$ are then deduced from those for $\mathrm{R}(\tilde{v}_{d,\varepsilon})$ and the fact that the transformation $\log : J^d \to I^d$ (understood componentwise) is real analytic. Hence it admits a holomorphic extension to an open neighbourhood of J^d in $\mathbb{C}^d$. Then Opschoor et al. [43, Theorem 3.6], combined with the affine transformation $T : [-1, 1]^d \to J^d$, implies that there are constants $C, \beta' > 0$ such that for every $\mathcal{N} \in \mathbb{N}$, there exists a ReLU DNN $\widetilde{\log}_{\mathcal{N}}$ such that we have

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

and the error bound

$$\|\log \circ T - \mathrm{R}(\widetilde{\log}_{\mathcal{N}}) \circ T\|_{W^{1,\infty}([-1,1]^d)} \leq C \exp(-\beta' \mathcal{N}^{\frac{1}{d+1}}). \tag{5.23}$$

For every $\mathcal{N} \in \mathbb{N}$, the set $\widetilde{I^d} = \mathrm{R}(\widetilde{\log}_{\mathcal{N}})(J^d) \cup \log(J^d) \subseteq (-\infty, \infty)^d$ is compact due to (5.23). For given $\varepsilon \in (0, 1/2]$, we choose $\mathcal{N} \in \mathbb{N}$ as before. Using that $\mathcal{N}^{\min\{\frac{1}{2\delta}, \frac{1}{\delta d+1}\}} \le \mathcal{N}^{\frac{1}{d+1}}$, this choice guarantees that $C \exp(-\beta' \mathcal{N}^{\frac{1}{d+1}}) \le \varepsilon$ in (5.23). Then we define $\widetilde{u}_d(\tau, \cdot) = \mathrm{R}(\widetilde{v}_{d,\varepsilon})(\cdot) \circ \mathrm{R}(\widetilde{\log}_{\mathcal{N}})(\cdot)$ and estimate

$$
\begin{aligned}
\sup_{s \in J^d} |u_d(\tau, s) - \widetilde{u}_d(\tau, s)| &= \sup_{s \in J^d} |v_d(\tau, \cdot) \circ \log s - \widetilde{u}_d(\tau, s)| \\
&\le \sup_{s \in J^d} |v_d(\tau, \cdot) \circ \log s - v_d(\tau, \cdot) \circ \mathrm{R}(\widetilde{\log}_{\mathcal{N}})(s)| \\
&\quad + \sup_{s \in J^d} |v_d(\tau, \cdot) \circ \mathrm{R}(\widetilde{\log}_{\mathcal{N}})(s) \\
&\qquad\qquad - \mathrm{R}(\widetilde{v}_{d,\varepsilon})(\cdot) \circ \mathrm{R}(\widetilde{\log}_{\mathcal{N}})(s)| \\
&\le \|v_d(\tau, \cdot)\|_{W^{1,\infty}(\widetilde{I^d})} \sup_{s \in J^d} |\log s - \mathrm{R}(\widetilde{\log}_{\mathcal{N}})(s)| \\
&\quad + \sup_{x \in \widetilde{I^d}} |v_d(\tau, x) - \mathrm{R}(\widetilde{v}_{d,\varepsilon})(x)| \\
&\le C\varepsilon.
\end{aligned}
$$

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

Remark 5.5 Some sufficient conditions on the characteristic triplet (A^d, γ^d, ν^d) that ensure (4.17) in the multivariate setting are as follows. Consider first the case when the diffusion component is non-degenerate, i.e., A^d is positive definite. Then

$$
\mathrm{Re}\,\psi(\xi) = \frac{1}{2}\xi^\top A^d \xi - \int_{\mathbb{R}^d} \left(\cos(\xi^\top y) - 1 \right) \nu^d(dy) \ge C_1 |\xi|^2,
$$

$$
|\psi(\xi)| \le \frac{1}{2}|\xi^\top A^d \xi| + |\xi^\top \gamma^d| + \int_{\{|y| \le 1\}} |e^{i\xi^\top y} - 1 - i\xi^\top y| \nu^d(dy)
$$

$$
+ 2\int_{\{|y| > 1\}} \nu^d(dy)
$$

$$
\le C_2 |\xi|^2 + C_3
$$

for suitable choices of $C_1, C_2, C_3 > 0$. In the case when A^d is not positive definite, we refer for instance to Eberlein and Glau [20, Sect. 7] and Hilber et al. [32, Lemma 14.5.1] for sufficient conditions.

5.4 Breaking the curse of dimensionality

The result in Theorem 5.1 gives an ε expression error for DNNs whose depth and size are bounded polynomially in terms of $\varepsilon^{-1}d$, for European-style options in multivariate exponential Lévy models. In particular, in Theorem 5.1, the curse of dimensionality is proved to be overcome for a market model with jumps: a DNN expression

rate is shown which is algebraic in terms of the target accuracy $\varepsilon > 0$ with constants that depend polynomially on the dimension d. The rates $\mathfrak{p}, \mathfrak{q} \in [0, \infty)$ can be read off from the proof of Theorem 5.1; however, these constants could be large, thereby affording only low DNN expression rates.

Theorem 5.4, on the other hand, states *exponential expressivity* of deep ReLU NNs, i.e., a 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 is, however, still prone to the curse of dimensionality (CoD).

In the present section, we further address alternative mathematical arguments on how DNNs can overcome the CoD in the presently considered Lévy models. Specifically, two mathematical arguments in addition to the probabilistic arguments in Sect. 5.1 are presented. Both exploit stationarity of the LP X^d which implies (5.15), (5.16), to obtain DNN expression rates free from the curse of dimensionality.

5.4.1 Barron space analysis

The first alternative approach to Theorem 5.1 is based on verifying, using (5.15), (5.16), regularity of option prices in the so-called *Barron space* introduced in the fundamental work of Barron [5]. It provides DNN expression error bounds with explicit values for $\mathfrak{p}$ and $\mathfrak{q}$, however, *in [5] only for DNNs with sigmoidal activation functions* ϱ; similar results for ReLU activations are asserted in E and Wojtowytsch [19]. For simplicity, we consider here a subset $\mathcal{B}$ of the Barron space. An integrable function $f : \mathbb{R}^d \to \mathbb{R}$ belongs to $\mathcal{B}$ if

$$\|f\|_{\mathcal{B}} = \int_{\mathbb{R}^d} |\xi| |\hat{f}(\xi)| d\xi < \infty. \tag{5.24}$$

Recall that $\hat{f}$ denotes the Fourier transform of f. The explicit appearance of $\hat{f}$ renders the norm $\|\cdot\|_{\mathcal{B}}$ in (5.24) particularly suitable for our purposes due to (5.15)–(5.17). As was pointed out in [5, 19], the relevance of the Barron norm $\|\cdot\|_{\mathcal{B}}$ stems from it being sufficient for dimension-robust DNN approximation rates. For $m \in \mathbb{N}$, consider the two-layer neural networks f_m given by

$$f_m : \mathbb{R}^d \to \mathbb{R}, \qquad x \mapsto \frac{1}{m} \sum_{i=1}^{m} a_i \varrho(w_i^\top x + b_i) \tag{5.25}$$

with parameters $(a_i, w_i, b_i) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$. Their relevance stems from the following result: If ϱ is sigmoidal, i.e., bounded, measurable and $\varrho(z) \to 1$ as $z \to \infty$, $\varrho(z) \to 0$ as $z \to -\infty$, then for $f \in \mathcal{B}$ and for every $R > 0$, every probability measure π on $[-R, R]^d$ and every $m \in \mathbb{N}$, there exist parameters $\{(a_i, w_i, b_i)\}_{i=1}^{m}$ such that for the corresponding DNN f_m in (5.25), we have

$$\|f - f_m\|_{L^2([-R,R]^d, \pi)} \leq \max\{1, R\} m^{-1/2} \|f\|_{\mathcal{B}}. \tag{5.26}$$

The bound (5.26) follows from Barron [5, Theorem 1], and was generalised in E and Wojtowytsch [19, Eq. (1.3)] to ReLU activation.

The bound in (5.26) is free from the CoD: the number N of parameters in the DNN grows as $\mathcal{O}(md)$ so that $m^{-1/2} \le C N^{-1/2} d^{1/2}$ with an absolute constant $C > 0$.

With (5.15), (5.16), for every $\tau \ge 0$, sufficient conditions for $x \mapsto v_d(\tau, x)$ to belong to $\mathcal{B}$ can be verified. With (5.26), DNN mean-square expression rate bounds of option prices that are free from the CoD follow.

Proposition 5.6 *Assume that $\varrho(z) = \mathrm{ReLU}(z) = \max\{z, 0\}$ for $z \in \mathbb{R}$. Assume furthermore that the payoff $v_d(0, \cdot)$ in log-variables belongs to $\mathcal{B}$. Then for every $\tau \ge 0$, $R > 0$ and probability measure π on $[-R, R]^d$, the price $x \mapsto v_d(\tau, x)$ can be expressed by a NN $x \mapsto \tilde{v}_d(\tau, x)$ of depth 2 and size $m(d + 2)$ with $m \in \mathbb{N}$ and error bound*

$$\|v_d(\tau, \cdot) - \tilde{v}_d(\tau, \cdot)\|_{L^2([-R,R]^d, \pi)} \le \max\{1, R\} m^{-1/2} \|v_d(0, \cdot)\|_{\mathcal{B}}.$$

Proof We observe that for every $\xi \in \mathbb{R}^d$, the identity (5.16) with $\tau = 1$ shows that

$$\exp\left(-\mathrm{Re}\,\psi_{X^d}(\xi)\right) = \left|\exp\left(-\psi_{X^d}(\xi)\right)\right| = |\mathbb{E}[\exp(i\xi^\top X_1^d)]| \le 1$$

and therefore $\mathrm{Re}\,\psi_{X^d}(\xi) \ge 0$. From (5.15), (5.16), we obtain for $\tau \ge 0$ that

$$|\hat{v}_d(\tau, \xi)| = \left|\exp\left(-\tau \psi_{X^d}(\xi)\right)\hat{v}_d(0, \xi)\right| = \exp\left(-\tau \mathrm{Re}\,\psi_{X^d}(\xi)\right)|\hat{v}_d(0, \xi)|, \quad \forall \xi \in \mathbb{R}^d.$$

The payoff $v_d(0, \cdot)$ in log-price belonging to $\mathcal{B}$ implies $\|v_d(0, \cdot)\|_{\mathcal{B}} < \infty$. Using $\mathrm{Re}\,\psi_{X^d}(\xi) \ge 0$, we find for every $\tau \ge 0$, $\xi \in \mathbb{R}^d$ that $|\hat{v}_d(\tau, \xi)| \le |\hat{v}_d(0, \xi)|$. This implies for every $\tau \ge 0$ that $\|v_d(\tau, \cdot)\|_{\mathcal{B}} \le \|v_d(0, \cdot)\|_{\mathcal{B}}$. The approximation bound (5.26) implies the assertion. $\qquad\square$

Pointwise, L^∞-norm error bounds can be obtained by using [19, Eq. (1.4)].

5.4.2 Parabolic smoothing and sparsity of chaos expansions

The second non-probabilistic approach to Theorem 5.1 towards DNN expression error rates not subject to the CoD is based on dimension-explicit derivative bounds of option prices, which allow in turn establishing summability bounds for *generalised polynomial chaos* (gpc for short) expansions of these prices. Good summability of gpc coefficient sequences is well known to imply high, dimension-independent rates of approximation by sparse, multivariate polynomials. This in turn implies corresponding expression rates by suitable DNNs; see Schwab and Zech [49, Theorem 3.9]. Key in this approach is to exploit *parabolic smoothing* of the Kolmogorov PDE. The corresponding dimension-independent expression rate results are in general higher than those based on probabilistic or Barron space analysis, but hold only *for sufficiently large $\tau > 0$*.

We start by discussing more precisely the dependence of the constants in the proof of Theorem 5.4 on the dimension d.

Remark 5.7 The constant $C(d) = C(\tau, d)$ in the derivative bound (5.19) need not be exponential in d. To see it, we bound (5.19) by the inverse Fourier transform and the

Cauchy–Schwarz inequality. For $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ with $|\boldsymbol{\alpha}| = \sum_{i=1}^d \alpha_i = k$, we find with the Cauchy–Schwarz inequality and the lower bound (4.17) that

$$
\sup_{x \in I^d} |(D_x^{\boldsymbol{\alpha}} v_d)(\tau, x)| = \sup_{x \in I^d} \left| \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} (i\xi)^{\boldsymbol{\alpha}} \exp(ix^\top \xi) \exp\big(-\tau \psi(\xi)\big) \hat{v}_d(0, \xi) d\xi \right|
$$

$$
\leq \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} |\xi|^k \big| \exp\big(-\tau \psi(\xi)\big) \big| |\hat{v}_d(0, \xi)| d\xi
$$

$$
\leq \frac{1}{(2\pi)^{d/2}} \left(\int_{\mathbb{R}^d} \exp(-2\tau C_1 |\xi|^{2\rho}) d\xi \right)^{1/2}
$$

$$
\times \left(\int_{\mathbb{R}^d} |\xi|^{2k} \exp(-2\tau C_1 |\xi|^{2\rho}) |\hat{v}_d(0, \xi)|^2 d\xi \right)^{1/2}.
$$

The last factor can be bounded precisely by the square root of the right-hand side of (5.18) (by using (4.23)). Using $k^k \leq k! \, e^k$, we obtain the bound (5.19) as

$$
\sup_{x \in I^d} |(D_x^{\boldsymbol{\alpha}} v_d)(\tau, x)| \leq C(d, \tau) \big(A(\tau, \rho) \big)^k (k!)^{1/\min\{1, 2\rho\}} \|v_d(0, \cdot)\|_{L^2(\mathbb{R}^d)} \quad (5.27)
$$

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

$$
C(d, \tau) = \frac{1}{(2\pi)^{d/2}} \left(\int_{\mathbb{R}^d} \exp(-2\tau C_1 |\xi|^{2\rho}) d\xi \right)^{1/2}
$$

$$
= \frac{1}{(2\pi)^{d/2}} \left(2 \frac{\pi}{d} \omega_d \int_0^\infty r^{d-1} \exp(-2\tau C_1 r^{2\rho}) dr \right)^{1/2}
$$

$$
= \frac{1}{(2\pi)^{d/2}} \left(\frac{\pi}{\rho d} \frac{1}{(2\tau C_1)^{d/(2\rho)}} \omega_d \Gamma\left(\frac{d}{2\rho}\right) \right)^{1/2}
$$

$$
= \frac{1}{(2\pi)^{d/2}} \left(\frac{\pi}{\rho d} \frac{1}{(2\tau C_1)^{d/(2\rho)}} \frac{\pi^{d/2} \Gamma(\frac{d}{2\rho})}{\Gamma(\frac{d}{2}+1)} \right)^{1/2}, \quad (5.28)
$$

where ω_d denotes the volume of the unit ball in $\mathbb{R}^d$. Inspecting the constant $C(d, \tau)$ in (5.28), 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 $C(d, \tau)$ is bounded independently of τ and d.

Remark 5.8 In certain cases, the parabolic smoothing implied by the ellipticity assumption (4.17) on the generator $\mathcal{A}$ entails that the constant C in the regularity estimates (5.19) grows only polynomially with respect to d. For instance, in Remark 5.7, we have provided sufficient conditions which ensure that the constant C in the regularity estimates (5.19) is even bounded with respect to d. This allows deriving an explicit and dimension-independent bound on the series of Taylor coefficients. This in turn allows obtaining bounds on the constant in (5.22) 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, $(2\tau C_1)^{1/2} \geq 1$ and $d A(\tau, 1) < 1$,

where $A(\tau, 1) = (2\tau C_1)^{-1/2}$ denotes the constant in large parentheses in (4.24). This holds if

$$\tau > \frac{d^2}{2C_1}. \tag{5.29}$$

With (5.29) and using $\sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^d, |\boldsymbol{\alpha}|=k} \binom{k}{\boldsymbol{\alpha}} = d^k$, we may estimate with the multinomial theorem that

$$\sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^d} \frac{\sup_{x\in I^d} |(D_x^{\boldsymbol{\alpha}} v_d)(\tau, x)|}{\boldsymbol{\alpha}!} \le C(d, \tau) \sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^d} \frac{A(\tau, 1)^{|\boldsymbol{\alpha}|} (|\boldsymbol{\alpha}|)!}{\boldsymbol{\alpha}!}$$

$$= C(d, \tau) \sum_{k=0}^{\infty} \big(d A(\tau, 1)\big)^k = C(d, \tau) \frac{1}{1 - d A(\tau, 1)}.$$

By Remark 5.7, (5.29) implies that $C(d, \tau)$ in (5.28) is bounded uniformly with respect to d. Thus in this case, one may obtain bounds on the constant in (5.22) which scale polynomially with respect to d. However, the DNN size still grows polylogarithmically with respect to the dimension d, in terms of $|\log \varepsilon|$ (i.e., at least as $\mathcal{O}(|\log \varepsilon|^d)$), so that the curse of dimensionality is not overcome.

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 (5.13), due to the reliance on the ReLU DNN expression rate analysis in Opschoor et al. [43]. Furthermore, the DNN size grows polylogarithmically with respect to the dimension d, in terms of $|\log \varepsilon|$. Considering exponential expression rate bounds, this exponential dependence on d in terms of $|\log \varepsilon|$ seems in general not avoidable, as can be seen from [43, Theorem 3.5]. Nevertheless, in Remark 5.8, we already hinted at parabolic smoothing implying sufficient regularity (under the d-dependent provision (5.29) on τ) for constants in DNN expression rate bounds which are polynomial with respect to d.

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 Schwab and Zech [49], Cohen et al. [15] and the references there. We develop a variation of the results in [49] in the present context.

We impose the following hypothesis, which takes the place of the lower bound in (4.17). We still impose $|\psi(\xi)| \le C_2 |\xi|^{2\rho} + C_3$, i.e., the second condition in (4.17) holds for each $d \in \mathbb{N}$ (but C_2, C_3 and ρ in that condition are allowed to depend on d).

Assumption 5.9 There exist a constant $C_1 > 0$ and $(\rho_j)_{j\in\mathbb{N}}$ with $\frac{1}{2} < \rho_j \le 1$ such that for each $d \in \mathbb{N}$, the symbol ψ_{X^d} of the LP X^d satisfies that

$$\operatorname{Re}\psi_{X^d}(\xi) \ge C_1 \sum_{j=1}^{d} |\xi_j|^{2\rho_j}, \qquad \forall \xi \in \mathbb{R}^d. \tag{5.30}$$

Furthermore,

$$\rho = \inf_{j \in \mathbb{N}} \rho_j > \frac{1}{2}.$$

The payoff function φ_d in (5.13) is such that $v_d(0, \cdot) = \varphi_d \circ \exp \in L^2(\mathbb{R}^d)$.

In comparison to the lower bound in (4.17), the condition (5.30) is restricted to the case $\rho > \frac{1}{2}$. On the other hand, different exponents ρ_j are allowed along each component. Furthermore, note that Assumption 5.9 imposes that C_1 does not depend on the dimension d.

Remark 5.10 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 $\psi_{X^d} : \mathbb{R}^d \to \mathbb{R}, \xi \mapsto \xi^\top A^d \xi$. A sufficient condition for assumption (5.30) to hold is that the eigenvalues $(\lambda_i^d)_{i=1,\ldots,d}$ of A^d should be lower bounded away from zero, i.e.,

$$C_1 = \inf_{i,d} \lambda_i^d > 0. \tag{5.31}$$

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 $\xi \in \mathbb{R}^d$ that

$$\psi_{X^d}(\xi) = \xi^\top A^d \xi = (\xi^\top) Q D Q^\top \xi = \sum_{i=1}^{d} \lambda_i (Q^\top \xi)_i^2$$

$$\geq \left(\min_i \lambda_i \right) |Q^\top \xi|^2 = \left(\min_i \lambda_i \right) |\xi|^2.$$

Therefore (5.30) is satisfied with C_1 as in (5.31) 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 5.11 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 (5.31) from Remark 5.10. Then the real part of the Lévy symbol ψ_{X^d} of X^d satisfies for all $\xi \in \mathbb{R}^d$ that

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

with C_1 as in (5.31). Hence Assumption 5.9 is satisfied also in this more general situation. Further examples of LPs satisfying Assumption 5.9 are based on stable-like processes and copula-based constructions as e.g. in Farkas et al. [24].

As we shall see below, Assumption 5.9 ensures good "separation" and "anisotropy" properties of the symbol (5.17) of the corresponding Lévy process X^d.

For $\tau > 0$ satisfying (5.29), we now analyse the regularity of $x \mapsto v_d(\tau, x)$. From Assumption 5.9, we find that for every $\tau > 0$, $x \mapsto v_d(\tau, x)$ is in $L^2(\mathbb{R}^d)$ and that its Fourier transform has the explicit form

$$\hat{v}_d(\tau, \xi) = F_{x \to \xi} v_d(\tau, \cdot) = \exp\left(-\tau \psi_{X^d}(\xi)\right) \hat{v}_d(0, \xi). \tag{5.32}$$

For a multi-index $\boldsymbol{v} = (v_1, \ldots, v_d) \in \mathbb{N}_0^d$, denote by $\partial_x^{\boldsymbol{v}}$ the mixed partial derivative of total order $|\boldsymbol{v}| = v_1 + \cdots + v_d$ with respect to $x \in \mathbb{R}^d$. Formula (5.32) and Assumption 5.9 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 5.9, we calculate for any $\boldsymbol{v} \in \mathbb{N}_0^d$ at $x = 0$ (by stationarity, the same bounds hold for the Taylor coefficients at any $x \in \mathbb{R}^d$) that

$$(2\pi)^{d/2} |\partial_x^{\boldsymbol{v}} v_d(\tau, 0)| = \left| \int_{\xi \in \mathbb{R}^d} (i\xi)^{\boldsymbol{v}} \hat{v}_d(\tau, \xi) d\xi \right|$$

$$\leq \int_{\xi \in \mathbb{R}^d} |\hat{v}_d(0, \xi)| \prod_{j=1}^d |\xi_j|^{v_j} \exp(-\tau C_1 |\xi_j|^{2\rho_j}) d\xi.$$

We use (4.23) with $m := v_j$, $\kappa := C_1 \tau$, $\mu = 2\rho_j$ to bound the product as

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

For the Taylor coefficient of order $\boldsymbol{v} \in \mathbb{N}_0^d$ of $v_d(t, \cdot)$ at $x = 0$, we arrive at the bound

$$|t_{\boldsymbol{v}}| = \left| \frac{1}{\boldsymbol{v}!} \partial_x^{\boldsymbol{v}} v_d(\tau, x)|_{x=0} \right|$$

$$\leq \frac{1}{(2\pi)^{d/2}} \|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)} \prod_{j=1}^d \frac{1}{v_j!} \left(\frac{v_j}{2\rho_j \tau C_1 e} \right)^{v_j/(2\rho_j)}. \tag{5.33}$$

Stirling's inequality

$$n! \geq n^n e^{-n} \sqrt{2\pi n} \geq n^n e^{-n}, \qquad \forall n \in \mathbb{N},$$

implies in (5.33) the bound

$$|t_{\boldsymbol{v}}| \leq \frac{1}{(2\pi)^{d/2}} \|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)} \left((\boldsymbol{v}!)^{-1} b^{\boldsymbol{v}} \right)^{\rho'}, \qquad \forall \boldsymbol{v} \in \mathbb{N}_0^d. \tag{5.34}$$

Here, $\rho' = 1 - \frac{1}{2\rho} > 0$ and the positive weight sequence $b = (b_j)_{j \geq 1}$ is given by $b_j = (2\rho_j \tau C_1)^{-1/(2\rho_j \rho')}$, $j = 1, 2, \ldots$, and multi-index notation is employed: we write $\boldsymbol{v}^{-\boldsymbol{v}} = (v_1^{v_1} v_2^{v_2} \cdots)^{-1}$, $b^{\boldsymbol{v}} = b_1^{v_1} b_2^{v_2} \cdots$ and $\boldsymbol{v}! = v_1! v_2! \ldots$, with the convention

$0! = 1$ and $0^0 = 1$. We raise (5.34) to a power $q > 0$ with $q < 1/\rho'$ and sum the resulting inequality over all $\boldsymbol{v} \in \mathbb{N}_0^d$ to estimate (generously)

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

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

To obtain the estimate (5.34), one could also use the L^2-bound with the explicit constant derived in (5.27), (5.28).

Under hypothesis (5.30) and for $\tau > 0$ satisfying (5.29), q-summability of the Taylor coefficients follows; indeed,

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

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

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

Using that $|\boldsymbol{v}|! \geq \boldsymbol{v}!$ and that $1 \geq q\rho' > 0$, the multinomial theorem yields

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

Hence, provided that

$$\tau > \tau_0(d) := \frac{d^{2\rho/q}}{2\rho C_1}, \tag{5.35}$$

it follows that

$$\|(t_{\boldsymbol{v}})\|_{\ell^q(\mathbb{N}_0^d)}^q = \sum_{\boldsymbol{v} \in \mathbb{N}_0^d} |t_{\boldsymbol{v}}|^q \leq \frac{\|\hat{v}_d(0, \cdot)\|_{L^1(\mathbb{R}^d)}^q}{(2\pi)^{dq/2}} \frac{1}{1 - d(2\rho\tau C_1)^{-q/(2\rho)}}. \tag{5.36}$$

Therefore, we have proved q-summability of the Taylor coefficients of the map $x \mapsto v_d(\tau, x)$ at $x = 0$ for any $\tau > \tau_0(d)$ as in (5.35). The q-norm $\|(t_{\boldsymbol{v}})\|_{\ell^q(\mathbb{N}_0^d)}$ 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 (5.36) of the Taylor coefficients of $x \mapsto v_d(\tau, x)$ at $x = 0$ with $q = 1$ implies for $\tau > \tau_0(d)$ absolute, pointwise convergence in the cube $[-1, 1]^d$ of

$$v_d(\tau, x) = \sum_{\boldsymbol{v} \in \mathbb{N}_0^d} t_{\boldsymbol{v}} x^{\boldsymbol{v}}, \qquad x^{\boldsymbol{v}} = x_1^{v_1} x_2^{v_2} \cdots. \qquad (5.37)$$

Furthermore, as was shown in Schwab and Zech [49, Lemma 2.8], the fact that the sequence $(t_{\boldsymbol{v}})$ is q-summable for some $0 < q < 1$ and the coefficient bound (5.34) imply that for $\tau > \tau_0(d)$ with $\tau_0(d)$ as defined in (5.35), there exists a sequence $(\Lambda_n)_{n \geq 1} \subseteq \mathbb{N}_0^d$ of *nested, downward closed* (i.e., if $\boldsymbol{e}_j \in \Lambda_n$, then $\boldsymbol{e}_i \in \Lambda_n$ for all $0 \leq i \leq j$) *multi-index sets* $\Lambda_n \subseteq \mathbb{N}_0^d$ with $\#(\Lambda_n) \leq n$ such that general polynomial chaos (gpc) approximations given by the partial sums

$$v_d^{\Lambda_n}(\tau, x) = \sum_{\boldsymbol{v} \in \Lambda_n} t_{\boldsymbol{v}} x^{\boldsymbol{v}}$$

converge at the dimension-independent rate $r = 1/q - 1$ (see e.g. Cohen et al. [15, Lemma 5.5]), i.e.,

$$\sup_{x \in [-1,1]^d} |v_d(\tau, x) - v_d^{\Lambda_n}(\tau, x)| \leq \sum_{\boldsymbol{v} \in \mathbb{N}_0^d \setminus \Lambda_n} |t_{\boldsymbol{v}}| \leq n^{-(1/q-1)} \|(t_{\boldsymbol{v}})\|_{\ell^q(\mathbb{N}_0^d)}.$$

The summability (5.36) of the coefficients in the Taylor gpc expansion (5.37) also implies quantitative bounds on the expression rates of ReLU DNNs. With [49, Theorem 2.7, (ii)], we find that there exists a constant $C > 0$ independent of d such that

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

We now refer to [49, Theorem 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 $(t_{\boldsymbol{v}})$ was used, we conclude that for $\tau > 0$ satisfying (5.35), there exists a constant $C > 0$ that is independent of d and for every $n \in \mathbb{N}$, there exists a ReLU DNN $\tilde{v}_d^n$ with input dimension d such that

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

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

$$\sup_{x \in [-1,1]^d} |v_d(\tau, x) - \mathrm{R}(\tilde{v}_d^n)(x)| \leq C n^{-(1/q-1)}. \qquad (5.38)$$

6 Conclusion and generalisations

We have proved that prices of European-style derivative contracts on baskets of $d \geq 1$ assets 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 is based on probabilistic arguments

and provides expression rate bounds that scale algebraically in terms of the DNN size. We have then also provided an alternative, analytic argument that allows proving *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 the DNN size growing polynomially with respect to $\log \varepsilon$ to achieve accuracy $\varepsilon > 0$. For sufficiently large $t > 0$, based on analytic arguments involving parabolic smoothing and sparsity of generalised polynomial chaos expansions, we have established in (5.38) a second, algebraic expression rate bound for ReLU DNNs that is free from the curse of dimensionality. In forthcoming work (Gonon and Schwab [28]), we address PIDEs (5.13) with non-constant coefficients. In addition, the main result of the present paper, Theorem 5.1, could be extended in the following 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 $\nu^d = 0$, we have established in Elbrächter et al. [22] for particular payoff functions a spectral expression rate in terms of the DNN size, free from the curse of dimensionality.

Next, solving Hamilton–Jacobi partial integro-differential equations (HJPIDEs for short) by DNNs: It is classical that the Kolmogorov equation for the exponential LP X^d in Sect. 2.2 is in fact a special case of an HJPIDE (see e.g. Barles et al. [2], Barles and Imbert [3]). In forthcoming work [28], we aim at proving that the expression rate bounds obtained in Sect. 5 imply corresponding expression rate bounds for ReLU DNNs which are free from the curse of dimensionality for viscosity solutions of general HJPIDEs associated to the LP X^d and for its exponential counterparts.

Barriers: We have considered payoff functions corresponding to European-style contracts. Here, the stationarity of the LP X^d and exponential Lévy modelling have allowed to reduce our analysis to Cauchy problems of the Kolmogorov equations of X^d in $\mathbb{R}^d$. In the presence of barriers, option prices in Lévy models in general exhibit singularities at the barriers. More involved versions of the Fourier-transform-based representations are available (involving a so-called Wiener–Hopf factorisation of the Fourier symbol; see e.g. Boyarchenko and Levendorskiĭ [13]). For LPs X^d with bounded exponential moments, the present regularity analysis may be localised to compact subsets, well separated from the barriers, subject to an exponentially small localisation error term; see Hilber et al. [32, 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 have assumed throughout that contracts do not pay dividends. However, including a dividend stream (with constant rate over $(0, T]$) on the underlying does not change the mathematical arguments; we refer to Lamberton and Mikou [38, Sect. 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 Becker et al. [8]. For the particular case of American-style contracts in exponential Lévy models, [38] 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 [38] 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 analysed in Opschoor et al. [42, Sect. 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 Barrios et al. [4].

References

1. Applebaum, D.: Lévy Processes and Stochastic Calculus, 2nd edn. Cambridge University Press, Cambridge (2009)
2. Barles, G., Buckdahn, R., Pardoux, E.: Backward stochastic differential equations and integral-partial differential equations. Stoch. Stoch. Rep. **60**, 57–83 (1997)
3. Barles, G., Imbert, C.: Second-order elliptic integro-differential equations: viscosity solutions' theory revisited. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **25**, 567–585 (2008)
4. Barrios, B., Figalli, A., Ros-Oton, X.: Free boundary regularity in the parabolic fractional obstacle problem. Commun. Pure Appl. Math. **71**, 2129–2159 (2018)
5. Barron, A.R.: Universal approximation bounds for superpositions of a sigmoidal function. IEEE Trans. Inf. Theory **39**, 930–945 (1993)
6. Bartlett, P.L., Mendelson, S.: Rademacher and Gaussian complexities: risk bounds and structural results. J. Mach. Learn. Res. **3**(Spec. Issue Comput. Learn. Theory), 463–482 (2002)
7. Beck, C., Gonon, L., Jentzen, A.: Overcoming the curse of dimensionality in the numerical approximation of high-dimensional semilinear elliptic partial differential equations. Working paper (2020). Available online at arXiv:2003.00596
8. Becker, S., Cheridito, P., Jentzen, A.: Deep optimal stopping. J. Mach. Learn. Res. **20**, 74 (2019).
9. Berner, J., Grohs, P., Jentzen, A.: Analysis of the generalisation error: empirical risk minimisation over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. SIAM J. Math. Data Sci. **2**, 631–657 (2020)
10. Bertoin, J.: Lévy Processes. Cambridge University Press, Cambridge (1996)
11. Böttcher, B., Schilling, R., Wang, J.: Lévy Matters. III. Lévy-Type Processes: Construction, Approximation and Sample Path Properties. Lecture Notes in Mathematics, vol. 2099. Springer, Cham (2013)
12. Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford University Press, Oxford (2013)
13. Boyarchenko, S., Levendorskiĭ, S.: Barrier options and touch-and-out options under regular Lévy processes of exponential type. Ann. Appl. Probab. **12**, 1261–1298 (2002)
14. Buehler, H., Gonon, L., Teichmann, J., Wood, B.: Deep hedging. Quant. Finance **19**, 1271–1291 (2019)
15. Cohen, A., DeVore, R., Schwab, C.: Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. Found. Comput. Math. **10**, 615–646 (2010)

16. Cont, R., Tankov, P.: Financial Modelling with Jump Processes. Chapman & Hall/CRC, Boca Raton (2004)
17. Cont, R., Voltchkova, E.: A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. SIAM J. Numer. Anal. **43**, 1596–1626 (2005)
18. Cont, R., Voltchkova, E.: Integro-differential equations for option prices in exponential Lévy models. Finance Stoch. **9**, 299–325 (2005)
19. E, W., Wojtowytsch, S.: Some observations on partial differential equations in Barron and multi-layer spaces. Working paper (2021). Available online at arXiv:2012.01484
20. Eberlein, E., Glau, K.: Variational solutions of the pricing PIDEs for European options in Lévy models. Appl. Math. Finance **21**, 417–450 (2014)
21. Eberlein, E., Kallsen, J.: Mathematical Finance. Springer, Cham (2019)
22. Elbrächter, D., Grohs, P., Jentzen, A., Schwab, C.: DNN expression rate analysis of high-dimensional PDEs: application to option pricing. Constr. Approx. (2021). https://doi.org/10.1007/s00365-021-09541-6, forthcoming
23. Esche, F., Schweizer, M.: Minimal entropy preserves the Lévy property: how and why. Stoch. Process. Appl. **115**, 299–327 (2005)
24. Farkas, W., Reich, N., Schwab, C.: Anisotropic stable Lévy copula processes—analytical and numerical aspects. Math. Models Methods Appl. Sci. **17**, 1405–1443 (2007)
25. Gerber, H.U., Shiu, E.S.W.: Option pricing by Esscher transforms. Trans.—Soc. Actuar. **46**, 99–191 (1994)
26. Glau, K.: A Feynman–Kac-type formula for Lévy processes with discontinuous killing rates. Finance Stoch. **20**, 1021–1059 (2016)
27. Gonon, L., Grohs, P., Jentzen, A., Kofler, D., Šiška, D.: Uniform error estimates for artificial neural network approximations for heat equations. IMA J. Numer. Anal. (2021, forthcoming). https://doi.org/10.1093/imanum/drab027. Available online at arXiv:1911.09647
28. Gonon, L., Schwab, C.: Deep ReLU neural networks overcome the curse of dimensionality for partial integro-differential equations. Working paper (2021). Available online at arXiv:2102.11707
29. Grohs, P., Hornung, F., Jentzen, A., von Wurstemberger, P.: A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. Mem. Am. Math. Soc. (2021, forthcoming). Available online at arXiv:1809.02362
30. Grohs, P., Hornung, F., Jentzen, A., Zimmermann, P.: Space-time error estimates for deep neural network approximations for differential equations. Working paper (2019). Available online at arXiv:1908.03833
31. Hilber, N., Reich, N., Schwab, C., Winter, C.: Numerical methods for Lévy processes. Finance Stoch. **13**, 471–500 (2009)
32. Hilber, N., Reichmann, O., Schwab, C., Winter, C.: Computational Methods for Quantitative Finance. Finite Element Methods for Derivative Pricing. Springer, Heidelberg (2013)
33. Hörmander, L.: An Introduction to Complex Analysis in Several Variables. Van Nostrand, Princeton (1966)
34. Ito, K., Reisinger, C., Zhang, Y.: A neural network based policy iteration algorithm with global H^2-superlinear convergence for stochastic games on domains. Found. Comput. Math. **21**, 331–374 (2021)
35. Jeanblanc, M., Klöppel, S., Miyahara, Y.: Minimal f^q-martingale measures of exponential Lévy processes. Ann. Appl. Probab. **17**, 1615–1638 (2007)
36. Kallsen, J., Tankov, P.: Characterisation of dependence of multidimensional Lévy processes using Lévy copulas. J. Multivar. Anal. **97**, 1551–1572 (2006)
37. Krantz, S.G., Parks, H.R.: A Primer of Real Analytic Functions. Birkhäuser, Basel (1992)
38. Lamberton, D., Mikou, M.: The critical price for the American put in an exponential Lévy model. Finance Stoch. **12**, 561–581 (2008)
39. Ledoux, M., Talagrand, M.: Probability in Banach Spaces. Isoperimetry and Processes. Springer, Berlin (2011). Reprint of the 1991 edition
40. Matache, A.M., von Petersdorff, T., Schwab, C.: Fast deterministic pricing of options on Lévy driven assets M2AN. Math. Model. Numer. Anal. **38**, 37–71 (2004)
41. Nualart, D., Schoutens, W.: Backward stochastic differential equations and Feynman–Kac formula for Lévy processes, with applications in finance. Bernoulli **7**, 761–776 (2001)
42. Opschoor, J.A.A., Petersen, P.C., Schwab, C.: Deep ReLU networks and high-order finite element methods. Anal. Appl. **18**, 715–770 (2020)
43. Opschoor, J.A.A., Schwab, C., Zech, J.: Exponential ReLU DNN expression of holomorphic maps in high dimension. Constr. Approx. **18** (2021). https://doi.org/10.1007/s00365-021-09542-5

44. Petersen, P., Voigtlaender, F.: Optimal approximation of piecewise smooth functions using deep ReLU neural networks. Neural Netw. **108**, 296–330 (2018)

45. Reisinger, C., Zhang, Y.: Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. Anal. Appl. **18**, 951–999 (2020)

46. Rodino, L.: Linear Partial Differential Operators in Gevrey Spaces. World Scientific, River Edge (1993)

47. Ruf, J., Wang, W.: Neural networks for option pricing and hedging: a literature review. J. Comput. Finance **24**, 1–45 (2020)

48. Sato, K.: Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press, Cambridge (1999)

49. Schwab, C., Zech, J.: Deep learning in high dimension: neural network expression rates for generalized polynomial chaos expansions in UQ. Anal. Appl. **17**, 19–55 (2019)