



# Calibration of Lévy Processes Using Optimal Control of Kolmogorov Equations with Periodic Boundary Conditions

Mario Annunziato<sup>a</sup> and Hanno Gottschalk<sup>b</sup>

<sup>a</sup>*Dipartimento di Matematica, Università degli Studi di Salerno*  
Via Giovanni Paolo II, 132 - 84084 Fisciano, Italia

<sup>b</sup>*School of Mathematics and Science, University of Wuppertal*  
Gaußstr. 20, 42119 Wuppertal, Germany

E-mail(*corresp.*): [mannunzi@unisa.it](mailto:mannunzi@unisa.it)

E-mail: [hanno.gottschalk@uni-wuppertal.de](mailto:hanno.gottschalk@uni-wuppertal.de)

Received April 26, 2017; revised May 4, 2018; accepted May 5, 2018

**Abstract.** We present an optimal control approach to the problem of model calibration for Lévy processes based on a non-parametric estimation procedure of the measure. The optimization problem is related to the maximum likelihood theory of sieves [25] and is formulated with the Fokker-Planck-Kolmogorov approach [3, 4].

We use a generic spline discretization of the Lévy jump measure and select an adequate size of the spline basis using the Akaike Information Criterion (AIC) [12]. The first order necessary optimality conditions are derived based on the Lagrange multiplier technique in a functional space. The resulting Partial Integral-Differential Equations (PIDE) are discretized, numerically solved using a scheme composed of Chang-Cooper, BDF2 and direct quadrature methods, jointly to a non-linear conjugate gradient method. For the numerical solver of the Kolmogorov's forward equation we prove conditions for non-negativity and stability in the  $L^1$  norm of the discrete solution.

**Keywords:** optimal control of PIDE, Kolmogorov-Fokker-Planck equation, Lévy processes, non-parametric maximum likelihood method, IMEX numerical method.

**AMS Subject Classification:** 93E10; 49K20; 60G51; 62G05.

## 1 Introduction

Lévy processes play a large role in contemporary mathematical finance [15], but also in many areas of physics, see e.g. [6, 37]. A real valued Lévy process

---

Copyright © 2018 The Author(s). Published by VGTU Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

is a stochastic process  $Y(t)$  that has increments  $Y(t) - Y(s)$ ,  $t \geq s$ , that are independent of the past. The increments are also stationary in the sense that the probability distribution of the increment only depends on the time difference  $t - s$ . Furthermore,  $Y(0) = 0$  and a stochastic continuity condition for  $t = 0$  holds, see e.g. [6]. Under the given conditions, the characteristic function of  $Y(t)$  is given by the Lévy-Khinchine representation

$$\mathbb{E} \left[ e^{iY(t)k} \right] = e^{t\psi(k)},$$

where  $\mathbb{E}[\cdot]$  stands for the expected value.  $\psi(k)$  is a conditionally positive definite function [9] that has the following representation in terms of the canonical triplet  $(b, \sigma^2, \nu)$ :

$$\psi(k) = ibk - \frac{\sigma^2}{2}k^2 + \int_{\mathbb{R} \setminus \{0\}} (e^{isk} - 1 - isk1_{\{|s| \leq 1\}}(s)) d\nu(s), \tag{1.1}$$

$b, \sigma^2 \in \mathbb{R}$  are constants,  $\sigma^2 \geq 0$ , and the Lévy measure  $\nu$  is a positive measure on  $\mathbb{R} \setminus \{0\}$  such that

$$\int_{\mathbb{R} \setminus \{0\}} \min(1, s^2)d\nu(s) < \infty.$$

In (1.1),  $1_{\{|s| \leq 1\}}(s)$  is the characteristic function of the set  $\{|s| \leq 1\}$  which takes the value 1 on this set and 0 otherwise.

The calibration problem for Lévy processes consists of the estimation of the canonical triplet  $(b, \sigma^2, \nu)$  given the observation  $Y(t_j)$  of the process' trajectory  $Y(t)$  at some prescribed times  $t_j$ ,  $j = 1, \dots, L$ .  $Y(t)$  can also be understood as the solution to the Stochastic Differential Equation (SDE) of jump-diffusion type

$$dY(t) = bdt + \sigma dB(t) + \int_{\{|y| \leq 1\}} y\tilde{N}(dt, dy) + \int_{\mathbb{R} \setminus \{|y| > 1\}} yN(dt, dy), \quad Y(0) = 0. \tag{1.2}$$

Here  $B(t)$  is a standard Brownian motion and  $N((t, t + \Delta t], A) \sim \text{Po}(\nu(A)\Delta t)$  is the random counting measure of jumps of height in the set  $A \subseteq \mathbb{R}$  in the time interval  $(t, t + \Delta t]$ .  $\text{Po}(\lambda)$  stands for the Poisson distribution with intensity  $\lambda$  and  $\tilde{N}((t, t + \Delta t], A) = N((t, t + \Delta t], A) - \nu(A)\Delta t$  is the compensated or martingale jump measure for small jumps, where we require  $A \subset \mathbb{R} \setminus \{|x| \leq \varepsilon\}$  for some  $\varepsilon > 0$ , see [6] for further details.

The calibration problem for Lévy processes, i.e. the solution of (1.2), unfortunately is ill posed: The collection of all Lévy measures  $\nu$  is infinite dimensional, while only  $L$  observations are available. Direct application of the maximum likelihood principle [21] in this situation leads to severe over-fitting issues [25]. In many applications, one chooses families of Lévy measures  $\nu(\bar{\alpha})$  that depend only on a finite dimensional parameter vector  $\bar{\alpha}$ , see e.g. [29]. This assumption might however not be justified and give rise to modelling errors. Notice that in this paper we do not deal with the option pricing problems, as this is mostly done with the public quotes of option prices, e.g. using butterfly spreads or second order derivatives of the European Calls with respect to the

strike [1, 15]. Anyway, if one would like to calculate option prices based on historical quotes of the underlying stock, our method can be adapted as described in Remark 2.

As a non-parametric alternative, one can use generic parametrizations for the density of the Lévy measure  $\nu$  that can be refined depending on the amount of data available. This gives rise to a hierarchy – or sieve [25] – of maximum likelihood (ML) problems with a finite number of parameters. The solution of the ML-problem then gives rise to high dimensional optimization problems.

Non parametric estimation of the Lévy measure has been studied previously, see e.g. [14, 30]. These works rely on fits of expressions involving the (empirical) Fourier transform and directly estimate the Lévy characteristics. Important features like consistency and rate optimality of these estimators are proven as well. Spectral methods have been studied in [7]. There also rate optimal convergence of the estimators has been proven. Also, these methods postulate the global solution of a non-convex optimization problem, as it is also the case for the ML-estimate. Rate optimality is however less strong than optimality in the Cramér-Rao sense [21, 31], that is reached by the maximum likelihood method, if the true law is within the parametrization scheme. Also, the scope of this work is different, as we focus on computational aspects which have the potential for generalization to SDE with non constant coefficients (although this is not realized in the present paper), while the spectral method is limited to pure Lévy distributions.

The maximum likelihood method requires representation of the probability density functions (PDF). The PDF can be obtained as a solution to the Kolmogorov forward equation (Fokker-Planck equation). The parameters  $\bar{\alpha}$  then enter in this equation via coefficients in the generator of the semigroup [6]. The Kolmogorov equation for Lévy processes involves an integral operator of convolution type. This places the model calibration problem in the framework of optimal control problems with partial integral differential equations (PIDE) constraints.

This problem can be framed as a constrained PDE optimization problem, where the PDE is the Fokker-Planck, i.e. Kolmogorov forward equation [3, 4, 5, 24]. Following this framework, the solution of the ML-problem is found by solving the first order optimality conditions in a functional space, that is the optimality system consisting of two PIDEs, named forward and backward (or adjoint) equations, plus an optimality condition.

The first difficulty to numerically solve this optimality system is the integral operator. In fact, in the case of using a fully implicit method, it would lead to solve a dense system of equation, for this reason implicit-explicit (IMEX) or operator splitting methods can be applied to bypass this problem (see [11, 16, 20, 23]). The solution of the Kolmogorov forward equation is a probability density function that is non negative with constant integral over the domain. Such properties must be owned from the discrete solution too. The Chang-Cooper (CC) is a non-negative and conservative numerical method that has been used to solve the classical Fokker-Planck equation [4, 13, 33]. For alternative numerical scheme, such a based on multiresolution and discontinuous Galerkin see, e.g., [32] and [36] respectively. We use the CC method with

an implicit time difference scheme for the differential operators of our PIDE, and evaluate the integral operator at the previous time step solution, i.e. in an explicit way. We prove for the resulting numerical solver: conservativeness, non-negative preserving and stability in the  $L^1$ -norm. The solver for the adjoint equation is obtained from the solver for the forward equation by the “discretize then optimize” approach. The second problem, the infinite computational domain in the spatial direction, is resolved by periodic boundary conditions. On stochastic process side, this corresponds to a mapping of sampled values from  $\mathbb{R}$  to a large but finite torus.

Finally, for related work with vanishing Lévy measure we quote [4, 10], and for estimation procedures based on non parametric approximations of the empirical characteristic function [7, 8]. An approach based on the method of moments and asymptotic expansions of Lévy densities can be found in [27] and one using multiple curve interest rate models in [28].

The article is organised as follows: In Section 2, we set up the maximum likelihood estimation problem for a given parametrization and derive Kolmogorov’s forward (Fokker-Planck) equation and its adjoint (Kolmogorov backward) equation with terminal conditions set by the log-likelihood objective functional. In Section 3 the discretization for Kolmogorov’s equations and the optimal control scheme is derived following a Chang-Cooper and IMEX approach and structural properties are proven. Section 4 gives numerical tests of the proposed procedure based on simulated data. We propose to use Akaike’s information criterion (AIC) [12] to choose an adequate parametrization from the hierarchy of parametrizations for density of the Lévy measure. As a final application, we fit a Lévy process to financial data from the German stock exchange DAX in a period between April 1998 and March 2002.

## 2 Kolmogorov equations and optimality for the log-likelihood

Let the  $L$  independent sample random values  $X_1, \dots, X_L$  be given and  $X_l \in \Omega$ ,  $l = 1, \dots, L$ , where  $\Omega = [\Omega_a, \Omega_b]$ . These values can e.g. be obtained as  $X_l = \phi(Y_l)$ , where  $\phi(\cdot)$  is the group homomorphism from  $\mathbb{R}$  to  $\Omega = [\Omega_a, \Omega_b]$  with periodic boundary conditions, from the Lévy process  $Y(t)$  on  $\mathbb{R}$ . We restrict our analysis to the bounded domain  $\Omega$ . This is a numerical necessity and is not motivated by the application. For practical purposes the size of  $\Omega$  can be determined using the Čebyšev’s inequality, e.g.  $P(|X(t)| > \Omega_a) \leq P(|X(t) - E[X(t)]| > \Omega_a - E[X(t)]) \leq \text{Var}(X(t))/(\Omega_a - E[X(t)])^2$  in order to determine  $\Omega_a$  ( $\Omega_b$  can be estimated analogously). In such a way is possible to calculate the size of the domain so large such that the probability of  $X(t)$  “being wrapped” by the cyclic conditions is sufficiently small. Expected value and variance can be efficiently estimated from the historical data.

We deal with the problem to find the PDF of  $X(T)$  such that it best fits with the sample values. For this purpose we consider the maximum likelihood problem in the framework of PIDE-constrained optimization: we have to find

the maximum likelihood estimator

$$\max_{\bar{\alpha}} J(f, \bar{\alpha})$$

with respect to the parametrization of the measure given by  $\bar{\alpha}$ , where

$$J(f, \bar{\alpha}) = \frac{1}{L} \sum_{l=1}^L \log(f(X_l, T, \bar{\alpha})) \tag{2.1}$$

is the (normalized) log-likelihood with the constraint given by the following Kolmogorov forward (Fokker-Planck) equation for the Lévy process  $X(t)$ . In the present case we define the process on the torus  $\Omega$  with Lévy data  $(b, \sigma^2, \nu_{\bar{\alpha}})$  and use the parametrization  $d\nu_{\bar{\alpha}}(s) = \sum_{j=1}^{N_{\Theta}} \alpha_j \Theta_j(s) ds$ , hence the Kolmogorov equation reads as:

$$\begin{cases} \partial_t f(x, t) + b \partial_x f(x, t) - \frac{\sigma^2}{2} \partial_x^2 f(x, t) \\ - \int_{\Omega} \sum_{j=1}^{N_{\Theta}} \alpha_j (f(x + s, t) - f(x, t)) \Theta_j(s) ds = 0, \\ f(x, 0) = f_0(x), \\ f(\Omega_a, t) = f(\Omega_b, t), \quad \partial_x f(\Omega_a, t) = \partial_x f(\Omega_b, t), \end{cases} \tag{2.2}$$

where  $f(x, t)$  represents the PDF of the process at time  $t$  and  $f_0(x)$  that one initial. This PIDE is defined in the interval of time  $t \in [0, T]$  and with periodic boundary conditions on  $\Omega_a$  and  $\Omega_b$ . Here  $\Theta_j(s)$  is a set of triangular shaped basis for the set of continuous functions that are linear on  $(\theta_{j-1}, \theta_j)$

$$\begin{aligned} \Theta_j(s) &= 1 + (s - \theta_j) / \Delta, & s \in [\theta_j - \Delta, \theta_j], \\ \Theta_j(s) &= 1 - (s - \theta_j) / \Delta, & s \in [\theta_j, \theta_j + \Delta], \end{aligned}$$

where  $\theta_j = \Omega_a + \Delta(j - 1)$  for  $j = 1, \dots, N_{\Theta} + 1$  are the points of a discrete uniform mesh of step size  $\Delta = (\Omega_b - \Omega_a) / N_{\Theta}$  defined on the domain. The periodicity  $\Theta_1(s) \equiv \Theta_{N_{\Theta}+1}(s)$  is assumed. We notice that the spline approximation can approximate also the case of Lévy singular measure. In fact, this choice corresponds in Equation (2.2) to a compound Poisson measure, whose set of distribution is dense in the set of all Lévy distributions w.r. to weak convergence in law [9], hence providing a pointwise convergence of the characteristic functions. Further, the triangular shaped base function gives a very simple positivity constraint  $\alpha_j \geq 0$  and continuity of the density for the Lévy measure such those from application in finance.

The existence and uniqueness of the solution of the Fokker-Planck-Kolmogorov equation (2.2) is well established, also for initial conditions belonging to the class of measures [9].

If we write the mapping  $\bar{\alpha} \rightarrow f(\bar{\alpha})$  between the maximization parameters and the PDF, then we introduce the so-called reduced cost functional  $\hat{J}(\bar{\alpha}) = J(f(\bar{\alpha}), \bar{\alpha})$ , so that the maximization problem becomes

$$\max_{\bar{\alpha}} \hat{J}(\bar{\alpha}) = \max_{\bar{\alpha}} J(f(\bar{\alpha}), \bar{\alpha}). \tag{2.3}$$

A local maxima  $\bar{\alpha}^*$  for  $\hat{J}$  can be found by solving the optimality system obtained by vanishing the variations of the following Lagrangian functional

$$\begin{aligned} \mathcal{L}(f, p, \bar{\alpha}, \bar{\pi}) &= \frac{1}{L} \sum_{l=1}^L \log(f(X_l, T, \bar{\alpha})) + \int_0^T \int_{\Omega} \left[ \partial_t f(x, t) + b \partial_x f(x, t) \right. \\ &\quad \left. - \frac{\sigma^2}{2} \partial_x^2 f(x, t) - \sum_{j=1}^{N_{\Theta}} \alpha_j \int_{\Omega} (f(x + s, t) - f(x, t)) \Theta_j(s) ds \right] p(x, t) dx dt - \sum_{j=1}^{N_{\Theta}} \pi_j \alpha_j, \end{aligned} \tag{2.4}$$

where  $\bar{\pi} = (\pi_1, \dots, \pi_{N_{\Theta}})$  and  $\bar{\alpha}$  fulfil the usual Karush-Kuhn-Tucker (KKT) conditions  $\pi_j \alpha_j = 0$  and  $\pi_j \geq 0$ . These are important to include the non-negativity constraints for the control variables. Note that if the condition  $\alpha_j \geq 0$  is violated for some  $j \in \{1, \dots, N_{\Theta}\}$ , the density of the measure  $d\nu_{\bar{\alpha}}(s)$  is negative in a neighbourhood of  $\theta_j$  and thus is not a Lévy measure any more. The sum  $\sum_{j=1}^{N_{\Theta}} \pi_j \alpha_j$  should be extended only on the active constraints, i.e. when  $\alpha_j^* = 0$ . For those values of  $j$  on the maximum where  $\alpha_j^* > 0$  we have  $\pi_j^* = 0$ .

First we calculate the variation  $\mathcal{L}(f + \delta f) - \mathcal{L}(f)$  for the adjoint equation. In the following the variations are calculated separately for each addend of the r.h.s. We get

$$\begin{aligned} & - \sum_{j=1}^{N_{\Theta}} \alpha_j \int_0^T \int_{\Omega} \int_{\Omega} (-\delta f(x, t)) \Theta_j(s) p(x, t) ds dx dt \\ &= \sum_{j=1}^{N_{\Theta}} \alpha_j \int_0^T \int_{\Omega} \left( \int_{\Omega} \Theta_j(s) ds \right) \delta f(x, t) p(x, t) dx dt. \end{aligned}$$

For the term  $-\sum_j \alpha_j \int_0^T \int_{\Omega} \left( \int_{\Omega} \delta f(x + s, t) \Theta_j(s) ds \right) p(x, t) dx dt$  we apply the substitution  $y = x + s$ , then exchange  $x \leftrightarrow y$ , so that it recasts to  $-\sum_j \alpha_j \int_0^T \int_{\Omega} \left( \int_{\Omega} p(y, t) \Theta_j(x - y) dy \right) \delta f(x, t) dx dt$ . Then, again, we substitute  $s = x - y$  and, by inserting also the former term, we get

$$- \int_0^T \int_{\Omega} \left[ \sum_{j=1}^{N_{\Theta}} \alpha_j \int_{\Omega} (p(x - s, t) - p(x)) \Theta_j(s) ds \right] \delta f(x, t) dx dt.$$

For the variation of the time derivative, integrating by parts, one obtains

$$\int_{\Omega} \delta f(x, t) p(x, t) \Big|_0^T dx - \int_0^T \partial_t p(x, t) \delta f(x, t) dx dt.$$

The variation  $\delta f(x, 0) = 0$  holds because of the Cauchy initial condition, while the variation in  $T$  can be defined in some points  $X_l$ . Next, we integrate by parts the term with the first order derivative in  $x$  and obtain  $b \int_0^T (\delta f(x, t) p(x, t)) \Big|_{\Omega_a}^{\Omega_b} dt - b \int_{\Omega} \partial_x p \delta f dx dt$ . Due to periodicity in the first term,  $\delta f(\Omega_a, t) (p(\Omega_b, t) - p(\Omega_a, t))$  has to be zero, hence  $p(\Omega_b, t) = p(\Omega_a, t)$ .

From the diffusive term we get

$$-\frac{\sigma^2}{2} \int_0^T \int_{\Omega} \partial_x^2 \delta f p(x, t) dx dt = -\frac{\sigma^2}{2} \int_0^T [\partial_x \delta f p - \delta f \partial_x p]_{\Omega_a}^{\Omega_b} dt + \int_{\Omega} \partial_x^2 p(x, t) dx dt.$$

The first boundary term is zero because of the periodic condition of the variation of the derivative of  $f$  at the boundaries, and because of the previous periodic condition on  $p$ . The second is analogous and has to vanish, we therefore get the continuity condition  $\partial_x p(\Omega_a, t) = \partial_x p(\Omega_b, t)$ .

By collecting all the terms under double integral, we get the adjoint equation. The remaining boundary term  $\int_{\Omega} \delta f(x, T) p(x, T) dx$  will be considered below.

To calculate the variation on  $f$  in the functional  $J$  we perform an additional integration in space, so that

$$\frac{1}{L} \sum_{l=1}^L \int_{\Omega} \log(f(x, T)) \delta(x - X_l) dx, \tag{2.5}$$

where  $\delta(\cdot)$  is the  $\delta$ -Dirac measure, then variate  $f(x, T) + \delta f(x, T)$ , hence

$$\begin{aligned} & \int_{\Omega} \log(f(x, T) + \delta f(x, T)) \delta(x - X_l) dx \\ &= \int_{\Omega} (\log(f(x, T)) + \delta f(x, T)/f(x, T)) \delta(x - X_l) dx, \end{aligned}$$

so that the first order terms plus the remaining boundary, give

$$\frac{1}{L} \sum_{l=1}^L \int_{\Omega} \frac{\delta f(x, T)}{f(x, T)} \delta(x - X_l) dx + \int_{\Omega} p(x, T) \delta f(x, T) dx.$$

This expression have to be zero for each  $\delta f(x, T)$ . It represents the terminal condition for the adjoint equation: that is  $p(X_l, T) = -1/(L f(x_l, T))$ , and  $p(x, T) = 0$ , if  $x \neq \{X_1, \dots, X_L\}$ . In case of multiplicity of  $X_l$  the condition becomes  $p(X_l, T) = -1/(L \sum_{l'} f(X_{l'}, T))$ , with  $l'$  running on the multiplicity value.

Summarizing, the adjoint equation (Kolmogorov's backward equation) is defined as follows:

$$\begin{cases} -\partial_t p(x, t) - b \partial_x p(x, t) - \frac{\sigma^2}{2} \partial_x^2 p(x, t) \\ - \int_{\Omega} \sum_{j=1}^{N_{\Theta}} \alpha_j (p(x - s, t) - p(x, t)) \Theta_j(s) ds = 0, \\ p(x, T) = -\frac{1}{L} \sum_{l=1}^L \delta(x - X_l) / f(X_l, T), \\ p(\Omega_a, t) = p(\Omega_b, t), \quad \partial_x p(\Omega_a, t) = \partial_x p(\Omega_b, t). \end{cases} \tag{2.6}$$

We note that by reverting the sign of the time we get the same PIDE as the forward equation (up to a reflection of the drift and jump direction), hence this equation has a unique solution, also for the non regular final value problem [9].

Second, we variate in (2.4) the fitting parameters  $\mathcal{L}(\alpha_j + \delta\alpha_j) - \mathcal{L}(\alpha_j)$ , from which we found the optimality equations (i.e. a vanishing gradient condition)

$$-\pi_{j'} - \int_0^T \int_{\Omega} \int_{\Omega} (f(x+s, t) - f(x, t))p(x, t)\Theta_j(s) ds dx dt = 0, \quad j = 1, \dots, N_{\Theta}, \tag{2.7}$$

where  $j'$  runs on the set of values where  $\alpha_{j'}^* = 0$ . Note that the active  $\pi_{j'}$  do not change the gradient, but simply balance non-zero gradient components that point to the directions where the inequality constraint  $\alpha_{j'} \geq 0$  is violated. As in our case we deal with simple box-constraints on the  $\alpha_j$  themselves, we can set those components of the negative gradient equal to zero that correspond to an active index  $j'$  and are negative, when determining the update. This then accounts for the effect of the  $\pi_{j'}$ , see e.g. [22, 34].

The forward equation (2.2), can be written in flux form:

$$\partial_t f(x, t) = \partial_x \mathcal{F}(x, t),$$

where  $\mathcal{F}(x, t)$  is the flux defined as

$$\mathcal{F}(x, t) = -bf(x, t) + \frac{\sigma^2}{2} \partial_x f(x, t) + \sum_{j=1}^{N_{\Theta}} \alpha_j \int_{\Omega} \left( \int_{-s}^0 f(x-y, t) dy \right) \Theta_j(s) ds. \tag{2.8}$$

By using  $\partial_x \int_{-s}^0 f(x-y) dy = \int_{-s}^0 f'(x-y) dy = \int_x^{x+s} f'(z) dz = f(x+s) - f(x)$ , it is easy to verify that Equation (2.8) is equivalent to Equation (2.2). Further, from the conservation of the total probability, it follows that the flux has the periodic boundary condition  $\mathcal{F}(\Omega_a, t) = \mathcal{F}(\Omega_b, t)$ . From this we immediately get the periodic condition on the first derivative  $\partial_x f(\Omega_a, t) = \partial_x f(\Omega_b, t)$ .

The 1<sup>st</sup> order necessary optimality system consists of the Equations (2.2), (2.6) and (2.7). Its solution gives values  $\alpha_1^*, \dots, \alpha_{N_{\Theta}}^*$  that are candidates for maximizing the functional (2.1). We remark that global minima can not be easily found when using a maximum likelihood method.

*Remark 1.* The optimization problem that emerges from maximizing the likelihood in the parameters  $\alpha_j$  is non convex. Therefore solutions might get stuck in local minima. This is typical for the maximum likelihood method in general. On the other hand, maximum likelihood provides error rates that asymptotically approach the Cramèr-Rao minimal error bound. It would be therefore desirable to maintain the optimal rates from maximum-likelihood theory and avoid sub optimal local minima by a clever initial guess for the starting point  $\alpha_j$ , such that the starting point is in the domain of attraction of the global maximum. We suggest that this can be achieved by an hybrid method that combines spectral methods, based on a direct estimation of the Lévy characteristic  $\hat{\psi}$  by taking the complex logarithm of the the empirical characteristic function on a slowly increasing domain [17], and, thereafter, solving a convex optimization problem by minimizing the distance between  $\hat{\psi}$  and the Lévy characteristic parametrized by the parameters  $\alpha_j$  subject to the constraint  $\alpha_j \geq 0$ . We intend to come back to this problem in future research.



*Remark 2.* If we consider option pricing with historical data (less frequently used in practice), one should introduce the constraint for the exponential martingale condition  $\psi_\alpha(-i) = 0$ , which is in explicit form using the Laplace transform of the triangular distribution

$$b + \sigma^2 + \Delta \left[ \sum_{j=1}^N \alpha_j e^{\theta_j} \left( \frac{\cosh(\Delta) - 1}{\Delta^2} \right) - \sum_{j=1}^N \alpha_j \right] = 0.$$

This equality constraint can now be added into the Lagrangian with a Lagrange multiplier. As the negative log-likelihood, up to a constant that does not depend on the parameters  $\alpha$ , almost surely converges to the Kullback-Leibler divergence uniformly on compact sets in parameter space (also called relative entropy). It is clear that up to the error of replacing the Kullback-Leibler divergence with the log-likelihood we minimize the relative entropy and thus approximate the minimal entropy martingale measure.

### 3 Numerical scheme

The numerical solution of the optimality system is found by a non linear gradient conjugate iterative procedure [4, 26, 39]. At each iteration the solution of two PIDEs, the forward and the adjoint one, must be found. In particular the structural properties of the PDF solution must be satisfied, as well as a stability condition of the PIDEs numerical scheme solver.

For the numerical discretization of the Kolmogorov forward equation we use the Chang-Cooper scheme (CC) [13], joint to a 2nd order backward differentiation formula (BDF2) for the discrete time operator. The CC method was proposed for a Fokker-Planck resp. Kolmogorov equation [4] without the integral term. It is stable, second-order accurate, non-negative, and conservative numerical scheme [4, 33].

The CC method is used for the differential operators, the integral term is treated separately according to an IMEX methodology. We denote the following  $B = -b$  and  $C = \sigma^2/2$ , then the Kolmogorov forward equation reads as follows

$$\partial_t f(x, t) = \partial_x F(x, t) + \int_{\Omega} \sum_{j=1}^{N_\Theta} \alpha_j (f(x + s, t) - f(x, t)) \Theta_j(s) ds, \tag{3.1}$$

where  $F(x, t) = Bf(x, t) + C\partial_x f(x, t)$ . Consider a uniform grid of size  $h$  on the space domain  $\{\Omega_h\}_{h>0}$  given by  $\Omega_h = \{x \in \mathbb{R} : x_i = i h + \Omega_a, i = 0, \dots, N, h = (\Omega_b - \Omega_a)/N\}$  and a uniform grid on the time domain  $I_{\delta t} = \{t \in [0, T] : t_m = m \delta t, m = 0, \dots, N_T, \delta t = T/N_T\}$ . Let  $f_i^m \approx f(x_i, t_m)$  denote the approximated values of the continuous solution of the FPE. We employ the following discretization of (2.8)

$$\partial_{BD}^- f_i^{m+1} = \frac{1}{h} (F_{i+1/2}^{m+1} - F_{i-1/2}^{m+1}) + Q(f_i^m; \bar{\alpha}), \tag{3.2}$$

where

$$\partial_{BD}^- f_i^m = \frac{3f_i^m - 4f_i^{m-1} + f_i^{m-2}}{2\delta t},$$

is the BDF2 operator.  $Q(f_i^m; \bar{\alpha})$  is the sum of the integrals of Equation (2.8) calculated with the mid-point scheme

$$Q(f_i^m; \bar{\alpha}) = h \sum_{j=1}^{N_\Theta} \alpha_j \sum_{k=1}^N \hat{f}_{ik}^m \theta_{jk} - a f_i^m,$$

where  $\bar{\alpha} = (\alpha_1, \dots, \alpha_{N_\Theta})$ ,  $\hat{f}_{ik}^m \approx (f(x_i + s_k, t_m))_{\Omega_h}$  represents the translated  $f_i^m$  by the value  $s_k \in \Omega_h$  and continued by periodicity,  $\theta_{jk} = \Theta_j(s_k)$ ,

$$a = h \sum_{j=1}^{N_\Theta} \alpha_j \sum_{k=1}^N \theta_{jk} \approx \sum_{j=1}^{N_\Theta} \alpha_j \int_{\Omega} \Theta_j(s) ds. \tag{3.3}$$

Note also that the summation starts from  $k = 1$ , because the point  $k = 0$  is the same of that  $k = N$ . Therefore, the solution at a new time step is calculated by solving the following equation for the unknown  $f_i^{m+1}$

$$3f_i^{m+1} - \frac{2\delta t}{h} (F_{i+1/2}^{m+1} - F_{i-1/2}^{m+1}) = 4f_i^m - f_i^{m-1} + 2\delta t Q(f_i^m; \bar{\alpha}) \tag{3.4}$$

with the initial condition

$$f_i^0 = f_{0,i}. \tag{3.5}$$

This scheme is based on the fluxes at  $N$  cell boundaries. The partial flux at the position  $x_{i+h/2}$  is computed as follows

$$F_{i+1/2}^{m+1} = \left[ (1 - \delta)B + \frac{1}{h}C \right] f_{i+1}^{m+1} - \left( \frac{1}{h}C - \delta B \right) f_i^{m+1}. \tag{3.6}$$

This formula results from the following linear convex combination of  $f$  at the points  $i$  and  $i + 1$ :

$$f_{i+1/2}^{m+1} = (1 - \delta) f_{i+1}^{m+1} + \delta f_i^{m+1}, \quad \delta \in [0, 1].$$

The idea of implementing this combination was proposed by Chang and Cooper in [13] and it was motivated with the need to guarantee positive solutions that preserve the equilibrium configuration. Indeed, the CC method is related to exponential fitting methods, such as that one proposed by Allen and Southwell [2], and by the Scharfetter-Gummel discretization scheme [38]. The value of the parameter  $\delta$  is  $\delta = 1/w - 1/(\exp(w) - 1)$ , where  $w = hB/C$ , which can be shown to be monotonically decreasing from 1 to 0 as  $w$  goes from  $-\infty$  to  $\infty$ . Notice that with the choice of  $\delta$  given above, the numerical scheme shares the same properties of the continuous FP equation that guarantee positiveness and conservativeness. This is a special case of the CC scheme because in the general one, the functions  $B$  and  $C$  may depends on  $(x, t)$ , hence also  $\delta$  may depend on  $(x, t)$ , too. Both the CC scheme [33] and the mid-point are second order accurate, then a second order numerical scheme results. For vanishing diffusion coefficient the differential operator of the Kolmogorov equation becomes hyperbolic. As can be seen by a straight calculation of the matrix coefficients of Equation (3.7), in the limit they correspond to the first order upwind scheme

that is the base scheme used for hyperbolic PDE. Hence, provided that the algorithm deals correctly with the vanishing diffusion coefficient, the numerical scheme continues working, although with a lower convergence order.

Let  $f^m = (f_1^m, \dots, f_N^m)^\dagger$  be the discrete solution at the time  $t_m$ , with  $f_0^m$  omitted due to periodicity, and  $\beta = C/h - \delta B$ . The action of the finite difference operator for  $F^m$  in Equation (3.2) reads as matrix  $A$  whose elements are defined by

$$\begin{aligned} A_{i,i} &= -\beta(1 + \omega)/h, & A_{i,i-1} &= \beta/h, & A_{i,i+1} &= \omega\beta/h, \\ A_{1,N} &= \beta/h, & A_{N,1} &= \omega\beta/h, \end{aligned} \tag{3.7}$$

where  $\beta = B/(\omega - 1)$ ,  $\omega = \exp(hB/C)$ . Hence,  $Af^m := (F_{i+1/2}^m - F_{i-1/2}^m)/h$ , and then the Equation (3.4) can be written in matrix form, as follows

$$Mf^{m+1} = 4f^m - f^{m-1} + 2\delta t Q(f^m; \bar{\alpha}), \tag{3.8}$$

where  $M := 3I - 2\delta t A$  is the matrix coefficients related to Equations (3.4) and (3.6). We note that this method needs of a second starting point, that can be calculated by using a first order Euler scheme with a smaller time step size than  $\delta t$ . The implicit Euler scheme for the Equations (3.1) and (3.6) is

$$(I - \delta t A)f^m = f^{m-1} + \delta t Q(f^{m-1}; \bar{\alpha}). \tag{3.9}$$

These two numerical schemes own some properties that can be easily proved, but we list here as remarks.

*Remark 3.* The Euler-CC scheme (3.9) to Equations (3.1) and (3.6), defined in the periodic domain  $\Omega_h$ , is conservative.

In fact,  $\sum_{i=1}^N A_{i,j} = 0, \forall j$ , and  $\sum_{i=1}^N Q(f_i^m, \bar{\alpha}) = 0$  because the set of values of  $f_i^m$  are the same as  $\hat{f}_{ik}^m$ , being the last only translated by  $k$ . Hence,  $\sum_{i=1}^N f_i^m = \sum_{i=1}^N f_i^{m-1}$ .

*Remark 4.* Provided that  $\sum_{i=1}^N f_i^m = \sum_{i=1}^N f_i^{m-1}$ , then the BDF2-CC scheme (3.8) to Equations (3.1) and (3.6) defined on the periodic domain  $\Omega_h$ , is conservative. In fact for the same constraints on  $A$  and  $Q$  as above, we get the identity  $3 \sum_{i=1}^N f_i^{m+1} = 4 \sum_{i=1}^N f_i^m - \sum_{i=1}^N f_i^{m-1} = 3 \sum_{i=1}^N f_i^m$ .

The positive preserving property of the numerical scheme is proved by using the theorem for the class of  $M$ -matrix [35]. Given a positive matrix  $E, E_{ij} \geq 0$ , we say that  $M = sI - E$  is a non singular  $M$ -matrix if  $s > \rho(E)$ , where  $\rho(E)$  is the spectral radius of  $E$ . A non singular  $M$ -matrix has the important property:  $M$  is non singular  $M$ -matrix  $\Rightarrow M^{-1} \geq 0$ .

**Theorem 1.** *Let  $\delta t \leq 1/a$ , with  $a$  defined in (3.3), then the Euler scheme (3.9) to Equation (3.1) defined in the periodic domain  $\Omega_h$ , is positive preserving.*

*Proof.* The argument is as follows: let  $R$  the matrix operator such that  $Rf^m = h \sum_{j=1}^N \alpha_j \sum_{k=1}^N \theta_{jk} \hat{f}_{ik}^m$ . Such a matrix is non negative because  $\alpha_j$  and  $\theta_{jk}$  are. The numerical scheme (3.9) can be recast as

$$\left( \left( 1 + \frac{\delta t \beta}{h} (1 + \omega) \right) I - \delta t \tilde{A} \right) f^m = (1 - a \delta t) f^{m-1} + \delta t R f^{m-1}, \tag{3.10}$$

where  $\tilde{A} = A - \text{diag}(A)$  is a positive matrix. Provided that  $f^{m-1} \geq 0$  and  $\delta t \leq 1/a$  the r.h.s. is a non negative vector. We observe that the matrix on the l.h.s is always diagonal dominant, hence it has a convergent regular splitting and consequently is an  $M$ -matrix [35]. Therefore,  $(I - \delta tA)^{-1}$  is non negative and  $f^m$  will be too.  $\square$

In order to prove the positive preserving property of the BDF2 numerical scheme (3.8), we need of the following Lemma that provides a lower bound to the velocity of decreasing of the solution.

**Lemma 1.** *Let the vector  $f^m \in \mathbb{R}^N$  be given non negative. Take a number  $\xi > 1$ , then the solution  $f^{m+q}$  calculated with the Euler scheme of Equation (3.9) after  $q$  time steps satisfies the following inequality*

$$f^{m+q} \geq f^m / \xi^q,$$

provided that  $\delta t < \frac{\xi - 1}{a\xi + \beta(1 + \omega)/h}$ , with parameters defined in Equations (3.7) and (3.3).

*Proof.* A proof is given for a particular case in [33] (see also References therein). Here we prove it as follows. Given  $f^m$  and  $f^{m+1}$  calculated with (3.9), let define  $v = \xi f^{m+1} - f^m$ . By applying the operator  $I - \delta tA$ , we get  $(I - \delta tA)v = (\xi I - (I - \delta tA))f^m + \delta t\xi Q(f^m; \bar{\alpha})$ , i.e.

$$(I - \delta tA)v = ((\xi - 1 - \delta t(\beta(1 + \omega)/h + a\xi))I + \delta t\tilde{A} + \delta t\xi R)f^m,$$

where  $\tilde{A} = A - \text{diag}(A)$  is a positive matrix. Now provided the bound for  $\delta t$ , then the r.h.s. is positive and from Theorem 1 we get that  $v \geq 0$ . By iterating that inequality  $q$  times, we get the thesis.  $\square$

*Remark 5.* The upper bound on  $\delta t$  in Lemma 1 results to be  $\delta t < 1/a$  for  $\xi > 1$ , hence the condition on the Lemma is stricter than those on non negativity of Theorem 1.

Now we show a Lemma similar to Lemma 1 valid for the BDF2 scheme.

**Lemma 2.** *Let  $1 < \xi < 3$  and  $2\delta t \leq h(\xi - 1)(3 - \xi)/(a\xi h + \beta(1 + \omega))$  be the time step size of the numerical scheme of Equation (3.8) that generates the sequence of vectors  $f^m$  for  $m = 2, 3, \dots$  from the starting vectors  $f^0, f^1$ . If there exists  $m^*$  such that  $\xi f^{m^*+1} - f^{m^*} \geq 0$  and  $f^{m^*} \geq 0$ , then  $\xi f^{m+1} - f^m \geq 0$  for all  $m > m^*$ .*

*Proof.* We apply the operator  $(3I - 2\delta tA)$  to  $v = \xi f^{m+2} - f^{m+1}$ ,

$$(3I - 2\delta tA)v = \xi(3I - 2\delta tA)f^{m+2} - (3I - 2\delta tA)f^{m+1}$$

and use Equation (3.8) to the first term on the r.h.s. to get

$$(3I - 2\delta tA)v = [4\xi - 3 - 2\delta t(a\xi + \beta(1 + \omega)/h)]f^{m+1} - \xi f^m + 2\delta t(\tilde{A} + \xi R)f^{m+1},$$

where  $\tilde{A} = A - \text{diag}(A)$  is a positive matrix. We know that  $(3I - 2\delta tA)$  is an  $M$ -matrix and its inverse is always non-negative. Also  $\tilde{A} + \xi R$  is non negative. Hence, we can prove non negativity of  $v$ , provided that

$$4\xi - 3 - 2\delta t(a\xi + \beta(1 + \omega)/h) \geq \xi^2$$

for a value  $m = m^*$ , because of the hypothesis  $\xi f^{m^*+1} - f^{m^*} \geq 0$ , that also states that  $f^{m^*+1} \geq 0$ . The last inequality is just the bound on  $\delta t$  in the assertion that gives a positive value for  $\delta t$  only when  $1 < \xi < 3$ .  $\square$

Indeed, this Lemma proves positivity of the numerical solution of Equation (3.8), provided that  $f^0 \geq 0$ , and  $\xi f^1 - f^0 \geq 0$ .  $f^1$  is the second starting value of the numerical scheme, that can be calculated with the Euler scheme (3.9).

**Theorem 2.** *Let  $f^0 \geq 0$  the discrete initial condition (3.5), and let  $f^1$  the second starting value calculated with the Euler scheme (3.9) with an appropriate time step, such that  $\xi f^1 - f^0 \geq 0$ , for  $1 < \xi < 3$ . Then, the BDF2 scheme (3.8) to Equation (3.1), defined in the periodic domain  $\Omega_h$ , is positive preserving for the solution  $f^m$ , with  $m > 1$ .*

*Proof.* The proof is an application of the Lemmas 1 and 2.  $\square$

In order to establish the stability of the discrete numerical schemes of Equations (3.8) and (3.9), we need inequalities of the form  $\|f^{m+1}\| \leq K\|f^m\|$  evaluated in a suitable norm with  $K$  possibly less or equal than 1. We prove that it realizes for the 1-norm with  $K = 1$ .

**Theorem 3.** *Let the positivity condition of Theorem 1 be fulfilled, i.e.  $\delta t \leq 1/a$ . Then, the Euler scheme (3.9) is stable in the 1-norm, that is  $\|f^m\|_1 \leq \|f^{m-1}\|_1$  for all  $m$ .*

*Proof.* Let  $r = \delta t\beta/h$  and invert the matrix operator at l.h.s., then Equation (3.10) reads as

$$f^m = \frac{\left(I - \frac{\delta t\tilde{A}}{1 + r(1 + \omega)}\right)^{-1}}{1 + r(1 + \omega)} [(1 - a\delta t)f^{m-1} + \delta tRf^{m-1}].$$

Now we observe that

$$\left\| \left(I - \frac{\delta t\tilde{A}}{1 + r(1 + \omega)}\right)^{-1} \right\|_1 \leq \left(1 - \left\| \frac{\delta t\tilde{A}}{1 + r(1 + \omega)} \right\|_1\right)^{-1} = 1 + r(1 + \omega).$$

Hence,

$$\|f^m\|_1 \leq \|(1 - a\delta t)f^{m-1} + \delta tRf^{m-1}\|_1.$$

Since  $\delta t \leq 1/a$ , all the components of the vectors inside the norm at the r.h.s. are positive, so that the modulus for the evaluation of the 1-norm can be removed. Using  $\sum_{i=1}^N Q(f_i^m, \bar{\alpha}) = 0$  as in Remark 3, we get the statement of the theorem.  $\square$

Now we can prove the stability of the numerical scheme with BDF2 integration of Equation (3.8).

**Theorem 4.** *Let the positivity condition of the Theorem 2 be fulfilled, i.e. let  $\delta t$  be the time step size of the numerical scheme of Equation (3.8),  $f^0 \geq 0$  the discrete initial condition (3.5) and  $f^1$  the second starting value evaluated at the time  $\delta t$ . If there exists a real number  $\xi$  such that  $f^1 \geq f^0/\xi$  with  $1 < \xi < 3$  and  $2\delta t \leq h(\xi - 1)(3 - \xi)/(a\xi h + \beta(1 + \omega))$ , then the BDF2 scheme (3.8) is stable in the 1-norm, that is  $\|f^m\|_1 \leq \|f^{m-1}\|_1$  for all  $m$ .*

*Proof.* The numerical scheme (3.8) can be written as

$$M f^{m+1} = (4 - \zeta) f^m + (\zeta - 2a\delta t) f^m - f^{m-1} + 2\delta t R f^m,$$

where  $R$  is defined as in Theorem 1. We apply  $M^{-1}$  and evaluate the 1-norm to both sides. Following the same calculations as in Theorem 3, we get that  $\|M^{-1}\|_1 = 1/3$ .

From the bound on  $\delta t$ , we note that

$$2a\delta t < (\xi - 1)(3 - \xi)/\xi \leq 4 - 2\sqrt{3} < 0.536.$$

This means that for all  $\zeta$  in the interval  $5 - 2\sqrt{3} < \zeta < 3$ , it is  $\zeta - 2a\delta t = \xi$  with  $\xi \in (1, 3)$ . Now we have that  $f^m \geq 0$  by virtue of the positivity condition,  $(\zeta - 2a\delta t) f^m - f^{m-1} \geq 0$  by our assumptions, and  $4 - \zeta > 0$ , hence is guaranteed that the sum in the r.h.s. is a non negative vector and the modulus in the calculation of the 1-norm can be removed. Using the property given in Remark 3, we conclude that  $\|f^{m+1}\|_1 \leq \|f^m\|_1$ .  $\square$

*Remark 6.* Indeed, in the stability Theorem 4 the equality  $\|f^{m+1}\|_1 = \|f^m\|_1$  holds. Because of the conservativeness from Remark 4 we have  $\sum_{i=1} f_i^{m+1} = \sum_{i=1} f_i^m$ , and under the non negativity condition of Theorem 2 all the components of the vectors  $f_i^{m+1}, f_i^m$  are non negative, so that the previous conservativeness identity corresponds to the 1-norm equivalence. Further, we can state that for these numerical schemes the conservativeness and the non negativity imply the stability of the discrete operator.

*Remark 7.* We can finally conclude from the Lax equivalence theorem, that for regular solutions of the Kolmogorov forward equation  $f(x, t), (x, t) \in [\Omega, T]$ , provided that the hypothesis of Theorem 4 are fulfilled, the numerical scheme (3.8) yields discrete solutions that are second order convergent in time and space.

The discrete adjoint equation can be found by discretizing the Lagrangian function (2.4) and then performing the variations on the discrete variables. This is know as the *discretize-then-optimize* approach (see [4] for details). This technique yields the following discrete adjoint equation

$$M^\dagger p^m = 4p^{m+1} - p^{m+2} + 2\delta t \tilde{Q}(p^{m+1}; \bar{\alpha}), \tag{3.11}$$

where  $M^\dagger$  is the transpose of  $M$ , and  $\tilde{Q}(p^{m+1}; \bar{\alpha}) = h \sum_{j=1}^{N_e} \alpha_j \sum_{k=1}^N \tilde{p}_{ik}^m \theta_{jk} - a p_i^m$ , with  $\tilde{p}_{ik}^m \approx (p(x_i - s_k, t_m))_{\Omega_h}$ .

The numerical stability is given by the same condition for the forward equation, since the transpose of the operator  $M$  has the same eigenvalues, but in this case the non negativity and conservativeness properties are not required.

Care has to be taken for the discrete terminal condition, since it can not be defined through the Equation (2.6) for the presence of the  $\delta$ -Dirac measure. For this purpose we discretize the term (2.5) as follows

$$\begin{aligned} & \frac{1}{L} \sum_{l=1}^L \sum_{i=1}^N \int_{x_i-1/2}^{x_i+1/2} \log(f(x, T)) \delta(x - X_l) dx \\ &= \frac{1}{L} \sum_{l=1}^L \sum_{i=1}^N \log(f(\hat{x}_i, T)) 1_{\{X_l \in [x_i-1/2, x_i+1/2)\}}, \end{aligned}$$

where  $\hat{x}_i$  are the points of the integral average theorem. Then we use the approximation  $f(\hat{x}_i, T) \approx f_i^{N_T}$ , so that, by performing the variation  $\delta f_i^{N_T}$  on this discrete functional, we get the discrete terminal condition

$$p_i^{N_T} = p_{T,i} = -\frac{1}{L} \sum_{l=1}^L 1_{\{X_l \in [x_i-1/2, x_i+1/2)\}} / f_i^{N_T}, \quad i = 1, \dots, N. \quad (3.12)$$

According to (2.6) it completes the formulation of the discrete adjoint problem.

The discrete approximation of the reduced gradient related to the optimality condition (2.7) is calculated with the mid-point quadrature formula. Each component  $j$  is given by

$$(D_{\bar{\alpha}} \hat{J})_j := -\delta t h^2 \sum_{m=0}^{N_T} \sum_{i=1}^N \sum_{k=1}^N (\hat{f}_{ik}^m - f_i^m) p_i^m \theta_{jk}, \quad (3.13)$$

where  $(D_{\bar{\alpha}} \hat{J})_j \approx (\nabla_{\bar{\alpha}} \hat{J})_j$ , for  $j = 1, \dots, N_{\Theta}$ .

The availability of the discrete gradient allows us to implement a non linear conjugate gradient scheme (NLCG) in order to solve the optimization problem (2.3). The NLCG represents an extension of the linear conjugate gradient method to non-quadratic problems [4, 26, 39].

The optimality system is solved by implementing the discrete gradient evaluation given by the following algorithm

*Algorithm 1* [Evaluation of the Gradient at  $\bar{\alpha}$ ].

- Input: model parameters, space step  $h$ , time step  $\delta t$  according to the restriction of Theorem 4.
  1. Solve the discrete FP equation (3.4) with given initial condition (3.5);
  2. Solve the discrete adjoint FP equation (3.11) with terminal condition (3.12);
  3. Compute the approximated discrete gradient  $D_{\bar{\alpha}} \hat{J}$  by using (3.13);
  4. End.

in a NLCG scheme. This is based on search directions  $d_k$  of the minimum that are recursively computed as

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{3.14}$$

where  $k = 0, 1, 2, \dots$  in this paragraph stands for the iteration index,  $g_k = D_{\bar{\alpha}} \hat{J}(\bar{\alpha}_k)$  is the numerical gradient, with  $d_0 = -g_0$ . Let  $\bar{\alpha}_k$  an estimation of the best rates at the iteration  $k$ , the next one for a minimum point are given by

$$\bar{\alpha}_{k+1} = \bar{\alpha}_k + \xi_k d_k, \tag{3.15}$$

where here  $\xi_k > 0$  is a steplength obtained with a line-search that satisfies the Armijo condition of sufficient decrease of  $\hat{J}$ 's value as follows

$$\hat{J}(\bar{\alpha}_k + \xi_k d_k) \leq \hat{J}(\bar{\alpha}_k) + \delta \xi_k (\nabla \hat{J}(\bar{\alpha}_k), d_k)_U, \tag{3.16}$$

where  $0 < \delta < 1/2$ ; see [34]. Notice that we use the inner product of the  $U \equiv \mathbb{R}^{N\theta}$  space. For  $\beta_k$  we use the Dai and Yuan formulation [19]

$$\beta_k^{DY} = (g_{k+1}, g_{k+1})_U / (d_k, y_k)_U, \tag{3.17}$$

where  $y_k = g_{k+1} - g_k$ .

Summarizing, the NLCG scheme is implemented as follows

*Algorithm 2* [NLCG Scheme].

- Input: initial approx.  $\bar{\alpha}_0$ , maximum  $\bar{\alpha}_M$ ,  $g_0 = \nabla \hat{J}(\bar{\alpha}_0)$ ,  $d_0 = -g_0$ , index  $k = 0$ , maximum  $k_{max}$ , tolerance  $tol$ .
  1. While ( $k < k_{max}$  &&  $\|g_k\|_{\mathbb{R}^\ell} > tol$ ) do
  2. Search the steplength  $\xi_k > 0$ , by sequentially shrinking, along  $d_k$  satisfying (3.16); if it can not be found, then the algorithm end.
  3. Set  $\bar{\alpha}_{k+1} = \bar{\alpha}_k + \xi_k d_k$ . i.e. Equation (3.15), according to the KKT condition, the eventually negative components of  $\bar{\alpha}_{k+1}$  are set to 0. Components  $\bar{\alpha}_{k+1} > \bar{\alpha}_M$  are set to  $\bar{\alpha}_M$ .
  4. Compute  $g_{k+1} = \nabla \hat{J}(\bar{\alpha}_{k+1})$  using Algorithm 1;
  5. Compute  $\beta_k^{DY}$  given by (3.17);
  6. Let  $d_{k+1} = -g_{k+1} + \beta_k^{DY} d_k$ , i.e. Equation (3.14);
  7. Set  $k = k + 1$ ;
  8. End while

The numerical evaluation of the functional (2.1) has the problem of the logarithm in the points  $X_l$  where the PDF at the final time has vanishing values. Hence, the functional is replaced as follows

$$J_\epsilon(f, \bar{\alpha}) = \frac{1}{L} \sum_{l=1}^L \log(\max(\epsilon, f(X_l, T, \bar{\alpha}))),$$



with  $\epsilon = 10^{-12}$ .

The discrete PDF is defined on the mesh grid  $\Omega_h$ , the sample values  $X_l$  used for the evaluation of the PDF are approximated to the nearest values of the space mesh grid  $\Omega_h$ . This approximation affects both the value of the functional and the terminal condition for the adjoint equation.

The initial PDF  $f_0(x)$  of Equation (2.2) is set as the following von Mises density distribution

$$\rho(x; \mu, \kappa) = \frac{e^{\kappa \cos(2\pi(x-\mu-\Omega_a)/(\Omega_b-\Omega_a)-\pi)}}{2\pi I_0(\kappa)},$$

where  $I_0(\cdot)$  is the modified Bessel function of order 0, and  $\kappa$  is the concentration parameter that should be taken large in order to approximate the  $\delta$ -Dirac function in zero as initial data for the forward PIDE.

The computational time cost of the Algorithm 1 scales as  $N^4$ . For a  $d$ -dimensional Lévy process, it would be  $N^{2(d+1)}$ . This is due to the evaluation of the numerical gradient (3.13). The calculation of the NLCG algorithm consists of many repetitions of the Algorithm 1. An estimation of the computational cost is difficult since it depends, e.g., on the initial search point  $\bar{\alpha}_0$  and on parameters of the line search of the Armijo condition. Further, also the choice of the formula for  $\beta_k$  affects the efficiency of the optimization algorithm. Anyway, in general, NLCG has linear convergence order [18]. We can roughly assume that the lower bound of the computational cost scales at least as  $N^{2d+3}$ . For high dimensional problems the computation requires high performance facilities, indeed for the following tests, executed on an Intel® Core™ Duo CPU, the time for computation took only some minutes.

## 4 Numerical tests

In this section we perform the non parametric estimation of Lévy density distribution functions, that is to find the value  $\bar{\alpha} = (\alpha_1, \dots, \alpha_{N_\Theta})$  such that best fits with the given data  $\{X_l\}$ . We present two validation test cases and one application case to finance.

However note that in our case, cyclic boundary conditions have to be taken into account. The data setting for our test case is as follows: the space domain  $\Omega = [-\pi, \pi)$ , the final time  $T = 1$ , the initial von Mises distribution has center  $\mu = 0$  and wideness  $\kappa = 400$ , the drift of the stochastic process is  $b = 0$  and the Gaussian volatility is  $\sigma = \sqrt{0.02}$ . The setting for the numerical solution is: space grid size  $N = 420$ , time grid size  $N_T = 250$ .

First, we perform a fit for a set of  $L = 10^5$  values generated by a Monte Carlo algorithm for a simulated Lévy process on the torus, with the following five values of the jump rates:  $\hat{\alpha} = \{3, 2, 1, 0.5, 0.25\}$ . We solve the fitting problem, i.e. calculating the estimates to  $\alpha_1, \dots, \alpha_{N_\Theta}$ , for different numbers of interpolatory functions:  $N_\Theta = 3, \dots, 7$ . The initial approximation is set to  $\bar{\alpha}_0 = \{1, \dots, 1\}$ . The center  $\theta_1, \dots, \theta_{N_\Theta}$  of the basis functions  $\Theta_j(x)$  are equally spaced in the domain  $(-1, 1)$  at the places  $\theta_j = -1 + j\Delta$ ,  $j = 1, \dots, N_\Theta$ ,  $\Delta = 2/(N_\Theta + 1)$ , this means the basis functions do not cover all the domain

$\Omega$ . In the following table the calculated value of  $\{\alpha_j\}$  for each problem are reported versus  $N_\Theta$

	$N_\Theta = 3$	$N_\Theta = 4$	$N_\Theta = 5$	$N_\Theta = 6$	$N_\Theta = 7$
$\alpha_1$	3.4502	3.1771	2.9746	2.8580	2.8452
$\alpha_2$	1.1089	1.5577	1.8100	1.9922	2.1003
$\alpha_3$	0.4505	0.6576	1.0198	1.3025	1.5083
$\alpha_4$		0.3362	0.4951	0.7607	1.0077
$\alpha_5$			0.2490	0.3946	0.6137
$\alpha_6$				0.2042	0.3428
$\alpha_7$					0.1847

We see the good match for  $N_\Theta = 5$  with the original rates  $\hat{\alpha}$ . In Figures 1, 2 and 3 we can also appreciate the good data fitting between the calculated PDF and the histograms of the simulated Monte Carlo data, for the proposed optimization problem with  $N_\Theta = 3, 5, 6$ .

Another interesting problem is the selection of the number of parameters  $N_\Theta$  and the corresponding basis functions  $\Theta_j$  for the best data fit. In Figure 4 we depict the result of the Akaike’s Information Criterion (AIC) [12], given by

$$AIC(N_\Theta) = LJ(f, \bar{\alpha}^*) - \log(N_\Theta).$$

A common choice in statistics is to pick that parametrization that maximises the AIC. Other choices such as Ridge regression or Lasso methods could not be so efficient when applied to our optimization technique. The first for its inability to automatically select an adequate number of parameters, the second for difficulties in the adjoint method raising from the non-differentiability of the regularisation term. We can see that criterion gives the value  $N_{\Theta,opt} = 6$ , while the correct value is 5. The difference in the AIC is however rather small for  $N_\Theta$  between 5 and 7.

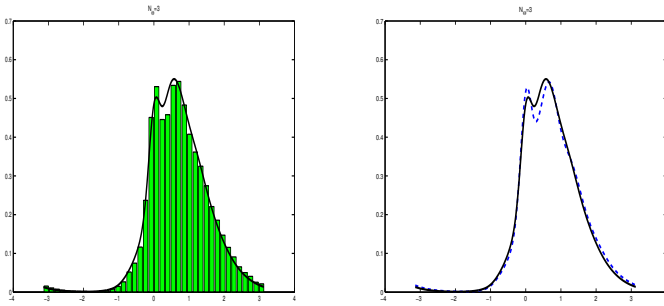
In the second test we fit the final position at  $T = 1$  of  $10^5$  samples of a stochastic process with the jumps distributed according to a bi-directional gamma process with Lévy measure  $\nu$  on  $\mathbb{R}$  given by the density [6]

$$d\nu(s) = A \frac{e^{-\beta|s|}}{|s|} ds.$$

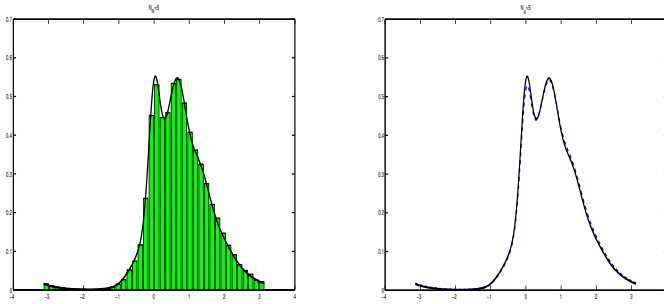
Here  $A > 0$  is the so-called *shape* parameter and  $\beta$  is the *rate* parameter. Note that this is not a finite measure, so we are out of the compound Poisson class, and the trajectory of the bi-directional gamma process as infinitely many (small) jumps. If we project this Lévy process to the torus  $[-\pi, \pi]$ , the effect on the projected Lévy measure  $\nu$  is

$$d\nu(s) = A \left( \frac{e^{-\beta|s|}}{|s|} + 2 \sum_{n=1}^{\infty} \frac{s \sinh(\beta s) + 2n\pi \cosh(\beta s)}{4n^2\pi^2 - s^2} e^{-2n\pi} \right) ds$$

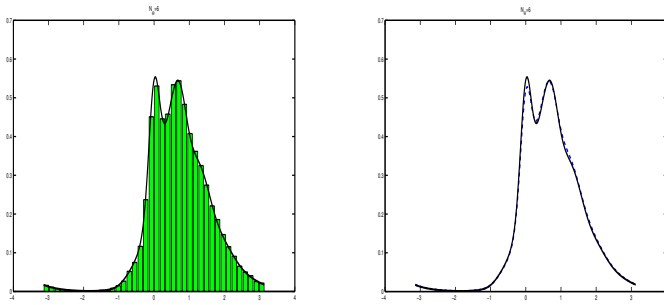
as the Lévy measure on the torus for the projected bi-directional Gamma process.



**Figure 1.** Left, result of the data fitting with  $N_{\Theta} = 3$  rates. Histograms: experimental Lévy data collected in 40 bins. Solid line calculated PDF. Right, dashed line calculated PDF with the original 5 rates.



**Figure 2.** Left, result of the data fitting with  $N_{\Theta} = 5$  rates. Histograms: experimental Lévy data collected in 40 bins. Solid line calculated PDF. Right, dashed line calculated PDF with the original 5 rates.



**Figure 3.** Left, result of the data fitting with  $N_{\Theta} = 6$  rates. Histograms: experimental Lévy data collected in 40 bins. Solid line calculated PDF. Right, dashed line calculated PDF with the original 5 rates.

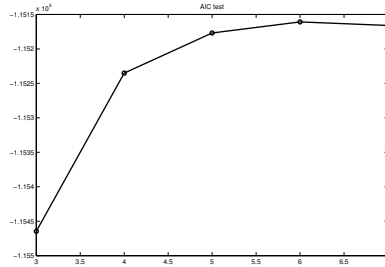


Figure 4. Test for the appropriate regularisation with Akaike’s Information Criterion

In the simulations data are generated by scaling the rate parameter such that  $1/\beta = 1$ , and setting the shape to  $A = 0.5$ . Finally, the data  $\{Y_l\}$  are projected to the torus  $[-\pi, \pi)$ . For this test the starting values of the unknowns are set to  $\bar{\alpha}_0 = \{0.1, \dots, 0.1\}$ . The drift and the Gaussian volatility have the same values of the previous test. In Figure 5 we report the result of the AIC test and the fit with  $N_\Theta = 9$  basis functions centered to  $\theta_j = (-1 + 2(j - 1)/9)\pi, j = 1, \dots, 9$ . We conclude that our procedure results in high quality fits, even for Lévy distributions that can only be approximated by the hierarchy of parametrizations.

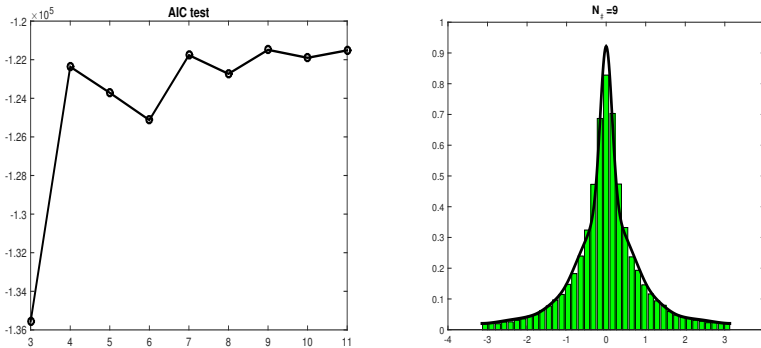
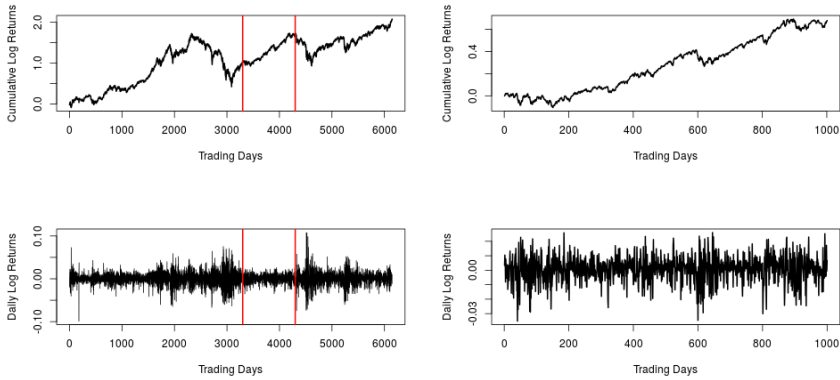


Figure 5. Left, AIC test for the Gamma test. Right, fit with  $N_\Theta = 9$  basis functions.

As an example from a real world problem we report the result of fitting of the German Stock Exchange (DAX) index see Figure 6. Within the data of all closing quotes between April 1998 and March 2015, there are several periods of volatility bursts. In order to avoid the pitfalls of time dependent (or stochastic [15, 29]) volatility, we identify a period of comparatively stable volatility of 1000 trading days between April 1998 and February 2002, see the right panel of Figure 6. The obvious absence of axial symmetry prohibits a Gaussian (Black-Scholes) market model from the outset. Our goal is to find a suitable description of this sample with an exponential Lévy market model from our hierarchy of parametrizations. The data has been mapped to the

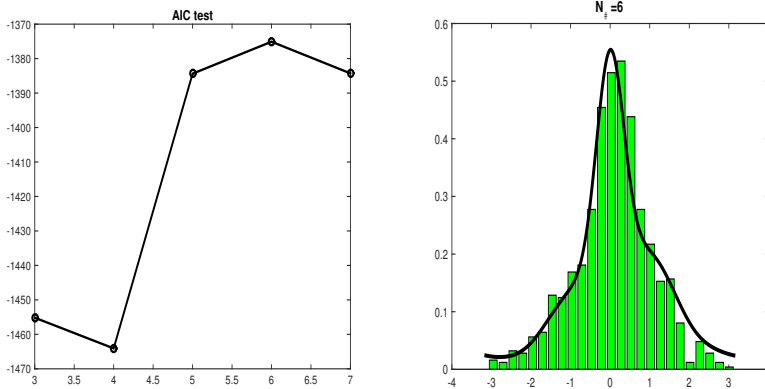


**Figure 6.** Cumulative and daily log-returns of the DAX index between 26/11/1990 and 05/03/2015 (left panel) 16/4/1998 and 4/02/2002 (right panel). The vertical lines on the left panel correspond to the time period on the right panel, downloaded from [finance.yahoo.com](http://finance.yahoo.com)

$[-\pi, \pi]$  torus, by rescaling and “wrapping” the daily log-returns below/above 3%, i.e.  $[\Omega_a, \Omega_b] = [-0.03, 0.03]$ . Three data sets, all of them negative, were situated outside this band. We calculated the fitting of the distribution, with equally spaced basis functions in the interval  $[-\pi, \pi)$ . The starting values of the unknown rates are the same of the previous test. The drift and Gaussian volatility are set  $b = 6.787 \cdot 10^{-4}$  and  $\sigma = 0.3444$ . In Figure 7 (left panel) we report the result of the AIC test and the fit with  $N_\Theta = 6$  (right panel). The selected basis functions are centred at  $\theta_j = (-1 + (j-1)/3)\pi, j = 1, \dots, 6$ , whose the calculated rates are  $\bar{\alpha} = (0, 0, 0.484, 0.223, 0.304, 0)$ . Although only three parameters are different from zero, the AIC is maximized at  $N_\Theta = 6$ . That the AIC at  $N_\Theta = 3$  is lower is explained by the fact, that the more localized basis functions in the  $N_\Theta = 6$ -basis are more adequate to fit the data. It is also a misinterpretation that the chosen parametrization misses an effective description with three parameters, since the position of the grid points are additional parameters. Note that the zero entries of the 1st, 2nd and 6th slot  $\bar{\alpha}$  actually correspond to small positive values and only represented as zero when rounded to the 3rd digit.

## 5 Conclusions

We have shown that our non-parametric method for calibration, grounded on a Fokker-Planck-Kolmogorov constrained optimization framework, works for spline discretizations of the density of the Lévy measure with symmetric boundary conditions for up to 11 parameters. The results consistently fit simulated data from the family of discretizations itself. The same turns out to be true from Lévy processes that only can be approximated by such discretizations, if the number of parameters goes to infinity, like the gamma process. Here the AIC provides an effective mechanism to choose an adequate discretization at a



**Figure 7.** Left, AIC for the maximum likelihood estimate as a function of parameters for the DAX data (left). Right density for the maximum-likelihood fit with  $N_{\Theta} = 6$  basis functions, corresponding to the maximal AIC.

given sample size. Finally, we have demonstrated that also real-world, financial data can be effectively fitted using our strategy.

### Acknowledgement

We would like to thank Professor Alfio Borzì for interesting discussions and hospitality at the University of Würzburg.

### References

- [1] Y. Aït-Sahalia and A. W. Lo. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, **53**(2):499–547, 1998. <https://doi.org/10.1111/0022-1082.215228>.
- [2] D.N. Allen and R.V. Southwell. Relaxation methods applied to determine the motion, in 2-D, of a viscous fluid past a fixed cylinder. *Quart.-J. Mech. Appl.*, **VIII**(2):129–145, 1955. <https://doi.org/10.1093/qjmam/8.2.129>.
- [3] M. Annunziato and A. Borzì. Optimal control of probability density functions of stochastic processes. *Mathematical Modelling and Analysis*, **15**(4):393–407, 2010. <https://doi.org/10.3846/1392-6292.2010.15.393-407>.
- [4] M. Annunziato and A. Borzì. A Fokker-Planck control framework for multidimensional stochastic processes. *Journal of Computational and Applied Mathematics*, **237**(1):487–507, 2013. ISSN 0377-0427. <https://doi.org/10.1016/j.cam.2012.06.019>.
- [5] M. Annunziato, A. Borzì, F. Nobile and R. Tempone. On the connection between the Hamilton-Jacobi-Bellman and the Fokker-Planck control frameworks. *Appl. Mathematics*, **5**:2476–2484, 2014. <https://doi.org/10.4236/am.2014.516239>.
- [6] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2 edition, 2009. <https://doi.org/10.1017/CBO9780511809781>.

- [7] D. Belomestny and M. Reiß. Spectral calibration of exponential Lévy models. *Finance Stoch.*, **10**:449–474, 2006. <https://doi.org/10.1007/s00780-006-0021-5>.
- [8] D. Belomestny and J. Schoenmakers. A jump-diffusion Libor model and its robust calibration. *Quantitative Finance*, **11**(4):529–546, 2011. <https://doi.org/10.1080/14697680903295176>.
- [9] C. Berg and G. Forst. *Potential theory on locally compact Abelian groups*, volume 87. Springer, Berlin - Heidelberg - New York, 1975. <https://doi.org/10.1007/978-3-642-66128-0>.
- [10] A. Borzì and V. Schulz. *Computational Optimization of Systems Governed by Partial Differential Equations*. SIAM, Philadelphia, 2012. <https://doi.org/10.1137/1.9781611972054>.
- [11] M. Briani, R. Natalini and G. Russo. Implicit–explicit numerical schemes for jump–diffusion processes. *Calcolo*, **44**(1):33–57, 2007. <https://doi.org/10.1007/s10092-007-0128-x>.
- [12] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, 2002. <https://doi.org/10.1007/b97636>.
- [13] J.S. Chang and G. Cooper. A practical difference scheme for Fokker-Planck equations. *J. Comput. Phys.*, **6**:1–16, 1970. [https://doi.org/10.1016/0021-9991\(70\)90001-X](https://doi.org/10.1016/0021-9991(70)90001-X).
- [14] F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'Institut Henri Poincaré*, **46**(3), 2010. <https://doi.org/10.1214/09-AIHP323>.
- [15] R. Cont and P. Tankov. *Financial Modeling with Jump Processes*. Chapman & Hall, Boca Raton - London - New York - Washington D.C., 2004.
- [16] R. Cont and E. Voltchkova. A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. *SIAM Journal on Numerical Analysis*, **43**(4):1596–1626, 2005. <https://doi.org/10.1137/S0036142903436186>.
- [17] S. Csörgö and V. Totil. On how long intervals is the empirical characteristic function uniformly consistent? *Acta Sci. Math.*, **45**:141–149, 1983.
- [18] Y.-H. Dai. Nonlinear conjugate gradient methods. In *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., 2010. ISBN 9780470400531. <https://doi.org/10.1002/9780470400531.eorms0183>.
- [19] Y.-H. Dai and Y. Yuan. A nonlinear conjugate gradient with a strong global convergence property. *SIAM J. Optim.*, **10**:177–182, 1999. <https://doi.org/10.1137/S1052623497318992>.
- [20] D. J. Duffy. Numerical analysis of jump diffusion models: A partial differential equation approach. *Wilmott Magazine*, pp. 68–73, 2009.
- [21] T. S. Fergusson. *A Course in Large Sample Theory*. Chapman & Hall, Boca Raton - London - New York - Washington D.C., 1996.
- [22] T. L. Friesz. *Dynamic Optimization and Differential Games*. Springer, New York - Dordrecht - Heidelberg - London, 2010. <https://doi.org/10.1007/978-0-387-72778-3>.
- [23] B. Gaviraghi, M. Annunziato and A. Borzì. Analysis of splitting methods for solving a partial integro-differential Fokker-Planck equation. *Applied Mathematics and Computation*, **294**:1–17, 2017. <https://doi.org/10.1016/j.amc.2016.08.050>.

- [24] B. Gaviraghi, A. Schindele, M. Annunziato and A. Borzi. On optimal sparse-control problems governed by jump-diffusion processes. *Applied Mathematics*, **7**:1978–2004, 2016. <https://doi.org/10.4236/am.2016.716162>.
- [25] S. Geman and C.-H. Hwang. Non-parametric maximum likelihood estimation by the method of sieves. *Ann. of Statistics*, **10**(2):401–414, 1982. <https://doi.org/10.1214/aos/1176345782>.
- [26] J.C. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.*, **2**:21–42, 1992. <https://doi.org/10.1137/0802003>.
- [27] H. Gottschalk, B. Smii and H. Thaler. The Feynman graph representation for the convolution semigroup and its applications to Lévy statistics. *Bernoulli*, **14**(2):322–351, 2008. <https://doi.org/10.3150/07-BEJ106>.
- [28] Z. Grbac, A. Papapantoleon, J. Schoenmakers and D. Skovmand. Affine LIBOR models with multiple curves: Theory examples and calibration. *SIAM J. Finan. Math.*, **6**:984–1025, 2015. <https://doi.org/10.1137/15M1011731>.
- [29] S. Iacus. *Option Pricing and Estimation of Financial Models with R*. Wiley, Chichester, 2011. <https://doi.org/10.1002/9781119990079>.
- [30] J. Kappus and M. Reiß. Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Statistica Neerlandica*, **64**(3):314–328, 2010. <https://doi.org/10.1111/j.1467-9574.2010.00461.x>.
- [31] K. Knight. *Mathematical Statistics*. Chapman & Hall, 1999. <https://doi.org/10.1201/9781584888567>.
- [32] D. Marazzina, O. Reichmann and C. Schwab. hp-DGFEM for Kolmogorov-Fokker-Planck equations of multivariate Lévy processes. *Math. Mod. and Meth. in Appl. Sci.*, **22**:1150005–1–37, 2012.
- [33] M. Mohammadi and A. Borzi. Analysis of the Chang-Cooper discretization scheme for a class of Fokker-Planck equations. *Journal of Numerical Mathematics*, **23**:271–288, 2015. <https://doi.org/10.1515/jnma-2015-0018>.
- [34] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, 1999. <https://doi.org/10.1007/b98874>.
- [35] R.J. Plemmons. M-matrix characterizations. I-nonsingular M-matrices. *Linear Algebra and its Applications*, **18**(2):175–188, 1977. [https://doi.org/10.1016/0024-3795\(77\)90073-8](https://doi.org/10.1016/0024-3795(77)90073-8).
- [36] O. Reichmann and C. Schwab. Numerical analysis of additive, Lévy and Feller processes with applications to option pricing Lévy matters. In *Lévy Matters I: Recent Progress in Theory and Applications: Foundations, Trees and Numerical Issues in Finance, Lecture Notes in Mathematics 2001*, pp. 137–196. Springer-Verlag, Berlin - Heidelberg, 2010.
- [37] L. A. Sakhonovich. *Lévy Processes, Integral Equations, Statistical Physics: Connections and Interactions*. Birkhäuser, Basel, 2012. <https://doi.org/10.1007/978-3-0348-0356-4>.
- [38] D. L. Scharfetter and H. K. Gummel. Large signal analysis of a silicon Read diode oscillator. *IEEE Trans. Electron. Dev.*, **16**:64–77, 1969. <https://doi.org/10.1109/T-ED.1969.16566>.
- [39] D.F. Shanno. Conjugate gradient methods with inexact searches. *Math. Oper. Res.*, **3**:244–256, 1978. <https://doi.org/10.1287/moor.3.3.244>.