The Lévy-Khintchine type operators with variable Lipschitz continuous coefficients generate linear or nonlinear Markov processes and semigroups *

Vassili N. Kolokoltsov[†]

July 18, 2009

Abstract

Ito's construction of Markovian solutions to stochastic equations driven by a Lévy noise is extended to nonlinear distribution dependent integrands aiming at the effective construction of linear and nonlinear Markov semigroups and the corresponding processes with a given pseudo-differential generator. It is shown that a conditionally positive integro-differential operator (of the Lévy-Khintchine type) with variable coefficients (diffusion, drift and Lévy measure) depending Lipschitz continuously on its parameters (position and/or its distribution) generates a linear or nonlinear Markov semigroup, where the measures are metricized by the Wasserstein-Kantorovich metrics. This is a nontrivial but natural extension to general Markov processes of a long known fact for ordinary diffusions.

Key words. Stochastic equations driven by Lévy noise, nonlinear integrators, Wasserstein-Kantorovich metric, pseudo-differential generators, linear and nonlinear Markov semigroups.

Running Head: Lévy-Khintchine operators with Lipschitz coefficients

1 Introduction and formulation of main results

By $C(\mathbf{R}^n)$ (respectively $C_\infty(\mathbf{R}^n)$) we denote the Banach space of continuous bounded functions on $\mathbf{R}^n$ (respectively its subspace of functions vanishing at infinity) with the sup-norm denoted by $\|\cdot\|$, and $C^k(\mathbf{R}^n)$ (resp. $C^k_c(\mathbf{R}^n)$) denotes the Banach space of k times continuously differentiable functions with bounded derivatives on $\mathbf{R}^n$ (resp. its subspace of functions with a compact support) with the norm being the sum of the sup-norms of a function and all its partial derivative up to and including the order k).

For an $f\in C^1(\mathbf{R}^n)$ the gradient will be denoted by

$$\nabla f=(\nabla_1 f,...,\nabla_n f)=(\frac{\partial f}{\partial x_1},...,\frac{\partial f}{\partial x_1}).$$

*submitted to PTRF September 2008

†Department of Statistics, University of Warwick, Coventry CV4 7AL UK, Email: v.kolokoltsov@warwick.ac.uk

For a measure ν and a mapping F we denote by ν^F the push forward of ν with respect to F defined as $\nu^F(A) = \nu(F^{-1}(A))$.

Further basic notations: $\mathbf{1}_M$ is the indicator function of a set M, $\mathcal{M}(\mathbf{R}^d)$ is the set of finite positive Borel measures on $\mathbf{R}^d$, B_r is the ball of radius r centered at the origin, and the pairing (f, μ) for $f \in C(\mathbf{R}^d)$, $\mu \in \mathcal{M}(\mathbf{R}^d)$ denotes the usual integration. The bold letters $\mathbf{E}$ and $\mathbf{P}$ will denote expectation and probability. A positive number in the square bracket, say $[x]$, will denote the integer part of it. By the small letter c we shall denote various constants indicating in brackets (when appropriate) the parameters on which they depend.

It is well known (the Courrège theorem, see e.g. [7]) that the generator L of a conservative (i.e. preserving constants) Feller semigroup in $\mathbf{R}^d$ is conditionally positive ($f \geq 0, f(x) = 0 \implies Lf(x) \geq 0$) and if its domain contains the space $C_c^2(\mathbf{R}^d)$, then it has the following Lévy-Khintchine form with variable coefficients:

$$Lf(x) = \frac{1}{2}(G(x)\nabla, \nabla)f(x) + (b(x), \nabla f(x)) + \int (f(x+y) - f(x) - (\nabla f(x), y)\mathbf{1}_{B_1}(y))\nu(x, dy), \tag{1}$$

where $G(x)$ is a symmetric non-negative matrix and $\nu(x, .)$ a Borel measure on $\mathbf{R}^d$ (called Lévy measure) such that

$$\int_{\mathbf{R}^n} \min(1, |y|^2)\nu(x; dy) < \infty, \quad \nu(\{0\}) = 0. \tag{2}$$

The inverse question on whether a given operator of this form (or better to say its closure) actually generates a Feller semigroup is nontrivial and attracted lots of attention. One can distinguish analytic and probabilistic approaches to this problem. The existence results obtained by analytic techniques require certain non-degeneracy condition on ν, e.g. a lower bound for the symbol of pseudo-differential operator L (see e.g. [2], [3], [7]- [12] and references therein), and for the construction of the processes via usual stochastic calculus one needs to have a family of transformations F_x of $\mathbf{R}^d$ preserving the origin, regularly depending on x and pushing a certain Lévy measure ν to the Lévy measures $\nu(x, .)$, i.e. $\nu(x, .) = \nu^{F_x}$ (see e.g. [1], [4], [19]). Of course yet more nontrivial is the problem of constructing the so called nonlinear Markov semigroups solving the weak equations of the form

$$\frac{d}{dt}(f, \mu_t) = (L_{\mu_t} f, \mu_t), \quad \mu_t \in \mathcal{P}(\mathbf{R}^d), \quad \mu_0 = \mu, \tag{3}$$

that should hold, say, for all $f \in C_c^2(\mathbf{R}^d)$, where L_μ has form (1), but with all coefficients additionally depending on μ, i.e.

$$L_\mu f(x) = \frac{1}{2}(G(x, \mu)\nabla, \nabla)f(x) + (b(x, \mu), \nabla f(x))$$

$$+ \int (f(x+y) - f(x) - (\nabla f(x), y)\mathbf{1}_{B_1}(y))\nu(x, \mu, dy). \tag{4}$$

Equations of type (3) play indispensable role in the theory of interacting particles (mean field approximation) and exhaust all positivity preserving evolutions on measures subject to certain mild regularity assumptions (see e.g. [13], [19]). A resolving semigroup $U_t : \mu \mapsto \mu_t$ of the Cauchy problem for equation (3) specified a so called generalized or nonlinear

Markov process $X(t)$, whose distribution μ_t at time t can be determined by the formula $U_{t-s}\mu_s$ from its distribution μ_s at any previous moment s.

In the case of diffusions (when ν vanishes in (1) or (4)) the theory of the corresponding semigroups is well developed, see [17] and more recent achievements in [6]. Also well developed is the case of pure jump processes, see e.g. the treatment of the Boltzmann equation (spatially trivial) in [20].

The goal of the present paper is to exploit the idea of nonlinear integrators (see [5], [16]) combined with a certain coupling of Lévy processes in order to push forward the probabilistic construction in a way that allows the natural Lipschitz continuous dependence of the coefficients G, b, ν on x, μ with measures equipped with their Wasserstein metric (see the definition below). Thus obtained extension of the standard SDEs with Lévy noise represent a probabilistic counterpart of the celebrated extension of the Monge mass transformation problem to the generalized Kantorovich one. To streamline the exposition we shall use Ito's approach (as exposed in detail in [19]) for constructing the solutions of stochastic equations directly via Euler approximation scheme bypassing the theory of stochastic integration itself. Roughly speaking the idea is to approximate a process with a given (formal) generator (or pre-generator) by processes with piecewise Lévy paths.

For a random variable X we shall denote by $\mathcal{L}(X)$ the distribution (probability law) of X. Recall that the so called Wasserstein-Kantorovich metrics W_p, $p \geq 1$, on the set of probability measures $\mathcal{P}(\mathbf{R}^d)$ on $\mathbf{R}^d$ are defined as

$$W_p(\nu_1, \nu_2) = \left(\inf_{\nu} \int |y_1 - y_2|^p \nu(dy_1 dy_2) \right)^{1/p}, \tag{5}$$

where inf is taken over the class of probability measures ν on $\mathbf{R}^{2d}$ that couple ν_1 and ν_2, i.e. that satisfy

$$\int \int (\phi_1(y_1) + \phi_2(y_2)) \nu(dy_1 dy_2) = (\phi_1, \nu_1) + (\phi_2, \nu_2) \tag{6}$$

for all bounded measurable ϕ_1, ϕ_2. It follows directly from the definition that

$$W_p^p(\mu, \mu') \leq \mathbf{E}\|X - X'\|^p \tag{7}$$

whenever $\mu = \mathcal{L}(X)$ and $\mu' = \mathcal{L}(X')$.

For random variable x, z we shall write sometimes shortly $W_p(x, z)$ for $W_p(\mathcal{L}(x), \mathcal{L}(z))$ (with some obvious abuse of notation).

It is well known (see e.g. [22]) that $(\mathcal{P}(\mathbf{R}^d), W_p)$, $p \geq 1$ is a complete metric space and that the convergence in this metric space is equivalent to the convergence in the weak sense combined with the convergence of the pth moments. In case $p = 1$ the celebrated Monge-Kantorovich theorem states that

$$W_1(\mu_1, \mu_2) = \sup_{f \in Lip} |(f, \mu_1) - (f, \mu_2)|,$$

where Lip is the set of continuous functions f such that $|f(x) - f(y)| \leq \|x - y\|$ for all x, y.

We shall need also the Wasserstein distances between the distributions in the Skorohod space $D([0,T],\mathbf{R}^d)$ of cadlag paths in $\mathbf{R}^d$ defined of course as

$$W_{p,T}(X_1,X_2)=\inf\left(\mathbf{E}\sup_{t\le T}|X_1(t)-X_2(t)|^p\right)^{1/p}, \qquad (8)$$

where inf is taken over all couplings of the distributions of the random paths X_1,X_2. Notice that this distance is linked with the uniform (and not Skorohod) topology on the path space.

To compare the Lévy measures, we shall need an extension of these distances to unbounded measures. Namely, let $\mathcal{M}_p(\mathbf{R}^d)$ denote the class of Borel measures μ on $\mathbf{R}^d\setminus\{0\}$ (not necessarily finite) with a finite p-th moment (i.e. such that $\int|y|^p\mu(dy)<\infty$). For a pair of measures ν_1,ν_2 from $\mathcal{M}_p(\mathbf{R}^d)$ we define the distance $W_p(\nu_1,\nu_2)$ by (5), where inf is now taken over all $\nu\in\mathcal{M}_p(\mathbf{R}^{2d})$ such that (6) holds for all ϕ_1,ϕ_2 satisfying $\phi_i(.)/|.|^p\in C(\mathbf{R}^d)$. It is easy to see that for finite measures this definition coincides with the previous one and that if measures ν_1 and ν_2 are infinite, the distance $W_p(\nu_1,\nu_2)$ is finite. [1]

Moreover, by the same argument as for finite measures (see [18] or [22]) one shows that whenever the distance $W_p(\nu_1,\nu_2)$ is finite, the infimum in (5) is achieved, i.e. there exists a measure $\nu\in\mathcal{M}_p(\mathbf{R}^{2d})$ such that

$$W_p(\mu_1,\mu_2)=\left(\int|y_1-y_2|^p\nu(dy_1dy_2)\right)^{1/p}. \qquad (9)$$

Theorem 1.1 *Let an operator L have form* (1), *where*

$$\|\sqrt{G(x_1)}-\sqrt{G(x_2)}\|+|b(x_1)-b(x_2)|+W_2(\mathbf{1}_{B_1}(.)\nu(x_1;.),\mathbf{1}_{B_1}(.)\nu(x_2;.))\le\kappa\|x_1-x_2\| \qquad (10)$$

with a certain constant κ, and

$$\sup_x\left(\sqrt{G(x)}+|b(x)|+\int_{B_1}|y|^2\nu(x,dy)\right)<\infty. \qquad (11)$$

Let the family of finite measures $\{\mathbf{1}_{\mathbf{R}^d\setminus B_1})(.)\nu(x;.)\}$ be uniformly bounded, tight and depend weakly continuous on x. Then L extends to the generator of a conservative Feller semigroup.

Remarks. 1. The boundedness condition (11) is not essential and can be dispensed with by the usual localization arguments, see [15]. 2. Once the well posed-ness of the equations generated by L is obtained, it implies various extensions of the results on the corresponding boundary value problems, problems with unbounded coefficients, fractional dynamics or Malliavin calculus (see [21], [12], [14], [4]) obtained earlier for particular cases.

For example, assumption on ν is satisfied if one can decompose the Lévy measures $\nu(x;.)$ in the countable sums $\nu(x;.)=\sum_{n=1}^{\infty}\nu_n(x;.)$ of probability measures so that $W_2(\nu_i(x;.),\nu_i(z;.))\le a_i|x-z|$ and the series $\sum a_i^2$ converges. It is well known that

[1]Let a decreasing sequence of positive numbers ϵ_n^1 be defined by the condition that ν_1 can be decomposed into the sum $\nu_1=\sum_{n=1}^{\infty}\nu_1^n$ of the probability measures ν_1^n having the support in the closed shells $\{x\in\mathbf{R}^d:\epsilon_n^1\le|x|\le\epsilon_{n-1}^1\}$ (where $\epsilon_0^1=\infty$). Similarly ϵ_2^n and ν_2^n are defined. Then the sum $\nu=\sum_{n=1}^{\infty}\nu_1^n\otimes\nu_2^n$ is a coupling of ν_1 and ν_2 with a finite $\int|y_1-y_2|^p\nu(dy_1dy_2)$.

the optimal coupling of probability measures (Kantorovich problem) can not always be realized via a mass transportation (a solution to the Monge problem), thus leading to the examples when the construction of the process via standard stochastic calculus would not work. On the other hand, no non-degeneracy is build in this example leading to serious difficulties when trying to apply analytic techniques in these circumstances.

Another important particular situation is that of a common star shape of the measures $\nu(x;.)$, i.e. if they can be represented as

$$\nu(x;dy) = \nu(x,s,dr)\,\omega(ds), \quad y \in \mathbf{R}^d, r = |y| \in \mathbf{R}_+, s = y/r \in S^{d-1}, \tag{12}$$

with a certain measure ω on S^{d-1} and a family of measures $\nu(x,s,dr)$ on $\mathbf{R}_+$. This allows to reduce the general coupling problem to a much more easily handled one-dimensional one, because evidently if $\nu_{x,y,s}(dr_1 dr_2)$ is a coupling of $\nu(x,s,dr)$ and $\nu(y,s,dr)$, then $\nu_{x,y,s}(dr_1 dr_2)\omega(ds)$ is a coupling of $\nu(x;.)$ and $\nu(y;.)$. If one-dimensional measures have no atoms, their coupling can be naturally organized via pushing along a certain mapping. Namely, the measure ν^F is the pushing forward of a measure ν on $\mathbf{R}_+$ by a mapping $F : \mathbf{R}_+ \mapsto \mathbf{R}_+$ whenever

$$\int f(F(r))\nu(dr) = \int f(u)\nu^F(du)$$

for a sufficiently rich class of test functions f, say for the indicators of intervals. Suppose we are looking for a family of monotone continuous bijections $F_{x,s} : \mathbf{R}_+ \mapsto \mathbf{R}_+$ such that $\nu^{F_{x,s}} = \nu(x,s,.)$. Choosing $f = \mathbf{1}_{[F(z),\infty)}$ as a test function in the above definition of pushing yields

$$G(x,s,F_{x,s}(z)) = \nu([z,\infty)) \tag{13}$$

for $G(x,s,z) = \nu(x,s,[z,\infty)) = \int_z^\infty \nu(x,s,dy)$. Clearly if all $\nu(x,s,.)$ and ν are unbounded, but bounded on any interval separated from the origin, have no atoms and do not vanish on any open interval, then this equation defines a unique continuous monotone bijection $F_{x,s} : \mathbf{R}_+ \mapsto \mathbf{R}_+$ with also continuous inverse. Hence we arrive to the following criterion.

Proposition 1.1 *Suppose the Lévy measures $\nu(x;.)$ can be represented in the form* (12) *and ν is a Lévy measure on $\mathbf{R}_+$ such that all $\nu(x,s,.)$ and ν are unbounded, have no atoms and do not vanish on any open interval. Then the family $\nu(x;.)$ depends Lipshitz continuous on x in W_2 whenever the unique continuous solution $F_{x,s}(z)$ to* (13) *is Lipschitz continuous in x with a constant $\kappa_F(z,s)$ enjoying the condition*

$$\int_{\mathbf{R}_+}\int_{S^{d-1}} \kappa_F^2(r,s)\omega(ds)\nu(dr) < \infty. \tag{14}$$

Proof. By the above discussion the solution F specifies the coupling $\nu_{x,y}(dr_1 dr_2 ds_1 ds_2)$ of $\nu(x;.)$ and $\nu(y;.)$ via

$$\int f(r_1,r_2,s_1,s_2)\nu_{x,y}(dr_1 dr_2 ds_1 ds_2) = \int f(F_{x,s}(r), F_{y,s}(r), s, s)\omega(ds)\nu(dr),$$

so that for Lipschitz continuity of the family $\nu(x;.)$ it is sufficient to have

$$\int_{\mathbf{R}_+}\int_{S^{d-1}} (F_{x,s} - F_{y,s})^2 \omega(ds)\nu(dr) \le c(x-y)^2,$$

which is clearly satisfied whenever (14) holds.

The particular case of $\nu(x,s,.)$ above having densities with respect to the Lebesgue measure on $\mathbf{R}_+$ is discussed in much detail in [19].

The point to make here is that a coupling for the sum of Lévy measures can be organized separately for each term allowing to use the above statement for star shape components and, say, some discrete methods for discrete parts.

Theorem 1.1 is a straightforward corollary of our main theorem that we shall formulate now. To make our exposition more transparent we shall present the main arguments in the case of L_μ having the form

$$L_\mu f(x) = \frac{1}{2}(G(x,\mu)\nabla,\nabla)f(x)+(b(x,\mu),\nabla f(x))+\int(f(x+z)-f(x)-(\nabla f(x),z))\nu(x,\mu;dz) \tag{15}$$

with $\nu(x,\mu;.)\in\mathcal{M}_2(\mathbf{R}^d)$. Let $Y_\tau(z,\mu)$ be a family of Lévy processes depending measurably on the points z and probability measures μ in $\mathbf{R}^d$ and specified by their generators

$$L[z,\mu]f(x) = \frac{1}{2}(G(z,\mu)\nabla,\nabla)f(x)+(b(z,\mu),\nabla f(x))$$
$$+\int(f(x+y)-f(x)-(\nabla f(x),y))\nu(z,\mu;dy) \tag{16}$$

where $\nu(z,\mu)\in\mathcal{M}_2(\mathbf{R}^d)$. Under the conditions of Theorem 1.2 given below, the existence of such a family follows from the well known randomization lemma [2] (see e.g. [8], Lemma 3.22), because by Proposition A.1 the mapping from z, μ to the law of the Lévy process $Y_\tau(z,\mu)$ is continuous, hence measurable, and consequently, by this Lemma (with Z being the complete metric space $D(\mathbf{R}_+,\mathbf{R}^d)$, and hence a Borel space) one can define all $Y_\tau(z,\mu)$ on the single standard probability space $[0,1]$. Let us stress for clarity that the processes $Y_\tau(x,\mu)$ depend on x,μ only via the parameters of the generator, i.e., say, the random variable $\xi = x+Y_\tau(x,\mathcal{L}(x))$ has the characteristic function

$$\mathbf{E}e^{ip\xi} = \int \mathbf{E}e^{ip(x+Y_\tau(x,\mathcal{L}(x))}\mu(dx).$$

Our approach to solving (3) is via the solution to the following nonlinear distribution dependent stochastic equation with nonlinear Lévy type integrators:

$$X(t) = X+\int_0^t dY_s(X(s),\mathcal{L}(X(s))), \quad \mathcal{L}(X)=\mu, \tag{17}$$

with a given initial distribution μ and a random variable X independent of $Y_\tau(z,\mu)$.

We shall define the solution through the Euler type approximation scheme, i.e. by means of the approximations X_μ^τ:

$$X_\mu^\tau(t) = X_\mu^\tau(l\tau)+\Delta Y^l_{t-l\tau}(X_\mu^\tau(l\tau),\mathcal{L}(X_\mu^\tau(l\tau))), \quad \mathcal{L}(X_\mu^\tau(0))=\mu, \tag{18}$$

where $l\tau < t \le (l+1)\tau$, $l=0,1,2,...$, and $\Delta Y^l_\tau(x,\mu)$ is a collection (depending on l) of independent families of the Lévy processes $Y_\tau(x,\mu)$ introduced above. Clearly these approximation processes are cadlag.

[2]It states that if $\mu(x,dz)$ is a probability kernel from a measurable space X to a Borel space Z, then there exists a measurable function $f: X\times[0,1]\to Z$ such that if θ is uniformly distributed on $[0,1]$, then $f(X,\theta)$ has distribution $\mu(x,.)$ for every $x\in X$.

For $x \in \mathbf{R}^d$ we shall write shortly $X_x^\tau(k\tau)$ for $X_{\delta_x}^\tau(k\tau)$.

By the weak solution to (17) we shall mean the weak limit of $X_\mu^{\tau_k}$, $\tau_k = 2^{-k}$, $k \to \infty$, in the sense of the distributions on the Skorohod space of cadlag paths (which is of course implied by the convergence of the distributions in the sense of the distance (8)). Alternatively one could define it as a solution to the corresponding nonlinear martingale problem (see below the proof of the main theorem) or directly via the construction of the corresponding stochastic integral. This issue is addressed in detail in [15], our purpose here being the construction of a Markov process with a given generator.

The following is our main result.

Theorem 1.2 *Let an operator L_μ have form* (15)*. Moreover*

$$\|\sqrt{G(x,\mu)} - \sqrt{G(z,\eta)}\| + |b(x,\mu) - b(z,\eta)| + W_2(\nu(x,\mu;.), \nu(z,\eta;.)) \leq \kappa(|x-z| + W_2(\mu,\eta)), \tag{19}$$

holds true with a constant κ and

$$\sup_{x,\mu} \left(\sqrt{G(x,\mu)} + |b(x,\mu)| + \int |y|^2 \nu(x,\mu,dy) \right) < \infty. \tag{20}$$

Then

(i) for any $\mu \in \mathcal{P}(\mathbf{R}^d) \cap \mathcal{M}_2(\mathbf{R}^d)$ the approximations $X_\mu^{\tau_k}$ converge to a process $X_\mu(t)$ in the sense that

$$\sup_\mu \sup_{t \in [0,t_0]} W_2^2 \left(X_\mu^{\tau_k}([t/\tau_k]\tau_k, X_\mu(t) \right) \leq c(t_0)\tau_k \tag{21}$$

for any t_0, and even stronger

$$\sup_\mu W_{2,t_0}^2 \left(X_\mu^{\tau_k}, X_\mu \right) \leq c(t_0)\tau_k; \tag{22}$$

(ii) the distributions $\mu_t = \mathcal{L}(X_\mu(t))$ depend $1/2$-Hölder continuous on t in the metric W_2 and $X_\mu(t)$ depend Lipschitz continuously on the initial condition in the following sense:

$$\sup_{t \in [0,t_0]} W_2^2(X_\mu(t), X_\eta(t)) \leq c(t_0) W_2^2(\mu,\eta); \tag{23}$$

(iii) the processes

$$M(t) = f(X_\mu(t)) - f(X_\mu^0) - \int_0^t (L_{\mathcal{L}(X_\mu(s))} f(X_\mu(s)) \, ds \tag{24}$$

are martingales for any $f \in C^2(\mathbf{R}^d)$; in other words, the process $X_\mu(t)$ solves the corresponding (nonlinear) martingale problem;

(iv) the distributions $\mu_t = \mathcal{L}(X_\mu(t))$ satisfy the weak nonlinear equation (3) *(that holds for all $f \in C^2(\mathbf{R}^d)$);*

(v) the resolving operators $U_t : \mu \mapsto \mu_t$ of the Cauchy problem (3) *form a nonlinear Markov semigroup, i.e. they are continuous mappings from $\mathcal{P}(\mathbf{R}^d) \cap \mathcal{M}_2(\mathbf{R}^d)$ (equipped with the metric W_2) to itself such that U_0 is the identity mapping and $U_{t+s} = U_t U_s$ for all $s,t \geq 0$. If $L[z,\mu]$ do not depend explicitly on μ the operators $T_t f(x) = \mathbf{E} f(X_x(t))$ form a conservative Feller semigroup preserving the space of Lipschitz continuous functions.*

This theorem is proved in the next section. In Sections 3 we obtain some regularity criteria for the Markov semigroups constructed.

A simple meaningful example is given by the nonlinear kinetic equations

$$\frac{d}{dt}(f, \mu_t) = (Lf, \mu_t) + \int (K(x,y), \nabla f(x)) \mu_t(dx)\mu_t(dy), \tag{25}$$

with L being of form (1) with Lipschitz continuous coefficients and K being a bounded Lipschitz continuous mapping $R^{2d} \mapsto R^d$, which arise as the mean-field limit for potentially interacting Feller processes.

Theorem 1.1 follows now from Theorem 1.2 by the standard perturbation theory, since dividing the generator into two parts, where the first part is the integral term with the Lévy measure reduced to $\mathbf{R}^d \setminus B_1$, one gets a sum of two generators, one of which is bounded in $C_\infty(\mathbf{R}^d)$ (as follows from the assumed tightness) and the other satisfies Theorem 1.2.

It is worth noting that in a simpler case of generators of up to the first order the continuity of Lévy measures with respect to a more easy handled metric W_1 is sufficient, as shows the following result, whose proof is omitted (as being a simplified version of the proof of Theorem 1.2).

Theorem 1.3 *Let an operator L_μ have the form*

$$L_\mu f(x) = (b(x,\mu), \nabla f(x)) + \int (f(x+z) - f(x))\nu(x,\mu; dz), \quad \nu(x,\mu;.) \in \mathcal{M}_1(\mathbf{R}^d). \tag{26}$$

and

$$\|b(x,\mu) - b(z,\eta)\| + W_1(\nu(x,\mu;.), \nu(z,\eta;.)) \le \kappa(\|x-z\| + W_1(\mu,\eta)) \tag{27}$$

holds true with a constant κ. Then for any $\mu \in \mathcal{P}(\mathbf{R}^d) \cap \mathcal{M}_1(\mathbf{R}^d)$ there exists a process $X_\mu(t)$ solving (17) *(with analogously defined $Y_\tau(z,\mu)$) such that*

$$\sup_\mu W_{1,t_0}\left(X_\mu^{\tau_k}, X_\mu\right) \le c(t_0)\tau_k, \tag{28}$$

the distributions $\mu_t = \mathcal{L}(X(t))$ depend $1/2$-Hölder continuous on t in the metric W_1 and $X_\mu(t)$ depend Lipschitz continuously on the initial condition in the following sense:

$$W_1(X_\mu(t), X_\eta(t)) \le c(t_0) W_1(\mu, \eta). \tag{29}$$

Moreover, the processes (24) *are martingales for any $f \in C^1(\mathbf{R}^d)$ and the distributions $\mu_t = \mathcal{L}(X_\mu(t))$ satisfy the weak nonlinear equation* (3) *(that holds for all $f \in C^1(\mathbf{R}^d)$). If $L[z,\mu]$ do not depend explicitly on μ the operators $T_t f(x) = \mathbf{E} f(X_x(t))$ form a conservative Feller semigroup.*

In Appendix we describe a coupling of Lévy processes that is crucial for our purposes.

2 Proof of Theorem 1.2

Step 1 (*uniform continuity of the approximations with respect to initial data*).

One has

$$W_2^2(x_1 + Y_\tau(x_1, \mathcal{L}(x_1)), x_2 + Y_\tau(x_2, \mathcal{L}(x_2))) \le \mathbf{E}(\xi_1 - \xi_2)^2$$

for any random variable (ξ_1, ξ_2) with the projections $\xi_i = x_i + Y_\tau(x_i, \mu_i)$, $i = 1, 2$. Let us choose the coupling described by the characteristic function

$$\mathbf{E}e^{i(p_1\xi_1 + p_2\xi_2)} = \int_{\mathbf{R}^{4d}} e^{ip_1(x_1+y_1)+ip_2(x_2+y_2)}\mu(dx_1dx_2)P^\tau_{x_1,x_2,\mu_1,\mu_2}(dy_1dy_2),$$

where $\mu_i = \mathcal{L}(x_i)$ and μ is an arbitrary coupling of the random variables x_1, x_2 and P^τ is the coupling of the Lévy processes $Y_\tau(x_i, \mu_i)$ given by Proposition A.1. Consequently,

$$\mathbf{E}(\xi_1 - \xi_2)^2 = -\frac{d^2}{dp^2} \mid_{p=0} \mathbf{E}e^{ip(\xi_1-\xi_2)}$$

$$= \int_{\mathbf{R}^{4d}} [(x_1 + y_1) - (x_2 + y_2)]^2 \mu(dx_1dx_2)P^\tau_{x_1,x_2,\mu_1,\mu_2}(dy_1dy_2),$$

which by (44) does not exceed

$$\int_{\mathbf{R}^{2d}} \left((x_1 - x_2)^2 + c\tau[(x_1 - x_2)^2 + W_2^2(\mathcal{L}(x_1), \mathcal{L}(x_2))]\right) \mu(dx_1dx_2).$$

Consequently, by (7),

$$\mathbf{E}(\xi_1 - \xi_2)^2 \le \int_{\mathbf{R}^{2d}} (1 + 2c\tau)(x_1 - x_2)^2\mu(dx_1dx_2). \tag{30}$$

Hence, taking infimum over all coupling, yields

$$W_2^2(x_1 + Y_\tau(x_1, \mathcal{L}(x_1)), x_2 + Y_\tau(x_2, \mathcal{L}(x_2))) \le (1 + 2c\tau)W_2^2(\mathcal{L}(x_1), \mathcal{L}(x_2)). \tag{31}$$

Applying this inequality inductively, yields

$$W_2^2(X^\tau_\mu(k\tau), X^\tau_\eta(k\tau)) \le e^{1+2ck\tau}W_2^2(\mu, \eta). \tag{32}$$

Step 2 (*subdivision and the existence of the limit*).
We want to estimate the W_2 distance between the random variables

$$\xi_1 = x + Y_\tau(x, \mu) = x' + \Delta Y_{\tau/2}(\tau/2, x, \mu), \quad \xi_2 = z' + \Delta Y_{\tau/2}(\tau/2, z', \eta'),$$

where

$$x' = x + Y_{\tau/2}(x, \mu), \quad z' = z + Y_{\tau/2}(z, \eta),$$

and $\mu = \mathcal{L}(x)$, $\eta = \mathcal{L}(z)$, $\eta' = \mathcal{L}(z')$. We shall couple ξ_1 and ξ_2 using sequentially Proposition A.1. Namely, we shall define it by the equation

$$\mathbf{E}f(\xi_1, \xi_2) = \int_{\mathbf{R}^{6d}} f(x + v_1 + y_1, z + v_2 + y_2)\mu(dxdz)P^{\tau/2}_{x,z,\mu,\eta}(dv_1dv_2)P^{\tau/2}_{x,z',\mu,\eta'}(dy_1dy_2)$$

for $f \in C(\mathbf{R}^{2d})$, where, say, $P^{\tau/2}_{x,z',\mu,\eta'}$ is the coupling of the Lévy processes $Y_{\tau/2}(x, \mu)$ and $Y_{\tau/2}(z', \eta')$ given by Proposition A.1 (note that the probability law η' is the function of z, η).

Now by (44)

$$W_2^2(\xi_1,\xi_2) \le \mathbf{E}(\xi_1-\xi_2)^2$$
$$\le \mathbf{E}(x'-z')^2 + c\tau[\mathbf{E}(x'-z')^2 + \mathbf{E}(x-z')^2 + W_2^2(\mu,\eta')].$$

Hence, by (30) and (7) $W_2^2(\xi_1,\xi_2)$ does not exceed

$$W_2^2(x,z)(1+2c\tau)(1+c\tau) + 2c\tau\mathbf{E}(x-z')^2$$

and consequently also

$$W_2^2(x,z)(1+2c\tau) + 4c\tau\mathbf{E}(Y_{\tau/2}(z,\eta))^2$$

(with another constant c) so that

$$W_2^2(\xi_1,\xi_2) \le W_2^2(x,z)(1+c\tau) + c\tau^2$$

(with yet another c), because the second moments of our processes Y_τ are bounded due to assumption (20). Consequently

$$W_2^2(X_\mu^\tau(k\tau), X_\mu^{\tau/2}(k\tau)) \le c\tau^2 + (1+c\tau)W_2^2(X_\mu^\tau((k-1)\tau), X_\mu^{\tau/2}((k-1)\tau)). \qquad (33)$$

By induction one estimates the l.h.s. of this inequality by

$$\tau^2[1 + (1+c\tau) + (1+c\tau)^2 + ... + (1+c\tau)^{(k-1)}] \le c^{-1}\tau(1+c\tau)^k \le c(t_0)\tau.$$

Repeating this subdivision and using the triangle inequality for distances yields

$$W_2^2(X_\mu^\tau(k\tau), X_\mu^{\tau/2^m}(k\tau)) \le c(t_0)\tau.$$

This implies the existence of the limit $X_x^{\tau_k}([t/\tau_k]\tau_k)$, as $k \to \infty$, in the sense of (21).

Observe now that (32) implies (23). Moreover, the mapping $T_t f(x) = \mathbf{E}f(X_x(t))$ preserves the set of Lipschitz continuous functions. In fact, if f is Lipschitz with the constant h, then

$$|\mathbf{E}f(X_x^\tau([t/\tau]\tau)) - \mathbf{E}f(X_z^\tau([t/\tau]\tau))| \le h\mathbf{E}\|X_x^\tau([t/\tau]\tau) - (X_z^\tau([t/\tau]\tau)\|$$

$$\le h\left(\mathbf{E}\|X_x^\tau([t/\tau]\tau) - (X_z^\tau([t/\tau]\tau)\|^2\right)^{1/2}.$$

for any coupling of the processes X_x^τ and X_z^τ. Hence by (32)

$$|\mathbf{E}f(X_x^\tau([t/\tau]\tau)) - \mathbf{E}f(X_z^\tau([t/\tau]\tau))| \le hc(t_0)W_2(x,z).$$

In particular, T_t preserves constant functions. Similarly one shows (first for Lipschitz continuous f and then for all $f \in C_\infty(\mathbf{R}^d)$ via standard approximation) that

$$\sup_{t\in[0,t_0]} \sup_x |\mathbf{E}f(X_x^{\tau_k}([t/\tau_k]\tau_k)) - \mathbf{E}f(X_x(t))| \to 0, \quad k \to \infty, \qquad (34)$$

for all $f \in C_\infty(\mathbf{R}^d)$. Moreover, as the dynamics of averages of the approximation processes clearly preserve the space $C_\infty(\mathbf{R}^d)$, the same holds for the limiting mappings T_t.. Consequently $T_t f = \mathbf{E}f(X_x(t))$ is a positivity preserving family of contractions in $C(\mathbf{R}^d)$ that preserve constants and the space $C_\infty(\mathbf{R}^d)$. Hence the mappings $U_t : \mu \mapsto \mu_t$ form a (nonlinear) Markov semigroup, and if $L[z,\mu]$ do not depend explicitly on μ, the operators $T_t f(x) = \mathbf{E} f(X_x(t))$ form a conservative Feller semigroup. The Markov (or semigroup) property of the solutions follows from the construction (a detailed discussion of this fact in a similar situation is given in [19]).

From the inequality

$$W_2^2(\mathcal{L}(X_\mu^\tau(l\tau)), \mathcal{L}(X_\mu^\tau((l-1)\tau))) \le \mathbf{E}\left[\Delta Y_\tau((l-1)\tau, X_\mu^\tau((l-1)\tau), \mathcal{L}(X_\mu^\tau((l-1)\tau)))\right]^2 \le c\tau$$

it follows that the curve μ_t depends 1/2-Hölder continuous on t in W_2.

Step 3 (*improving convergence and solving the martingale problem*)

The processes

$$M_\tau(t) = f(X_\mu^\tau(t)) - f(X) - \int_0^t L[X_\mu^\tau([s/\tau]\tau), \mu^\tau_{[s/\tau]}] f(X_\mu^\tau(s))\, ds, \quad \mu = \mathcal{L}(X), \tag{35}$$

where $\mu_l^\tau = \mathcal{L}(X_\mu^\tau(l\tau))$, are martingales by Dynkin's formula, applied to Lévy processes $Y_\tau(z,\mu)$. Our aim is to pass to the limit $\tau_k \to 0$ to obtain the martingale characterization of the limiting process. But let us first strengthen our convergence result.

Observe that the step by step inductive coupling of the trajectories X_μ^τ and X_η^τ used above to prove (32) actually defines the coupling between the distributions of these random trajectories in the Skorohod space $D([0,t_0],\mathbf{R}^d)$ for any t_0, i.e. a random trajectory $(X_\mu^\tau, X_\eta^\tau)$ in $D([0,t_0],\mathbf{R}^{2d})$. One can construct the Dynkin martingales for this coupled process in the same way as above for X_μ^τ. Namely, for a function f of two variables with bounded second derivatives the process

$$M_\tau(t) = f(X_\mu^\tau(t), X_\eta^\tau(t)) - \int_0^t \tilde{L}_s f(X_\mu^\tau(s), X_\eta^\tau(s))\, ds, \quad \mu = \mathcal{L}(x_\mu),\ \eta = \mathcal{L}(x_\eta),$$

is a martingale, where $\tilde{L}_t$ is the coupling operator (42) constructed from the Lévy processes Y with parameters $X_\mu^\tau([t/\tau]\tau), \mu^\tau_{[t/\tau]}$ and $X_\eta^\tau([t/\tau]\tau), \eta^\tau_{[t/\tau]}$. Choosing $f(x,y) = x-y$ leads to the martingale of the form

$$\tilde{M}_\tau(t) = X_\mu^\tau(t) - X_\eta^\tau(t) + \int_0^t O(1)|X_\mu^\tau(s) - X_\eta^\tau(s)|\, ds.$$

Using (32) in conjunction with Doob's maximal inequality implies

$$\mathbf{E} \sup_{s\le t} |\tilde{M}_\tau(s)|^2 \le c(t) W_2^2(\mu,\eta),$$

which in turn implies

$$\mathbf{E} \sup_{s\le t} (X_\mu^\tau(s) - X_\eta^\tau(s))^2 \le c(t) W_2^2(\mu,\eta).$$

This allows to improve (32) to the estimate of the distance on paths:

$$W_{2,T}^2(X_\mu^\tau, X_\eta^\tau)^2 \le c(T) W_2^2(\mu,\eta). \tag{36}$$

Similarly one can strengthen the estimates for subdivisions leading to the convergence of the distributions on paths (22).

Using the Skorohod theorem for the weak converging sequence of random trajectories $X_\mu^{\tau_k}$ (let us stress again that the convergence with respect to the distance (8) implies the weak convergence of the distributions in the sense of the Skorohod topology), one can put them all on a single probability space forcing the processes $X_\mu^{\tau_k}$ to converge to X_μ almost surely in the sense of the Skorohod topology.

Passing to the limit $\tau = \tau_k \to 0$ in (35), using the continuity and boundedness of f and Lf and the dominated convergence theorem allows to conclude that the martingales $M_\tau(t)$ converge almost surely and in L^1 to the martingale

$$M(t) = f(X_\mu(t)) - f(X) - \int_0^t (L_{\mathcal{L}(X_\mu(s))} f)(X_\mu(s))\, ds,$$

in other words that the process $X_\mu(t)$ solves the corresponding (nonlinear) martingale problem.

Step 4 (*completion*)

To prove (3) one writes using the martingale properties of $M(t)$:

$$\frac{d}{dt}(f, \mu_t) = \lim_{s\to 0} \frac{1}{s} \mathbf{E}(f, X_\mu(t+s) - X_\mu(t)) = \lim_{s\to 0} \frac{1}{s} \mathbf{E} \int_t^{t+s} (L_{\mathcal{L}(X_\mu(s))} f)(X_\mu(s))\, ds$$

$$= (L_{\mu_t} f, \mu_t)) + \lim_{s\to 0} \frac{1}{s} \mathbf{E} \int_t^{t+s} [(L_{\mathcal{L}(X_\mu(s))} f)(X_\mu(s)) - L_{\mu_t} f(X_\mu(t)]\, ds,$$

implying (3) by the continuity of μ_t.

3 Regularity and uniqueness

Discussing regularity we reduce our attention for simplicity to Feller processes, where uniqueness follows from the sufficient regularity. It is also known (see e.g. [13]) that from the regularity of nonhomogeneous versions of these Feller processes one can naturally deduce the uniqueness and regularity for the corresponding nonlinear problems.

By C^k_{Lip} (respectively C^k_∞) we shall denote the subspace of functions from $C^k(\mathbf{R}^d)$ with a Lipschitz continuous derivative of order k (respectively with all derivatives up to order k vanishing at infinity).

We shall discuss in detail only the first derivative.

Theorem 3.1 *Assume the conditions of Propositions A.4 and A.5 hold. Then the spaces C^1_{Lip} and $C^1_{Lip} \cap C^1_\infty$ are invariant under the semigroup T_t constructed above from the generator*

$$Lf(x) = \frac{1}{2}(G(x)\nabla, \nabla) f(x) + (b(x), \nabla f(x)) + \int (f(x+y) - f(x) - (\nabla f(x), y))\nu(x, dy), \quad (37)$$

and for any $f \in C^1_{Lip}$, $\phi \in L_1 \cap C_\infty(\mathbf{R}^d)$

$$\frac{d}{dt}(T_t f, \phi) = (L T_t f, \phi), \quad t \geq 0. \quad (38)$$

Proof. First let us calculate $\nabla_j \mathbf{E} g(X_x^\tau(k\tau))$ for an arbitrary k and $g \in C^1_{Lip}(\mathbf{R}^d)$. One has

$$\mathbf{E} g(X_x^\tau(k\tau)) = \int g(x + z_1 + ... + z_k) P_x^\tau(dz_1) ... P^\tau_{x+\sum_{m=1}^{k-1} z_m}(dz_k).$$

As

$$\nabla_j \mathbf{E} g(X_x^\tau(k\tau)) = \lim_{h\to 0} \frac{1}{h}(\mathbf{E} g(X^\tau_{x+he_j}(k\tau)) - \mathbf{E} g(X_x^\tau(k\tau)))$$

does not depend on coupling, one can write

$$\nabla_j \mathbf{E} g(X_x^\tau(k\tau)) = \lim_{h\to 0} \frac{1}{h} \int (g(x + he_j + w_1 + ... + w_k) - g(x + v_1 + ... + v_k))$$

$$P^\tau_{x+he_j,x}(dw_1 dv_1) ... P^\tau_{x+he_j+\sum_{m=1}^{k-1} w_m, x+\sum_{m=1}^{k-1} v_m}(dw_k dv_k)$$

$$= \lim_{h\to 0} \int \big[(\nabla g(x + v_1 + ... + v_k), e_j + \frac{w_1 - v_1}{h} ... + \frac{w_k - v_k}{h})$$

$$+ O(1)\frac{1}{h}(he_j + w_1 - v_1 + ... + w_k - v_k)^2\big]$$

$$P^\tau_{x+he_j,x}(dw_1 dv_1) ... P^\tau_{x+he_j+\sum_{m=1}^{k-1} w_m, x+\sum_{m=1}^{k-1} v_m}(dw_k dv_k).$$

The term with $O(1)$ vanishes as it can be rewritten by Proposition A.1 as

$$\lim_{h\to 0} O(1)\frac{1}{h} \int (he_j + w_1 - v_1 + ... + w_{k-1} - v_{k-1})^2 (1 + c\tau)$$

$$P^\tau_{x+he_j,x}(dw_1 dv_1) ... P^\tau_{x+he_j+\sum_{m=1}^{k-2} w_m, x+\sum_{m=1}^{k-2} v_m}(dw_{k-1} dv_{k-1}),$$

and consequently, iterating this procedure as

$$\lim_{h\to 0} \frac{1}{h} O(1) h^2 (1 + c\tau)^k = 0.$$

Hence

$$\nabla_j \mathbf{E} g(X_x^\tau(k\tau)) = \lim_{h\to 0} \frac{1}{h} \int (\nabla g(x + v_1 + ... + v_k), he_j + w_1 - v_1 + ... + w_k - v_k)$$

$$P^\tau_{x+he_j,x}(dw_1 dv_1) ... P^\tau_{x+he_j+\sum_{m=1}^{k-1} w_m, x+\sum_{m=1}^{k-1} v_m}(dw_k dv_k).$$

Assume now first that g is from the Schwartz space $S(\mathbf{R}^d)$ so that Proposition A.5 applies and one can write

$$\int (\nabla g(x + v_1 + ... + v_k), he_j + w_1 - v_1 + ... + w_k - v_k) P^\tau_{x+he_j+\sum_{m=1}^{k-1} w_m, x+\sum_{m=1}^{k-1} v_m}(dw_k dv_k)$$

$$= \int \sum_{j_k, j_{k-1}=1}^{d} \nabla_{j_k} g(x + v_1 + ... + v_k)(he_j + w_1 - v_1 + ... + w_{k-1} - v_{k-1})^{j_{k-1}} (\delta^{j_k}_{j_{k-1}} + z_k^{j_k})$$

$$Q^\tau_{D^{j_{k-1}}\nu(x+\sum_{m=1}^{k-1} v_m)}(dz_k dv_k) + O(\tau)(w_{k-1} - v_{k-1})^2.$$

Consequently, as the last term does not contribute to the limit $h \to 0$, and iterating this procedure one obtains

$$\nabla_j \mathbf{E} g(X_x^\tau(k\tau)) = \int \sum_{j_1,...,j_k=1}^d \nabla_{j_k} g(x+v_1+...+v_k)(\delta^j_{j_1}+z_1^{j_1})(\delta^{j_1}_{j_2}+z_2^{j_2})...(\delta^{j_k}_{j_{k-1}}+z_k^{j_k})$$

$$Q^\tau_{D^j\nu(x)}(dz_1dv_1)Q^\tau_{D^{j_1}\nu(x+v_1)}(dz_2dv_2)...Q^\tau_{D^{j_{k-1}}\nu(x+\sum_{m=1}^{k-1}v_m)}(dz_kdv_k), \tag{39}$$

which is the rigorous explicit form of the (a priori not clearly defined but intuitively appealing) expression

$$\mathbf{E}\nabla g(x+Y_\tau(x)+\Delta Y_\tau(\tau, X(\tau))+...+\Delta Y_\tau((k-1)\tau, X((k-1)\tau))$$

$$\left(1+\frac{\partial \Delta Y_\tau((k-1)\tau, X((k-1)\tau))}{\partial X((k-1)\tau)}\right)...\left(1+\frac{\partial Y_\tau(x)}{\partial x}\right).$$

Approximating arbitrary g by functions from the Schwartz space one can conclude that (39) holds for all $g \in C^1_{Lip}(\mathbf{R}^d)$.

We want to show now that these derivatives are Lipschitz continuous. To shorten the formulas let us do it for the case of $d=1$ only. In this case

$$\nabla \mathbf{E} g(X_x^\tau(k\tau)) = \int \nabla g(x+v_1+...+v_k)(1+z_1)...(1+z_k)$$

$$Q^\tau_{D\nu(x)}(dz_1dv_1)Q^\tau_{D\nu(x+v_1)}(dz_2dv_2)...Q^\tau_{D\nu(x+\sum_{m=1}^{k-1}v_m)}(dz_kdv_k), \tag{40}$$

and by Proposition A.4 one can write

$$\nabla \mathbf{E} g(X_{x_1}^\tau(k\tau)) - \nabla \mathbf{E} g(X_{x_2}^\tau(k\tau))$$

$$= \int [\nabla g(x_1+v_1+...+v_k)(1+z_1)...(1+z_k) - \nabla g(x_2+\tilde{v}_1+...+\tilde{v}_k)(1+\tilde{z}_1)...(1+\tilde{z}_k)]$$

$$Q^\tau_{D\nu(x_1,x_2)}(dz_1d\tilde{z}_1dv_1d\tilde{v}_1)Q^\tau_{D\nu(x_1+v_1,x_2+\tilde{v}_1)}(dz_2d\tilde{z}_2dv_2d\tilde{v}_2)...Q^\tau_{D\nu(x_1+\sum_{m=1}^{k-1}v_m,x_2+\sum_{m=1}^{k-1}\tilde{v}_m)}(dz_kd\tilde{z}_kdv_kd\tilde{v}_k).$$

Writing

$$\nabla g(x_1+v_1+...+v_k)(1+z_1)...(1+z_k) - \nabla g(x_2+\tilde{v}_1+...+\tilde{v}_k))(1+\tilde{z}_1)...(1+\tilde{z}_k)$$

$$= (\nabla g(x_1+v_1+...+v_k) - \nabla g(x_2+\tilde{v}_1+...+\tilde{v}_k))(1+z_1)...(1+z_k)$$

$$+\nabla g(x_2+\tilde{v}_1+...+\tilde{v}_k)[(1+z_1)...(1+z_k)-(1+\tilde{z}_1)...(1+\tilde{z}_k)],$$

and applying the Hölder inequality to estimate the integral over each of these two terms yields the estimate

$$|\nabla \mathbf{E} g(X_{x_1}^\tau(k\tau)) - \nabla \mathbf{E} g(X_{x_2}^\tau(k\tau))| \le \kappa[I_0^2(k,x_1,x_2)+I_1^2(k,x_1,x_2)]$$

with κ depending on the norm and the Lipschitz constant of ∇g, where

$$I_0^2(k,x_1,x_2) = \int [x_1-x_2+v_1-\tilde{v}_1+...+v_k-\tilde{v}_k]^2|(1+z_1)^2...(1+z_k)^2$$

$$Q^\tau_{D\nu(x_1,x_2)}(dz_1d\tilde{z}_1dv_1d\tilde{v}_1)...Q^\tau_{D\nu(x_1+\sum_{m=1}^{k-1}v_m,x_2+\sum_{m=1}^{k-1}\tilde{v}_m)}(dz_kd\tilde{z}_kdv_kd\tilde{v}_k).$$

and

$$I_1^2(k,x_1,x_2)=\int[(1+z_1)...(1+z_k)-(1+\tilde z_1)...(1+\tilde z_k)]^2$$

$$Q^\tau_{D\nu(x_1,x_2)}(dz_1d\tilde z_1dv_1d\tilde v_1)...Q^\tau_{D\nu(x_1+\sum_{m=1}^{k-1}v_m,x_2+\sum_{m=1}^{k-1}\tilde v_m)}(dz_kd\tilde z_kdv_kd\tilde v_k).$$

By (52)

$$I_0^2(k,x_1,x_2)\le(1+c\tau)I_0^2(k-1,x_1,x_2)\le...\le(1+c\tau)^k(x_1-x_2)^2.$$

It remains to estimate I_1^2. It would be convenient here to introduce special notations for the products:

$$Z_k=(1+z_1)...(1+z_k),\quad \tilde Z_k=(1+\tilde z_1)...(1+\tilde z_k).$$

Now one can write

$$Z_k-\tilde Z_k=Z_{k-1}(1+z_k)-\tilde Z_{k-1}(1+\tilde z_k)=(1+z_k)(Z_{k-1}-\tilde Z_{k-1})+(z_k-\tilde z_k)\tilde Z_{k-1},$$

so that

$$[Z_k-\tilde Z_k]^2=(1+z_k)^2(Z_{k-1}-\tilde Z_{k-1})^2+(z_k-\tilde z_k)^2\tilde Z_{k-1}^2+2(1+z_k)(z_k-\tilde z_k)(Z_{k-1}-\tilde Z_{k-1})\tilde Z_{k-1}.$$

Plugging this into the expression for I_1^2 yields

$$I_1^2(k,x_1,x_2)\le(1+c\tau)I_1^2(k-1,x_1,x_2)+c\tau I_0^2(k-1,x_1,x_2)+c\tau\int\Omega(Z_{k-1}-\tilde Z_{k-1})\tilde Z_{k-1}$$

$$Q^\tau_{D\nu(x_1,x_2)}(dz_1d\tilde z_1dv_1d\tilde v_1)...Q^\tau_{D\nu(x_1+\sum_{m=1}^{k-2}v_m,x_2+\sum_{m=1}^{k-2}\tilde v_m)}(dz_{k-1}d\tilde z_{k-1}dv_{k-1}d\tilde v_{k-1}),$$

where Ω in the last integral is a function of $x_1,x_2,v_j,\tilde v_j$ such that

$$|\Omega|\le c\|x_1-x_2+v_1-\tilde v_1+...+v_{k-1}-\tilde v_{k-1}\|.$$

Hence, applying to this last integral again the Hölder inequality yields

$$I_1^2(k,x_1,x_2)\le(1+c\tau)I_1^2(k-1,x_1,x_2)+c\tau I_0^2(k-1,x_1,x_2),$$

which taking into account the above bound for I_0^2 rewrites as

$$I_1^2(k,x_1,x_2)\le(1+c\tau)I_1^2(k-1,x_1,x_2)+c\tau(x_1-x_2)^2$$

with yet another c as long as $t=\tau k$ remains bounded. Using this formula recursively implies

$$I_1^2(k,x_1,x_2)\le c\tau(x_1-x_2)^2(1+(1+c\tau)+...+(1+c\tau)^k)\le c(k\tau)(x_1-x_2)^2.$$

Consequently one obtains the uniform estimate

$$|\nabla\mathbf{E}g(X^\tau_{x_1}(k\tau))-\nabla\mathbf{E}g(X^\tau_{x_2}(k\tau))|\le\kappa c(k\tau)\|x_1-x_2\|.$$

Hence from the sequence of the uniformly Lipschitz continuous functions $\nabla\mathbf{E}f(X_x^{\tau_k}(s))$, $k=1,2,...$, one can choose a convergent subsequence the limit being clearly $\nabla\mathbf{E}f(X_x(t))$, showing that $\mathbf{E}f(X_x(t))\in C^1_{Lip}$. The uniform convergence implies $\mathbf{E}f(X_x(t))\in C^1_{Lip}\cap C^1_\infty$ whenever the same holds for f.

To complete the proof of the theorem it remains equation (38). But this is easy: for $t=0$ it follows by approximating f with $f_n\in C^2(\mathbf{R}^d)$ and then for arbitrary t it follows by the invariance of the class C^1_{Lip} under T_t.

Second derivative is analyzed similarly. Namely, if the 'second derivative' of the Lévy measure is well defined by (56) and satisfies the continuity assumptions similar to those of Propositions A.4 and A.5 for the first one, the invariance of the space $C^2_{Lip}\cap C^1_\infty$ under T_t follows. The importance of this second order regularity lies in the well known fact that it implies uniqueness of the semigroup (see e.g. [19] or [10]).

A Coupling of Lévy processes

We describe here the natural coupling of Lévy processes leading in particular to the analysis of their weak derivatives with respect to a parameter. Recall that by C^k_{Lip} we denote the subspace of functions from $C^k(\mathbf{R}^d)$ with a Lipschitz continuous derivative of order k.

Proposition A.1 *Let Y^i_s, $i = 1, 2$, be two Lévy processes in $\mathbf{R}^d$ specified by their generators*

$$L_i f(x) = \frac{1}{2}(G_i\nabla, \nabla)f(x) + (b_i, \nabla f(x)) + \int (f(x+y) - f(x) - (\nabla f(x), y))\nu_i(dy) \quad (41)$$

with $\nu_i \in \mathcal{M}_2(\mathbf{R}^d)$. Let $\nu \in \mathcal{M}_2(\mathbf{R}^{2d})$ be a coupling of ν_1, ν_2, i.e. (6) *holds for all ϕ_1, ϕ_2 satisfying $\phi_i(.)/|.|^2 \in C(\mathbf{R}^d)$. Then the operator*

$$Lf(x_1, x_2) = \left[\frac{1}{2}(G_1\nabla_1, \nabla_1) + \frac{1}{2}(G_2\nabla_2, \nabla_2) + (\sqrt{G_2}\sqrt{G_1}\nabla_1, \nabla_2)\right] f(x_1, x_2)$$

$$+(b_1, \nabla_1 f(x_1, x_2)) + (b_2, \nabla_2 f(x_1, x_2))$$

$$+\int [f(x_1 + y_1, x_2 + y_2) - f(x_1, x_2) - ((y_1, \nabla_1) + (y_2, \nabla_2))f(x_1, x_2)]\nu(dy_1 dy_2) \quad (42)$$

(where ∇_i means the gradient with respect to x_i) specifies a Lévy process Y_s in $\mathbf{R}^{2d}$ with the characteristic exponent

$$\eta_{x_1,x_2}(p_1, p_2) = -\frac{1}{2}\left[\sqrt{G(x_1)}p_1 + \sqrt{G(x_2)}p_2\right]^2 + ib(x_1)p_1 + ib(x_2)p_2$$

$$+\int (e^{iy_1p_1+iy_2p_2} - 1 - i(y_1p_1 + y_2p_2))\nu(dy_1dy_2),$$

that is a coupling of Y^1_s, Y^2_s in the sense that the components of Y_s have the distribution of Y^1_s and Y^2_s respectively. Moreover, if $f(x_1, x_2) = h(x_1 - x_2)$ with a function $h \in C^2(\mathbf{R}^d)$, then

$$Lf(x_1, x_2) = \frac{1}{2}((\sqrt{G_1} - \sqrt{G_2})^2\nabla, \nabla)h(x_1 - x_2) + (b_1 - b_2, \nabla h)(x_1 - x_2)$$

$$+\int [h(x_1 - x_2 + y_1 - y_2) - h(x_1 - x_2) - (y_1 - y_2, \nabla h(x_1 - x_2)]\nu(dy_1dy_2). \quad (43)$$

Finally

$$\mathbf{E}(\xi + Y^1_t - Y^2_t)^2 = (\xi + t(b_1 - b_2))^2 + t\left(Tr(\sqrt{G_1} - \sqrt{G_2})^2 + \int\int (y_1 - y_2)^2\nu(dy_1dy_2)\right). \quad (44)$$

Proof. Straightforward. In fact, clearly Y_s couples Y^1_s, Y^2_s, because say $\eta_{x_1,x_2}(p_1, 0)$ is the characteristic exponent of Y^1_s. Equation (43) follows from (42). The second moment (44) is found either by twice differentiating the characteristic function, or by the Dynkin formula in conjunction with (43).

Similarly one obtains

Proposition A.2 *Let Y_s^i, $i=1,2$, be two Lévy processes in $\mathbf{R}^d$ specified by their generators*

$$L_i f(x) = (b_i, \nabla f(x)) + \int (f(x+y) - f(x))\nu_i(dy) \tag{45}$$

with $\nu_i \in \mathcal{M}_1(\mathbf{R}^d)$. Let $\nu \in \mathcal{M}_1(\mathbf{R}^{2d})$ be a coupling of ν_1, ν_2, i.e. (6) *holds for all ϕ_1, ϕ_2 satisfying $\phi_i(.)/|.| \in C(\mathbf{R}^d)$. Then the operator*

$$Lf(x_1,x_2) = (b_1, \nabla_1 f(x_1,x_2)) + (b_2, \nabla_2 f(x_1,x_2)) + \int [f(x_1+y_1, x_2+y_2) - f(x_1,x_2)]\nu(dy_1 dy_2) \tag{46}$$

specifies a Lévy process Y_s in $\mathbf{R}^{2d}$ that is a coupling of Y_s^1, Y_s^2 such that for all t

$$\mathbf{E}\|\xi + Y_t^1 - Y_t^2\| \le \|\xi\| + t\left(\|b_1 - b_2\| + \int\int \|y_1 - y_2\|\nu(dy_1 dy_2)\right). \tag{47}$$

Proof. One approximates $|y|$ by a smooth function, applies Dynkin's formula and then passes to the limit.

Next, let $Y_t(z)$ be a family of Lévy processes in $\mathbf{R}^d$ parametrized by points $z \in \mathbf{R}^d$ and specified by their generators

$$L[z]f(x) = \frac{1}{2}(G(z)\nabla, \nabla)f(x) + (b(z), \nabla f(x)) + \int (f(x+y) - f(x) - (\nabla f(x), y))\nu(z; dy) \tag{48}$$

where $\nu(z; .) \in \mathcal{M}_2(\mathbf{R}^d)$. We are interested in defining the process $\frac{\partial}{\partial z} Y_t(z)$.

We shall describe this process via a certain derivative type operator on Lévy measures connected with a coupling. Namely, let $\nu_{x_1,x_2}(dy_1 dy_2)$ be a family of $\mathcal{M}_2$-couplings of $\nu(x_1; .), \nu(x_2; , .)$ (in the sense that $\nu_{x_1,x_2} \in \mathcal{M}_2(\mathbf{R}^{2d})$ and (6) holds for all ϕ_1, ϕ_2 such that $\phi_i(.)/|.|^2 \in C(\mathbf{R}^d)$). For instance, these could be optimal couplings with respect to the cost function $(y_1 - y_2)^2$, i.e. those couplings, where the infinum in the definition of $W_2(\nu(x_1, .), \nu(x_2, .))$ is attained.

Let $T_h(y_1, y_2) = ((y_1 - y_2)/h, y_2)$ and the measure $\nu^{T_h}_{T_h^{-1}(\xi,x)}$ on $\mathbf{R}^{2d}$ be defined as the push forward of $\nu_{x+h\xi,x} = \nu_{T_h^{-1}(\xi,x)}$ by T_h, i.e.

$$\int\int f(z,y)\nu^{T_h}_{T_h^{-1}(\xi,x)}(dzdy) = \int\int f(\frac{y_1 - y_2}{h}, y_2)\nu_{x+h\xi,x}(dy_1 dy_2).$$

Clearly $\nu^{T_h}_{T_h^{-1}(\xi,x)}$ is a Lévy measure with a finite second moment whenever this is the case for $\nu_{x+h\xi,x}$. The relevant smoothness of ν will be defined now as the existence of the weak limit

$$D_\xi \nu_x = \lim_{h\to 0} \nu^{T_h}_{T_h^{-1}(\xi,x)},$$

i.e.

$$\lim_{h\to 0} \int\int g(y_1,y_2)\nu^{T_h}_{T_h^{-1}(\xi,x)}(dy_1 dy_2) = \int\int g(y_1,y_2) D_\xi \nu_x(dy_1 dy_2), \quad \frac{g(y_1,y_1)}{y_1^2 + y_2^2} \in C(\mathbf{R}^{2d}).$$

To see the rational behind this definition observe that if $\nu(x, .) = \nu^{F_x}$ with a given ν and a family of transformations $F_x(.)$, then

$$D_\xi \nu_x = \nu^{(\xi, \nabla F_x(.)), F_x(.)}$$

is the push forward of ν with respect to $y \mapsto (\xi, \nabla F_x(y)), F_x(y)$ (∇ is the derivative with respect to x). On the other hand, if $\nu_{z,x}$ has a density, i.e. $\nu_{z,x}(dy_1 dy_2) = \nu_{z,x}(y_1, y_2) dy_1 dy_2$. then $D_\xi \nu_x$ has the density

$$\lim_{h \to 0} h^d \nu_{x+h\xi,x}(y + hz, y).$$

If the coupling is optimal (is given by minimizers in the definition of the W_2- distance) this derivative is connected with the derivative of W_2 via the formula

$$\int |z|^2 D_\xi \nu_x(dzdy) = \left(\frac{d}{dh} |_{h=0} W_2(\nu(x + h\xi; .), \nu(x; .)) \right)^2.$$

We shall need further only the partial derivatives $D^i \nu_x = D_{e_i} \nu_x$ in the directions of the co-ordinate vectors e_i. The reason for introducing these derivatives lies in the observation that its action on Lévy measures corresponds to the derivation of Lévy processes. More precisely, the following holds.

Proposition A.3 *Let $Y_t(z)$ be the family of the Lévy processes in $\mathbf{R}^d$, $z \in \mathbf{R}^d$, specified by their generators (48). Suppose $G(x), b(x) \in C^1(\mathbf{R}^d)$ and $\nu(x, .)$ is smooth in the above sense (i.e. $D^j \nu$ are well defined with respect to a certain coupling). (i) Then the coupled random variables in $\mathbf{R}^{2d}$*

$$(h^{-1}(Y_t(x + he_j) - Y_t(x)), Y_t(x))$$

in $\mathbf{R}^{2d}$ has a weak limit that we denote $(\nabla_j Y_t(x), Y_t(x))$ and that has the distribution $Q^t_{D^j \nu(x)}$ of the Lévy process at time t with the characteristic exponent

$$\eta^j_x(q, p) = -\frac{1}{2} \left[\nabla_j \sqrt{G(x)} q + \sqrt{G(x)} p \right]^2 + i(\nabla_j b(x), q) + i(b(x), p)$$

$$+ \int (e^{iqz+ipy} - 1 - ipy - iqz)) D^j \nu_x(dzdy). \tag{49}$$

(ii) Moreover, if $g \in C^1_{Lip}(\mathbf{R}^{2d})$, then the partial derivatives $\nabla_j \mathbf{E} g(x, Y_t(x))$ exist and

$$\nabla_j \mathbf{E} g(x, Y_t(x)) = \int \left(\nabla_j g(x, y) + (\frac{\partial g}{\partial y}(x, y), z) \right) Q^t_{D^j \nu(x)}(dzdy) \tag{50}$$

(∇_j means the derivative with respect to the variable x).

Proof. (i) The characteristic exponent of the Lévy process $T_h(Y_t(T_h^{-1}(e_j, x))$ is

$$\eta^{j,h}_x(q, p) = -\frac{1}{2} \left[\sqrt{G(x + he_j)} \frac{q}{h} + \sqrt{G(x)}(p - \frac{q}{h}) \right]^2 + i(b(x + he_j), \frac{q}{h}) + i(b(x), (p - \frac{q}{h})$$

$$+ \int (e^{iy_1 q/h + iy_2(p-q/h)} - 1 - i(y_1 - y_2)\frac{q}{h} - ipy_2)) \nu_{T_h^{-1}(e_j,x)}(dy_1 dy_2),$$

which clearly converges to (49).

(ii) One has

$$\frac{1}{h} \left[\mathbf{E} g(x + he_j, Y_t(x + he_j)) - \mathbf{E} g(x, Y_t(x)) \right] = \frac{1}{h} \mathbf{E} \left[g(x + he_j, Y_t(x + he_j)) - g(x, Y_t(x)) \right],$$

where the last expectation can be taken with respect to any coupling of $Y_t(x+he_j)$ and $Y_t(x)$. Hence it can be written as

$$\int\int \left(\nabla_j g(x,y_2) + (\frac{\partial g}{\partial y}(x,y_2), \frac{y_1-y_2}{h}) + O(1)\frac{1}{h}(h^2\xi^2+(y_1-y_2)^2\right) P^t_{x+he_j,x}(dy_1dy_2).$$

By the property of the coupling (Proposition A.1) the term with $O(1)$ tends to zero as $h\to 0$. Consequently

$$\frac{d}{dh}\mid_{h=0} \mathbf{E}g(x+he_j, Y_t(x+he_j)) = \int \nabla_j g(x,y)P^t_x(dy)$$

$$+\lim_{h\to 0}\int\int (\frac{\partial g}{\partial y}(x,y_2), \frac{y_1-y_2}{h})P^t_{x+he_j,x}(dy_1dy_2),$$

implying (50) due to statement (i).

It is worth noting that statement (ii) implies that the distributions of the derivatives actually do not depend on coupling.

So far we have got only partial derivatives. We are now interested in their continuity which clearly is linked to the continuity of the measures $D^i\nu_x$. It turns out that the relevant notion of continuity is a bit finer than the W_2-continuity used above. Next two statements reveal two 'crucial bits' of this continuity.

Proposition A.4 *Under the assumptions of Proposition A.3 assume additionally that $G(x), b(x) \in C^1_{Lip}(\mathbf{R}^d)$ and that the Lévy measures $D^j\nu_x$ are Lipschitz continuous in the following sense: for any $x_1, x_2 \in \mathbf{R}^d$ and $j=1,...,d$ there exists a Lévy coupling $D^j(x_1,x_2)$ of the Lévy measures $D^j\nu_{x_1}$, $D^j\nu_{x_2}$ such that*

$$\int_{\mathbf{R}^{4d}} \left[(y_1-y_2)^2(1+z_1^2+z_2^2)+(z_1-z_2)^2\right] D^j(x_1,x_2)(dz_1dz_2dy_1dy_2) \le \kappa(x_1-x_2)^2 \quad (51)$$

with a constant κ. Let $Q^t_{D^j(x_1,x_2)}$ denote the distribution at time t of the Lévy process that couples $(\nabla_j Y_t(x_1), Y_t(x_1))$ and $(\nabla_j Y_t(x_2), Y_t(x_2))$ according to Proposition A.1, i.e. the Lévy process in $\mathbf{R}^{4d}$ specified by the characteristic exponent

$$\eta^j_{x_1,x_2}(q_1,q_2,p_1,p_2) = -\frac{1}{2}\left[\nabla_j\sqrt{G(x_1)}q_1 + \nabla_j\sqrt{G(x_2)}q_2 + \sqrt{G(x_1)}p_1 + \sqrt{G(x_2)}p_2\right]^2$$

$$+i(\nabla_j b(x_1)q_1 + \nabla_j b(x_2)q_2 + b(x_1)p_1 + b(x_2)p_2)$$

$$+\int (e^{iy_1p_1+iy_2p_2+iz_1q_1+iz_2q_2} - 1 - i(y_1p_1+y_2p_2+z_1q_1+z_2q_2)D^j_{x_1,x_2}(dz_1dz_2dy_1dy_2).$$

Then for any $\xi \in \mathbf{R}^d$

$$\int_{\mathbf{R}^{4d}} \left[(\xi+y_1-y_2)^2(1+z_1)^2+(z_1-z_2)^2\right] Q^t_{D^j(x_1,x_2)}(dz_1dz_2dy_1dy_2) \le \xi^2 + ct(\xi^2+(x_1-x_2)^2) \quad (52)$$

with a constant c uniformly for finite times, and for any $g \in C^1_{Lip}(\mathbf{R}^{2d})$ the function $\mathbf{E}g(x, Y_t(x))$ belongs to $C^1_{Lip}(\mathbf{R}^d)$ (also uniformly for finite times).

Proof. The moment estimates (52) are obtained directly from the derivatives of the characteristic function as in Proposition A.1. For the second statement we write

$$|\nabla_j \mathbf{E} g(x_1, Y_t(x_1)) - \nabla_j \mathbf{E} g(x_1, Y_t(x_1))| \le \int_{\mathbf{R}^{4d}} Q^t_{D^j(x_1,x_2)}(dz_1 dz_2 dy_1 dy_2)$$

$$\times |\nabla_j g(x_1, y_1) - \nabla_j g(x_2, y_2) + (\frac{\partial g}{\partial y}(x_1, y_1), z_1) + (\frac{\partial g}{\partial y}(x_2, y_2), z_2)|$$

(the derivative ∇_j with respect to x), which does not exceed

$$\int_{\mathbf{R}^{4d}} ((|x_1 - x_2| + |y_1 - y_1|(1 + |z_1| + |z_2|) + |z_1 - z_2|) Q^t_{D^j(x_1,x_2)}(dz_1 dz_2 dy_1 dy_2),$$

and which in turn does not exceed $\sqrt{t}|x_1 - x_2|$ due to (52) and the Hölder inequality.

In case $\nu(x, .) = \nu^{F_x}$ for a family of transformations $F_x(.)$ the coupling $D^j(x_1, x_2)$ can be obtained as

$$\int f(z_1, z_2, y_1, y_2) D^j(x_1, x_2)(dz_1 dz_2 dy_1 dy_2)$$

$$= \int f(\nabla_j F(x_1, y), \nabla_j F(x_2, y), F(x_1, y), F(x_2, y)) \nu(dy),$$

and the condition (51) is fulfilled whenever the derivatives $\frac{\partial}{\partial x} F(x, y)$ are bounded and Lipschitz continuous.

By $D\nu_x$ we shall denote the vector $\{D^j \nu_x\}$ and by $Q^t_{D\nu(x)}$ the vector $\{Q^t_{D^j \nu(x)}\}$, $j = 1, ..., d$.

Proposition A.5 *Under the assumptions of Proposition A.3 assume additionally that* $G(x), b(x) \in C^1_{Lip}(\mathbf{R}^d)$ *and that the function*

$$\int \int (y_1 - y_2, e_j)(e^{iy_2 p} - 1) \nu_{x,z}(dy_1 dy_2)$$

is differentiable in x *around* $x = z$ *with uniform estimates, more precisely that*

$$\int \int (y_1 - y_2, e_j)(e^{iy_2 p} - 1) \nu_{x,z}(dy_1 dy_2)$$

$$= \left(\frac{\partial}{\partial x} |_{x=z} \int \int (y_1 - y_2, e_j)(e^{iy_2 p} - 1) \nu_{x,z}(dy_1 dy_2), x - z \right) + O(1 + |p|)(x - z)^2. \quad (53)$$

Then for a continuous function g *represented via the inverse Fourier transform as*

$$g(y) = \int e^{iyp} \hat{g}(p)\, dp, \quad (1 + |p|)\hat{g}(p) \in (L^1(\mathbf{R}^d))^d,$$

one has the estimate

$$\mathbf{E}(Y_t(x) - Y_t(z), g(Y_t(z))) = \int (y_1 - y_2, g(y_2)) P^\tau_{x,z}(dy_1 dy_2)$$

$$= \int \int (w, g(y))(Q^t_{D\nu(z)}(dw dy), x - z) + O(t)(x - z)^2 \int (1 + |p|)|\hat{g}(p)|\, dp. \quad (54)$$

Proof. Comparing the r.h.s of (53) with the definition of $D\nu_x$ yields

$$\int\int (y_1-y_2,e_j)(e^{iy_2p}-1)\nu_{x,z}(dy_1dy_2)$$

$$=\left(\int\int (w,e_j)(e^{iyp}-1)D\nu_x(dwdy),x-z\right)+O(1+|p|)(x-z)^2. \qquad (55)$$

Now one has

$$\int (y_1-y_2,e_j)e^{iy_2p}P^t_{x,z}(dy_1dy_2)$$

$$=-i\frac{\partial}{\partial q^j}\mid_{q=0}\mathbf{E}\exp\{i(Y_t(x)-Y_t(z))q+iY_t(z)p\}=-i\frac{\partial}{\partial q^j}\mid_{q=0}\exp\{t\eta_{x,z}(q,p-q)\}$$

$$=t\left[i(\sqrt{G(z)}(\sqrt{G(x)}-\sqrt{G(z)})p)^j+(b(x)-b(z))^j+\int (y_1-y_2)^j(e^{ipy_2}-1)\nu_{x,z}(dy_1dy_2)\right]\mathbf{E}e^{iY_t(z)p}$$

$$=t\left(\frac{i}{2}(\nabla(G(z)p)^j+\nabla b^j(z)+\int\int w^j(e^{ipy}-1)(D\nu_x(dwdy),x-z\right)\mathbf{E}e^{iY_t(z)p}+O(t)(1+|p|)(x-z)^2.$$

Consequently,

$$\int (y_1-y_2,g(y_2))P^t_{x,z}(dy_1dy_2)$$

$$=t\int\left(\frac{i}{2}\nabla(G(z)p,\hat g(p))+\nabla(b(z),\hat g(p))+\int\int (w,\hat g(p))(e^{ipy}-1)D_{\nu_x}(dwdy),x-z\right)\mathbf{E}e^{iY_t(z)p}\,dp$$

$$+O(t)\int(1+|p|)\hat g(p)\,dp(x-z)^2.$$

Similarly

$$\int\int w^je^{ipy}Q^t_{D\nu(z)}(dwdy)=-i\frac{\partial}{\partial q^j}\mid_{q=0}$$

$$\exp\{t[-\frac{1}{2}(\nabla\sqrt{G(z)}q+\sqrt{G(z)}p]^2+i(\nabla b(z)q+b(z)+\int\int (e^{iqw+ipy}-1-ipy-iqw)D\nu_x(dwdy)\}$$

$$=t\left(\frac{i}{2}\nabla(G(z)p)^j+\nabla b^j(z)+\int\int w^j(e^{ipy}-1)D\nu_x(dwdy)\right)\mathbf{E}e^{iY_t(z)p},$$

implying (54).

To differentiate the Lévy process for the second time, one needs of course the 'second derivative' of the Lévy measure defined similarly to the first one. Namely, one needs the existence of the limit

$$\lim_{h\to0}\int f(\frac{z_1-z_2}{h},\frac{y_1-y_2}{h},z_2,y_2)D^j(x+he_k,x)(dz_2dz_1dy_2dy_1)=\int f(w,z_k,z_j,y)D^{kj}_x(dwdz_jdz_kdy) \qquad (56)$$

whenever $f(w,z_k,z_j,y)/(w^2+z_j^2+z_k^2+y^2)\in C(\mathbf{R}^{4d})$ with $D^{kj}_x(dwdz_jdz_kdy)$ belonging to $\mathcal{M}_2(\mathbf{R}^{4d})$. The following is a straightforward analog of Proposition A.3.

Proposition A.6 *Under the assumptions of Proposition A.3 assume that* $G(x), b(x) \in C^2(\mathbf{R}^d)$ *and the measures* $D_x^{kj} \in \mathcal{M}_2(\mathbf{R}^{4d})$ *are well defined by* (56). *(i) Then for any* j, k *the process*

$$(\nabla_k \nabla_j Y_t(x), \nabla_k Y_t(x), \nabla_j Y_t(x), Y_t(x))$$

is defined weakly in $\mathbf{R}^{4d}$ *and has the distribution* $Q^t_{D^{jk}\nu(x)}$ *of the Lévy process at time* t *with the characteristic exponent*

$$\eta_x^{jk}(r, q_k, q_j, p) = -\frac{1}{2}\left[\nabla_k \nabla_j \sqrt{G(x)}q + \nabla_k \sqrt{G(x)}q_k + \nabla_j \sqrt{G(x)}q_j + \sqrt{G(x)}p\right]^2$$

$$+i[(\nabla_k \nabla_j b(x), q) + (\nabla_k b(x), q_k) + (\nabla_j b(x), q_j) + (b(x)p)]$$

$$+\int [e^{irw+iq_k z_k+iq_j z_j+ipy} - 1 - i(rw + q_k z_k + q_j z_j + py)] D_x^{kj}(dw dz_k dz_j dy). \qquad (57)$$

(ii) Moreover, if $g \in C^2_{Lip}(\mathbf{R}^d)$*, the partial derivatives* $\nabla_k \nabla_j \mathbf{E} g(x + Y_t(x))$ *exist and*

$$\nabla_k \nabla_j \mathbf{E} g(x + Y_t(x)) = \int Q^t_{D^{kj}\nu(x)}(dw dz_k dz_j dy)$$

$$\left[\sum_{l,m=1}^d \nabla_m \nabla_l g(x+y)(\delta_k^m + z_k^m)(\delta_j^l + z_j^l) + \sum_{l=1}^d \nabla g(x+y) w^l\right] \qquad (58)$$

(∇ means the derivative with respect to the variable x).

Acknowledgments. The author is grateful to the referee whose constructive critique of the first draft helped to improve essentially the quality of the exposition.

References

[1] D. Applebaum. Lévy Processes and Stochastic Calculus. Cambridge studies in advanced mathematics, v. 93. Cambridge Univ. Press, 2004.

[2] R.F. Bass. Uniqueness in law for pure jump type Markov processes. Probab. Theory Related Fields **79** (1988), 271-287.

[3] R.F. Bass, Z.-Q. Chen. Systems of equations driven by stable processes. Probab. theory Relat. Fields **134** (2006), 175-214.

[4] K. Bichteler, J.-B. Gravereaux, J. Jacod. Malliavin Calculus for Processes with Jumps. Stochastic Monographs v. 2. Gordon and Breach 1987.

[5] R.A. Carmona, D.Nualart. Nonlinear Stochsatic Integrators, Equations and Flows. Stochatic Monographs, v. 6, Gordon and Breach, 1990.

[6] H. Guérin, S. Méléard, E. Nualart. Estimates for the density of a nonlinear Landau process. Journal of Functional Analysis **238** (2006), 649-677.

[7] N. Jacob. Pseudo-differential Operators and Markov Processes. Vol. I,II,III. London: Imperial College Press, 2001, 2002, 2005.

[8] O. Kallenberg. Foundations of Modern Probability. Second Ed., Springer 2002.

[9] A.N. Kochubei. Parabolic pseudo-differentiable equation, supersingular integrals and Markov processes. Izvestia Akad. Nauk, Ser. Matem. **52:5** (1988), 909-934 (Russian), Math.-USSR Izv. **33:2** (1989), 233-259 (English).

[10] V. N. Kolokoltsov. Symmetric Stable Laws and Stable-Like Jump-Diffusions. Proc. London Math. Soc. **3:80** (2000), 725-768.

[11] V. N. Kolokoltsov. Semiclassical Analysis for Diffusions and Stochastic Processes. Springer Lecture Notes in Math. v. 1724, 2000.

[12] V. Kolokoltsov. On Markov processes with decomposable pseudo-differential generators. Stochastics and Stochastics Reports **76:1** (2004), 1-44.

[13] V. Kolokoltsov. Nonlinear Markov Semigroups and Interacting Lévy Type Processes. Journ. Stat. Physics **126:3** (2007), 585-642.

[14] V. N. Kolokoltsov. Generalized Continuous-Time Random Walks (CTRW), Subordination by Hitting Times and Fractional Dynamics. arXiv:0706.1928v1[math.PR] 13 June 2007. To appear in Prob. Theory Appl.

[15] V. Kolokoltsov. Nonlinear Markov Processes and Kinetic Equations. Monograph. To appear in CUP.

[16] H. Kunita. Stochastic Flows and Stochastic Differential Equations. Cambridge studies in advanced mathematics, v. 24. Cambridge Univ. Press, 1990.

[17] H.P. McKean. A class of Markov processes associated with nonlinear parabolic equations. Proc. Nat. Acad. Sci. **56** (1966), 1907-1911.

[18] S. Rachev, L. Rüschendorf. Mass Transportation Problems. V. I: Theory, V. II: Applications. Springer-Verlag, N.Y. 1988.

[19] D.W. Stroock. Markov Processes from K. Ito's Perspective. Annals of Mathematics Studies. Princeton University Press 2003.

[20] A.-S. Sznitman. Topics in Propagation of Chaos. In: Ecole d'Eté de Probabilités de Saint-Flour XIX-1989. Springer Lecture Notes Math. 1464 (1991), 167-255.

[21] K. Taira. On the Existence of Feller Semigroups with Boundary Conditions. Memoirs of the American Mathematical Society, v. 99, N. 475, 1992.

[22] C. Villani. Topics in Optimal Transportation. Graduates Studies in Mathematics 58, AMS 2003.