SOME CONSIDERATIONS ON THE STRUCTURE OF TRANSITION DENSITIES OF SYMMETRIC LÉVY PROCESSES

LEWIS J. BRAY AND NIELS JACOB

Abstract. For a class of symmetric Lévy processes \((Y_t)_{t \geq 0}\) with characteristic exponent \(\psi\) we show that 
\[ \nu_t = e^{-\frac{1}{t} \psi(\cdot)/p_t(0)} \] 
\(t > 0\), gives rise to an additive process \((X_t)_{t \geq 0}\) with \(t\)-dependent characteristic exponent 
\[ -\frac{d}{dt} \ln(p_t(\xi)/p_t(0)) \] 
where \((p_t)_{t > 0}\) are the transition densities of \((Y_t)_{t \geq 0}\).

We estimate (from above and below) \(p_t\) in terms of two metrics \(\delta_{\psi,t}\) and \(d_{\psi,t}\) controlling \(p_t(0)\) and \(\delta_{\psi,t}\) the spatial decay, and we prove that the transition density \(\pi_{t,0}\) of \(P_{X_t-X_0}\) is controlled by \(\delta_{\psi,t}\) and \(d_{\psi,t}\) now with \(\delta_{\psi,t}\) controlling \(\pi_{t,0}(0)\) and \(d_{\psi,t}\) the spatial decay.

1. Introduction

In the mid 1980’s it became clear that heat kernels of second order elliptic partial differential operators are best understood in terms of the underlying Riemannian geometry, see [4] and [5] as seminal contributions. Further investigation led to the concept of metric measure spaces associated with (local) Dirichlet forms, and now geometry is used to construct corresponding diffusions and their generators, see [21], and for more recent developments [7, 8, 9].

We bypass the work of the E.M.Stein school on the corresponding sub-elliptic problem which led to the emergence of sub-Riemannian geometry as we want to emphasise the efforts in [3] to extend the programme to non-local generators of Markovian semigroups. The crucial role of the carré du champ operator, see in particular [16], was highlighted, however to our best knowledge so far this programme has not led to the desired results.

In studying concrete transition densities of symmetric Lévy processes as well as the observation that the characteristic exponent of such a process often led to a metric measure space, the second named author suggested to use this metric to study the transition densities. In [15] for the diagonal term a first result could be proved, and in [14] a suggestion of a theory was outlined to interpret the transition density of certain symmetric Lévy processes in terms of two time-dependent metrics \(d_{\psi,t}\) and \(\delta_{\psi,t}\) where \(\psi\) is the characteristic exponent of the Lévy process. Assuming volume doubling for the metric measure space \((\mathbb{R}^n, d_{\psi,t}, \lambda^{\alpha})\) the result

Received 2016-7-13; Communicated by D. Applebaum.

2010 Mathematics Subject Classification. 60G51; 60J75; 47D07; 31E05; 51F99.

Key words and phrases. Lévy processes, additive processes, transition density, metric measure spaces.
reads as,
\[ p_t(x - y) \asymp \lambda^{(n)}(B^{d_{\psi,t}}(0,1))e^{-\delta_{\psi,t}(x,y)}, \]
where \( B^{d_{\psi,t}}(0,1) \) denotes the unit ball with respect to \( d_{\psi,t} \) and \( a_t \asymp b_t \) means for two constants \( 0 < \gamma_0 < \gamma_1 \) that \( \gamma_0 a_t \leq b_t \leq \gamma_1 a_t \). While in [14] the existence of \( \delta_{\psi,t} \) could only be proved for some classes, the estimate,
\[ p_t(0) \asymp \lambda^{(n)}(B^{d_{\psi,t}}(0,1)), \]
holds always under the doubling condition and we can write,
\[ p_t(x - y) = p_t(0) \frac{p_t(x - y)}{p_t(0)} = p_t(0)e^{-\left(-\ln \frac{p_t(x - y)}{p_t(0)}\right)}, \]
In the case we can show that,
\[ x \mapsto -\ln \frac{p_t(x)}{p_t(0)}, \]
is a continuous negative definite function, see [13, Vol I.] for the definition (but note that the continuous negative definite function is just another name for the characteristic exponent of a Lévy process), our conditions will imply that,
\[ \delta_{\psi,t}(x,y) = \left(-\ln \frac{p_t(x - y)}{p_t(0)}\right)^{1/2}, \]
is a metric. So far we do not know a general result of this type, but plenty of examples, see [2] or [14]. However even for the transition density of the relativistic Hamiltonian process associated with \( \psi(\xi) = (|\xi|^2 + m)^{1/2} - m \), see [10], the problem is still open. On the other hand, using subordination in the sense of Bochner, see as a general reference [1] or [19], new examples can be constructed. It is helpful to note in this context that for subordination a good functional calculus is available [18], and that certain functional inequalities are stable under subordination, see [20], since functional inequalities are useful tools to handle transition densities. We want to also mention that in [6] the suggested approach was tested for \( Q \)-matrices with state space \( \mathbb{Z}^n \) relying much on commutative harmonic analysis. It would be of interest to extend these ideas to locally compact groups since they allow a corresponding harmonic analysis and are well studied objects in probability theory, see H. Heyer [11].

In this paper we start with an obvious observation. If \((\mu_t)_{t \geq 0}\) is a convolution semigroup of probability measures on \( \mathbb{R}^n \) with Fourier transform,
\[ \hat{\mu}_t(\xi) = (2\pi)^{-n/2}e^{-t\psi(\xi)}, \]
then by,
\[ \rho_t(dx) := \frac{e^{-t\psi(x)}}{(2\pi)^n p_t(0)} \, dx, \]
a family of probability measures \((\rho_t)_{t \geq 0}\) is given. Using a ratio limit result as proved in [14] it turns out that the family,
\[ \nu_t := \rho_t^+, \quad t > 0, \]
is for \( t \to 0 \) weakly continuous and it gives rise to a family of strongly continuous convolution operators \((S_t)_{t \geq 0}\) which are contractions in either \( (C_\infty(\mathbb{R}^n), \| \cdot \|_\infty) \) or
(\mathcal{L}^2(\mathbb{R}^n), \| \cdot \|_0)$ and are either positivity preserving or sub-Markovian. However, in general $(S_t)_{t \geq 0}$ is not a semigroup. We can show that with,

$$q(t, \xi) := -\frac{\partial}{\partial t} \ln \frac{p_{\xi t}(0)}{p_{\xi 0}(0)},$$

for $u \in \mathcal{S}(\mathbb{R}^n)$ we get,

$$\frac{\partial}{\partial t} S_{t} u + q(t, D) S_{t} u = 0, \quad \lim_{t \to 0} S_{t} u = u,$$

where $q(t, D)$ is the pseudo-differential operator with symbol $q(t, \xi)$. We give examples for $\xi \mapsto q(t, \xi)$ being a continuous negative definite function. Under the assumption that $q : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}$ is continuous and $\xi \mapsto q(t, \xi)$ is a negative definite function we prove further that we can associate with $q(t, \xi)$ an additive process $(X_t)_{t \geq 0}$ with,

$$p_{(X_t - X_s)} = \gamma_{t,s}, \quad 0 \leq s < t,$$

where,

$$\hat{\gamma}_{t,s}(\xi) = (2\pi)^{-n/2}e^{-\int_t^s q(\tau, \xi) d\tau}.$$  

Studying the density $\pi_{t,0}$ of $P_{X_t - X_0}$ in comparison with the density $p_t$ of $(Y_t)_{t \geq 0}$ we obtain (under some additional assumptions, see Theorem 5.1) our main result:

$$p_t(x - y) \asymp \lambda^n(B_{\psi,t}(0,1)) e^{-\delta_{\psi,t}(x,y)},$$

and,

$$\pi_{t,0}(x - y) \asymp \lambda^n(B_{\psi,t}(0,1)) e^{-\left(\frac{d_{\psi,t}(x,y)}{\sqrt{t\psi(x-y)}}\right)^{1/2}},$$

with $d_{\psi,t}(x,y) = \left( -\ln \frac{p_{\psi t}(0)}{p_{t}(0)} \right)^{1/2}$.

We refer to [2], but also [14], where attempts were made to extend the results to processes generated by pseudo-differential operators with symbol $q(t, x, \xi)$ and the $x$-dependence is subjected to oscillation conditions with respect to a reference function as in [12].

In general, our notations are the ones used in [13].

2. Families of Measures Associated with Convolution Operators

Let $(\mu_t)_{t \geq 0}$ be a symmetric convolution semigroup of probability measures, i.e., each $\mu_t$ is a probability measure on $\mathbb{R}^n$, $\mu_0 = \epsilon_0$, $\mu_s \ast \mu_t = \mu_{s+t}$, and $\mu_t \to \epsilon_0$ vaguely, hence weakly, for $t \to 0$, with Fourier transform,

$$\hat{\mu}_t(\xi) = (2\pi)^{-n/2}e^{-t\psi(\xi)},$$

where $\psi : \mathbb{R}^n \to \mathbb{R}$ is a continuous negative definite function. Since all measures $\mu_t$ are assumed to be probability measures it follows that $\psi(0) = 0$. We add the following assumptions on $\psi$:

Ai) $\psi(\xi) = 0$ if and only if $\xi = 0$;
Aii) $\lim\inf_{|\xi| \to \infty} \psi(\xi) > 0$;
Aiii) $e^{-\psi}, \psi e^{-\psi} \in L^1(\mathbb{R}^n)$ for all $t > 0$. 


Note that if $f$ is a Bernstein function growing as a power at infinity, e.g. $f(s) = s^\alpha, 0 < \alpha < 1$, then $f(|\xi|^2)$ as well as $f(|\xi|^\alpha_1 + |\xi|^\alpha_2), \alpha_j \in (0, 2]$, will satisfy these assumptions. More examples are given in [14]. From condition Ai) we deduce, compare Lemma 3.6.21 in [13, Vol I.], that by,

$$d_\psi(\xi, \eta) := \psi^{1/2}(\xi - \eta), \quad (2.2)$$

a metric is given on $\mathbb{R}^n$ and Ai) assures, see Lemma 3.2 in [14], that the metric $d_\psi$ generates on $\mathbb{R}^n$ the Euclidean topology. Furthermore, by Aii) the measures $\mu_t$ have a density $p_t \in C^\infty(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ with respect to the Lebesgue measure $\lambda^{(n)}$ given by,

$$p_t(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} e^{-t\psi(\xi)} d\xi, \quad (2.3)$$

and of course we have,

$$\int_{\mathbb{R}^n} p_t(x) dx = \int_{\mathbb{R}^n} 1 d\mu_t = 1. \quad (2.4)$$

Furthermore we find that

$$\frac{\partial p_t}{\partial t}(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{ix\cdot\xi} \left(-\psi(\xi)\right) e^{-t\psi(\xi)} d\xi,$$

exists. Note that in the case that $\psi$ has at least power growth for $|\xi| \to \infty$ the condition $\psi e^{-t\psi} \in L^1(\mathbb{R}^n)$ is trivial. Thus with $\psi$ (or the convolution semigroup $(\mu_t)_{t \geq 0}$ or the corresponding canonical Lévy process $(X_t)_{t \geq 0}$) we can associate a metric measure space $(\mathbb{R}^n, d_\psi, \lambda^{(n)})$, see [14]. In the next section we will employ this metric measure space to study $p_t$. Observe that,

$$p_t(0) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-t\psi(\xi)} d\xi, \quad (2.5)$$

and therefore it follows that,

$$\rho_t := \frac{e^{-t\psi(\cdot)}}{(2\pi)^n p_t(0)} \lambda^{(n)}; \quad (2.6)$$

is for $t > 0$ a symmetric probability measure on $\mathbb{R}^n$ with Fourier transform,

$$\hat{\rho}_t(y) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-iy\cdot\xi} \frac{e^{-t\psi(\xi)}}{(2\pi)^n p_t(0)} d\xi = (2\pi)^{-n/2} \frac{p_t(y)}{p_t(0)}. \quad (2.7)$$

We now introduce the family of measures $(\nu_t)_{t > 0}$ by,

$$\nu_t := \rho_t^\perp = \frac{e^{-\frac{t}{2}\psi(\cdot)}}{(2\pi)^n p_t^\perp(0)} \lambda^{(n)}. \quad (2.8)$$

First we note that,

$$\nu_t(\mathbb{R}^n) = \int_{\mathbb{R}^n} 1 \nu_t(d\xi) = \int_{\mathbb{R}^n} e^{-\frac{t}{2}\psi(\xi)} \lambda^{(n)}(d\xi) = \frac{(2\pi)^{-n} \int_{\mathbb{R}^n} e^{-\frac{t}{2}\psi(\xi)} \lambda^{(n)}(d\xi)}{p_t^\perp(0)} = \frac{p_t^\perp(0)}{p_t^\perp(0)} = 1,$$

i.e., $\nu_t$ is a probability measure on $\mathbb{R}^n$. Further we have,
Proposition 2.1. For \( t \to 0 \) the family of measures \( (\nu_t)_{t>0} \) converges weakly to the Dirac measure \( \varepsilon_0 \), i.e.,

\[
\lim_{t \to 0} \nu_t = \varepsilon_0 \quad \text{(weak limit)}.
\]

Proof. Let \( \varphi \in C_0^\infty(\mathbb{R}^n) \). By Plancherel’s theorem we find,

\[
\int_{\mathbb{R}^n} \varphi(x-y) \nu_t(dy) = \int_{\mathbb{R}^n} \frac{e^{-\frac{1}{2} \psi(y)}}{(2\pi)^n p^+_t(0)} \varphi(x-y) dy = \frac{1}{(2\pi)^{n/2} p^+_t(0)} \int_{\mathbb{R}^n} F^{-1}((2\pi)^{-n/2} e^{-\frac{1}{2} \psi(1)}(\xi)) F^{-1}(\varphi(x-\cdot))(\xi) d\xi,
\]

and since \( F^{-1}(\varphi(x-\cdot))(\xi) = e^{-ix\cdot \xi} F^{-1}(\varphi(\xi)) \) we obtain,

\[
\int_{\mathbb{R}^n} \varphi(x-y) \nu_t(dy) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \frac{p^+_t(\xi)}{p^+_t(0)} e^{-ix\cdot \xi} (F^{-1}(\varphi))(\xi) d\xi.
\]

By the ratio limit theorem, Theorem 5.7 in [15], it holds for the transition density \( \pi_t \) of a Lévy process on \( \mathbb{R}^n \) that \( \lim_{t \to \infty} \frac{\pi_t(x)}{\pi_t(0)} = 1 \) for all \( x \in \mathbb{R}^n \). Passing in (2.10) to the limit \( t \to 0 \) we get,

\[
\lim_{t \to 0} \int_{\mathbb{R}^n} \varphi(x-y) \nu_t(dy) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-ix\cdot \xi} \left( \lim_{t \to 0} \frac{p^+_t(\xi)}{p^+_t(0)} \right) (F^{-1}(\varphi))(\xi) d\xi = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-ix\cdot \xi} \left( \lim_{t \to \infty} \frac{p^+_t(\xi)}{p^+_t(0)} \right) (F^{-1}(\varphi))(\xi) d\xi = \varphi(x).
\]

The density of \( C_0^\infty(\mathbb{R}^n) \) in \( C_\infty(\mathbb{R}^n) \) implies \( \lim_{t \to 0} \nu_t = \varepsilon_0 \) vaguely and since \( \lim_{t \to 0} \nu_t(\mathbb{R}^n) = 1 \) it follows that (2.9) holds.

Hence the family \( (\nu_t)_{t>0} \) is a family of probability measures converging weakly to \( \varepsilon_0 \). An open question is when they form a projective family or when we can associate with \( (\nu_t)_{t>0} \) a stochastic process. To investigate the situation further we want to switch from \( (\nu_t)_{t>0} \) to the corresponding family of operators \( (S_t)_{t>0} \) defined on \( C_b(\mathbb{R}^n) \) by,

\[
S_t u(x) := (u * \nu_t)(x) = \int_{\mathbb{R}^n} u(x-y) \nu_t(dy).
\]

For \( u \in \mathcal{S}(\mathbb{R}^n) \) we find by the convolution theorem when noting that \( S_t u = F^{-1}(F(u * \nu_t)) \) that,

\[
S_t u(x) = \int_{\mathbb{R}^n} e^{ix\cdot \xi} \hat{\nu_t}(\xi) \hat{u}(\xi) d\xi = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{ix\cdot \xi} \frac{p^+_t(\xi)}{p^+_t(0)} \hat{u}(\xi) d\xi,
\]
or with,
\[ \sigma_t(\xi) := \frac{p_{1t}(\xi)}{p_{1t}(0)}, \] (2.12)
we have on \( S(\mathbb{R}^n) \),
\[ S_t u(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{ix \cdot \xi} \sigma_t(\xi) \hat{u}(\xi) \, d\xi, \] (2.13)
i.e., \( S_t \) is a pseudo-differential operator with symbol \( \sigma_t(\xi) \). Of course \((S_t)_{t>0}\) is in general not a semigroup of linear operators. However from Proposition 2.1 and (2.11) we deduce that \((S_t)_{t>0}\) is a family of contractions on \( C_\infty(\mathbb{R}^n) \), i.e.,
\[ \|S_t u\|_\infty \leq \|u\|_\infty, \] (2.14)
which is strongly continuous, i.e.,
\[ \lim_{t \to 0} \|S_t u - u\|_\infty = 0, \] (2.15)
for \( u \in S(\mathbb{R}^n) \) and hence for \( u \in C_\infty(\mathbb{R}^n) \). Furthermore,
\[ u \geq 0 \text{ implies } S_t u \geq 0, \] (2.16)
i.e., \((S_t)_{t>0}\) is on \( C_\infty(\mathbb{R}^n) \) a strongly continuous family of positivity preserving contractions. Moreover, using the density of \( S(\mathbb{R}^n) \) in \( L^2(\mathbb{R}^n) \), Plancherel’s theorem and \( 0 \leq \sigma_t(\xi) \leq 1 \) we deduce for \( u \in L^2(\mathbb{R}^n) \),
\[ \|S_t u\|_0 \leq \|u\|_0, \] (2.17)
\[ \lim_{t \to 0} \|S_t u - u\|_0 = 0, \] (2.18)
and,
\[ 0 \leq u \leq 1 \text{ a.e. implies } 0 \leq S_t u \leq 1 \text{ a.e.}, \] (2.19)
which means that \((S_t)_{t>0}\) is a strongly continuous family of sub-Markovian contraction on \( L^2(\mathbb{R}^n) \).

Proposition 2.2. Let
\[ q(t, \xi) = -\frac{\partial}{\partial t} \ln \frac{p_{1t}(\xi)}{p_{1t}(0)}. \] (2.20)
Then it follows for \( u \in S(\mathbb{R}^n) \) that,
\[ \frac{\partial}{\partial t} S_t u(x) + q(t, D) S_t u(x) = 0, \] (2.21)
anid,
\[ \lim_{t \to 0} S_t u = u, \] (2.22)
where \( q(t, D) \) is the pseudo-differential operator with the time-dependent symbol \( q(t, \xi) \), and the limit in (2.22) can be taken in \( C_\infty(\mathbb{R}^n) \), hence also pointwise for \( u \in C_\infty(\mathbb{R}^n) \), or in \( L^2(\mathbb{R}^n) \).

Proof. It remains to prove (2.21) which follows for \( u \in S(\mathbb{R}^n) \) by differentiating (2.13). Note that by (2.3) \( t \mapsto p_{1t}(\xi) \) is differentiable for \( t > 0 \) and every \( \xi \in \mathbb{R}^n \). \( \square \)
Here we encounter a further open problem: As the Fourier transform of the measure $\sigma_t$ is for every $t > 0$ a continuous positive definite function, we are searching for conditions implying that $q(t, \cdot)$ is a continuous negative definite function. Note that formally we expect $\sigma_t(\xi) = ce^{-\int_0^t q(\tau, \xi) d\tau}$ to hold, hence for $q(\tau, \cdot)$ negative definite we would obtain $\sigma_t(\cdot)$ positive definite and $\int_0^t q(\tau, \cdot) d\tau$ would be a type of characteristic exponent.

**Example 2.3.**

A. For Brownian motion in $\mathbb{R}^n$ we have $\psi_B(\xi) = \frac{1}{2}|\xi|^2$ with

$$p_t^B(x) = (2\pi t)^{-n/2} e^{-\frac{|x|^2}{2t}}$$

which yields $q_B(t, \xi) = \frac{1}{2}|\xi|^2$.

B. For the Cauchy process in $\mathbb{R}^n$ we have $\psi_C(\xi) = |\xi|$ with

$$p_t^C(x) = \pi^{-\frac{n+1}{2}} \Gamma\left(\frac{n}{2}\right) t^{-\frac{n}{2}} e^{-\frac{|x|^2}{2t} + \frac{|\xi|^2}{2t} + \frac{1}{t}}$$

which yields,

$$q_C(t, \xi) = -\frac{\partial}{\partial t} \ln \frac{p_t^C(\xi)}{p_t^C(0)} = \frac{\partial}{\partial t} \ln \left(1 + t^2|\xi|^2\right)^{\frac{n+1}{2}}$$

$$= \frac{n+1}{t} \left|\xi|^2 \right|^2 + \frac{n+1}{t} f_{t}(|\xi|^2),$$

where $f_{t}(s) = \frac{s}{e^s + 1}$ is a Bernstein function, hence $q_C(t, \cdot)$ is a continuous negative definite function.

**Example 2.4.** (See [2]) The symmetric Meixner process on $\mathbb{R}$ has the symbol $\psi_M(\xi) = \ln \cosh \xi$ and the transition density $p_t^M(x) = 2^{t^{-1}} \Gamma\left(\frac{t+i\xi}{2}\right)^2$ and we find, see [14],

$$\frac{p_t^M(\xi)}{p_t^M(0)} = \left|\Gamma\left(\frac{t+i\xi}{2}\right)\Gamma\left(\frac{t-i\xi}{2}\right)\right|^2 = \prod_{j=0}^{\infty} \left(1 + \frac{\xi^2}{(t + 2j)^2}\right),$$

which implies,

$$-\ln \frac{p_t^M(\xi)}{p_t^M(0)} = \sum_{j=0}^{\infty} \ln \left(1 + \frac{\xi^2}{(t + 2j)^2}\right),$$

and eventually,

$$q_M(t, \xi) = -\frac{\partial}{\partial t} \ln \frac{p_t^M(\xi)}{p_t^M(0)} = \sum_{j=0}^{\infty} \frac{2}{t^2(t + 2j)} \left(\frac{\xi^2}{(t + 2j)^2} + \xi^2\right).$$

This series converges for $t > 0$ locally uniformly with respect to $\xi$ and since $\frac{2}{t^2(t + 2j)^2} \xi^2 = f_{t}(t+2j)^2(\xi^2)$ with the Bernstein function as in Example 2.3.B, we conclude that $q_M(t, \cdot)$ is a continuous negative definite function.
Example 2.5. For \( \psi_H(\xi) = (m^2 + \xi^2)^{1/2} - m, m > 0 \), it is known, see [10],
that \( p^H_t(x) = \frac{m e^{-mt}}{\pi} K_1(m \sqrt{t^2 + \xi^2}) \) where \( K_1 \) is the modified Bessel function of the second kind with index 1. It is an open problem whether \( q_H(t, \cdot) \) is a continuous negative definite function.

In order to clarify the situation further, we need to introduce additive processes.

3. Additive Processes and Fundamental Solutions

Let \( q : (0, \infty) \times \mathbb{R}^n \to \mathbb{R} \) be a continuous function such that \( q(t, 0) = 0 \) and \( q(t, \cdot) : \mathbb{R}^n \to \mathbb{R} \) be a continuous negative definite function. For \( 0 < s < t \) it follows that \( \xi \mapsto \int_s^t q(\tau, \xi) \, d\tau \) is again a continuous negative definite function. The continuity follows directly from our assumptions and since the the pointwise limit of negative definite functions is again negative definite, approximating the integral by a sequence of Riemann sums will yield the negative definiteness. Consequently the function \( \xi \mapsto e^{-\int_s^t q(\tau, \xi) \, d\tau} \) is a continuous positive definite function and hence for \( 0 < s < t \) we can define a family of bounded measures \( (\gamma_{t,s})_{t>s>0} \) by,
\[
\hat{\gamma}_{t,s}(\xi) = (2\pi)^{-n/2} e^{-\int_s^t q(\tau, \xi) \, d\tau}.
\]
From \( q(t, 0) = 0 \) we deduce that \( \gamma_{t,s} \) is a probability measure. Moreover, using results for the Fourier transform of measures, we find that,

Mi) \( \gamma_{s,s} = \delta_0 \) for \( 0 \leq s \);
Mii) \( \gamma_{t,r} * \gamma_{r,s} = \gamma_{t,s} \) for \( 0 < s \leq r \leq t < \infty \);
Miii) \( \gamma_{t,s} \to \delta_0 \) weakly for \( s \to t, s < t \);
Miv) \( \gamma_{t,s} \to \delta_0 \) weakly for \( t \to s, s < t \).

According to K. Sato [17, Theorem 9.7], we can associate with \( (\gamma_{t,s})_{0<s<t<\infty} \) a canonical additive process in law \( (X_t)_{t \geq 0} \) with state space \( \mathbb{R}^n \), i.e., \( P_{X_t-X_s} = \gamma_{t,s} \), \( t > s \).

Theorem 3.1. Let \( \psi : \mathbb{R}^n \to \mathbb{R} \) be a continuous negative definite function satisfying Ai) - Aiii). Denote by \( p_t \) the density of \( \mu_t \) where \( (\mu_t)_{t \geq 0} \) is the convolution semigroup associated with \( \psi \). If \( q(t, \xi) := -\frac{\partial}{\partial t} \ln (p_t(\xi)/p_t(0)) \) is with respect to \( \xi \) a continuous negative definite function, then we can associate with \( q \) a canonical additive process in law \( (X_t)_{t \geq 0} \) by the relation \( P_{X_t-X_s} = \gamma_{t,s} \) where \( t > s > 0 \) and \( \gamma_{t,s} \) is defined by (3.1).

With the help of the probability measures \( \gamma_{t,s}, 0 < s < t < \infty \), we can define on \( C_{\infty}(\mathbb{R}^n) \) or \( L^2(\mathbb{R}^n) \) the operators,
\[
V(t, s)u(x) = \int_{\mathbb{R}^n} u(x-y) \gamma_{t,s}(dy),
\]
and from Mi) - Miv) we deduce (for either \( u \) in \( C_{\infty}(\mathbb{R}^n) \) or in \( L^2(\mathbb{R}^n) \) and convergence is meant accordingly),

- \( V(s, s)u = u \);
- \( (V(t, r) \circ V(r, s))u = V(t, s)u = V(t, s)u, r < s < t \);
- \( V(t, s)u \to u \) as \( s \to t, s < t \);
- \( V(t, s)u \to u \) as \( t \to s, s < t \).
Since \( \gamma_{t,s} \) can be viewed as an element in \( \mathcal{S}'(\mathbb{R}^n) \) the convolution theorem yields for \( u \in \mathcal{S}(\mathbb{R}^n) \) that,

\[
(V(t, s)u)\hat{}(\xi) = e^{-\int_t^s q(\tau, \xi) \, d\tau} \hat{u}(\xi),
\]

which gives,

\[
\frac{\partial}{\partial t} (V(t, s)u)\hat{}(\xi) = -q(t, \xi)e^{-\int_t^s q(\tau, \xi) \, d\tau} \hat{u}(\xi),
\]

and,

\[
\frac{\partial}{\partial s} (V(t, s)u)\hat{}(\xi) = q(s, \xi)e^{-\int_t^s q(\tau, \xi) \, d\tau} \hat{u}(\xi).
\]

Therefore we deduce (at least as equations in \( \mathcal{S}'(\mathbb{R}^n) \), given \( u \in \mathcal{S}(\mathbb{R}^n) \)),

\[
\frac{\partial}{\partial t} V(t, s)u + q(t, D)V(t, s)u = 0,
\]

and,

\[
\frac{\partial}{\partial s} V(t, s)u - q(s, D)V(t, s)u = 0.
\]

Depending on properties of \( q(t, \cdot) \) we can identify \( V(t, s), \ t > s \), as a fundamental solution in the form of [22] for the initial value problem,

\[
\begin{aligned}
&\frac{\partial u}{\partial t}(t, x) - A(t)u(t, x) = f(t, x), \\
&u(0, x) = u_0(x),
\end{aligned}
\]

in \( L^2([0, T]; L^2(\mathbb{R}^n)) \) or \( C_0([0, T]; C_\infty(\mathbb{R}^n)) \), we refer to [2] or [23] for more details. For the purposes of this note we do not need the details.

Since \( q(t, \xi) = -\frac{\partial}{\partial t} \ln \left( \frac{p_\gamma(\xi)}{p_\gamma(0)} \right) \) we observe that,

\[
\int_s^t q(\tau, \xi) \, d\tau = -\ln \frac{p_\gamma(\xi)}{p_\gamma(0)} + \ln \frac{p_\gamma(\xi)}{p_\gamma(0)},
\]

or,

\[
e^{-\int_t^s q(\tau, \xi) \, d\tau} = \frac{p_\gamma(\xi)}{p_\gamma(0)} \frac{p_\gamma(0)}{p_\gamma(\xi)},
\]

i.e., using the definition of \( S_t \) we arrive at,

\[
V(t, 0) = S_t,
\]

a relation which even holds when \( \xi \mapsto q(t, \xi) \) is not a continuous negative definite function but just given by (2.20). In addition we see that with (2.12),

\[
P_{X_t - X_0} = \gamma_{t,0}, \quad \hat{\gamma}_{t,0} = (2\pi)^{-n/2} \sigma_t.
\]

Assuming that the additive process \( (X_t)_{t > s > 0} \) associated with \( q(t, \xi) \) given by (2.20) exists and denoting the Lévy process associated with \( \psi \) by \( (Y_t)_{t > 0} \) we find,

\[
P_{Y_t - Y_0} = \mu_t = F^{-1}(e^{-t\psi})(\cdot) \lambda^{(n)} = p_t(\cdot) \lambda^{(n)},
\]

and,

\[
P_{X_t - X_0} = \gamma_{t,0} = F^{-1} \left( \frac{p_\gamma}{p_\gamma(0)} \right)(\cdot) \lambda^{(n)} = \frac{e^{-\int_0^t \psi(\cdot)} - 1}{2\pi^n p_\gamma(0)} \lambda^{(n)}.
\]
In the next section we will use (3.8) and (3.9) to obtain a geometric interpretation of $p_t$. For later purposes we define,

$$\pi_{t,0}(x) := \frac{e^{-\frac{1}{2}t\psi(x)}}{(2\pi)^n p^\perp(0)}.$$  \hspace{1cm} (3.10)

4. Transition Functions and Geometry I. The Diagonal Term

Let $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous negative definite function satisfying Ai) - Aiii) with associated metric $d_\psi(\xi, \eta) := \sqrt{\psi(\xi - \eta)}$, and denote the corresponding metric measure space by $(\mathbb{R}^n, d_\psi, \lambda^{(n)})$. In this case $p_t$, the transition density as defined by (2.3), belongs to $L^1(\mathbb{R}^n) \cap C^\infty(\mathbb{R}^n)$ and,

$$p_t(x) \leq p_t(0).$$ \hspace{1cm} (4.1)

For later purposes it is helpful to note that $p_t(x) < p_t(0)$ for all $x \in \mathbb{R}^n$. Indeed,

$$p_t(0) - p_t(x) = (2\pi)^{-n} \int_{\mathbb{R}^n} \left( e^{i0^0 \xi} - e^{ix \xi} \right) e^{-t\psi(\xi)} d\xi$$

$$= 2(2\pi)^{-n} \int_{\mathbb{R}^n} (1 - \cos x \cdot \xi) e^{-t\psi(\xi)} d\xi,$$

and the function $\xi \mapsto (1 - \cos x \cdot \xi) e^{-t\psi(\xi)}$ is for every $x \in \mathbb{R}^n$ non-negative and continuous. Hence $p_t(0) = p_t(x_0)$ for some $x_0 \in \mathbb{R}^n \setminus \{0\}$ is impossible. From (2.5) we obtain immediately, see [14] or [15], that,

$$p_t(0) = (2\pi)^{-n} \int_{\mathbb{R}^n} e^{-t\psi(\xi)} d\xi = (2\pi)^{-n} \int_0^\infty \lambda^{(n)}(B^{d_\psi}(0, \sqrt{t/r})) e^{-r} dr,$$ \hspace{1cm} (4.2)

where $B^{d_\psi}(x_0, r) = \{x \in \mathbb{R}^n : d_\psi(x, x_0) < r\}$. We assume further that the metric measure space $(\mathbb{R}^n, d_\psi, \lambda^{(n)})$ has the doubling property, i.e., for all $x \in \mathbb{R}^n$ and all $r > 0$ it holds,

$$\lambda^{(n)}(B^{d_\psi}(x, 2r)) \leq c\lambda^{(n)}(B^{d_\psi}(x, r)).$$ \hspace{1cm} (4.3)

In this case, as shown in [14, Theorem 4.1], it follows that,

$$p_t(0) \asymp \lambda^{(n)}(B^{d_\psi}(0, 1/\sqrt{t})),$$ \hspace{1cm} (4.4)

recall that $a_t \asymp b_t$ means that $\gamma_0 a_t \leq b_t \leq \gamma_1 a_t$ holds with constants $0 < \gamma_0 \leq \gamma_1$ independent of $t$. Switching to the $t$-dependent metric $d_{\psi,t}(\xi, \eta) = \sqrt{t\psi(\xi - \eta)}$ we can re-write (4.4) as,

$$p_t(0) \asymp \lambda^{(n)}(B^{d_{\psi,t}}(0, 1)),$$ \hspace{1cm} (4.5)

Hence $p_t(0)$, the “diagonal term” of the (translation invariant) transition function $p_t(x, y) = p_t(x - y)$ is controlled by the volume of the unit ball in the (volume doubling) metric measure space $(\mathbb{R}^n, d_{\psi,t}, \lambda^{(n)})$.

Our first observation is that such a result carries over to additive processes. For ease of notation we set,

$$Q_{t,s}(\xi) := \int_s^t q(\tau, \xi) d\tau = -\ln \frac{p^\perp(\xi)}{p^\perp(0)} + \ln \frac{p^\perp(\xi)}{p^\perp(0)}, \hspace{1cm} t > s > 0,$$ \hspace{1cm} (4.6)

where $q$ is as in Theorem 3.1. Since $Q_{t,s}(\cdot)$ is a real-valued continuous negative definite function and $Q_{t,s}(\xi) = 0$ implies $q(\tau, \xi) = 0$ for all $\tau \in [s, t]$, note
that \( q(\tau, \xi) \geq 0 \) and by assumption \( \tau \mapsto q(\tau, \xi) \) is continuous, we conclude that \( Q_{t,s}(\xi) = 0 \) if and only if \( \xi = 0 \), hence by,

\[
d_{Q_{t,s}}(\xi, \eta) := Q_{t,s}^{1/2}(\xi - \eta),
\]

(4.7)
a metric is defined on \( \mathbb{R}^n \). Further, we know that if,

\[
d_q(\tau, \cdot)(\xi, \eta) = q^{1/2}(\tau, \xi - \eta),
\]

(4.8)
generates the Euclidean topology, then \( \liminf_{|\xi| \to \infty} q(\tau, \xi) > 0 \), see [14, Lemma 3.2]. Now Fatou’s lemma yields,

\[
\liminf_{|\xi| \to \infty} Q_{t,s}(\xi) \geq \int_t^s \left( \liminf_{|\xi| \to \infty} q(\tau, \xi) \right) d\tau > 0,
\]

(4.9)
i.e., \( d_{Q_{t,s}} \) generates the Euclidean topology too.

In the case that \( e^{- \int_t^s q(\tau, \cdot) d\tau} \in \mathcal{L}^1(\mathbb{R}^n) \) we denote the density of the measure \( \gamma_{t,s} \) by \( \pi_{t,s} \), i.e.,

\[
\pi_{t,s}(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{ix \cdot \xi} e^{- \int_t^s q(\tau, \cdot) d\tau} d\xi.
\]

(4.10)

**Theorem 4.1.** Assume that for every \( \tau > 0 \) the metric (4.8) generates the Euclidean topology and that \( e^{- \int_t^s q(\tau, \cdot) d\tau} \in \mathcal{L}^1(\mathbb{R}^n) \). Then it holds,

\[
\pi_{t,s}(0) = (2\pi)^{-n} \int_{\mathbb{R}^n} \lambda(n) \left( B^{d_{Q_{t,s}}}(0, \sqrt{\tau}) \right) e^{-r} dr.
\]

(4.11)

If we have in addition,

\[
\beta_0 q(t_0, \xi) \leq q(t, \xi) \leq \beta_1 q(t_0, \xi),
\]

(4.12)
for some \( t_0 > 0 \), all \( t > 0 \) and \( \xi \in \mathbb{R}^n \) with constants \( 0 < \beta_0 \leq \beta_1 \), and if the metric measure space \((\mathbb{R}^n, d_q(t_0, \cdot), \lambda(n))\) has the volume doubling property then we get,

\[
\pi_{t,s}(0) \asymp \lambda(n) \left( B^{d_{Q_{t,s}}}(0, \sqrt{\beta_1/\beta_0}) \right).
\]

(4.13)

**Remark 4.2.** Note that (4.12) implies \( e^{- \int_t^s q(\tau, \cdot) d\tau} \in \mathcal{L}^1(\mathbb{R}^n) \).

Further note that we can always use the examples from [14] to construct examples for Theorem 4.1 provided we introduce a \( t \)-dependence respecting (4.12).

**Proof.** Since,

\[
(2\pi)^n \pi_{t,s}(0) = \int_{\mathbb{R}^n} e^{-Q_{t,s}(\xi)} d\xi
\]

\[
= \int_0^\infty \lambda(n) \left( \{ \xi \in \mathbb{R}^n : e^{-Q_{t,s}(\xi)} \geq \rho \} \right) d\rho
\]

\[
= \int_0^1 \lambda(n) \left( \{ \xi \in \mathbb{R}^n : Q_{t,s}(\xi) \leq -\ln \rho \} \right) d\rho,
\]
we get,

\[
(2\pi)^n \pi_{t,s}(0) = -\int_0^\infty \lambda^n(\{\xi \in \mathbb{R}^n : Q_{t,s}(\xi) \leq r\}) e^{-r} \, dr \\
= \int_0^\infty \lambda^n(\{\xi \in \mathbb{R}^n : Q_{t,s}(\xi) \leq r\}) e^{-r} \, dr,
\]

and (4.11) is proved. Next, since \((\mathbb{R}^n, d_{q(t_0,\cdot)}, \lambda^{(n)})\) has the volume doubling property, by [14, Corollary 3.10] we get \(e^{-\beta_0(t-s)q(t_0,\cdot)} \in L^1(\mathbb{R}^n)\) for all \(t > 0\), hence \(e^{-\beta_0(t-s)q(t_0,\cdot)} \in L^1(\mathbb{R}^n)\) for all \(t > s \geq 0\). Now, for all \(\xi \in \mathbb{R}^n\) we have,

\[
\beta_0(t-s)q(t_0,\xi) \leq \int_s^t q(\tau, \xi) \, d\tau,
\]

or,

\[
e^{-\beta_0(t-s)q(t_0,\xi)} \geq e^{-\int_s^t q(\tau, \xi) \, d\tau},
\]

i.e., \(e^{-\int_s^t q(\tau, \xi) \, d\tau} \in L^1(\mathbb{R}^n)\) for all \(t > s \geq 0\). Using the monotonicity of \(r \mapsto \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{r}))\) we find,

\[
(2\pi)^n \pi_{t,s}(0) \geq \int_{\beta_1/\beta_0}^{\infty} \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{r})) e^{-r} \, dr \\
\geq \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{\beta_1/\beta_0})) \int_{\beta_1/\beta_0}^{\infty} e^{-r} \, dr \\
= \frac{1}{e^{\beta_1/\beta_0}} \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{\beta_1/\beta_0})).
\]

For the upper estimate we split the integral according to,

\[
(2\pi)^n \pi_{t,s}(0) = \int_{0}^{\beta_1/\beta_0} \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{r})) e^{-r} \, dr \\
+ \int_{\beta_1/\beta_0}^{\infty} \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{r})) e^{-r} \, dr,
\]

and note that,

\[
\int_{0}^{\beta_1/\beta_0} \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{r})) e^{-r} \, dr \leq \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{\beta_1/\beta_0})) \int_{0}^{\beta_1/\beta_0} e^{-r} \, dr \\
= \left(1 - \frac{1}{e^{\beta_1/\beta_0}}\right) \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{\beta_1/\beta_0})).
\]

On the other hand, by (4.12) we get,

\[
\int_{\beta_1/\beta_0}^{\infty} \lambda^n(B^{d_{q(t_0,\cdot)}}(0, \sqrt{r})) e^{-r} \, dr \\
\leq \int_{\beta_1/\beta_0}^{\infty} \lambda^n(\left\{\xi \in \mathbb{R}^n : \left(\int_{s}^{t} \beta_0 q(t_0, \xi) \, d\tau \leq \sqrt{r}\right)^{1/2}\right\}) e^{-r} \, dr \\
\leq \int_{\beta_1/\beta_0}^{\infty} \lambda^n(\left\{\xi \in \mathbb{R}^n : \sqrt{\beta_0(t-s)q(t_0, \xi)} \leq \sqrt{r}\right\}) e^{-r} \, dr.
\]

By the volume doubling property it follows for \(k \geq 1\) that,

\[
\lambda^n(B^{d_{q(t_0,\cdot)}}(0, r)) \leq c(t_0, r) \lambda^n(B^{d_{q(t_0,\cdot)}}(0, 1)),
\]
where \(c(t_0, r) \leq r^{\alpha(t_0)}c_0(t_0, 1)\) for all \(k \geq 1\) and some \(\alpha(t_0) \geq 0\). This implies,

\[
\int_{\beta_1/\beta_0}^{\infty} \lambda^{(n)}(B^{d_{\lambda,t},s}(0, \sqrt{r})) e^{-r} \, dr
\]

\[
\leq \int_{\beta_1/\beta_0}^{\infty} c(t_0, r) \lambda^{(n)}(\{\xi \in \mathbb{R}^n : \sqrt{\beta_0(t-s)}q(t_0, \xi) \leq 1\}) e^{-r} \, dr
\]

\[
\leq c(t_0, 1) \int_{\beta_1/\beta_0}^{\infty} \lambda^{(n)}(\{\xi \in \mathbb{R}^n : \sqrt{(t-s)}q(t_0, \xi) \leq \sqrt{1/\beta_0}\}) r^{\alpha(t_0)/2} e^{-r} \, dr
\]

\[
= d(t_0) \lambda^{(n)}(B^{d_{\lambda,t},s}(0, \sqrt{1/\beta_0}))
\]

where \(d(t_0) = c_0(t_0, 1) \int_{\beta_1/\beta_0}^{\infty} r^{\alpha(t_0)/2} e^{-r} \, dr < \infty\), and finally we obtain,

\[
(2\pi)^{n} \pi_{t,s}(0) \leq \left(1 - \frac{1}{e^{\beta_1/\beta_0} + d(t_0)}\right) \lambda^{(n)}(B^{d_{\lambda,t},s}(0, \sqrt{1/\beta_0})).
\]

\[\square\]

5. Transition Functions and Geometry II. The Off-Diagonal Term

Suppose that we are given a symmetric Lévy process \((Y_t)_{t \geq 0}\) associated with a continuous negative definite function \(\psi : \mathbb{R}^n \to \mathbb{R}\) satisfying \(A_1, A_3, A_{11}\). We denote by \(d_{\psi}\) or \(d_{\psi,t}\) the corresponding metrics \(d_{\psi}(\xi, \eta) = \psi^{1/2}(\xi - \eta)\) or \(d_{\psi,t}(\xi, \eta) = \sqrt{t}\psi(\xi - \eta)\). We can now write the corresponding transition density as,

\[
\tilde{p}_t(x, y) := p_t(x - y) = p_t(0) \frac{p_t(x - y)}{p_t(0)}, \quad (5.1)
\]

and further as,

\[
\tilde{p}_t(x, y) = p_t(0) e^{\ln \frac{p_t(x - y)}{p_t(0)}} = p_t(0) e^{-\left(-\ln \sigma_t(x - y)\right)}, \quad (5.2)
\]

where we used (2.12). We need the following observation: Let \(q(t, \xi)\) be defined as in (2.20) and \(\sigma_t\) as in (2.12). If the function \(\xi \mapsto q(t, \xi)\) is for all \(\tau > 0\) a continuous negative definite function then the function \(\xi \mapsto -\ln \sigma_t(\xi)\) is also for all \(\tau > 0\) a continuous negative definite function. We have seen that \(\xi \mapsto \int_s^t q(\tau, \xi) \, d\tau\) is a continuous negative definite function if \(\xi \mapsto q(\tau, \xi)\) is. Since,

\[
-\int_s^t \frac{\partial}{\partial \tau} \ln \frac{p_{\tau}(x)}{p_{\tau}(0)} \, d\tau = -\ln \frac{p_{\tau}(x)}{p_{\tau}(0)} + \ln \frac{p_{\tau}(x)}{p_{\tau}(0)},
\]

a further application of the ratio limit theorem yields,

\[
\int_t^\infty q(\tau, \xi) \, d\tau = -\ln \frac{p_{\tau}(x)}{p_{\tau}(0)},
\]

i.e., the negative definiteness of \(-\ln \sigma_t(\cdot)\) for all \(t > 0\). Thus we can write,

\[
\tilde{p}_t(x, y) = p_t(0) e^{-\left(\ln \sigma_t(x - y)\right)}, \quad (5.3)
\]
where \( \eta \mapsto -\ln \sigma_\psi(\eta) \) is a continuous negative definite function and from previous considerations it follows that if \( p_\psi^+(0) = p_\psi^+(0) \) then \( \eta_0 = 0 \), hence \( \delta_{\psi,t}(x,y) = ( -\ln \sigma_\psi(x,y) )^{1/2} \) is a further metric on \( \mathbb{R}^n \). So, if \( \psi : \mathbb{R}^n \to \mathbb{R} \) is a continuous negative definite function satisfying A(i)-A(iii) for which \( \xi \mapsto -\frac{\psi}{\delta} \ln (p_\psi^+(\xi)/p_\psi^+(0)) \) is also continuous negative definite, then the transition density \( p_t \) is constructed by two families of (time-dependent) metrics on \( \mathbb{R}^n \) namely \( d_{\psi,t} \) and \( \delta_{\psi,t} \). In particular, if \( (\mathbb{R}^n, d_{\psi,t}, \lambda^{(n)}) \) has the volume doubling property then it holds that,

\[
p_t(x-y) \asymp \lambda^{(n)}(B^{d_{\psi,t}}(0,1))e^{-\delta_{\psi,t}^2(x,y)}. \tag{5.4}
\]

Now we assume that we can associate with \( \psi \) also the additive process \((X_t)_{t \geq 0}\) constructed in Section 3. The density of the distribution \( P_{X_t-X_0} \) is given by,

\[
\pi_{t,0}(x-y) = \frac{e^{-\frac{1}{2}\psi(x-y)}}{(2\pi)^n p_\psi^+(0)}, \tag{5.5}
\]

compare with (4.10). Assuming in addition (4.12) we arrive at,

\[
\pi_{t,0}(x-y) \asymp \lambda^{(n)}(B^{d_{Q_{t,0}}}(0,\sqrt{\beta_1/\beta_0}))e^{-\delta_{Q_{t,0}}^2(x,y)},
\]

where \( \delta_{Q_{t,0}} \) is the metric,

\[
\delta_{Q_{t,0}}(x,y) = \sqrt{\frac{1}{t}\psi(x-y)} = d_{\psi,t}(x,y),
\]

and for \( d_{Q_{t,0}} \) we find,

\[
Q_{t,0}(x,y) = -\ln \frac{p_\psi^+(x-y)}{p_\psi^+(0)} = -\ln \sigma_t(x-y),
\]

thus,

\[
d_{Q_{t,0}}(x,y) = ( -\ln \sigma_t(x-y) )^{1/2} = \delta_{\psi,t}(x,y),
\]

which yields,

\[
\pi_{t,0}(x-y) \asymp \lambda^{(n)}(B^{\delta_{\psi,t}}(0,\sqrt{\beta_1/\beta_0}))e^{-\delta_{\psi,t}^2(x,y)}. \tag{5.6}
\]

Summing our considerations up we have proved,

**Theorem 5.1.** Let \( \psi : \mathbb{R}^n \to \mathbb{R} \) be a continuous negative definite function satisfying A(i)-A(iii) and assume that \( q(t,\xi) = -\frac{\psi}{\delta} \ln (p_\psi^+(\xi)/p_\psi^+(0)) \) is with respect to \( \xi \) a continuous negative definite function. Moreover assume that the metric measure space \( (\mathbb{R}^n, d_{\psi}, \lambda^{(n)}) \) has the volume doubling property and for \( q(t,\xi) \) we have that (4.12) holds. Denote by \((Y_t)_{t \geq 0}\) the Lévy process associated with \( \psi \) and by \((X_t)_{t \geq 0}\) the additive process constructed in Section 3. With

\[
d_{\psi,t}(\xi,\eta) = \sqrt{\psi(\xi-\eta)}
\]

and

\[
\delta_{\psi,t}(x,y) = ( -\ln \frac{p_\psi^+(x-y)}{p_\psi^+(0)} )^{1/2} = ( -\ln \sigma_t(x-y) )^{1/2}
\]

we find,

\[
p_t(x-y) \asymp \lambda^{(n)}(B^{d_{\psi,t}}(0,1))e^{-\delta_{\psi,t}^2(x,y)}, \tag{5.6}
\]
and,
\[ \pi_{t,0}(x - y) \simeq \lambda(n) \left( B^{\psi} \ast (0, \sqrt{\beta_1/\beta_0}) \right) e^{-d^2 \psi \ast (x,y)}, \tag{5.7} \]
hold.

Acknowledgment. The authors would like to thank the referee for the careful reading and his helpful suggestions. Financial support of Swansea University and her College of Science in form of a PhD-scholarship for the first named author is gratefully appreciated.

References


LEWIS J. BRAY: MATHEMATICS DEPARTMENT, SWANSEA UNIVERSITY, SWANSEA, SA2 8PP, UNITED KINGDOM
  E-mail address: lewisbray@hotmail.co.uk

NIELS JACOB: MATHEMATICS DEPARTMENT, SWANSEA UNIVERSITY, SWANSEA, SA2 8PP, UNITED KINGDOM
  E-mail address: n.jacob@swansea.ac.uk
  URL: http://www.swansea.ac.uk/staff/science/maths/n.jacob/
BROWNIAN MANIFOLDS, NEGATIVE TYPE AND GEO-TEMPORAL COVARIANCES

N. H. BINGHAM, ALEKSANDAR MIJATOVIĆ, AND TASMIN L. SYMONS

Abstract. We survey Brownian manifolds – manifolds that can parametrise Brownian motion – and those that cannot. We consider covariances of space-time processes, particularly those when space is the sphere – geo-temporal processes. There are connections with functions of negative type.

1. Brownian Manifolds and Negative Type

1.1. Spatio-temporal processes. A stochastic process, \( X = \{ X_t \} \) say, is generally a mathematical model for a random phenomenon evolving with time \( t \) – a temporal process; sometimes the relevant parameter is a point in space – a spatial process, or a random field; sometimes one needs both time and space – a spatio-temporal process. Our main interest here is the case when space is a sphere, which we think of as the Earth; in this case we speak of a geo-temporal process.

The simplest manifold that might be used for the space variable is Euclidean space, of dimension \( n \) say, \( \mathbb{R}^n \), when one might speak of a process with multi-dimensional (\( n \)-dimensional) time. The next simplest space manifold is the sphere, \( S \) – the Earth, say, though we shall take the \( n \)-sphere \( S^n \), an \( n \)-dimensional manifold embedded in \( \mathbb{R}^{n+1} \).

The most useful class of random fields for modelling purposes is the Gaussian random fields. In this as in any other context, the prototypical Gaussian process is Brownian motion.

Processes will be real-valued, unless otherwise stated.

1.2. Lévy’s Brownian motion with multi-dimensional time. One can define Brownian motion \( B = (B_t : t \in \mathbb{R}) \) on the line (on \( (\Omega, \mathcal{F}, P) \), say) as the centred Gaussian process with incremental variance

\[
\text{var}(B_t - B_s) = |t - s|
\]

and (say) \( B_0 = 0 \). One can regard \( B \) as a map \( t \mapsto B_t \) from \( \mathbb{R} \) to the Hilbert space \( H = L^2(\Omega, \mathcal{F}, P) \), and then

\[
\|B_t - B_s\|^2 = |t - s|,
\]

Received 2016-10-6; Communicated by D. Applebaum.
2010 Mathematics Subject Classification. Primary 60D05; Secondary 60J65.
Key words and phrases. Negative type, geo-temporal covariances, incremental variance, Lévy Brownian motion, random fields, spherical transform.
the left being the incremental variance. The covariance is then given by the inner product
\[ c(t, s) := (B_t, B_s) = \frac{1}{2}(|t| + |s| - |t-s|), \]
and as independence is just zero correlation in the Gaussian case, this gives independent increments as usual.

Generalising this viewpoint, Paul Lévy [62] showed that one can define multi-parameter Brownian motion (or Brownian motion with multi-dimensional time) as the real-valued centred Gaussian process \( B = (B_t : t \in \mathbb{R}^n) \) with incremental variance
\[ i(t, s) := \mathbb{E}[(B_s - B_t)^2] = \|B_t - B_s\|^2 = |t-s| \]
(assuming \( |\cdot| \) for Euclidean distance); see also [63, 64] for later treatments.

One has
\[ i(s, t) = c(s, s) + c(t, t) - 2c(s, t), \quad (i - c) \]
and as \( c(s, s) = \mathbb{E}[B_s^2] = i(s, 0) \),
\[ c(s, t) = \frac{1}{2}(i(s, 0) + i(t, 0) - i(s, t)). \quad (c - i) \]
Thus either of \( c, i \) determines the other; \( i \) is more convenient here.

Lévy also showed that Brownian motion can be defined so as to be parametrised by the sphere \( \mathbb{S}^n \), in addition to \( \mathbb{R}^n \) as above. Now the incremental variance is given by the geodesic distance \( d \) on the sphere (from the North Pole \( o \), which plays the role of the origin above):
\[ i(s, t) = \|B_s - B_t\|^2 = d(s,t). \quad (*) \]
Thus \( \sqrt{d}(s, t) = \|B_t - B_s\| \); one calls \( \sqrt{d} \) a Hilbertian distance.

A word on terminology: our incremental variance is also known by several other names: the variogram (a term due to Matheron, arising in mining), the structure function (Yaglom), mean-squared difference (Jowett), etc.; see e.g. Cressie ([22], 2.3.1).

1.3. Brownian and non-Brownian manifolds. For \( M \) a manifold with geodesic distance \( d \), or more generally with \( (M, d) \) a metric space, one can proceed as above and call \( B = (B_x : x \in M) \) a Brownian motion parametrised by \( M \) if the \( B_x \) are centred Gaussian, the incremental variance is the geodesic distance,
\[ \text{var}(B_x - B_y) = d(x, y), \]
and the finite-dimensional distributions are Gaussian (that is, linear combinations \( \sum c_i B_{t_i} \) are Gaussian). Then as before, (*) above is satisfied with \( d \) the geodesic distance on \( M \). Call such a manifold, or metric space, Brownian. Thus Euclidean space and spheres are Brownian, by Lévy’s results above. Further examples are given by the real or complex hyperbolic spaces, a result due to Faraut and Harzallah ([28], III.3, [31], Prop. 7.3) (and implicit in Gangolli [33]). By contrast, quaternionic hyperbolic spaces are not Brownian ([28], Cor. IV.2, or [31]), and nor is the octonion (Cayley) projective plane.

The question of whether a space \( M \) is Brownian is thus purely geometric, as it
depends on whether a map \( B \) exists satisfying (*)). See e.g. Cartier ([19], Th. 1 d) for this viewpoint, and for background on Gaussian Hilbert spaces, Janson [50].

1.4. Spaces and kernels of negative type. Call a metric space \((M, d)\) of negative type if
\[
\sum_{i,j=1}^n d(x_i, x_j)u_iu_j \leq 0
\]
for all \( n = 2, 3, \cdots \), all points \( x_i \in M \) and all real \( u_i \) with \( \sum u_i = 0 \) (the term conditionally negative definite is also used, reflecting the condition \( \sum u_i = 0 \)). Call \( M \) of strictly negative type if the sum above is negative for all such \( u_i \) not all zero. Such spaces are important in a variety of contexts, and have been studied at length in the books by Blumenthal [18] and Deza and Laurent [25].

A kernel \( k : M \times M \to \mathbb{R}^+ \) is of negative type if
\[
\sum_{i,j=1}^n k(x_i, x_j)u_iu_j \leq 0
\]
for all \( n = 2, 3, \cdots \), all points \( x_i \in M \) and all real \( u_i \) with \( \sum u_i = 0 \), and of positive type (or positive definite) if
\[
\sum_{i,j=1}^n k(x_i, x_j)u_iu_j \geq 0
\]
for all \( n = 2, 3, \cdots \) and all points \( x_i \in M \); similarly for strictly positive type.

Covariances \( c \) are of positive type. So, incremental variances \( i \) are of negative type: the first two terms on the right of \( (i - c) \) contribute 0 to the relevant summation, as \( \sum u_i = 0 \), so the sum is \( \leq 0 \) as \( c \) is of positive type.

For negative type on locally compact groups, see Heyer ([45], Ch. 5), Berg and Forst ([6], II).

1.5. Schoenberg’s theorems. It was shown by Schoenberg ([80, 81]) in 1937-8 that a metric space \((M, d)\) is of negative type if and only if there is a map \( \phi : M \to H \) for some Hilbert space \( H \) with
\[
d(x, y) = \|\phi(x) - \phi(y)\|^2.
\]
Thus, when \( H = L^2(\Omega, \mathcal{F}, P) \) as before, \( M \) is Brownian if and only if it is of negative type, and then Brownian motion \( B \) on (parametrised by) \( M \) is the map \( \phi \) above. Then \( \phi : (M, \sqrt{d}) \to H \) is called the Brownian embedding (or just, embedding). See Lyons [65] for a short proof of Schoenberg’s theorem.

The other classical result of Schoenberg relevant here [81] is that a kernel \( k \) is of negative type if and only if \( e^{-tk} \) is of positive type for every \( t \geq 0 \). This, of course, suggests the Lévy–Khintchine formula, and was part of Gangolli’s motivation for his theory of Lévy-Schoenberg kernels [33]; see also [6, 45].

1.6. The Kazhdan property. The geometrical property of being Brownian has an algebraic interpretation in the case \( M = G/K \) of a symmetric space (see Section 2 for these and other related terms).

Kazhdan [54] defined a locally compact group to have Property (T), now called the Kazhdan property, if the unit representation is isolated in the space of unitary representations. Groups with the Kazhdan property – Kazhdan groups – have proved to be important in many areas; for a monograph treatment, see Bekka et
al. [5]. (We note that a locally compact group is compact if and only if it is amenable and Kazhdan; see e.g. Paterson [74] for background on amenability.)

The irreducible unitary representations are in bijection with the positive definite spherical functions (Section 2 below); the set of these is called the spherical dual. In the rank-one case considered below, this can be identified with a set \( \Lambda \subset \mathbb{R} \), where if \( M \) is compact, \( \Lambda \) is a discrete set tending to infinity, while if \( M \) is Euclidean, or is real or complex hyperbolic space, \( \Lambda = [0, \infty) \). By contrast, if \( M \) is quaternionic hyperbolic space, \( \Lambda = \{0\} \cup [\lambda_0, \infty) \), where \( \lambda_0 > 0 \) (Faraut [28, 31]; cf. Kostant ([58], 428)). Thus \( M \) is Kazhdan in this case. Here \( M = G/K \), \( G = \text{Sp}(n, 1) \), \( K = \text{Sp}(n) \times \text{Sp}(1) \) (see e.g. ([5], 3.3) for a full treatment), and \( \text{Sp}(n, 1) \) is Kazhdan. So too is the octonion (Cayley) projective plane. By contrast, real and complex hyperbolic space are not Kazhdan, this being most easily seen as a consequence of Schoenberg’s theorem ([5], 2.11). This explains the Faraut–Harzallah results above.

2. Symmetric Spaces, Spherical Functions and Weights

2.1. Symmetric spaces. A symmetric space (Helgason [40, 41, 42], Wolf [94]) is a Riemannian manifold \( M \) whose curvature tensor is invariant under parallel translation. These may also be described as spaces where at each point \( x \) the geodesic symmetry exists: this fixes \( x \) and reverses the (direction of) geodesics through \( x \), an involutive automorphism ([94], Ch. 11). Then \( M \) is a Riemannian homogeneous space \( M = G/K \), where \( G \) is a closed subgroup of the isometry group of \( M \) containing the transvections, and \( K \) is the isotropy subgroup of \( G \) fixing the base-point \( x \). Here \( (G, K) \) is called a Riemannian symmetric pair. The Banach algebra \( L_1(K \backslash G/K) \) of (Haar) integrable functions on \( G \) bi-invariant under \( K \) is commutative. Pairs with this property are called Gelfand pairs, and such Banach algebras are called commutative spaces [94] (a misnomer, as it is the algebra, rather than the space, which is commutative).

Symmetric spaces may be split into compact, Euclidean and non-compact parts; there is a duality between the compact and non-compact cases, with the Euclidean case being self-dual ([40], V.2). We confine ourselves here to symmetric spaces of rank one ([40], V.6). These are two-point homogeneous spaces; they may be classified, as spheres, Euclidean and hyperbolic spaces, of constant curvature \( \kappa > 0 \), \( \kappa = 0 \) and \( \kappa < 0 \) respectively ([40] Sections IX.5, X.3 p.401, and [93]).

2.2. Spherical functions. For harmonic analysis in this context, one needs the analogue of the classical Fourier transform in Euclidean space and the Gelfand transform for Banach algebras. This involves spherical measures and spherical functions ([94], Ch. 8) and the spherical transform ([94], Ch. 9); cf. Applebaum [2].

For \( (G, K) \) a Gelfand pair, a spherical measure \( m \) is a \( K \)-bi-invariant multiplicative linear functional on \( C_c(K \backslash G/K) \); a spherical function is a continuous function \( \omega : G \to \mathbb{C} \) such that the measure \( m_\omega(f) := \int_G f(x)\omega(x^{-1})d\mu_G(x) \) is spherical. The map
\[
 f \mapsto \hat{f}(\omega) := m_\omega(f) = \int_G f(x)\omega(x^{-1})d\mu_G(x)
\]
is called the spherical transform for \((G, K)\). The positive definite spherical functions \(\phi\) on \((G, K)\) are in bijection with the irreducible unitary representations \(\pi\) of \(G\) with a \(K\)-fixed unit vector \(u\) via

\[
\phi(g) = \langle u, \pi(g)u \rangle.
\]

These form the spherical dual (called \(\Lambda\) in [28], \(\Omega\) in [30]).

2.3. Weights. When \(G\) is compact, the \(\pi\) here are in bijection with the dominant weights, in the sense of the Cartan–Weyl theory of weights; see e.g. Applebaum ([1], Ch. 2), or Wolf ([94], 6.3). In the rank-one case, the dominant weights are a subset \(\Lambda \subset \mathbb{R}\), as in Section 1; here \(\Lambda\) is specified by the Cartan–Helgason theorem ([94], 11.4B), ([42], V.1.1, 534-538, 550), ([38], 549-550).

The spherical functions satisfy an integral equation due to Harish–Chandra (see e.g. [40], X), which subsumes a number of the addition formulae of classical special-function theory (see e.g. ([92], XI) for Bessel functions and Gegenbauer polynomials).

2.4. Compact Lie groups. Compact connected Lie groups are themselves symmetric spaces ([40], IV.6). It was shown by Baldi and Rossi [4] that \(SU(2)\) is Brownian, but that \(SO(n)\) is not for \(n \geq 3\) (the question is decided by the signs of the coefficients in the Peter-Weyl expansion). This is despite \(SU(2)\) and \(SO(3)\) being locally isomorphic: \(SO(3) \cong SU(2)/\{\pm e\}\); cf. [29], Chapter 7 and 8.

3. Geo-temporal Covariances

3.1. Sphere cross line. For modelling purposes in the earth sciences and climatology, one needs both a space coordinate on the sphere and a time coordinate on the line (or half-line); thus the space \(M = S \times \mathbb{R}\) (or \(M = S \times \mathbb{R}_+\)) is needed. The most basic process one might wish to model on \(M\) is Brownian motion. But the product can be taken in several different senses, and it turns out that the question of existence of Brownian motion depends on which kind of product we take. Recall that by Lévy’s results of §1, Brownian motion exists on both \(S\) and \(\mathbb{R}\) (or \(\mathbb{R}_+\)), since both are of negative type.

First, take the product of metric spaces, under Hamming distance (“city-block metric”, for those who know Manhattan), under which distances \(s\) add:

\[
s := s_1 + s_2,
\]

in the obvious notation. From the definition of negative type, this property is preserved under such products; see e.g. [18], §3.2. So Brownian motion on the sphere cross line exists, with the product taken in this sense.

Next, one can take the product under the ordinary cartesian (or pythagorean) rule:

\[
s^2 := s_1^2 + s_2^2.
\]

Here again, Brownian motion exists. McKean [67] gives a thorough study of the white-noise case (from which the Brownian case follows by integration), starting from the work of Chentsov [20] on white noise in this setting. McKean’s construction moves between Euclidean space \(\mathbb{R}^{d+1}\) and “sphere cross half-line”, \(S^d \times \mathbb{R}_+\).
By contrast, if one takes the cartesian product of two Riemannian manifolds, distance is given by the differential cartesian rule:

$$ds^2 := ds_1^2 + ds_2^2,$$

again in the obvious notation. It turns out that $M = S \times R$ is no longer of negative type – so is no longer Brownian – viewed as a manifold in this way. The same holds for any product of manifolds with at least one spherical factor – or even a factor with two pairs of antipodal points. See [47] for background and details, [61, 75] for Riemannian manifolds.

Thus Brownian motion exists on $M = S \times R$, regarded as a product of metric spaces in both the above senses, though not of Riemannian manifolds. This provides a starting-point for geo-temporal modelling – but separates the effects of space and time.

### 3.2. Separable and non-separable covariances

The last result, however, does not take us very far. Real applications, e.g. to climate science, involve both spatial and temporal variation together rather than separately. The need thus arises for a range of examples of non-separable covariances on $M = S \times R$, that can be used flexibly for modelling. For detail here we refer to Cressie and Huang [23], Gneiting ([35] on $\mathbb{R}^n \times \mathbb{R}$, [36] on $S^n$, [37]). For space-time modelling in general, see Kyriakidis and Journel [59], Finkenstädt et al. [32], Porcu et al. [76, 77]. For applications to global climate data see the work of Jun and Stein [52, 53], and Jeong and Jun [51]. We note that progress was hampered in the past by lack of an adequate range of examples of covariances for modelling purposes – to the extent that practical statisticians and climatologists felt themselves forced to use as “covariances” functions that were not even positive definite. It was the modelling needs of climate scientists, together with the interest of the Brownian case, that prompted this study.

### 4. Complements

#### 4.1. Testing for independence

The ideas above found a new and powerful application in statistics, in the work of Székely, Rizzo and Bakirov [84] in 2007 and Székely and Rizzo [85] in 2009. See in particular the extensive commentary to the invited paper [85], and for further developments, [86, 87, 88, 89]; there are also applications to the theory of algorithms. Given a bivariate sample $((X_1, Y_1), \cdots, (X_n, Y_n))$, where each coordinate has finite mean, it turns out that one can test for independence of the $X$- and $Y$-coordinates, consistently against all alternatives (again, with finite means) by test statistics involving only distances between observations. The crux is the concept of distance covariance (equivalent to a related concept of Brownian covariance). See below.

#### 4.2. Distance covariance

The theory of distance covariance in metric spaces has been re-worked and simplified by Russell Lyons [65, 66]. It turns out that this area too belongs to distance geometry. The crux is for the distance covariance of $(X, Y)$ to be zero if and only if $X$ and $Y$ are independent. It turns out that this does not hold for general metric spaces, but does so exactly for those of strong negative type, a class that includes Euclidean spaces. As before, embeddability
into Hilbert space is crucial; other embeddings also occur (Riesz, Fourier, Crofton, Brownian). We refer to the excellent paper [65] for details. For applications to high-dimensional data, see Kosorok [57].

Regarding the link with Crofton’s formula: see the paper by Guyan Robertson [78]. The Crofton formula goes back to 1885 [24]; see Santaló [79] for background and details. It is a precursor of the Radon transform, for which see e.g. [43].

4.3. Other approaches. The first person to use white-noise integrals for Lévy’s Brownian motion was Chentsov [20], an approach later taken up by Lévy himself [64], and McKean [67]. For other approaches, see the work of Molchan [68, 69, 70, 71], Noda [72, 73], and Takenaka, Kubo and Urakawa [90].

4.4. Gaussian processes. One may construct Gaussian processes wherever one has a covariance, though these are necessarily more complicated than Brownian motion when working on a non-Brownian space. Covariance structure is always important, but is only fully informative in the case of Gaussianity – here, as always, a useful first approximation. For compact symmetric spaces (such as spheres), a detailed study was given by Askey and Bingham [3]. It would be interesting to extend this study to the geo-temporal context.

4.5. Gaussian fields. Gaussian random fields (this term is now more common for spatial processes) have been studied by Cohen and Lifshits [21] on spheres and hyperbolic spaces. For Gaussian free fields – which arise in physics (quantum field theory), but may be regarded as multi-parameter analogues of Brownian motion – see Sheffield [83]. For an extensive survey of Gaussian random fields and their links with physics, see Léandre [60]. For contours in this context (motivated by work of Pyke on Brownian sheets), see Kendall [55].

4.6. Positive definite functions on spheres. This subject, at the heart of the work here, goes back to seminal work by Schoenberg [82] in 1942. It has been considered further by the first author [8], Faraut [28] and Gneiting [36].

4.7. Manifold-valued Brownian motion. Dual (in the sense of harmonic analysis – see Gangolli [33]) to (real-valued) Brownian motion parametrised by a manifold is Brownian motion taking values in a manifold. Here there is a rich interplay between the geometry of the manifold and the probabilistic properties of Brownian motion on it. See for example Grigoryan [39], Varopoulos et al. [91], Elworthy [26]. One of the highlights in this context is that the radial part of a Brownian motion on a Riemannian manifold is a semimartingale, even though the smoothness of the distance function fails along the cut-locus; see Kendall [56]. For a probabilistic proof of the Atiyah–Singer index theorem using Brownian motion on manifolds, see Bismut [12].

4.8. Hypergroups. The theory of hypergroups is by now well established, but too extensive for us to discuss here. We refer to the standard work on the subject by Bloom and Heyer [13]; see also [14, 15, 16, 46]. Hypergroups make contact with the work studied here, for instance through our main example, the symmetric spaces of rank one; these have constant curvature $\kappa$. For the spherical case $\kappa > 0$, the relevant hypergroup here is the Bingham (or Bingham-Gegenbauer) hypergroup;
see [7], Bloom and Heyer [13], 3.4.23. For the Euclidean case $\kappa = 0$, it is the Kingman (or Kingman-Bessel) hypergroup ([13], 3.4.30). For the hyperbolic case $\kappa < 0$, it is the hyperbolic hypergroup ([13], Section 3.5.68, [95]).

4.9. Markov property. In one dimension, the Markov property is expressed by present time being a splitting time: past and future are conditionally independent given the present. In higher dimensions, the geometry is more complicated. In the plane, for example, one might have values within and without a contour conditionally independent given values on the contour. See for example Evstigneev [27] for background and references.

4.10. Time series. The work above is relevant to time series. For instance, Zhou [96] used distance correlation to study nonlinear dependence. Prediction theory may be extended to hypergroups – see e.g. Hösel and Lasser [48] – and can be applied on spheres, using the Bingham hypergroup. This provided the first author with a pleasing link to his recent work on prediction theory; see e.g. [9, 10, 11].

4.11. Fractional processes. Brownian motion is too smooth for some purposes, and may be usefully generalised to fractional Brownian motion, which has a parameter (essentially the Hurst parameter from hydrology) that governs the degree of smoothness. Such fractional Gaussian fields have been studied in contexts related to ours by Gelbaum [34] and Istas [49].

4.12. Higher dimensions. It is of interest to see what happens to the $n$-dimensional spheres and hyperbolic spaces considered here as the dimension $n \to \infty$. There has been much progress in this area in recent decades, due largely to Olshanski, Okounkov and Vershik. For background and details, see several recent papers by Jacques Faraut, e.g. [30]; cf. the paper by Bloom and Wildberger [17] in this volume.

5. Postscript

It is a pleasure to dedicate this survey to Herbert Heyer on the occasion of his eightieth birthday. The preface to his classic book [44] is preceded by a quote from Pierre Lelong: “... les probabilités sur les structures algébriques, sujet neuf et passionnant”. This captures well the lifelong dedication to the subject that Herbert has always shown. His work, his example and his friendship have enriched the lives of us, our fellow-contributors and many others; long may they continue to do so.

Acknowledgment. We are most grateful to the referee for many helpful comments and references, which have greatly improved the paper.

References
2. Applebaum, D.: Convolution semigroups of probability measures on Gelfand pairs, revisited. This volume.


N. H. Bingham: Mathematics Dept., Imperial College, London SW7 2AZ, England
E-mail address: n.bingham@ic.ac.uk
URL: http://wwwf.imperial.ac.uk/~bin06/

Aleksander Mijatović: Mathematics Dept., King’s College, London WC2R 2LS, England
E-mail address: aleksandar.mijatovic@kcl.ac.uk
URL: https://nms.kcl.ac.uk/mijatovic/

T. L. Symons: Mathematics Dept., Imperial College, London SW7 2AZ, England
E-mail address: tls111@ic.ac.uk
ON THE KOLMOGOROV-WIENER-MASANI SPECTRUM OF A
MULTI-MODE WEAKLY STATIONARY QUANTUM PROCESS

K. R. PARTHASARATHY AND RITABRATA SENGUPTA

ABSTRACT. We introduce the notion of a $k$-mode weakly stationary quantum process $\rho$ based on the canonical Schrödinger pairs of position and momentum observables in copies of $L^2(\mathbb{R}^k)$, indexed by an additive abelian group $D$ of countable cardinality. Such observables admit an autocovariance map $\tilde{K}$ from $D$ into the space of real $2k \times 2k$ matrices. The map $\tilde{K}$ admits a spectral representation as the Fourier transform of a $2k \times 2k$ complex Hermitian matrix-valued totally finite measure $\Phi$ on the compact character group $\hat{D}$, called the Kolmogorov-Wiener-Masani (KWM) spectrum of the process $\rho$. Necessary and sufficient conditions on a $2k \times 2k$ complex Hermitian matrix-valued measure $\Phi$ on $\hat{D}$ to be the KWM spectrum of a process $\rho$ are obtained. This enables the construction of examples. Our theorem reveals the dramatic influence of the uncertainty relations among the position and momentum observables on the KWM spectrum of the process $\rho$. In particular, the KWM spectrum cannot admit a gap of positive Haar measure in $\hat{D}$.

1. Introduction

In his celebrated little book “Osnovnye ponyatiya teorii veroyatnostei” [5], A. N. Kolmogorov introduced the notion of a stochastic process as a consistent family of finite dimensional probability distributions in $\mathbb{R}^n$, $n = 1, 2, \ldots$. In the same spirit a quantum process can be described as a consistent family of density operators or, equivalently, states in tensor products $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ of Hilbert spaces with $n = 1, 2, \ldots$. One can replace the ‘time set’ $\{1, 2, \ldots\}$ by an abstract countable set $D$ with the discrete topology and a family $\{\mathcal{H}_a : a \in D\}$ of Hilbert spaces. Then a quantum process yields a density operator $\rho_{a_1, a_2, \ldots, a_n}$ in $\mathcal{H}_{a_1} \otimes \cdots \otimes \mathcal{H}_{a_n}$ for every finite sequence $(a_1, \ldots, a_n)$ with distinct elements from $D$. All these density operators will obey natural consistency conditions. For example, the relative trace of $\rho_{a_1, a_2, \ldots, a_n}$ over $\mathcal{H}_{a_n}$ is $\rho_{a_1, a_2, \ldots, a_{n-1}}$. If $(b_1, \ldots, b_n)$ is a permutation of $a_1, \ldots, a_n$ then $\rho_{b_1, \ldots, b_n} = U \rho_{a_1, \ldots, a_n} U^{-1}$ where $U$ is the corresponding Hilbert space isomorphism from $\mathcal{H}_{a_1} \otimes \cdots \otimes \mathcal{H}_{a_n}$ onto $\mathcal{H}_{b_1} \otimes \cdots \otimes \mathcal{H}_{b_n}$ induced by the permutation. We denote the quantum process over $D$ by

$$\mathcal{Q} = \{(\mathcal{H}_{a_1, \ldots, a_n}, \rho_{a_1, \ldots, a_n}) : (a_1, \ldots, a_n) \in S_D\}. \quad (1.1)$$
where \( \mathcal{S}_D \) denotes the set of all finite-length sequences of distinct elements from the countable set \( D \).

In this paper we are interested in the special case where \( \mathcal{H}_a = L^2(\mathbb{R}^k) \) for all \( a \) in \( D \), \( k \) being a fixed positive integer, called the number of modes of the process. Each \( \mathcal{H}_a \) admits Schrödinger canonical pairs \( q_{ar}, p_{ar}, r = 1, 2, \cdots, k \), of position and momentum observables obeying the Heisenberg canonical commutation relations (CCR). We can look upon \( q_{ar}, p_{ar}, r = 1, 2, \cdots, k \), as observables in \( \mathcal{H}_{a_1} \otimes \cdots \otimes \mathcal{H}_{a_n} \) whenever the sequence \((a_1, a_2, \cdots, a_n)\) from \( \mathcal{S}_D \) contains the element \( a \) and denote such amplified observables by the same respective symbols.

With such a convention one obtains the algebra of all polynomials of all \( q_{ar}, p_{ar}, r = 1, 2, \cdots, k \), where \( a \in D \). Using the finite-partite states \( \rho_{a_1, \cdots, a_n} \), where \((a_1, \cdots, a_n) \in \mathcal{S}_D \), one can compute the expectations of the polynomials whenever they exist. Write

\[
(X_{a_1}, X_{a_2}, \cdots, X_{a_{(2k-1)}}, X_{a_{2k}}) = (q_{a_1}, p_{a_1}, \cdots, q_{a_k}, p_{a_k})
\]

and define the covariances

\[
\kappa_{rs}(a, b) = \frac{1}{2} \langle X_{a_r} X_{b_s} + X_{b_s} X_{a_r} \rangle - \langle X_{a_r} \rangle \langle X_{b_s} \rangle \tag{1.2}
\]

where \( \langle \cdot \rangle \) denotes expectation. To compute these quantities we need a knowledge of only the ‘bipartite’ states \( \rho_{a,b} \) for all \((a, b) \in \mathcal{S}_D \). Thus we obtain a \( 2^k \times 2^k \) real matrix-valued covariance kernel \( K = [[\kappa_{rs}(a, b)]] \) defined by

\[
K(a, b) = [[\kappa_{rs}(a, b)]], \quad r, s, \in \{1, 2, \cdots, 2k\} \tag{1.3}
\]

for any \( a, b \in D \).

Suppose \( D \) is a countable discrete additive abelian group with addition operation + and null element 0. Let the covariance kernel \( K \) of a \( k \)-mode quantum process over \( D \) be translation invariant in the sense that

\[
K(a + x, b + x) = K(a, b) \quad \forall a, b, x \in D. \tag{1.4}
\]

Then we say that the quantum process is second order weakly stationary, or, simply, weakly stationary. For such a process there exists a map \( \tilde{K} \) from \( D \) into the space of \( 2^k \times 2^k \) real matrices such that

\[
K(a, b) = \tilde{K}(b - a) \quad \forall a, b \in D. \tag{1.5}
\]

The map \( \tilde{K} \) is called the autocovariance map of the weakly stationary quantum process.

Owing to the matrix-positivity properties enjoyed by covariances between observables the autocovariance map \( \tilde{K} \) satisfies the matrix inequalities

\[
\sum_{i,j} \alpha_i \alpha_j \tilde{K}(a_j - a_i) \geq 0 \tag{1.6}
\]

for any \( a_1, a_2, \cdots, a_n \in D \) and real scalars \( \alpha_1, \alpha_2, \cdots, \alpha_n, n = 1, 2, \cdots \). Thanks to Bochner’s theorem for locally compact abelian groups there exists a complex Hermitian and positive \( 2^k \times 2^k \) matrix-valued measure \( \Phi \) on the compact dual
character group $\hat{D}$ of $D$ such that
$$\tilde{K}(a) = \int_{\hat{D}} \chi(a) \Phi(d\chi) \text{ for all } a \in D.$$ (1.7)

The matrix-valued measure $\Phi$ satisfies the conjugate symmetry property
$$\Phi(S^{-1}) = \overline{\Phi(S)}$$ (1.8)
for any Borel set $S \subset \hat{D}$. Furthermore, the Heisenberg uncertainty relations prevailing among the various position and momentum observables of the quantum process reveal their dramatic influence on the measure $\Phi$ through the matrix inequalities
$$\Phi(S) + \frac{i}{2} \lambda(S) J_{2k} \geq 0$$ (1.9)
for all Borel sets $S \subset \hat{D}$, where $\lambda$ is the normalised Haar measure of the compact group $\hat{D}$ and $J_{2k}$ is the fundamental symplectic matrix given by
$$J_{2k} = \bigoplus_k \text{copies } \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$ (1.10)
which is a diagonal block matrix with each diagonal block equal to $J_2$. The inequality (1.9) implies, in particular, that whenever $\Phi(S) = 0$, $\lambda(S)$ is also zero.

Borrowing from the extensive theory of linear least square prediction of real valued weakly stationary processes pioneered by A. N. Kolmogorov [3, 4] and N. Wiener [10], and multivariate weakly stationary processes by N. Wiener and P. Masani [11, 12] we call (1.7) the spectral representation of $\tilde{K}$ in $\hat{D}$ and the matrix-valued positive measure $\Phi$ the Kolmogorov-Wiener-Masani spectrum (or KWM spectrum) of the autocovariance map $\tilde{K}$ of the underlying quantum process.

As noted above, inequality (1.9) implies that whenever $\Phi(S) = 0$ for some Borel set $S \subset \hat{D}$, then $\lambda(S) = 0$. In other words, the KWM spectrum does not admit a ‘Haar gap’.

Conversely, given a complex Hermitian positive $2k \times 2k$ matrix-valued measure $\Phi$ on the Borel $\sigma$-algebra of $\hat{D}$ satisfying the conjugate symmetry condition (1.8), the spectral uncertainty relations (1.9), and the condition $\Phi(\hat{D}) = \tilde{K}(0)$, there exists a weakly stationary $k$-mode quantum process over $D$ with KWM spectrum $\Phi$. Indeed, such a process can be realized as a mean zero quantum Gaussian process in the sense that all its finite-partite states $\rho_{a_1, \ldots, a_n}$, where $(a_1, \ldots, a_n) \in S_D$, are mean zero Gaussian states.

The spectral representation of the autocovariance function and its converse enable us to construct interesting examples of weakly stationary quantum processes.

2. Quantum Processes

A quantum system in its most elementary form is determined by a pair $(\mathcal{H}, \rho)$, where $\mathcal{H}$ is a complex separable Hilbert space and $\rho$ is a density operator in $\mathcal{H}$, i.e., a positive operator with unit trace. The operator $\rho$ is called the state of the system. We shall deal with several quantum systems and assume that all the Hilbert spaces in this paper are complex and separable. Scalar products in Hilbert spaces will be expressed in the Dirac notation and adjoints of operators as well as
matrices will be indicated by the symbol \( \dagger \). By a positive operator \( X \) in a Hilbert space \( \mathcal{H} \) we mean that \( \langle u | X | u \rangle \geq 0 \) for all \( u \in \mathcal{H} \). By a positive \( n \times n \) matrix we mean an \( n \times n \) Hermitian matrix which is positive semidefinite.

If \( \mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n \) is the tensor product of Hilbert spaces \( \mathcal{H}_i \) for \( 1 \leq i \leq n \), \( \rho \) is a state in \( \mathcal{H} \) and \( F \subset \{ 1, 2, \ldots, n \} \) is the subset \( \{ i_1 < i_2 < \cdots < i_k \} \) then we write \( \mathcal{H}_F = \mathcal{H}_{i_1} \otimes \cdots \otimes \mathcal{H}_{i_k} \). One obtains a state \( \rho_F \) in \( \mathcal{H}_F \) by taking the relative trace of \( \rho \) successively in \( \mathcal{H}_i \), for all \( i \notin F \) in some order. The resulting state \( \rho_F \) is independent of the order in which the traces are taken. The system \(( \mathcal{H}_F, \rho_F )\) is called the \( F\)-marginal of \(( \mathcal{H}, \rho)\).

In the Hilbert space of any quantum system a bounded or unbounded self-adjoint operator \( X \) is called an observable of the system. Suppose \( \mathcal{F}_\mathbb{R} \) is the Borel \( \sigma \)-algebra of \( \mathbb{R} \) and \( P^X(\cdot) \) is the spectral measure of \( X \) on \( \mathcal{F}_\mathbb{R} \). Then the quantity \( \text{Tr} \rho P^X(E) \), where \( E \in \mathcal{F}_\mathbb{R} \), is interpreted as the probability that the observable takes a value in \( E \) in the state \( \rho \). Thus \( \text{Tr} \rho P^X(\cdot) \) is the distribution of \( X \) in the state \( \rho \). Such an interpretation enables the computation of all moments of \( X \). Indeed, the \( n \)-th moment of \( X \), if it exists, is denoted by \( \langle X^n \rangle \) and is given by

\[
\langle X^n \rangle = \text{Tr} \, X^n \, \rho.
\]

If \( X, \ Y \) are two observables such that \( XY + YX \) is also an observable then the covariance between \( X \) and \( Y \) in the state \( \rho \) is denoted by \( \text{Cov}(X, Y) \) and is defined as

\[
\text{Cov}(X, Y) = \left( \frac{1}{2} \langle XY + YX \rangle - \langle X \rangle \langle Y \rangle \right).
\]

The quantity \( \text{Cov}(X, Y) \) is called the variance of \( X \). If \( X_1, X_2, \ldots, X_n \) are observables with well-defined covariance between \( X_i \) and \( X_j \) for all \( i, j \) then the \( n \times n \) positive matrix

\[
\Sigma_n = \Sigma_n(X_1, \cdots, X_n) = \text{diag}([\text{Cov}(X_i, X_j)])
\]

is called the covariance matrix of the observables \( (X_1, X_2, \cdots, X_n) \) in the state \( \rho \).

Consider a composite quantum system \(( \mathcal{H}, \rho)\), where \( \mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n \). If the set \( \{ 1, 2, \ldots, n \} = E \cup F \) with \( E \cap F = \emptyset, E \neq \emptyset \), and \( F \neq \emptyset \) then \( \mathcal{H} \) can be viewed as the tensor product

\[
\mathcal{H} = \mathcal{H}_E \otimes \mathcal{H}_F
\]

and an observable in \( \mathcal{H}_E \) can be looked upon as the observable \( X_E \otimes I_F \) in \( \mathcal{H} \) with \( I_F \) being the identity operator in \( \mathcal{H}_F \). We call \( X_E \otimes I_F \) the ampliation of \( X_E \) in \( \mathcal{H} \) and denote it by the same symbol \( X_E \). If \( \rho_E \) is the \( E \)-marginal of \( \rho \) in \( \mathcal{H}_E \) then

\[
\langle X_E \rangle = \text{Tr} \, X_E \rho_E = \text{Tr} \, X_E \rho = \langle X_E \otimes I_F \rangle.
\]

We now introduce the notion of a quantum process over a countable index set \( D \). Let \( \{ \mathcal{H}_a : a \in D \} \) be a family of Hilbert spaces. Denote by \( \mathcal{S}_D \) the set of all finite sequences of distinct elements from \( D \). Suppose \( \rho_{a_1, a_2, \ldots, a_n} \) is a density operator in \( \mathcal{H}_{a_1, a_2, \ldots, a_n} \) as in (1.1) for each \( (a_1, a_2, \ldots, a_n) \) in \( \mathcal{S}_D \), satisfying the following properties:

1. If \( \{a_1, a_2, \ldots, a_n\} = \{b_1, b_2, \ldots, b_n\} \) as sets and \( \pi \) is a permutation of \( \{1, 2, \ldots, n\} \) such that \( a_{\pi(j)} = b_j, \forall j \) and \( U_\pi : \mathcal{H}_{a_1, a_2, \ldots, a_n} \rightarrow \mathcal{H}_{b_1, b_2, \ldots, b_n} \)

is the natural Hilbert space isomorphism induced by $\pi$ then
\[ \rho_{a_1, a_2, \cdots, a_n} = U_\pi \rho_{a_1, a_2, \cdots, a_n} U_\pi^{-1}. \]

(2) The $\{a_1, a_2, \cdots, a_n\}$-marginal of $\rho_{a_1, a_2, \cdots, a_{n+1}}$ is equal to $\rho_{a_1, a_2, \cdots, a_n}$ for all $(a_1, a_2, \cdots, a_{n+1}) \in S_D$, $n = 1, 2, \cdots.$

Then we say that $\{\rho_{a_1, a_2, \cdots, a_n} : (a_1, \cdots, a_n) \in S_D\}$ is a consistent family of states. The family $\mathfrak{g} = \{(H_{a_1, \cdots, a_n}, \rho_{a_1, \cdots, a_n}) : (a_1, \cdots, a_n) \in S_D\}$ of finite-partite quantum systems is called a quantum process over $D$.

One obtains interesting examples of discrete ‘time’ quantum processes with $D$ equal to $\mathbb{Z}$, $\mathbb{Z}^d$ or a general discrete abelian group. When $D$ is $\mathbb{Z}$, the element $a$ in $\mathbb{Z}$ can be interpreted as time. In general, $a$ in $D$ is interpreted as a site.

Suppose $D = \{0, 1, 2, \cdots\}$ and $\mathcal{H}_n = \mathcal{H}_0 \otimes \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$. Let $\rho_n$ be a density operator in $\mathcal{H}_n$ such that $\rho_{n-1}$ is the marginal in $\mathcal{H}_{n-1}$ obtained by tracing out $\rho_n$ over $\mathcal{H}_n$ for each $n$. Then $\{(\mathcal{H}_n, \rho_n) : n = 0, 1, 2, \cdots\}$ yields a quantum process. Denote by $\mathcal{B}_n$ the C* algebra of all bounded operators in $\mathcal{H}_n$. Then there is a natural C* embedding $\phi_n : \mathcal{B}_n \to \mathcal{B}_{n+1}$ with the property
\[ \phi_n(X) = X \otimes I \quad \forall X \in \mathcal{B}_n, \]
where $I$ is the identity operator in $\mathcal{H}_{n+1}$. This enables the construction of an inductive limit C* algebra $\mathcal{B}_\infty$ with a C* embedding $\pi_n : \mathcal{B}_n \to \mathcal{B}_\infty$ such that the sequence $\{\pi_n(\mathcal{B}_n)\}$ is increasing in $n$ and $\bigcup_n \pi_n(\mathcal{B}_n)$ is dense in $\mathcal{B}_\infty$. This yields a normalized positive linear functional $\omega$ in $\mathcal{B}_\infty$ such that
\[ \omega(\pi_n(X)) = \rho_n(X) \quad \forall X \in \mathcal{B}_n \text{ and } n = 0, 1, 2, \cdots. \]

In other words $(\mathcal{B}_\infty, \omega)$ is a C* probability space which may be considered as the analogue of Kolmogorov’s measure space constructed from a consistent family of finite dimensional probability distributions. However, there is no limiting Hilbert space in general with a density operator. A similar construction of a C* probability space is possible for a quantum process over any countable index set $D$.

**Definition 2.1.** Suppose $D$ is a countable abelian group with addition operation $+$, $\mathcal{H}_a = \mathcal{H}$ for all $a \in D$, and $\mathfrak{g}$ is a quantum process over $D$. Then it is said to be strictly stationary or translation invariant if
\[ \rho_{a_1 + x, a_2 + x, \cdots, a_n + x} = \rho_{a_1, a_2, \cdots, a_n} \quad \forall x \in D \text{ and } (a_1, \cdots, a_n) \in S_D. \]

Let $\{\rho_{a_1, \cdots, a_n}\}$ and $\{\sigma_{a_1, \cdots, a_n}\}$, where $(a_1, \cdots, a_n) \in S_D$, be a pair of consistent families of finite-partite states in $\{\mathcal{H}_{a_1, \cdots, a_n}\}$. Then, for any $0 < p < 1$, setting
\[ \tau_{a_1, \cdots, a_n} = p \rho_{a_1, \cdots, a_n} + (1 - p) \sigma_{a_1, \cdots, a_n} \quad \forall (a_1, \cdots, a_n) \in S_D \]
yields a consistent family of finite-partite states.

Suppose, $a \mapsto U_a$ is any map on $D$ where $U_a$ is a unitary operator in $\mathcal{H}_a$ for every $a \in D$. Then setting
\[ \rho_{a_1, \cdots, a_n} = (U_{a_1} \otimes \cdots \otimes U_{a_n}) \rho_{a_1, \cdots, a_n} (U_{a_1}^\dagger \otimes \cdots \otimes U_{a_n}^\dagger) \quad \forall (a_1, \cdots, a_n) \in S_D \]
also yields a consistent family of states. Indeed, this is a consequence of the following proposition.
Proposition 2.2. Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, $\rho$ a state in $\mathcal{H} \otimes \mathcal{K}$, and $U$ and $V$ unitary operators in $\mathcal{H}$ and $\mathcal{K}$ respectively. Then

$$\text{Tr}_{\mathcal{K}}(U \otimes V)\rho(U \otimes V)^\dagger = U (\text{Tr}_{\mathcal{K}} \rho) U^\dagger,$$

where $\text{Tr}_{\mathcal{K}}$ is relative trace over $\mathcal{K}$.

Proof. This is immediate from the fact that the relative trace over $\mathcal{K}$ can be computed by using any orthonormal basis $\{e_j\}$ in $\mathcal{K}$, and if $\{e_j\}$ is one such basis so is $\{V^\dagger e_j\}$.

Combining the two elementary remarks above we can construct new quantum processes over $D$ from a given quantum process $\{(\mathcal{H}_{a_1},\ldots,a_n,\rho_{a_1},\ldots,a_n) : (a_1,\ldots,a_n) \in S_D\}$ as follows: Start with a probability space $(\Omega, F, P)$ and a random process $\{U_a(\omega) : a \in D\}$ where $U_a(\omega)$ is a unitary operator in $\mathcal{H}_a$ for every $a$. Define

$$\rho'_{a_1,\ldots,a_n} = \int_{\Omega} P(d\omega) (U_{a_1} \otimes \cdots \otimes U_{a_n}) \rho_{a_1,\ldots,a_n} (U_{a_1} \otimes \cdots \otimes U_{a_n})^\dagger. \quad (2.1)$$

Then $\{\rho'_{a_1,\ldots,a_n} : (a_1,\ldots,a_n) \in S_D\}$ is also a consistent family of finite-partite states.

Remark 2.3. When $D$ is a countable additive abelian group and $\mathcal{H}_a = \mathcal{H}$ for all $a \in D$, $\rho$ is a strictly stationary quantum process and the random process $\{U_a(\omega) : a \in D\}$ is also strictly stationary, then the quantum process $\rho'$ determined by equation (2.1) is also strictly stationary.

3. Multi-mode Processes and Their Covariance Kernels

We now pass on to the definition of a $k$-mode quantum process over a countable index set $D$. Let $\mathcal{H}_a = L^2(\mathbb{R}^k)$ for each $a \in D$, where $k$ is a fixed positive integer called the number of modes. We view $\mathcal{H}_a$ as the $a$-th copy of $L^2(\mathbb{R}^k)$ and introduce the canonical Schrödinger pairs of position and momentum observables $q_{a_1}, p_{a_1}$, where $1 \leq j \leq k$, given by

$$(q_{a_1} f)(x) = x_j f(x),$$

$$(p_{a_1} f)(x) = \frac{1}{i} \frac{\partial}{\partial x_j} f(x)$$

on their respective maximal domains in $L^2(\mathbb{R}^k)$, $x$ denoting $(x_1,x_2,\cdots,x_k) \in \mathbb{R}^k$.

We arrange these $2k$ observables as

$$(X_{a_1},X_{a_2},\cdots,X_{a_{2k-1}},X_{a_{2k}}) = (q_{a_1},p_{a_1},\cdots,q_{a_k},p_{a_k}).$$

Let now $\rho$ be a quantum process over $D$. Then $X_{a_r}$ can be viewed as an amplified observable in $\mathcal{H}_{a_1,a_2,\cdots,a_n}$ whenever the element $a$ occurs in the sequence $a_1,a_2,\cdots,a_n$. We assume that all observables which are closures of polynomials of degree not exceeding 2 in $\{X_{a_r} : a \in D, r = 1,2,\cdots,2k\}$ have finite expectations under the process so that $X_{a_r}$ and $X_{b_s}$ have a well-defined covariance for any $a,b \in D$, and $r,s \in \{1,2,\cdots,2k\}$. We write for any $a,b \in D$,

$$\kappa_{r,s}(a,b) = \text{Cov}(X_{a_r},X_{b_s}) \quad \forall r,s \in \{1,2,\cdots,2k\},$$
where the covariances can be evaluated in any state \( \rho_{a_1, a_2, \ldots, a_n} \) when both \( a \) and \( b \) occur in \( (a_1, \ldots, a_n) \in S_D \). Indeed, this follows from consistency of the states occurring in the quantum process. We call \( \rho \) to occur in \((a, b, c)\) for all \( a, b, c \in D \). Thus, \( \rho \) is a \( 2 \times 2 \) matrix \( \begin{bmatrix} \kappa_{ab} & \kappa_{ac} \\ \kappa_{cb} & \kappa_{cc} \end{bmatrix} \) with \( a, b, c \in D \).

If \( D \) is an additive abelian group, \( (X_{a,b}) = 0 \) for all \( a \in D, 1 \leq j \leq 2k \), and \( K(a, b) = K(a + c, b + c) \) for all \( a, b, c \in D \) we then say that \( \rho \) is a mean zero second order weakly stationary or simply weakly stationary \( k \)-mode quantum process. In such a case, there exists a map \( \hat{K} \) from \( D \) into the space of \( 2k \times 2k \) real matrices, such that \( K(a, b) = \hat{K}(b - a) \). This map \( \hat{K} \) is called the autocovariance map of the weakly stationary process.

**Theorem 3.1.** Let \( K(a, b) = \begin{bmatrix} \kappa_{ab} \\ \kappa_{cb} \end{bmatrix} \) where \( a, b \in D \), be a family of \( 2k \times 2k \) real matrices satisfying the following conditions:

\[
\kappa_{ab} = \kappa_{ba} \quad \forall r, s \in \{1, 2, \ldots, 2k\} \text{ and } a, b \in D.
\]

Then there exists a \( k \)-mode quantum process \( \rho \) with covariance kernel \( K(\cdot, \cdot) \) if and only if for any sequence \( (a_1, \ldots, a_n) \in S_D \) the block matrix \( [[K(a_i, a_j)]] \) satisfies the matrix inequality

\[
[[K(a_i, a_j)]] + \frac{i}{2} J_{2kn} \geq 0.
\] (3.1)

**Proof.** Since \( \rho_{a_1, \ldots, a_n} \) is a \( k \)-mode state and \( [[K(a_i, a_j)]] \) is the covariance matrix of the position-momentum observables \( (X_{a_1,1}, X_{a_1,2}, \ldots, X_{a_1,2k}, X_{a_2,1}, \ldots, X_{a_2,2k}, \ldots, X_{a_n,1}, \ldots, X_{a_n,2k}) \) in \( L^2(\mathbb{R}^{kn}) \), necessity is immediate from the uncertainty relation fulfilled by such a covariance matrix [2, 1, 7]. From Theorem 3.1 of [7] and inequality (3.1), it follows that there exists a Gaussian state \( \{\rho_{a_1, \ldots, a_n}\} \) with covariance matrix \( [[K(a_i, a_j)]] \). Then \( \{\rho_{a_1, \ldots, a_n}\} \) is a consistent family of Gaussian states constituting the required quantum process. \(\square\)

**Definition 3.2.** A kernel \( \mathcal{K} = \begin{bmatrix} [K(a, b)] \end{bmatrix} \) \( \forall a, b \in D \), where \( K(a, b) \) are real \( 2k \times 2k \) matrices satisfying the conditions

1. \( K(a, b)^T = K(b, a) \) for all \( a, b \in D \),
2. \( [[K(a_i, a_j)]] \geq 0 \) for all \( (a_1, a_2, \ldots, a_n) \in S_D \),

is called a \( k \)-mode classical covariance kernel.

If, in addition, the inequality (3.1) is fulfilled, then it is called a \( (k, m) \)-mode quantum covariance kernel.

**Corollary 3.3.** If \( \mathcal{K} \) is a \( k \)-mode quantum covariance kernel and \( \mathcal{C} \) is a \( k \)-mode classical covariance kernel then \( \mathcal{K} + \mathcal{C} \) is a \( k \)-mode quantum covariance kernel.

**Proof.** Immediate. \(\square\)

Let \( \rho \) be a \( k \)-mode quantum process over \( D \) with quantum covariance kernel \( \mathcal{K} = [[K(a, b)]] \). Suppose \( \mathcal{C} = [[C(a, b)]] \) is the covariance kernel of a real \( 2k \)-variate classical stochastic process with index set \( D \) so that the matrix inequalities

\[
\sum_{i,j} \alpha_i \alpha_j C(a_i, a_j) \geq 0
\]
for all real scalars \( \alpha_1, \cdots, \alpha_n \) and all elements \( a_1, \cdots, a_n \in D \). Then the sum

\[
\mathcal{K} + \mathcal{C} = [[K(a, b) + C(a, b)]]
\]

is the covariance kernel of a \( k \)-mode quantum process \( \sigma \). We shall now realise such a process \( \sigma \) by an explicit construction which is an interaction between the quantum process \( \varphi \) and a family of unitary conjugations mediated by a classical process with covariance kernel \( \mathcal{C} \).

To this end we start with the 1-mode Hilbert space \( L^2(\mathbb{R}) \), its Schrödinger position-momentum pair \( q, p \), the associated annihilation-creation pair \( \hat{a}, \hat{a}^\dagger \) given by \( \hat{a} = 2^{- \frac{1}{2}}(q + ip) \), \( \hat{a}^\dagger = 2^{- \frac{1}{2}}(q - ip) \) and the unitary Weyl (displacement) operators \( W(z) = \exp(\hat{z}\hat{a}^\dagger - \hat{z}\hat{a}) \), where \( z \in \mathbb{C} \). These satisfy the relations

\[
W(z)\hat{a}W(z)^\dagger = \hat{a} - z \quad \forall z \in \mathbb{C},
\]

with the convention that \( z \) denotes the scalar as well as the operator \( zI \). This leads to the relations

\[
W(2^{- \frac{1}{2}z})qW(2^{- \frac{1}{2}z})^\dagger = q - x, \tag{3.2}
\]

\[
W(2^{- \frac{1}{2}z})pW(2^{- \frac{1}{2}z})^\dagger = p - y, \tag{3.3}
\]

where \( x = \text{Re} \, z \) and \( y = \text{Im} \, z \).

Now for \( a \in D \), let

\[
\begin{align*}
\mathbf{z}_a &= (z_{a1}, z_{a2}, \cdots, z_{ak})^T, \\
\mathbf{z}_{ar} &= (x_{ar} + iy_{ar}),
\end{align*}
\]

where \( x_{ar} = \text{Re} \, z_{ar} \) and \( y_{ar} = \text{Im} \, z_{ar} \). Viewing \( \mathcal{H} = L^2(\mathbb{R}) \) as the \( k \)-fold product \( L^2(\mathbb{R}) \otimes L^2(\mathbb{R}) \otimes \cdots \otimes L^2(\mathbb{R}) \), introduce the \( k \)-mode Weyl operators

\[
W(\mathbf{z}_a) = W(z_{a1}) \otimes \cdots \otimes W(z_{ak}).
\]

Then the relations (3.2, 3.3) yield the relations for the operators \( X_a1, \cdots, X_{a2k} \) defined above as

\[
W(2^{- \frac{1}{2}z_a}) \begin{bmatrix} X_{a1} \\ X_{a2} \\ \vdots \\ X_{a2k} \end{bmatrix} W(2^{- \frac{1}{2}z_a})^\dagger = \begin{bmatrix} X_{a1} - \alpha_{a1} \\ X_{a2} - \alpha_{a2} \\ \vdots \\ X_{a2k} - \alpha_{a2k} \end{bmatrix} \tag{3.4}
\]

where

\[
(\alpha_{a1}, \alpha_{a2}, \cdots, \alpha_{a2k}) = (x_{a1}, y_{a1}, x_{a2}, y_{a2}, \cdots, x_{ak}, y_{ak}). \tag{3.5}
\]

Let \( (\xi_{a1}, \eta_{a1}, \xi_{a2}, \eta_{a2}, \cdots, \xi_{ak}, \eta_{ak}) (\omega) \), where \( \omega \in \Omega \), and \( a \in D \), be a \( 2k \) real variate stochastic process defined on a probability space \( (\Omega, \mathcal{F}, P) \) with zero mean and covariance kernel \( \mathcal{C} = [[C(a, b)]] \), where

\[
C(a, b) = \mathbb{E} \begin{bmatrix} \xi_{a1} \\ \eta_{a1} \\ \xi_{a2} \\ \eta_{a2} \\ \vdots \\ \xi_{ak} \\ \eta_{ak} \end{bmatrix} \begin{bmatrix} \xi_{a1} & \eta_{a1} & \xi_{a2} & \eta_{a2} & \cdots & \xi_{ak} & \eta_{ak} \end{bmatrix} \quad \forall a, b \in D. \tag{3.6}
\]
Define the random unitary operators $U_a(\omega)$ in $\mathcal{H}_a$, where $a \in D$, by putting

\[
\zeta_a r = \xi_a r + \eta_a r \quad \text{for} \quad r = 1, 2, \ldots, k, 
\]

(3.7)

\[
U_a(\omega) = W(2^{-\frac{k}{2}} \zeta_a(\omega))^\dagger, 
\]

(3.8)

where $\zeta_a = (\zeta_a 1, \ldots, \zeta_a k) \in \mathbb{C}^k$. By following the remarks in §2 around equation (2.1), define the $k$-mode quantum process $\sigma$ by

\[
\sigma_{a_1, \ldots, a_n} = \int \Omega \rho a_1 U_a(\omega) \otimes \cdots \otimes U_a(\omega) \rho a_1, \ldots, a_n U_a(\omega)^\dagger \otimes \cdots \otimes U_a(\omega)^\dagger. 
\]

(3.9)

Then we have the following theorem:

**Theorem 3.4.** The covariance kernel of the $\sigma$ process determined by the finite-partite states (3.9) is equal to $K + C$.

**Proof.** Consider the observable $X_{a r}$. Its expectation under the $\sigma$ process is given by

\[
\text{Tr} X_{a r} \sigma_a = \int \rho a_1 U_a(\omega) \otimes \cdots \otimes U_a(\omega) \rho a_1, \ldots, a_n U_a(\omega)^\dagger \otimes \cdots \otimes U_a(\omega)^\dagger. 
\]

(3.10)

Going to second order moments

\[
\text{Tr} X_{a r} X_{b s} \sigma_{a b} 
\]

\[
= \int \rho a_1 U_a(\omega) \otimes U_b(\omega) \rho a_1 U_a(\omega)^\dagger \otimes U_b(\omega)^\dagger 
\]

\[
= \int \rho a_1 W(2^{-\frac{k}{2}} \zeta_a) \otimes W(2^{-\frac{k}{2}} \zeta_b) X_{a r} X_{b s} W(2^{-\frac{k}{2}} \zeta_a)^\dagger \otimes W(2^{-\frac{k}{2}} \zeta_b)^\dagger \rho a_1 U_a(\omega)^\dagger \otimes U_b(\omega)^\dagger 
\]

\[
= \int \rho a_1 U_a(\omega) \otimes \cdots \otimes U_a(\omega) \rho a_1, \ldots, a_n U_a(\omega)^\dagger \otimes \cdots \otimes U_a(\omega)^\dagger. 
\]

Let $a \neq b$. Then

\[
\text{Cov}_\sigma(X_{a r}, X_{b s}) = \langle X_{a r} X_{b s} \rangle - \langle X_{a r} \rangle \langle X_{b s} \rangle 
\]

\[
= \langle X_{a r} X_{b s} \rangle - \langle X_{a r} \rangle \langle X_{b s} \rangle = C_{r s}(a, b). 
\]
Let \( a = b \). Then \( C_{rs}(a,b) = C_{sr}(a,a) = C_{sr}(a,a) \), and so
\[
\Cov(\sigma X_{ar}, X_{as}) = \left\langle \frac{1}{2}(X_{ar}X_{as} + X_{as}X_{ar}) \right\rangle - \left\langle X_{ar} \right\rangle \left\langle X_{as} \right\rangle \\
= \left\langle \frac{1}{2}(X_{ar}X_{as} + X_{as}X_{ar}) \right\rangle + C_{rs}(a,a) \left\langle X_{as} \right\rangle \\
= \Cov(\sigma X_{ar}, X_{as}) + C_{rs}(a,a) \\
= K_{rs}(a,a) + C_{rs}(a,a).
\]
\[ \square \]

Remark 3.5. If \( \sigma \) is a weakly stationary quantum process and \((\xi_{a1}, \eta_{a1}, \cdots, \xi_{ak}, \eta_{ak})\) is a weakly stationary classical process with mean 0 such that
\[
K(a,b) = \tilde{K}(b-a) \quad \text{and} \\
C(a,b) = \tilde{C}(b-a) \quad \forall a, b \in D,
\]
then \( \sigma \) is also a weakly stationary quantum process. If in addition, \( \sigma \) is Gaussian then so is \( \sigma \).

4. The KWM Spectrum of a Weakly Stationary \( k \)-mode Quantum Process

Let \( \sigma \) be a weakly stationary \( k \)-mode quantum process over a countable discrete additive abelian group \( D \), with autocovariance map \( \tilde{K} \). Let \( \hat{D} \) be the compact dual multiplicative group of all characters of \( D \). Denote by \( \mathcal{F} \) the Borel \( \sigma \)-algebra on \( \hat{D} \).

Define
\[
L(a) = \tilde{K}(a) + \frac{i}{2} \mathbb{I}_{\{0\}}(a) J_{2k} \quad \forall a \in D.
\]

Then Theorem 3.1 yields the following proposition.

**Proposition 4.1.** A real \( 2k \times 2k \) matrix-valued map \( \tilde{K} \) is the autocovariance map of a second order weakly stationary \( k \)-mode quantum process on \( D \) if and only if the associated map \( L \) defined by (4.1) satisfies the following matrix inequalities
\[
[[L(a_s - a_r)]] \geq 0, \quad \text{where} \ r, s \in \{1, 2, \cdots, n\},
\]
for all \((a_1, a_2, \cdots, a_n) \in \mathcal{S}_D\), and \( n = 1, 2, \cdots \).

**Proof.** Immediate. \[ \square \]

**Theorem 4.2.** A real \( 2k \times 2k \) matrix-valued map \( \tilde{K} \) on \( D \) is the autocovariance map of a second order weakly stationary \( k \)-mode quantum process on \( D \) if and only if there exists a \( 2k \times 2k \) Hermitian positive matrix-valued measure \( \Phi \) on \( (\hat{D}, \mathcal{F}) \) satisfying the following conditions:
\begin{enumerate}
\item \( \Phi(\hat{D}) = \tilde{K}(0) \).
\item \( \tilde{K}(a) = \int_{\hat{D}} \chi(a) \Phi(d\chi) \).
\item \( \Phi(S) + \frac{i}{2} \lambda(S) J_{2k} \geq 0, \forall S \in \mathcal{F}, \) where \( \lambda \) is the normalized Haar measure of the compact group \( \hat{D} \). In particular, \( \lambda \) is absolutely continuous with respect to \( \text{Tr} \Phi \).
\end{enumerate}
(4) \( \Phi(S^{-1}) = \Phi(S)^T, \forall S \in \mathcal{F} \).

Under the conditions (1)-(4) the underlying quantum process can be chosen to be a strictly stationary \( k \)-mode quantum Gaussian process with mean zero.

Proof. Let \( \tilde{K} \) be the autocovariance map of a weakly stationary \( k \)-mode process. Define \( L \) by (4.1). By Proposition 4.1 the matrices \( [[L(a_s - a_r)]] \), where \( r, s \in \{1,2,\ldots,n\} \), are positive for all \( (a_1,a_2,\ldots,a_n) \in \mathcal{S}_D \). Hence for any vector \( u \in \mathbb{C}^{2k} \), the function

\[
a \mapsto \psi_u(a) = u^\dagger L(a) u \quad (4.2)
\]

is positive definite on the abelian group \( D \) in the sense of Bochner. By Bochner’s theorem there exists a totally finite measure \( \nu_u \) on \( \mathcal{F} \) satisfying the relations

\[
\psi_u(a) = \int_{\hat{D}} \chi(a) \nu_u(d\chi) \quad \forall a \in D, \quad (4.3)
\]

\[
\psi_u(0) = u^\dagger L(0) u = u^\dagger \left( \tilde{K}(0) + \frac{i}{2} J_{2k} \right) u. \quad (4.4)
\]

By (4.2) the left hand side of (4.3) is a quadratic form in \( u \) for each fixed \( a \) in \( D \). By the bijective correspondence between totally finite measures on \( \hat{D} \) and their Fourier transforms on \( D \) it follows that there exists a \( 2k \times 2k \) Hermitian positive matrix-valued measure \( \Psi \) on \( \mathcal{F} \) such that

\[
L(a) = \int_{\hat{D}} \chi(a) \Psi(d\chi) \quad \forall a \in D, \quad (4.5)
\]

\[
L(0) = \tilde{K}(0) + \frac{i}{2} J_{2k} = \Psi(\hat{D}) \geq 0. \quad (4.6)
\]

Now define

\[
\phi_u(a) = u^\dagger \tilde{K}(a) u \quad \forall a \in D \text{ and } u \in \mathbb{C}^{2k}.
\]

By (4.1), \( \tilde{K}(a) = \text{Re } L(a) \), where the real part is taken entry-by-entry, and hence \( [[\tilde{K}(a_s - a_r)]] \geq 0 \) for any \( (a_1,a_2,\ldots,a_n) \in \mathcal{S}_D \). In other words, \( \phi_u \) is also a positive definite function on \( D \) and by the same arguments as employed for \( L \) we have the relations

\[
\tilde{K}(a) = \int_{\hat{D}} \chi(a) \Phi(d\chi), \quad (4.7)
\]

\[
\tilde{K}(0) = \Phi(\hat{D}), \quad (4.8)
\]

where \( \Phi \) is again a \( 2k \times 2k \) Hermitian positive matrix-valued measure on \( \mathcal{F} \).

By (4.1)

\[
L(a) - \tilde{K}(a) = \frac{i}{2} \frac{1}{2} L(0) J_{2k}
\]

\[
= \left[ \frac{i}{2} \int_{\hat{D}} \chi(a) \lambda(d\chi) \right] J_{2k} \quad \forall a \in D. \quad (4.9)
\]

Subtracting (4.7) from (4.5) and using (4.9) we have

\[
\int_{\hat{D}} \chi(a)(\Psi - \Phi)(d\chi) = \left[ \frac{i}{2} \int_{\hat{D}} \chi(a) \lambda(d\chi) \right] J_{2k}
\]
for all $a \in D$. Thus by uniqueness of Fourier transforms we have
\[ \Psi(S) - \Phi(S) = \frac{i}{2} \lambda(S) J_{2k}, \quad \forall S \in \mathcal{F}. \]
Thus
\[ \Phi(S) + \frac{i}{2} \lambda(S) J_{2k} \geq 0, \quad \forall S \in \mathcal{F}. \quad (4.10) \]
Now property (1) follows from (4.8), property (2) from (4.7), and property (3) from (4.10). If $\text{Tr} \Phi(S) = 0$ then $\Phi(S) = 0$, by positivity, and (4.10) implies
\[ \frac{i}{2} \lambda(S) J_{2k} \geq 0 \]
which is positive only if $\lambda(S) = 0$, as may be seen by taking the determinant. In other words $\lambda \ll \text{Tr} \Phi$.

To prove property (4) of $\Phi$ we introduce the map $\tau : \hat{D} \to \hat{D}, \tau(\chi) = \chi^{-1}$ and observe that
\[ \tilde{K}(a) = \int_{\hat{D}} \chi(a) \Phi(d\chi) = \int_{\hat{D}} \chi(a) \Phi^{-1}(d\chi). \]
Since $\tilde{K}(a)$ has real entries, property (2) implies
\[ \Phi^{-1}(a) = \Phi(\overline{a}), \quad \forall S \in \mathcal{F}. \]
This completes the proof of necessity. To prove sufficiency consider a $2k \times 2k$ Hermitian positive matrix-valued measure $\Phi$ satisfying properties (3) and (4) of the theorem. Define
\[ \tilde{K}(a) = \int_{\hat{D}} \chi(a) \Phi(d\chi). \]
Then properties (2) and (1) hold. Property (3) implies that the function $L(a)$, on $D$ defined by setting
\[ L(a) = \tilde{K}(a) + \frac{i}{2} \mathbb{1}_{\{0\}}(a) J_{2k} \]
satisfies the matrix inequalities $[L(a_s - a_r)] \geq 0$ for any sequence $(a_1, \ldots, a_n) \in \mathcal{S}_D$. By Proposition 4.1, $\tilde{K}$ is the autocovariance function of a second order weakly stationary $k$-mode quantum process which, indeed, can be chosen to be a strictly stationary Gaussian process of mean zero. \[ \square \]

Remark 4.3. As already described in the introduction, we call equation (2) in Theorem 4.2 the spectral representation of the autocovariance map $\tilde{K}$ and say that $\Phi$ is the Kolmogorov-Wiener-Masani (KWM) spectrum of the $k$-mode weakly stationary quantum process. Theorem 4.2 enables us to construct a whole class of examples of KWM spectra and hence autocovariance maps as follows. Choose and
fix any Borel map $\chi \mapsto M(\chi)$, on $\hat{D}$ where $M(\chi)$ is a $k$-mode quantum covariance matrix of order $2k$, so that

$$M(\chi) + \frac{i}{2} J_{2k} \geq 0 \quad \text{for every } \chi \in \hat{D}.$$ 

Assume that $M(\cdot)$ is integrable with respect to the normalised Haar measure $\lambda$ on $\hat{D}$. Let $\Psi$ be any totally finite positive Hermitian $2k \times 2k$ matrix-valued measure on $(\hat{D}, \mathcal{F})$ satisfying the conjugate symmetry condition $\Psi(S^{-1}) = \overline{\Psi(S)}$ for any Borel set $S \in \mathcal{F}$. Define

$$\Phi(S) = \int_{\hat{D}} M(\chi) \lambda(d\chi) + \Psi(S), \quad S \in \mathcal{F}.$$ 

Then by Theorem 4.2, $\Phi$ is the KWM spectrum of a stationary quantum process over $D$ with autocovariance map $\tilde{K}$ given by equation (2) of the theorem.

**Remark 4.4.** The second part of property (3) of $\Phi$ in Theorem 4.2 implies that $\lambda(S) = 0$ whenever $\text{Tr} \Phi(S) = 0$. In other words the KWM spectrum of a weakly stationary $k$-mode quantum process over $D$ cannot admit a gap of positive Haar measure in $\hat{D}$. For example, when $D = \mathbb{Z}$ and $\hat{D}$ is identified with $[0, 2\pi]$, the KWM spectrum of a stationary $k$-mode quantum process over $\mathbb{Z}$ cannot admit an interval gap.

**Remark 4.5.** In Theorem 4.2, express the KWM spectrum $\Phi$ as $\Phi = [(\phi_{rs})]$, where $r, s \in \{1, 2, \ldots, 2k\}$, and write

$$\Phi_q = [(\phi_{2i-1,2j-1})], \quad \text{where } i, j \in \{1, 2, \ldots, k\},$$

$$\Phi_p = [(\phi_{2i,2j})], \quad \text{where } i, j \in \{1, 2, \ldots, k\}.$$ 

In the inductive limit $C^*$ probability space $(B_\infty, \omega)$ associated with the process $\varrho$ described in §2 the commuting family of position observables $\{q_{ar} : a \in D, r \in \{1, 2, \ldots, k\}\}$ affiliated to $B_\infty$ execute a classical weakly stationary process with spectrum $\Phi_q$. A similar property holds for the family $\{p_{ar} : a \in D, r \in \{1, 2, \ldots, k\}\}$.

In the quantum Gaussian case this raises the question that under what conditions on the spectrum $\Phi$ do these processes enjoy properties like ergodicity, weak mixing, strong mixing etc. The results of G. Maruyama [6] suggest that a minimum requirement would be the absence of atoms in the spectrum $\Phi$.

Suppose $\Phi$ has no atoms. For arbitrary real scalars $c_r$, where $1 \leq r \leq 2k$, consider the associated observables

$$Z_a = \sum_{r=1}^{2k} (c_{2r-1}q_{ar} + c_{2r}p_{ar}) \quad \forall a \in D.$$ 

Then $\{Z_a : a \in D\}$ executes a classical Gaussian process with values in the real line and autocovariance function $c^T \tilde{K}(\cdot)c$ with $c^T = (c_1, c_2, \ldots, c_{2k})$ and spectrum equal to $c^T \Phi(\cdot)c$, a measure in $\hat{D}$. Now let $D = \mathbb{Z}$ be the integer group. Then Maruyama’s theorem implies that this scalar-valued process is, indeed, weakly
mixing, and in particular ergodic. If \( \lim_{a \to \infty} e^{T} \tilde{K}(a)c = 0 \), then this scalar-valued process is also strongly mixing.

**Remark 4.6.** Following [9] one can introduce the observable

\[
N_{a,j} = \frac{1}{2} (q_{a,j}^2 + p_{a,j}^2 - 1), \quad \text{where } 1 \leq j \leq k,
\]

which is the number of particles (photons) in the \( j \)-th mode at the site \( a \). If the underlying process \( \rho \) is Gaussian with mean 0 then

\[
\langle N_{a,j} \rangle = \frac{1}{2} \left\{ \phi_{2j-1,2j-1}(\tilde{D}) + \phi_{2j,2j}(\tilde{D}) - \frac{1}{2} \right\}.
\]

This is a consequence of property (1) of \( \Phi \) in Theorem 4.2 and equation (3.4) from Corollary 3.1 in [9].

This shows that the relationships between photon numbers and KWM spectrum need a deeper exploration.

**Theorem 4.7.** Let \( D \) be a countable, discrete, additive abelian group with its compact character group \( \hat{D} \) and let \( \lambda \) be the normalized Haar measure of \( \hat{D} \) on its Borel \( \sigma \)-algebra \( F \). Suppose \( \Phi \) is a complex, totally finite \( 2^k \times 2^k \) positive matrix-valued measure on \( F \). Then \( \Phi \) is the KWM spectrum of a \( k \)-mode weakly stationary quantum process \( \rho \) over \( D \) if and only if \( \Phi \) admits the representation

\[
\Phi(S) = \int_{F} F(\chi) \lambda(d\chi) + \Psi(S) \quad \forall S \in F,
\]

where \( F \) is a \( 2k \times 2k \) positive Hermitian matrix-valued Borel function satisfying the matrix inequality

\[
F(\chi) + \frac{i}{2} J_{2k} \geq 0 \quad \forall \chi \in \hat{D},
\]

and

\[
S \mapsto \left[ [\psi_{rs}(S)] \right] \quad (4.13)
\]

is a \( 2k \times 2k \) positive matrix-valued measure on \( F \) with each \( \psi_{rs} \) being singular with respect to \( \lambda \).

In particular, \( F \) can be chosen to satisfy

\[
\det \text{Re} F(\chi) \geq \frac{1}{4kn} \quad \forall \chi \in \hat{D}
\]

and the absolutely continuous part of \( u^T \Phi(\cdot)u \) is equivalent to \( \lambda \) for every nonzero element \( u \) of \( \mathbb{C}^{2k} \).

**Proof.** Let \( \Phi = [[\phi_{rs}]] \), with \( r, s \in \{1, 2, \cdots, 2k\} \), be the KWM spectrum of a \( k \)-mode weakly stationary quantum process \( \rho \) over \( D \). By property (3) of Theorem 4.2 it follows that

\[
\begin{bmatrix}
\phi_{2r-1,2r-1}(S) & \phi_{2r-1,2r}(S) + \frac{1}{2} \lambda(S) \\
\phi_{2r,2r-1}(S) - \frac{1}{4} \lambda(S) & \phi_{2r,2r}(S)
\end{bmatrix} \geq 0
\]
for any $S \in \mathcal{F}$. Suppose $\phi_{2r-1,2r-1}(S) = 0$. Then positivity of $\Phi$ implies that $\phi_{2r-1,2r}(S) = \phi_{2r,2r-1}(S) = 0$ and hence

$$\begin{bmatrix} 0 & \frac{i}{2}\lambda(S) \\ -\frac{i}{2}\lambda(S) & \phi_{2r,2r}(S) \end{bmatrix} \geq 0.$$  

Since the determinant of the left hand side in the inequality is nonnegative we conclude that $\lambda(S) = 0$. In other words $\lambda \ll \phi_{2r-1,2r-1}$. By the same argument $\lambda \ll \phi_{2r,2r}$. In other words $\lambda$ is absolutely continuous with respect to every diagonal entry of $\Phi$.

Choose and fix an $S_0 \in \mathcal{F}$ such that $S_0 = S^{-1}$, $\lambda(S_0) = 1$, the measure $\mu_{rr}$ defined by setting $\mu_{rr}(S) = \phi_{rr}(S \cap S_0) \forall S \in \mathcal{F}$ is the part of $\phi_{rr}$ equivalent to $\lambda$ and the measure $\psi_{rr}$ defined by setting $\psi_{rr}(S) = \phi_{rr}(S \cap (\hat{D} \setminus S_0)) \forall S \in \mathcal{F}$ is the part of $\phi_{rr}$ singular with respect to $\lambda$, so that $\phi_{rr} = \mu_{rr} + \psi_{rr} \forall r, 1, 2, \cdots, 2k$.

Now define

$$\mu_{rs}(S) = \phi_{rs}(S \cap S_0) \quad \text{and} \quad \psi_{rs}(S) = \phi_{rs}(S \cap (\hat{D} \setminus S_0)) \forall S \in \mathcal{F}.$$  

If $\lambda(S) = 0$ then $\mu_{rr}(S) = 0$, so $\phi_{rr}(S \cap S_0) = 0$, whence $\phi_{rs}(S \cap S_0) = 0$ and therefore $\mu_{rs}(S) = 0$. In other words $\mu_{rs} \ll \lambda$. By definition all the measures $\psi_{rs}$ are singular with respect to $\lambda$ and

$$\phi_{rs} = \mu_{rs} + \psi_{rs} \forall r, s \in \{1, 2, \cdots, 2k\}.$$  

Define $f_{rs}$ to be the Radon Nykodym derivative of $\mu_{rs}$ with respect to $\lambda$ and put

$$F(\chi) = [f_{rs}(\chi)] \forall r, s \in \{1, 2, \cdots, 2k\}.$$  

Now $F$ is defined a.e. with respect to $\lambda$ and

$$\Phi(S) = \int_S F(\chi) \lambda(d\chi) + \Psi(S) \forall S \in \mathcal{F}, \quad \text{(4.15)}$$

where every entry $\psi_{rs}$ of $\Psi$ is singular with respect to $\lambda$. By the choice of $S_0$ it follows that

$$\int_S F(\chi) \lambda(d\chi) + \frac{i}{2} \lambda(S) J_{2k} \geq 0$$

for all $S \subset S_0$, such that $S \in \mathcal{F}$, so that

$$\int_S \left( F(\chi) + \frac{i}{2} J_{2k} \right) \lambda(d\chi) \geq 0.$$  

Thus

$$F(\chi) + \frac{i}{2} J_{2k} \geq 0, \quad \lambda \text{ a.e.,} \quad \text{(4.16)}$$

and also, by the symmetry of the Haar measure,

$$F(\chi^{-1}) + \frac{i}{2} J_{2k} \geq 0, \quad \lambda \text{ a.e..} \quad \text{(4.17)}$$
On the other hand
\[
\Phi(S^{-1}) = \int_S F(\chi^{-1}) \lambda(d\chi) + \Psi(S^{-1}),
\]
\[
\Phi(S) = \int_S F(\chi) \lambda(d\chi) + \Psi(S).
\]
The conjugate symmetry of \( \Phi \), property (4) of Theorem 4.2, and the fact that \( \Psi(S) = 0 \) if \( S \subseteq S_0 = S_0^{-1} \) imply
\[
F(\chi^{-1}) = \overline{F(\chi)}, \quad \lambda \text{ a.e.}
\]
Choosing \( F(\chi) = \frac{1}{2} I_{2k} \) whenever this fails we may assume that \( F(\chi^{-1}) = \overline{F(\chi)} \) for all \( \chi \). Now (4.16) and (4.17) imply that \( F \) can be altered on a set of \( \lambda \)-measure zero so that
\[
\text{Re} F(\chi) = \frac{F(\chi) + F(\chi^{-1})}{2} \geq -\frac{i}{2} J_{2k}
\]
holds for every \( \chi \). In other words \( F \) is a complex positive \( 2k \times 2k \) matrix whose real part is the quantum covariance matrix of position and momentum observables in a \( k \)-mode state in \( L^2(\mathbb{R}^k) \). It follows from [8] that
\[
\det \text{Re} F(\chi) \geq \frac{1}{4^k} \quad \forall \chi \in \hat{D}
\]
and the representation (4.11) holds.

The converse is already a part of Theorem 4.2. \( \square \)

**Remark 4.8.** Suppose the quantum process \( \varrho \) of Theorem 4.7 is Gaussian and symmetric under the reflection transformation \( a \mapsto -a \) in \( D \). Then the autocovariance function \( K \) of the process \( \varrho \) satisfies the condition \( K(a) = K(-a) \), \( a \in D \) and the KWM spectrum \( \Phi \) is real, i.e., \( \Phi = \overline{\Phi} \). Then (4.14) implies
\[
\int_{\hat{D}} \log \det \Phi(\chi) \lambda(d\chi) > -\infty.
\]
When \( D = \mathbb{Z} \) is the integer group it follows from the Wiener Masani theorem, in particular, that the position observables \( \{q_n, \cdots, q_{n,k}\} \) execute a purely indeterministic shift invariant \( k \)-variate Gaussian process. So do the momentum observables \( \{p_n, \cdots, p_{n,k}\} \).

When \( D = \mathbb{Z} \) is the integer group it follows from the Wiener Masani theorem, in particular, that the position observables \( \{q_n, \cdots, q_{n,k}\} \) execute a purely indeterministic shift invariant \( k \)-variate Gaussian process. So do the momentum observables \( \{p_n, \cdots, p_{n,k}\} \). It may be recalled that a \( d \)-variate, mean zero shift invariant Gaussian process of random variables \( \{\xi_n = (\xi_{n,1}, \xi_{n,2}, \cdots, \xi_{n,d}), -\infty < n < \infty\} \) defined on a probability space \( (\Omega, \mathcal{F}, P) \) is a purely indeterministic if, for any \( n \),
\[
\bigcap_{-\infty}^{n} \mathcal{H}_m = \{0\}
\]
where \( \mathcal{H}_m \) denotes the closed subspace spanned by the random variables \( \{\xi_r, -\infty < r \leq m, 1 \leq j \leq d\} \) in \( L^2(P) \).

**Acknowledgment.** We thank the referee for a careful reading of the manuscript and making valuable suggestions. RS acknowledges financial support from the National Board for Higher Mathematics, Govt. of India.
References


K. R. Parthasarathy: Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi Centre, 7 S J S Sansanwal Marg, New Delhi 110 016, India

E-mail address: krp@isid.ac.in

Ritabrata Sengupta: Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Delhi Centre, 7 S J S Sansanwal Marg, New Delhi 110 016, India

Ritabrata Sengupta (Present address): Department of Mathematics, Indian Institute of Science Education & Research (IISER), Berhampur, Transit campus, Govt. ITI Berhampur, National Highway 59, Engg. School Jn., Berhampur, Odisha 760 010, India.

E-mail address: rb@iiserbpr.ac.in, ritabrata.sengupta@gmail.com
SEMIMARTINGALES IN LOCALLY COMPACT ABELIAN GROUPS AND THEIR CHARACTERISTIC TRIPLES

M. S. BINGHAM

Abstract. The concepts of semimartingales and their characteristic triples are introduced for stochastic processes taking their values in a locally compact second countable abelian group. It is proved that the third characteristic always exists and that the first two characteristics always exist when the group is compact. Any continuous additive Gaussian process in a locally compact second countable abelian group is shown to be a semimartingale with characteristics that agree with its Lévy-Khinchine canonical triple.

1. Introduction

In order to motivate the new developments in this paper, we first recall the definitions of the characteristics of a real-valued semimartingale $X = \{X(t) : t \geq 0\}$ with a right continuous filtration and sample paths that are a.s. càdlàg. See [6] for full details. Choose and fix a continuous truncation function $h$ on the real line; that is, $h : \mathbb{R} \to \mathbb{R}$ is a continuous function with compact support and such that $h(x) = x$ for all $x$ in a neighbourhood $N$ of zero. Define the process $\hat{X}$ for $t \geq 0$ by

$$\hat{X}(t) := X(t) - \sum_{s \leq t} [\Delta X(s) - h(\Delta X(s))],$$

where $\Delta X(s) := X(s) - X(s^-)$ is the jump in $X$ at $s$. The subtracted sum has only finitely many non-zero terms because $X$ has càdlàg sample paths and in the interval $[0, t]$ a càdlàg function can have only a finite number of jumps outside the neighbourhood $N$ of zero. In going from $X$ to $\hat{X}$, each jump $\Delta X(s)$ that has a sufficiently large magnitude to differ from $h(\Delta X(s))$ has been replaced by $h(\Delta X(s))$. Therefore the jumps in $\hat{X}$ are uniformly bounded, which implies that $\hat{X}$ is a special semimartingale. Consequently $\hat{X}$ has a canonical decomposition

$$\hat{X}(t) = X(0) + M(t) + B(t), \quad (t \geq 0),$$

in which $M := \{M(t) : t \geq 0\}$ is a local martingale and $B := \{B(t) : t \geq 0\}$ is a predictable process with finite variation, the sample paths of $M$ and $B$ are almost surely càdlàg, $M(0) = 0$ and $B(0) = 0$. Moreover, $M$ and $B$ are uniquely determined up to indistinguishability. We also know that $X$ has a continuous martingale part $X^c$, which is uniquely determined up to indistinguishability. Let us recall the following.

Received 2016-9-7; Communicated by D. Applebaum.

2010 Mathematics Subject Classification. Primary 60B15; Secondary 60G48.

Key words and phrases. Semimartingale, characteristics, locally compact group.
Definition 1.1. The characteristics of the real-valued semimartingale $X$ are the three predictable a.s. c\`adl\`ag processes $B, C, \nu$, where:

- $B$ is the predictable finite variation process above;
- $C := \langle X^c, X^c \rangle$ is the compensator of the process $(X^c)^2$, so $C$ is the non-decreasing continuous process such that $(X^c)^2 - C$ is a local martingale;
- $\nu$ is the compensator of the jump measure $\mu$ of $X$.

$(B, C, \nu)$ is called the characteristic triple of the semimartingale $X$.

Whereas the process $B$ may depend on the choice of truncation function $h$, the processes $C$ and $\nu$ do not. The characteristics $B$, $C$ and $\nu$ play similar roles in semimartingale theory to those of the drift, the variance of the Gaussian part and the Lévy measure in the theory of processes with independent increments.

These ideas generalise in a natural way to semimartingales with values in $\mathbb{R}^d$ (for any positive integer $d$), and in the book of Jacod and Shiryaev [6] it is shown that, in suitable circumstances, conditions for the convergence in law of a sequence of $\mathbb{R}^d$-valued semimartingales to a limiting semimartingale can be given in terms of the characteristics of the semimartingales. It is natural to ask whether we can develop an analogous theory of semimartingales and their characteristics, together with a corresponding theory of convergence, for stochastic processes which take their values in a locally compact second countable abelian group $G$. Such a theory should be applicable, for example, to proving functional central limit theorems on $G$. This paper presents some first steps towards the development of such a theory by suggesting a way of defining $G$-valued semimartingales and their characteristics. It is proved that the third characteristic always exists and that the first two characteristics always exist if $G$ is compact. In the case when $G = \mathbb{R}$ the definitions are consistent with the classical definitions for real-valued semimartingales, provided that an appropriate local inner product is used. For any locally compact second countable abelian group $G$ that can support a Gaussian distribution, it is shown that any continuous additive Gaussian process on $G$ is a semimartingale with respect to a suitable stochastic basis and that its characteristics exist and coincide with the canonical triple in the Lévy-Khinchine representation.

Except where stated otherwise, the stochastic processes considered in this paper will be defined on the same stochastic basis $\mathcal{B} = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$, in which the filtration is right continuous. Accordingly, when no other basis is mentioned, properties like being ‘adapted’ or ‘predictable’, and concepts like ‘martingale’, ‘local martingale’ and ‘semimartingale’ should be understood as being relative to this stochastic basis. As is customary in order to avoid tedious repetition, the qualification “a.s.($P$)” is sometimes omitted and should therefore be mentally inserted wherever appropriate.

The following notation will be used throughout. $G$ is a locally compact second countable abelian group and $\widehat{G}$ is its dual group. Therefore $\widehat{G}$ is the group of all continuous homomorphisms (characters) of $G$ into the unit circle group $\mathbb{T}$ and is endowed with the topology of uniform convergence on compact subsets of $G$. It is well known that $\widehat{G}$ is also a locally compact second countable abelian group, and the Pontryagin duality theorem tells us that the dual group of $\widehat{G}$ can be identified with $G$. The value of the character $y \in \widehat{G}$ at the point $x \in G$ will be denoted by
Therefore \( \langle x, y \rangle \) and the identity element of \( G \) will be denoted by \( e \). The group operation will be denoted by \( + \) both in \( G \) and in \( \hat{G} \). All neighbourhoods that appear will be assumed without loss of generality to be Borel sets.

General information about locally compact abelian groups can be found in the books by Hewitt and Ross [4] and Rudin [11]. The books by Heyer [5] and Parthasarathy [9] are recommended for probability theory on such groups.

A key role in this paper will be played by a local inner product, which we now define.

**Definition 1.2.** By a local inner product on \( G \times \hat{G} \) we mean a continuous function \( g : G \times \hat{G} \to \mathbb{R} \) with the following properties:

1. \( g(x, y_1 + y_2) = g(x, y_1) + g(x, y_2) \) for all \( x \in G \), \( y_1, y_2 \in \hat{G} \);
2. \( g(-x, y) = -g(x, y) \) for all \( x \in G \), \( y \in \hat{G} \);
3. \( \sup_{x \in G} \sup_{y \in K} |g(x, y)| < \infty \) for every compact set \( K \subseteq \hat{G} \);
4. for every compact \( K \subseteq \hat{G} \) there exists a neighbourhood \( N \) of the identity \( e \) in \( G \) such that \( \langle x, y \rangle = \exp(ig(x, y)) \) for all \( x \in N \), \( y \in K \);
5. for every compact \( K \subseteq \hat{G} \), \( \sup_{y \in K} |g(x, y)| \to 0 \) as \( x \to e \).

Parthasarathy, Ranga Rao and Varadhan proved in [10] that local inner products exist on \( G \times \hat{G} \) for every locally compact second countable abelian group \( G \). Their proof can also be found in Parthasarathy [9]; see Lemma 5.3 in Chapter IV. See also Heyer [5], pages 340–343. Local inner products are not uniquely determined, as is illustrated by the following example.

**Example 1.3.** Consider the case in which \( G = \mathbb{R}^d \), where \( \mathbb{R} \) is the real line with its usual topology and additive group structure, and \( d \) is a positive integer. Then the dual group \( \hat{G} \) can be identified with \( \mathbb{R}^d \), and \( \langle x, y \rangle = \exp(ix \cdot y) \) holds for all \( x, y \in \mathbb{R}^d \), where \( x \cdot y \) is the usual inner product of \( x \) and \( y \). Let \( g \) be any local inner product on \( \mathbb{R}^d \times \mathbb{R}^d \). If \( x \in \mathbb{R}^d \), then \( g(x, \cdot) \) is continuous and satisfies the Cauchy functional equation:

\[
g(x, y_1 + y_2) = g(x, y_1) + g(x, y_2) \quad \text{for all } y_1, y_2 \in \mathbb{R}^d.
\]

Therefore \( g(x, y) = k(x) \cdot y \) for some \( k(x) \in \mathbb{R}^d \). Using the properties of the local inner product \( g \) given in Definition 1.2, it is easy to see that \( k \) is a bounded continuous function from \( \mathbb{R}^d \) into \( \mathbb{R}^d \) and \( k(-x) = -k(x) \) for all \( x \). Also, the validity for all \( y \) of

\[
\exp(k(x) \cdot y) = \exp(ig(x, y)) = \langle x, y \rangle = \exp(\text{ix} \cdot y)
\]

whenever \( x \) is sufficiently near to the zero element 0 of \( \mathbb{R}^d \) implies that \( k(x) = x \) whenever \( x \) is in a sufficiently small neighbourhood \( N_1 \) of 0. Conversely, any bounded continuous function \( k \) from \( \mathbb{R}^d \) into \( \mathbb{R}^d \) such that \( k(-x) = -k(x) \) for all \( x \) and \( k(x) = x \) for all \( x \) in a neighbourhood of 0 leads to a local inner product on \( \mathbb{R}^d \times \mathbb{R}^d \) given by \( g(x, y) = k(x) \cdot y \).
From the properties listed in Definition 1.2, it follows that any local inner product $g$ also satisfies the following condition: given any compact set $K \subseteq \hat{G}$, there is a neighbourhood $M$ of the identity in $G$ such that

$$g(x_1 + x_2, y) = g(x_1, y) + g(x_2, y) \quad \text{for all } x_1, x_2 \in M, \ y \in K.$$ (1.2)

To see this, let $K \subseteq \hat{G}$ be compact and use properties (iv) and (v) of $g$ in Definition 1.2 to deduce that there is a neighbourhood $N$ of $e$ in $G$ such that $\langle x, y \rangle = \exp(ig(x, y))$ and $|g(x, y)| < \frac{1}{2}$ both hold for all $x \in N$ and $y \in K$. Then let $M$ be a neighbourhood of $e$ such that $M + M \subseteq N$. Using the fact that $x \mapsto \langle x, y \rangle$ is a homomorphism of $G$ into $T$, we then have for $x_1, x_2 \in M$ and $y \in K$

$$\exp(i g(x_1 + x_2, y)) = \langle x_1 + x_2, y \rangle = \langle x_1, y \rangle \langle x_2, y \rangle = \exp(i g(x_1, y)) \exp(i g(x_2, y)) = \exp(i g(x_1, y) + g(x_2, y))$$

As $|g(x_1 + x_2, y) - g(x_1, y) - g(x_2, y)| < \frac{4}{9}$, equation (1.2) follows.

Another property of $g$ that will be used later is given by the following lemma, which is essentially the same as Lemma 2.4 in Bingham [1].

**Lemma 1.4.** Given a compact subset $K \subseteq \hat{G}$ the following exist: a constant $c_K \geq 0$, a finite set $F_K \subseteq \hat{G}$ and a neighbourhood $N_K$ of the identity element in $G$ such that

$$\sup_{y \in K} |g(x, y)| \leq c_K \max_{y \in F_K} |g(x, y)| \quad \text{for every } x \in N_K.$$  

2. Unwrapping a càdlàg Function

Denote by $\mathbb{D}$ the Skorokhod space of $G$-valued càdlàg functions defined on the half-line $[0, \infty)$. Choose and fix a fixed local inner product $g$ on $G \times \hat{G}$. For later use we shall begin by seeing how we can use $g$ to modify the jumps of an arbitrary function $\alpha \in \mathbb{D}$ such that $\alpha(0) = e$ to obtain a function $\beta \in \mathbb{D}$ such that, for each $y \in \hat{G}$, the càdlàg function $t \mapsto \langle \beta(t), y \rangle$ with values in the unit circle $T$ can be 'unwrapped' from the circle to the line to give a real-valued càdlàg function $t \mapsto w(\alpha, t, y)$ that is uniquely determined by $\alpha$, $g$ and $y$.

**Lemma 2.1.** Let $\alpha \in \mathbb{D}$ with $\alpha(0) = e$ and let $\Delta \alpha(s) := \alpha(s) - \alpha(s^-)$ for all $s > 0$, $\Delta \alpha(0) := e$. Then all of the following hold.

(a) There exists a uniquely determined $\gamma \in \mathbb{D}$ with $\gamma(0) = e$ such that

$$\langle \gamma(t), y \rangle = \prod_{0 \leq s \leq t} \langle \Delta \alpha(s), y \rangle \exp[-ig(\Delta \alpha(s), y)]$$ (2.1)

for all $t > 0$ and all $y \in \hat{G}$.

(b) If $\beta = \alpha - \gamma$, then, for each $y \in \hat{G}$, there is a uniquely determined real-valued càdlàg function $t \mapsto w(\alpha, t, y)$ such that

$$\langle \beta(t), y \rangle = \exp[iw(\alpha, t, y)] \quad \text{for all } t \geq 0,$$ (2.2)

$$w(\alpha, s, y) - w(\alpha, s^-, y) = g(\Delta \alpha(s), y) \quad \text{for all } s > 0,$$ (2.3)

and

$$w(\alpha, 0, y) = 0.$$ (2.4)
(c) For each \( t \geq 0 \), the mapping \( y \mapsto w(\alpha, t, y) \) is continuous and
\[
w(\alpha, t, y_1 + y_2) = w(\alpha, t, y_1) + w(\alpha, t, y_2)
\] (2.5)
for all \( y_1, y_2 \in \hat{G} \).

(d) If \( K \) is any compact subset of \( \hat{G} \), then
(i) for every \( t \geq 0 \), \( w(\alpha, s, y) \to w(\alpha, t, y) \) uniformly in \( y \in K \) as \( s \downarrow t \), and
(ii) for every \( t > 0 \), \( w(\alpha, s, y) \to w(\alpha, t-, y) \) uniformly in \( y \in K \) as \( s \uparrow t \).

Remark 2.2. Given \( y \in \hat{G} \) there is, by property (iv) of \( g \), a neighbourhood \( N \) of the identity in \( G \) such that \( (x, y) = \exp(ig(x, y)) \) for all \( x \in N \). As \( \alpha \) is càdlàg there are only finitely many points \( s \in [0, t] \) for which \( \Delta \alpha(s) \notin N \). Consequently, only a finite number of the factors in the product on the right hand side of (2.1) can be different from 1. Therefore the product is well defined.

For each jump \( \Delta \alpha(s) \) in \( \alpha \), property (i) of \( g \) and the continuity of \( g \) imply that the mapping \( y \mapsto \exp(ig(\Delta \alpha(s), y)) \) is a continuous character on \( \hat{G} \). Therefore, by the Pontryagin duality theorem, for each jump \( \Delta \alpha(s) \) in \( \alpha \), there is a unique element \( \Delta' \alpha(s) \) in \( G \) such that \( \langle \Delta' \alpha(s), y \rangle = \exp(ig(\Delta \alpha(s), y)) \) for all \( y \in \hat{G} \).

Using this notation, (2.1) can be rewritten as
\[
\langle \gamma(t), y \rangle = \prod_{0 \leq s \leq t} \langle \Delta \alpha(s) - \Delta' \alpha(s), y \rangle.
\]

From this it follows that, for \( s > 0 \), \( \gamma(s) - \gamma(s-) = \Delta \alpha(s) - \Delta' \alpha(s) \). The function \( \beta \) can be obtained from the function \( \alpha \) by subtracting \( \gamma \) from \( \alpha \), and the effect of this is to replace each jump \( \Delta \alpha(s) \) in \( \alpha \) by the corresponding element \( \Delta' \alpha(s) \) of \( G \).

Proof. Lemma 2.1 will be proved in a sequence of steps.

Step 1: First, let \( K \) be a compact subset of \( \hat{G} \). Using property (iv) of \( g \) and equation (1.2), let \( M \) be a neighbourhood of the identity element \( e \) in \( G \) such that
\[
\langle x, y \rangle = \exp(ig(x, y)) \quad \text{for all } x \in M, y \in K
\] (2.6)

and
\[
g(x_1 + x_2, y) = g(x_1, y) + g(x_2, y) \quad \text{for all } x_1, x_2 \in M, y \in K \] (2.7)

Let \( N \) be a symmetric neighbourhood of \( e \) in \( G \) such that \( N + N + N + N \subseteq M \).

Let \( \alpha \in \mathbb{D} \) with \( \alpha(0) = e \). Define \( w(\alpha, 0, y) = 0 \) for all \( y \in \hat{G} \). For now fix a positive number \( T \), and suppose that \( \Delta \alpha(s) \in N \) whenever \( 0 < s \leq T \), where \( \Delta \alpha(s) := \alpha(s) - \alpha(s-) \) is the jump in \( \alpha \) at the point \( s \). The right continuity of \( \alpha \) implies that for every \( s \in [0, T] \) there exists \( \delta(s) > 0 \) such that \( \alpha(s') - \alpha(s) \in N \) whenever \( s \leq s' < s + \delta(s) \). For \( s > 0 \) the existence of left limits and the assumption that \( \Delta \alpha(s) \in N \) imply that \( \delta(s) \) can be chosen so that, in addition, \( \alpha(s') - \alpha(s) = \alpha(s') - \alpha(s-) + \Delta \alpha(s) \in N + N \) whenever \( s - \delta(s) < s' < s \) and \( s' \geq 0 \). Hence \( \alpha(s') - \alpha(s) \in N + N \) whenever \( s - \delta(s) < s' < s + \delta(s) \) and \( s' \in [0, T] \). The collection of intervals \( \{(s - \delta(s), s + \delta(s)) : s \in [0, T]\} \) covers the compact set \([0, T]\), so there is a finite subcover, \( \{(s_j - \delta(s_j), s_j + \delta(s_j)) : j = 1, 2, \ldots, J\} \) say, of \([0, T]\). Let \( \delta = \min_{1 \leq j \leq J} \delta(s_j) \).

If \( t_1, t_2 \in [0, T] \) with \( |t_1 - t_2| < \delta \), then \( t_1 \in (s_j - \delta(s_j), s_j + \delta(s_j)) \) for some \( j \), so \( \alpha(t_1) - \alpha(s_j) \in N + N \). Also, for the same \( j \), \( |s_j - t_2| \leq |s_j - t_1| + |t_1 - t_2| \)
Now suppose that \( 0 < t \leq T \) and let the partition
\[
D = (t_0 = 0 < t_1 < \cdots < t_r = t)
\]
of \([0, t]\) have mesh less than \(\delta\); i.e., \(\max_{1 \leq j \leq r} (t_j - t_{j-1}) < \delta\). Fix \(y \in K\) and consider the sum
\[
S(D) := \sum_{j=1}^{r} g(\alpha(t_j) - \alpha(t_{j-1}), y).
\]

Suppose that we create a refinement \(D'\) of \(D\) by inserting an extra point \(s\). Then \(t_{j-1} < s < t_j\) for some \(j\) and
\[
S(D') - S(D) = g(\alpha(t_j) - \alpha(s), y) + g(\alpha(s) - \alpha(t_{j-1}), y) - g(\alpha(t_j) - \alpha(t_{j-1}), y).
\]

By the choice of \(\delta\) the elements \(\alpha(t_j) - \alpha(s)\) and \(\alpha(s) - \alpha(t_{j-1})\) are in \(M\). Equation (2.7) then implies that \(S(D') - S(D) = 0\). By a simple inductive argument, it therefore follows that \(S(D') = S(D)\) for every (finite) partition \(D'\) of \([0, t]\) that is a refinement of \(D\). By considering a common refinement of any two given partitions of mesh less then \(\delta\), it then follows also that \(S(D)\) does not depend on the choice of \(\delta\) if the mesh of \(D\) is less than \(\delta\). Therefore, for \(0 < t \leq T\), we can define
\[
w_N(\alpha, t, y) := \sum_{j=1}^{r} g(\alpha(t_j) - \alpha(t_{j-1}), y)
\]
for all (finite) partitions \(D = (0 = t_0 < t_1 < \cdots < t_r = t)\) such that \(\max_{j} (t_j - t_{j-1}) < \delta\). Note that \(\delta\) depends on \(T\) but not on the choice of \(t \in [0, T]\).

**Step 2:** Now let \(\alpha \in \mathbb{D}\) be arbitrary with \(\alpha(0) = e\), and let \(M, N, K, T, y\) be as above in Step 1. Define \(\alpha_N \in \mathbb{D}\) by
\[
\alpha_N(t) := \alpha(t) - \sum \{ \Delta \alpha(s) : 0 \leq s \leq t, \; \Delta \alpha(s) \notin N \} \quad (2.8)
\]
for all \(t \geq 0\). Because \(\alpha \in \mathbb{D}\), the set \(\{ s \in [0, t] : \Delta \alpha(s) \notin N \}\) is finite for each \(t \geq 0\); consequently \(\alpha_N\) is well defined. Moreover, for each \(s \in [0, t]\), we have \(\Delta \alpha_N(s) \in N\). By the argument in Step 1 applied to \(\alpha_N\) (taking \(0 < t \leq T\)), there exists \(\delta = \delta(T, N) > 0\) such that
\[
w_N(\alpha_N, t, y) = \sum_{j=1}^{r} g(\alpha_N(t_j) - \alpha_N(t_{j-1}), y)
\]
when the partition \(0 = t_0 < t_1 < \cdots < t_r = t\) satisfies \(\max (t_j - t_{j-1}) < \delta\).

Next define
\[
s_N(\alpha, t, y) := w_N(\alpha_N, t, y) + \sum \{ g(\Delta \alpha(s), y) : 0 \leq s \leq t, \; \Delta \alpha(s) \notin N \}.
\]
Again this is well defined because the set \(\{ s \in [0, t] : \Delta \alpha(s) \notin N \}\) is finite.

It is now claimed that \(s_N(\alpha, t, y)\) does not depend on the neighbourhood \(N\) of the identity, provided that \(N\) is sufficiently small. To prove this claim, it
is sufficient to prove that \( s_{N_t}(\alpha, t, y) = s_{N_1}(\alpha, t, y) \) whenever \( N_2 \) is a neighbourhood of \( e \) such that \( N_2 \subset N_1 = N \). Given such \( N_1, N_2 \) there exists \( \delta := \min(\delta(T, N_1), \delta(T, N_2)) > 0 \) such that

\[
w_{N_k}(\alpha, t, y) = \sum_{j=1}^{r} g(\alpha_{N_k}(t_j) - \alpha_{N_k}(t_{j-1}), y) \quad \text{for} \quad k = 1, 2
\]

whenever \( 0 = t_0 < t_1 < \cdots < t_r = t, r \in \mathbb{N} \) and \( \max_j(t_j - t_{j-1}) < \delta \). As \( \alpha \) is càdlàg, the set of points \( F := \{ s \in [0, t] : \Delta \alpha(s) \in N_1 \setminus N_2 \} \) must be finite. Choose a finite partition \( D = (0 = t_0 < t_1 < \cdots < t_r = t) \), such that \( \max_j(t_j - t_{j-1}) = \text{both less than} \ \delta \) and sufficiently small to ensure that no interval \((t_j - t_{j-1})\) in \( D \) contains more than one of the points that are in \( F \). Consider one of the intervals \((t_{j-1}, t_j)\) in \( D \). If this interval does not contain any point of \( F \), then \( \alpha_{N_1}(t_j) - \alpha_{N_1}(t_{j-1}) = \alpha_{N_2}(t_j) - \alpha_{N_2}(t_{j-1}) \) and consequently

\[
g(\alpha_{N_1}(t_j) - \alpha_{N_1}(t_{j-1}), y) = g(\alpha_{N_2}(t_j) - \alpha_{N_2}(t_{j-1}), y).
\]

By the choice of \( D \), the only other possibility is that \((t_{j-1}, t_j)\) contains exactly one point \( s \) that is in \( F \). In this case \( \alpha_{N_1}(t_j) - \alpha_{N_1}(t_{j-1}) = \alpha_{N_2}(t_j) - \alpha_{N_2}(t_{j-1}) + \Delta \alpha(s) \). Therefore, as \( \alpha_{N_1}(t_j) - \alpha_{N_1}(t_{j-1}) \) and \( \Delta \alpha(s) \) are in \( M \), (2.7) implies that

\[
g(\alpha_{N_1}(t_j) - \alpha_{N_1}(t_{j-1}), y) = g(\alpha_{N_2}(t_j) - \alpha_{N_2}(t_{j-1}) + \Delta \alpha(s), y) = g(\alpha_{N_2}(t_j) - \alpha_{N_2}(t_{j-1}), y) + g(\Delta \alpha(s), y).
\]

Summing over all the intervals in the partition \( D \) and using (2.9), we obtain

\[
w_{N_1}(\alpha, t, y) = w_{N_2}(\alpha, t, y) + \sum \{ g(\Delta \alpha(s), y) : 0 \leq s \leq t, \Delta \alpha(s) \in N_1 \setminus N_2 \}.
\]

Adding \( \sum \{ g(\Delta \alpha(s), y) : 0 \leq s \leq t, \Delta \alpha(s) \notin N_1 \} \) to both sides of the last equation, we conclude that \( s_{N_2}(\alpha, t, y) = s_{N_1}(\alpha, t, y) \) as required.

**Step 3:** Let \( K \) be a compact subset of \( \hat{G} \) and \( t > 0 \). By the previous steps, for any \( \alpha \in \mathbb{D} \) with \( \alpha(0) = e \) and all \( y \in K \), we can define

\[
w(\alpha, t, y) := s_{N}(\alpha, t, y)
\]

\[
:= \sum \{ g(\Delta \alpha(s), y) : 0 \leq s \leq t, \Delta \alpha(s) \notin N \} + \sum_{j=1}^{r} g(\alpha_{N}(t_j) - \alpha_{N}(t_{j-1}), y)
\]

(2.10)

for all sufficiently small symmetric neighbourhoods \( N \) of the identity \( e \) in \( G \) and all sufficiently fine partitions \( 0 = t_0 < t_1 < \cdots < t_r = t \) of \([0, t] \). In fact the quantity \( w(\alpha, t, y) \) does not depend on the compact set \( K \), provided that \( y \in K \), the neighbourhood \( N \) is sufficiently small and the partition is sufficiently fine. For, if \( K_1, K_2 \) are compact subsets of \( \hat{G} \) such that \( y \in K_1 \cap K_2 \), let \( N_1, N_2 \) be neighbourhoods of \( e \) in \( G \) that correspond to \( K_1, K_2 \) respectively in the same way that \( N \) corresponds to \( K \) in the construction above of \( w(\alpha, t, y) \). Then we can use \( N_1 \cap N_2 \) in place of \( N \) in (2.10), so the use of \( K_1 \) or \( K_2 \) leads to the same quantity \( w(\alpha, t, y) \).
If \( y_1, y_2 \in \hat{G} \), we can choose the neighbourhood \( N \) small enough and the partition fine enough for (2.10) to hold for \( y = y_1, y = y_2 \) and \( y = y_1 + y_2 \). The validity of (2.5) therefore follows from property (i) of \( g \) in Definition 1.2.

Let \( y_0 \in \hat{G} \) and suppose that the compact set \( K \) is a neighbourhood of \( y_0 \) in \( \hat{G} \). Let the neighbourhood \( N \) and the number \( \delta(T, N) \) correspond to \( K \) as in the construction of \( w(\alpha, t, y) \) above. Fix a partition such that \( \max_j (t_j - t_{j-1}) < \delta(T, N) \). Then (2.10) holds for all \( y \in K \). As the two sums in (2.10) have finitely many terms, each of which is continuous in \( y \), we deduce that \( w(\alpha, t, y) \to w(\alpha, t, y_0) \) as \( y \to y_0 \). Thus, \( y \mapsto w(\alpha, t, y) \) is continuous for each \( t \) and \( \alpha \).

The last two paragraphs prove that \( w(\alpha, t, \cdot) \) has the properties stated in part (c) of Lemma 2.1. Therefore, the mapping \( y \mapsto \text{exp}[i w(\alpha, t, y)] \) is a continuous character on \( \hat{G} \). Hence, by the Pontryagin duality theorem, for each \( t \geq 0 \) there exists \( \beta(t) \in G \) such that (2.2) holds for all \( y \in \hat{G} \). Define \( \gamma(t) := \alpha(t) - \beta(t) \).

For any \( y \in \hat{G} \) choose a compact subset \( K \) of \( \hat{G} \) with \( y \in K \), and choose a neighbourhood \( N \) of the identity in \( G \) that corresponds to \( K \) as before. Then, using (2.10) with a sufficiently fine partition of \([0, t]\), and also using (2.6),

\[
\text{exp}[i w(\alpha, t, y)] \quad = \quad \prod \{ \text{exp}[i g(\Delta \alpha(s), y)] : 0 \leq s \leq t, \Delta \alpha(s) \notin N \} \\
\quad \times \prod_{j=1}^r \text{exp}[i g(\alpha_N(t_j) - \alpha_N(t_{j-1}), y)] \\
= \quad \prod \{ \text{exp}[i g(\Delta \alpha(s), y)] : 0 \leq s \leq t, \Delta \alpha(s) \notin N \} \\
\quad \times \prod_{j=1}^r (\alpha_N(t_j) - \alpha_N(t_{j-1}), y) \\
= \quad \prod \{ \text{exp}[i g(\Delta \alpha(s), y)] : 0 \leq s \leq t, \Delta \alpha(s) \notin N \} \\
\quad \times \left( \sum_{j=1}^r (\alpha_N(t_j) - \alpha_N(t_{j-1}), y) \right) \\
= \quad \langle \alpha_N(t), y \rangle \prod \{ \text{exp}[i g(\Delta \alpha(s), y)] : 0 \leq s \leq t, \Delta \alpha(s) \notin N \}. \quad (2.11)
\]

In particular, (2.11) applied to \( \alpha_N \) in place of \( \alpha \) shows that

\[
\langle \alpha_N(t), y \rangle = \text{exp}[i w(\alpha_N, t, y)] \quad \text{for} \quad y \in K. \quad (2.12)
\]

By (2.8) and (2.11),

\[
\langle \alpha(t), y \rangle \quad = \quad \langle \alpha_N(t), y \rangle \prod \{ \langle \Delta \alpha(s), y \rangle : 0 \leq s \leq t, \Delta \alpha(s) \notin N \} \\
\quad = \quad \text{exp}[i w(\alpha, t, y)] \prod \{ \langle \Delta \alpha(s), y \rangle \text{exp}[i g(\Delta \alpha(s), y)] : 0 \leq s \leq t \} \\
= \quad \langle \beta(t), y \rangle \prod \{ \langle \Delta \alpha(s), y \rangle \text{exp}[i g(\Delta \alpha(s), y)] : 0 \leq s \leq t \}. 
\]
Consequently, (2.1) holds for all \( y \in \hat{G} \). As the right hand side of (2.1) is uniquely determined, it follows that \( \gamma(t) \) and \( \beta(t) \) are also unique.

**Step 4:** The construction described above defines \( w(\alpha, t, y) \) for \( \alpha \in \mathbb{D} \) with \( \alpha(0) = c, t > 0 \) and \( y \in \hat{G} \). Recall that we also defined \( w(\alpha, 0, y) := 0 \). It will now be demonstrated that the function \( t \mapsto w(\alpha, t, y) \) is càdlàg. The construction of \( w(\alpha, t, y) \) for \( t > 0 \) involved the selection of a neighbourhood \( N = N(K) \) which depended on the compact set \( K \), but did not depend on \( t \). The choice of \( \delta \) in that construction may, however, depend on the choice of a number \( T, t \geq 0 \), as well as on \( N \); recall the meaning of the notation \( \delta(T, N) \) in Step 2.

Let \( K \) be a compact subset of \( \hat{G} \) with \( y \in K \) and let \( N = N(K) \), as described above. Given \( t > 0 \), let \( T > t \), \( \delta := \delta(T, N) \) and let \( s \) satisfy \( t < s < \min(T, T + \delta) \). Then consider any partition of \([0, t]\) given by \( 0 = t_0 < t_1 < \cdots < t_r = t \) with \( \max_j(t_j - t_{j-1}) < \delta \) and the partition of \([0, s]\) obtained by adjoining to it the point \( s \). Using these partitions gives for any \( y \in K \)

\[
w_N(\alpha_N, t, y) = \sum_{j=1}^{r} g(\alpha_N(t_j) - \alpha_N(t_{j-1}), y)
\]

and

\[
w_N(\alpha_N, s, y) = \sum_{j=1}^{r} g(\alpha_N(t_j) - \alpha_N(t_{j-1}), y) + g(\alpha_N(s) - \alpha_N(t), y)
\]

respectively. Therefore, by the right continuity of \( \alpha_N \),

\[
w_N(\alpha_N, s, y) - w_N(\alpha_N, t, y) = g(\alpha_N(s) - \alpha_N(t), y) \to 0 \quad \text{as } s \downarrow t.
\]

Hence, \( s \mapsto w_N(\alpha_N, s, y) \) is right continuous at each \( t > 0 \). To prove the right continuity at \( t = 0 \), simply omit the partition of \([0, t]\).

With the same choice of \( t_0, t_1, \ldots, t_r \) as above, let \( t_{r-1} < s < t \). Then, as \( s \uparrow t \),

\[
w_N(\alpha_N, s, y)
= \sum_{j=1}^{r-1} g(\alpha_N(t_j) - \alpha_N(t_{j-1}), y) + g(\alpha_N(s) - \alpha_N(t_{r-1}), y)
\]

\[
\to \sum_{j=1}^{r-1} g(\alpha_N(t_j) - \alpha_N(t_{j-1}), y) + g(\alpha_N(t) - \alpha_N(t_{r-1}), y). \quad (2.14)
\]

In fact, if the neighbourhood \( N \) is sufficiently small, then the convergences in (2.13) and (2.14) occur uniformly in \( y \in K \). To prove this we use Lemma 1.4. Choose the neighbourhoods \( M \) and \( N \) with the properties specified in Step 1 in such a way that \( M \) is closed and \( M \subseteq N_K \), where \( N_K \) is as in Lemma 1.4. For \( t < s < \min(T, T + \delta) \), (2.13) then gives

\[
\sup_{y \in K} |w_N(\alpha_N, s, y) - w_N(\alpha_N, t, y)| = \sup_{y \in K} |g(\alpha_N(s) - \alpha_N(t), y)|
\]

\[
\leq c_K \max_{y \in F_K} |g(\alpha_N(s) - \alpha_N(t), y)|
\]

\[
\to 0 \quad \text{as } s \downarrow t.
\]
For \( t_{r-1} < s < t \), \( \alpha_N(s) - \alpha_N(t-) \) and \( \alpha_N(t-) - \alpha_N(t_{r-1}) \) are in \( M \) because \( M \) is closed. Therefore, using (2.14), (2.7), \( M \subseteq N_K \) and Lemma 1.4,

\[
\sup_{y \in K} |w_N(\alpha_N, s, y) - w_N(\alpha_N, t-, y)| \\
= \sup_{y \in K} |g(\alpha_N(s) - \alpha_N(t_{r-1}), y) - g(\alpha_N(t-), y)| \\
= \sup_{y \in K} |g(\alpha_N(s) - \alpha_N(t-), y)| \\
\leq c_K \max_{y \in F_K} |g(\alpha_N(s) - \alpha_N(t-), y)| \\
\to 0 \text{ as } s \uparrow t.
\]

**Step 5:** Continuing the development in Step 4, with \( K, y, N \) and \( \alpha \) as before, now consider the quantity \( w(\alpha, t, y) - w_N(\alpha_N, t, y) \) for \( t \geq 0 \). We have

\[
w(\alpha, t, y) - w_N(\alpha_N, t, y) = \sum \{g(\Delta \alpha(s), y) : 0 \leq s \leq t, \Delta \alpha(s) \notin N\}
= \int_{G \setminus N} g(x, y) \mu(\alpha, [0, t], dx)
\quad (2.15)
\]

where \( \mu(\alpha, \cdot, \cdot) \) is the jump measure associated with \( \alpha \): for each bounded interval \( I \subset [0, \infty) \) and each Borel subset \( B \) of \( G \) such that \( B \cap U = \emptyset \) for some neighbourhood \( U \) of the identity, \( \mu(\alpha, I, B) \) is the (finite) number of points \( s \in I \) such that \( \Delta \alpha(s) \in B \). For \( s > t \)

\[
\sup_{y \in K} \left| \int_{G \setminus N} g(x, y) \mu(\alpha, [0, s], dx) - \int_{G \setminus N} g(x, y) \mu(\alpha, [0, t], dx) \right| \\
\leq \sup_{x \in G} \sup_{y \in K} |g(x, y)| \mu(\alpha, (t, s], G \setminus N) \\
\to 0 \text{ as } s \downarrow t.
\]

For \( t > 0 \) and \( 0 \leq s < t \),

\[
\sup_{y \in K} \left| \int_{G \setminus N} g(x, y) \mu(\alpha, [0, t], dx) - \int_{G \setminus N} g(x, y) \mu(\alpha, [0, s], dx) \right| \\
\leq \sup_{x \in G} \sup_{y \in K} |g(x, y)| \mu(\alpha, (s, t], G \setminus N) \\
\to 0 \text{ as } s \uparrow t.
\]

Using equation (2.15), these results imply that \( t \mapsto w(\alpha, t, y) - w_N(\alpha_N, t, y) \) is right continuous uniformly in \( y \in K \) at each point \( t \geq 0 \) and has the left limit \( \int_{G \setminus N} g(x, y) \mu(\alpha, [0, t], dx) \) uniformly in \( y \in K \) at each point \( t > 0 \). Taking this together with what was proved in Step 4, we see that part (d) of Lemma 2.1 is proved.

**Step 6:** The properties (d) of the unwrapping \( w(\alpha, \cdot, \cdot) \), the validity of (2.2) for all \( t \geq 0 \) and \( y \in \hat{G} \), together with the Pontryagin duality theorem, imply that \( \beta \in \hat{D} \) and \( \beta(0) = e \). Therefore \( \gamma \in \hat{D} \) and \( \gamma(0) = e \). The uniqueness of \( \beta \) and \( \gamma \) were already noted in Step 3. To complete the proof of Lemma 2.1, it only remains to prove (2.3) and the uniqueness of \( w(\alpha, \cdot, \cdot) \).
Taking the partition in (2.10) and $t_{r-1} < u < t$, we have
\[
w(\alpha, u, y) = \sum \{g(\Delta\alpha(s), y) : 0 \leq s \leq u, \, \Delta\alpha(s) \notin N \} \\
\quad + \sum_{j=1}^{r-1} g(\alpha_N(t_j) - \alpha_N(t_{j-1}), y) + g(\alpha_N(u) - \alpha_N(t_{j-1}), y).
\]

Subtracting this from (2.10) and using (2.7) yields
\[
w(\alpha, t, y) - w(\alpha, u, y) = \sum \{g(\Delta\alpha(s), y) : u < s \leq t, \, \Delta\alpha(s) \notin N \} \\
\quad + g(\alpha_N(t) - \alpha_N(u), y)
\]
and then letting $u$ increase to $t$ yields (2.3).

Suppose that, for some $y \in \hat{G}$, there are two càdlàg functions, $w(\alpha, \cdot, y)$ and $w'(\alpha, \cdot, y)$, satisfying the conditions (2.2), (2.3) and (2.4) in the lemma. Then, by (2.3), they have exactly the same discontinuities and therefore their difference, $d(\alpha, \cdot, y) := w(\alpha, \cdot, y) - w'(\alpha, \cdot, y)$, is continuous. But equations (2.2) and (2.4) imply that $d(\alpha, \cdot, y)/(2\pi)$ is integer-valued and $d(\alpha, 0, y) = 0$. It follows that $d(\alpha, t, y) = 0$ for all $t \geq 0$. This completes the proof of Lemma 2.1. \qed

**Example 2.3.** Using the same notation as in Example 1.3, let us return to the case in which $G = \mathbb{R}^d$. In this situation consider applying Lemma 2.1 to a càdlàg function $\alpha : [0, \infty) \to \mathbb{R}^d$ such that $\alpha(0) = 0$. Let $0 \neq y \in \hat{G} = \mathbb{R}^d$ and $t > 0$. As in Step 1 of the proof of Lemma 2.1, choose neighbourhoods $M$, $N$ of $0$ in $\mathbb{R}^d$ with the extra requirement that $M \subseteq N_1$, where $N_1$ is as in Example 1.3. Then, by (2.10),
\[
w(\alpha, t, y) := \sum \{k(\Delta\alpha(s)) \cdot y : 0 \leq s \leq t, \, \Delta\alpha(s) \notin N \} \\
\quad + \sum_{j=1}^r k(\alpha_N(t_j) - \alpha_N(t_{j-1})) \cdot y
\]
for all sufficiently fine partitions $0 = t_0 < t_1 < \cdots < t_r = t$ of $[0, t]$. But, if the partition is fine enough, $\alpha_N(t_j) - \alpha_N(t_{j-1}) \in M \subseteq N_1$ and so $k(\alpha_N(t_j) - \alpha_N(t_{j-1})) = \alpha_N(t_j) - \alpha_N(t_{j-1})$ for all $j$. Therefore
\[
w(\alpha, t, y) = \left[ \sum \{k(\Delta\alpha(s)) : 0 \leq s \leq t, \, \Delta\alpha(s) \notin N \} + \alpha_N(t) \right] \cdot y \\
= \left[ \alpha(t) - \sum \{\Delta\alpha(s) - k(\Delta\alpha(s)) : 0 \leq s \leq t \} \right] \cdot y. \tag{2.16}
\]
The expression inside square brackets in (2.16) is $\beta(t)$ in this case and should be compared with the right hand side of (1.1).

### 3. $G$-valued semimartingales

We begin by defining $G$-valued semimartingales that start at the identity $e$.

**Definition 3.1.** Let $X = \{X(t) : t \geq 0\}$ be a $G$-valued adapted stochastic process on the given stochastic basis $\mathbb{B}$. Suppose also that the sample paths of $X$ are a.s. càdlàg and that $X(0) = e$, the identity of $G$. Then we call $X$ a $G$-valued
semimartingale (on the stochastic basis $\mathbf{B}$) if, for every $y \in \tilde{G}$, $\{(X(t), y) : t \geq 0\}$ is (in the usual classical sense) a complex-valued semimartingale (on the stochastic basis $\mathbf{B}$).

Corollary 3.4 below shows that this definition is consistent with the classical notion of an $\mathbb{R}^d$-valued semimartingale starting at the origin $0$. In this paper we consider only semimartingales that start at the identity; but, as in the classical situation, we could extend this to semimartingales such that $X(0) = X_0$, where $X_0$ is an $\mathcal{F}_0$-measurable $G$-valued random variable. This would mean that $\{X(t) - X_0 : t \geq 0\}$ is a semimartingale starting at $e$.

The next task is to define the characteristics of a $G$-valued semimartingale. Before we can give the definitions, however, we need some preparation. Let $X$ is an $\omega$-sense of Definition 3.1. By discarding an $W$ define a real-valued process $\alpha, t, y$ where $X$ of situation, we could extend this to semimartingales such that $X(0) = X_0$, where $X_0$ is an $\mathcal{F}_0$-measurable $G$-valued random variable. This would mean that $\{X(t) - X_0 : t \geq 0\}$ is a semimartingale starting at $e$.

For $\alpha, t, y$ Lemma 3.2.

Given $y \in \tilde{G}$, fix a closed (non-random) neighbourhood $N$ of the identity $e$ in $G$ that satisfies the requirements in Step 1 in the proof of Lemma 2.1 with $K = \{y\}$. Then $W(t, y)$ is the limit (pointwise with respect to $\omega$) of

$$\sum \{g(\Delta X(s), y) : 0 \leq s < t, \Delta X(s) \notin N\} + \sum_{j=1}^{r} g(X_N(t_j) - X_N(t_{j-1}), y) \quad (3.2)$$

as the partition $D = (0 = t_0 < t_1 < \cdots < t_r = t)$ passes through a fixed sequence of partitions of $[0, t]$ whose meshes decrease to 0. (Here, as with the definition of $\alpha_N(t), X_N(t) = X(t) - \sum_{s \leq t} (\Delta X(s) : 0 \leq s < t, \Delta X(s) \notin N,).)$ Therefore, to prove the lemma, it is sufficient to prove that the expression (3.2) is $\mathcal{F}_t$-measurable for each fixed partition $D$.

Define random times $(T_n)_{n=0,1,2,\ldots}$ inductively by $T_0 = 0$ and, for $n = 1, 2, \ldots$, $T_n = \inf\{s : s > T_{n-1}, \Delta X(s) \notin N\}$. Because the complement of $N$ is open and the filtration $(\mathcal{F}_t)_{t \geq 0}$ is right continuous, each $T_n$ is an $(\mathcal{F}_t)$-stopping time. For $T_1$ this is proved for example in Sokol [12] for the case of $\mathbb{R}$-valued processes. It can be proved for $G$-valued processes in the same way as in [12] by replacing the Euclidean metric on $\mathbb{R}$ by a metric for the topology of $G$. The result for general $n$ then follows by induction. Note that Sokol’s proof does not require the filtration to be $P$-complete, whereas to obtain the same result by applying the Début Theorem (Dellacherie and Meyer [3] IV.50, or Meyer [8] IV.52) to the set $\{(t, \omega) : \Delta X(t, \omega) \notin N\}$ would require $P$-completeness.
For given \( t > 0 \) and all \( n = 1, 2, \ldots \), define \( \Delta_n = \Delta X(T_n) \) when \( T_n \leq t \) and \( \Delta_n = e \) when \( T_n > t \). Then \( \Delta_n \) is \( \mathcal{F}_t \)-measurable, because the process \( \Delta X \) is progressively measurable and \( T_n \) is a stopping time. As \( X \) is càdlàg, we have \( \Delta_n \neq e \) for only finitely many \( n \). Therefore, in the identity

\[
\sum_{n=1}^{\infty} \{ g(\Delta X(s), y) : 0 \leq s \leq t, \Delta X(s) \notin N \} = \sum_{n=1}^{\infty} g(\Delta_n, y),
\]

the sum on the right has only a finite number of non-zero terms, so it converges for every \( \omega \in \Omega \). Its partial sums are all \( \mathcal{F}_t \)-measurable, and therefore the sum on the left is \( \mathcal{F}_t \)-measurable. By similar reasoning, the identity

\[
\sum_{n=1}^{\infty} \{ \Delta X(s) : 0 \leq s \leq t, \Delta X(s) \notin N \} = \sum_{n=1}^{\infty} \Delta_n
\]

implies that the sum on its left side is \( \mathcal{F}_t \)-measurable. This in turn implies that \( X_N(t) \) is \( \mathcal{F}_t \)-measurable for every \( t > 0 \), and hence that the expression (3.2) is \( \mathcal{F}_t \)-measurable. \( \square \)

**Lemma 3.3.** Let \( y \in \mathcal{G} \). Then \( W(\cdot, y) \) is a semimartingale.

**Proof.** Given \( y \in \mathcal{G} \), fix a closed neighbourhood \( N \) of the identity in \( G \) in the same way as in the proof of Lemma 3.2 and with the additional property that

\[
|g(x, y)| \leq \frac{1}{4} \quad \text{for all } x \in N. \tag{3.3}
\]

This is possible by property (v) of \( g \) in Definition 1.2. With \( w(\alpha, t, y) \) as in Lemma 2.1, let \( W_N(\cdot, y) \) be the stochastic process whose sample path is \( w(\alpha_N, t, y) \) whenever the sample path of \( X \) is \( \alpha \). As \( W(\cdot, y) \) differs from \( W_N(\cdot, y) \) by a finite variation process, which is given by the first sum in (3.2), the conclusion of the lemma will follow if we prove that \( W_N(\cdot, y) \) is a semimartingale.

We have, by applying (2.12) and (2.8) to the sample paths of the processes \( W_N(\cdot, y) \) and \( X_N \),

\[
\exp[iW_N(t, y)] = \langle X_N(t), y \rangle \quad = \langle X(t), y \rangle - \sum_{s \leq t, \Delta X(s) \notin N} \Delta X(s) - \sum_{s \leq t} \{ \Delta X(s) : 0 \leq s \leq t, \Delta X(s) \notin N \}, y \rangle.
\]

Because \( \langle X(\cdot), y \rangle \) is a semimartingale by hypothesis and the other factor in the last line of the above equation gives a finite variation process, which is therefore also a semimartingale, we conclude that \( \exp[iW_N(\cdot, y)] \) is a semimartingale.

Define a sequence of stopping times \( (T_n)_{n=0,1,2,\ldots} \) inductively by \( T_0 = 0 \) and, for \( n = 1, 2, \ldots \),

\[
T_n := \inf \{ t > T_{n-1} : |W_N(t, y) - W_N(T_{n-1}, y)| > \frac{1}{7} \},
\]

where \( \inf \emptyset = \infty \). Each path of \( W_N(\cdot, y) \) is càdlàg and therefore, for any \( \varepsilon > 0 \), it can have only finitely many \( \varepsilon \)-oscillations in any bounded interval. Consequently \( T_n \not\to \infty \) as \( n \to \infty \). For each \( n = 1, 2, \ldots \), the process with paths \( t \mapsto \exp[iW_N(t \wedge T_n, y)] \) is a non-vanishing semimartingale, being the semimartingale \( \exp[iW_N(\cdot, y)] \) stopped at \( T_n \). Therefore the complex-valued process with
Define the classical semimartingale for all $y$, $x \mapsto d = 1$. Accordingly, let $W_y$ be a semimartingale for every $y$. It is sufficient to prove that $X$ is c\`adl\`ag, the sum has only a finite number of non-zero terms and therefore defines a finite variation process. To prove that $X$ is c\`adl\`ag, the sum has only a finite number of non-zero terms and therefore defines a finite variation process. Therefore $W_y$ is a classical semimartingale if and only if $X \cdot \cdot \cdot$.

Because $X$ is a classical semimartingale if and only if $X \cdot$ is a classical semimartingale for every $y \in \mathbb{R}^d$, it is sufficient to prove the result for the case when $d = 1$. Accordingly, let $X$ be a c\`adl\`ag real-valued process such that $X(0) = 0$.

If $X$ is a classical semimartingale, then so is $\exp(iX \cdot)$ for all $y \in \mathbb{R}$, because $x \mapsto \exp(xy)$ is a smooth function. Conversely, suppose that $\exp(iX \cdot)$ is a classical semimartingale for all $y \in \mathbb{R}$. It is enough to assume this for $y = 1$.

Define the c\`adl\`ag process $\hat{X}$ by

$$\hat{X}(t) = X(t) - \sum_{k=1}^{n} \{ \Delta X(s) : 0 \leq s \leq t, \quad |\Delta X(s)| > \frac{1}{t} \}$$

for all $t \geq 0$, where $\Delta X(0) = 0$ and $\Delta X(s) = X(s) - X(s-)$ for $s > 0$. Because $X$ is c\`adl\`ag, the sum has only a finite number of non-zero terms and therefore defines a finite variation process. To prove that $X$ is a semimartingale, it is therefore sufficient to prove that $\hat{X}$ is a semimartingale.

On the right hand side of the identity

$$\exp(i\hat{X}(t)) = \exp(iX(t)) \cdot \exp \left[ -t \sum_{k=1}^{n} \{ \Delta X(s) : 0 \leq s \leq t, \quad |\Delta X(s)| > \frac{1}{t} \} \right]$$

the first factor defines a semimartingale (by hypothesis) and the second factor defines a finite variation process. Therefore $\exp(i\hat{X}(\cdot))$ is a semimartingale.

Corollary 3.4. Let $X = \{X(t) : 0 \leq t < \infty\}$ be an $\mathbb{R}^d$-valued c\`adl\`ag stochastic process such that $X(0) = 0$. Then $X$ is an $\mathbb{R}^d$-valued semimartingale (in the classical sense) if and only if $\exp(iX \cdot)$ is a complex-valued semimartingale (in the classical sense) for every $y \in \mathbb{R}^d$.

Proof. Because $X$ is a classical semimartingale if and only if $X \cdot$ is a classical semimartingale for every $y \in \mathbb{R}^d$, it is sufficient to prove the result for the case when $d = 1$. Accordingly, let $X$ be a c\`adl\`ag real-valued process such that $X(0) = 0$.

If $X$ is a classical semimartingale, then so is $\exp(iX \cdot)$ for all $y \in \mathbb{R}$, because $x \mapsto \exp(xy)$ is a smooth function. Conversely, suppose that $\exp(iX \cdot)$ is a classical semimartingale for all $y \in \mathbb{R}$. It is enough to assume this for $y = 1$.
same argument that was used in the proof of Lemma 3.3 to prove that $W_N(\cdot, y)$ is a semimartingale can now be used to conclude that $\hat{X}$ is a semimartingale. □

Lemma 3.3 shows that $W(\cdot, y)$ is a semimartingale for each $y \in \hat{G}$. But we also have $|\Delta W(\cdot, y)| \leq \sup_{x \in \hat{G}} |g(x, y)| < \infty$, so the jumps of $W(\cdot, y)$ have uniformly bounded magnitudes. Therefore $W(\cdot, y)$ is a special semimartingale (Jacod and Shiryaev [6], 4.24, page 44). Consequently $W(\cdot, y)$ has a canonical decomposition

$$W(\cdot, y) = M(\cdot, y) + B(\cdot, y) \tag{3.5}$$

where $M(\cdot, y)$ is a local martingale with $M(0, y) = 0$, $B(\cdot, y)$ is a predictable process of locally integrable variation with $B(0, y) = 0$ and each of $M(\cdot, y)$ and $B(\cdot, y)$ almost surely has càdlàg paths (Jacod and Shiryaev [6], 4.21, page 43). Moreover, the processes $M(\cdot, y)$ and $B(\cdot, y)$ are uniquely determined up to evanescence.

We can now define the semimartingale characteristics for $X$, in analogy with the real line case, as follows.

**Definition 3.5.** Let $G$ be a locally compact second countable abelian group and let $X$ be a $G$-valued semimartingale in the sense of Definition 3.1. For each $y \in \hat{G}$, let $W(\cdot, y) = w(X\cdot, y)$ be the real-valued special semimartingale given by equation (3.1) above, and let $B(\cdot, y)$ be as in the canonical decomposition in (3.5). Then stochastic processes $\tilde{B}$, $\Phi$ and $\nu$ with the following properties (1), (2) and (3) are called the **first, second, and third characteristics** of $X$ respectively:

1. $\tilde{B}$ is a predictable a.s. càdlàg $G$-valued process such that, for each $y \in \hat{G}$, the processes $(B, y)$ and $\exp \{i B(\cdot, y)\}$ are indistinguishable;
2. $\Phi$ is a non-decreasing continuous process of random continuous nonnegative quadratic forms $\{\Phi(t, \cdot) : t \geq 0\}$ on $\hat{G}$, such that, for each $y \in G$, $\Phi(\cdot, y)$ is the compensator of the square of the continuous martingale part of $W(\cdot, y)$;
3. $\nu$ is the predictable compensator of the jump measure of $X$.

It is conjectured, but not yet proved, that all three characteristics always exist for semimartingales with values in locally compact second countable abelian groups. Theorem 3.6 below gives the partial result that $\nu$ always exists and that $\tilde{B}$ and $\Phi$ always exist if $G$ is compact. Proposition 3.7 establishes some uniqueness properties and Theorem 3.8 shows that all three characteristics exist when $G = \mathbb{R}$.

**Theorem 3.6.** Let $G$ be a locally compact second countable abelian group and let $X$ be any $G$-valued semimartingale in the sense of Definition 3.1. Then the third characteristic of $X$ exists. If $G$ is compact, then the first and second characteristics of $X$ exist.

**Proof.** Let $\mu$ be the jump measure of $X$. By Proposition II 1.16 and Theorem II 1.8 in [6], replacing the Euclidean distance on $\mathbb{R}^d$ by a metric for the topology of $G$, we conclude that there is a predictable random measure $\nu$, which is the compensator of $\mu$ and is unique up to a $P$-null set. Thus, the third characteristic of $X$ exists and is unique (up to indistinguishability).

Now let $y_1, y_2 \in \hat{G}$. It follows from (2.5) that $W(\cdot, y_1 + y_2) = W(\cdot, y_1) + W(\cdot, y_2)$. By the uniqueness of canonical decompositions (3.5) up to indistinguishability, it
follows that
\[ M(t, y_1 + y_2) = M(t, y_1) + M(t, y_2) \quad \text{for all } t \geq 0 \]
holds a.s.\((P)\) for all \(y_1, y_2 \in \hat{G}\) (3.6)

and
\[ B(t, y_1 + y_2) = B(t, y_1) + B(t, y_2) \quad \text{for all } t \geq 0 \]
holds a.s.\((P)\) for all \(y_1, y_2 \in \hat{G}\). (3.7)

Similarly, the continuous martingale part \(M^c(\cdot, y)\) of \(M(\cdot, y)\) (which is also the continuous martingale part of \(W(\cdot, y)\)) is also uniquely determined up to indistinguishability and therefore
\[ M^c(t, y_1 + y_2) = M^c(t, y_1) + M^c(t, y_2) \quad \text{for all } t \geq 0 \]
holds a.s.\((P)\) for all \(y_1, y_2 \in \hat{G}\). (3.8)

The exceptional \(\omega\)-sets of \(P\)-measure zero in equations (3.6), (3.7) and (3.8) may, in general, depend on \(y_1\) and \(y_2\). This fact causes technical difficulties that have long prevented the construction of a complete general proof of the existence of the first two characteristics for locally compact second countable abelian groups that are not compact. If, however, \(G\) is compact, these difficulties do not arise, because there is then an exceptional \(\omega\)-set that is independent of \(y_1\) and \(y_2\).

For the rest of this proof assume that the locally compact second countable abelian group \(G\) is compact. Then the dual group \(\hat{G}\) is discrete and countable. We can therefore deduce from (3.7) that there is a set \(\Omega_1 \in \mathcal{F}\) with \(P(\Omega_1) = 1\) such that
\[ B(t, y_1 + y_2, \omega) = B(t, y_1, \omega) + B(t, y_2, \omega) \quad \text{for all } t \geq 0 \]
holds for all \(y_1, y_2 \in \hat{G}\) for all \(\omega \in \Omega_1\). (3.9)

Let \(t \geq 0\) and \(\omega \in \Omega_1\). Because every function on \(\hat{G}\) is continuous, we see from (3.9) that \(y \mapsto \exp{[iB(t, y, \omega)]}\) is a continuous homomorphism from \(\hat{G}\) into \(T\). By the Pontryagin duality theorem there is a uniquely determined \(\tilde{B}(t, \omega) \in G\) such that
\[ \langle \tilde{B}(t, \omega), y \rangle = \exp{[iB(t, y, \omega)]} \quad \text{for all } y \in \hat{G}. \]
(3.10)

Let \(\tilde{B}(t, \omega)\) be defined by condition (3.10) if \(\omega \in \Omega_1\) and let \(\tilde{B}(t, \omega) = e\) if \(\omega \in \Omega \setminus \Omega_1\). Because the set \(\Omega_1 \times [0, \infty)\) is predictable, this defines a \(G\)-valued stochastic process \(\tilde{B} := \{\tilde{B}(t) : t \geq 0\}\), such that the process \(\langle \tilde{B}, y \rangle\) is predictable for each \(y \in \hat{G}\). But the elements of \(\hat{G}\) generate the Borel \(\sigma\)-field of \(G\), so this implies that the process \(\tilde{B}\) is itself predictable; i.e., \(\tilde{B}\) is measurable with respect to the predictable \(\sigma\)-field in \(\Omega \times [0, \infty)\) and the Borel \(\sigma\)-field in \(G\).

We also know that the sample paths of all of the processes \(B(\cdot, y)\) for \(y \in \hat{G}\) are càdlàg with probability 1. Bearing in mind that uniform convergence on compact subsets of \(\hat{G}\) is the same as pointwise convergence, the Pontryagin duality theorem implies that the sample paths of \(\tilde{B}\) are also càdlàg with probability 1. Therefore \(\tilde{B}\) has all the properties of the first characteristic of \(X\) in Definition 3.5.
Next, for each $y \in \hat{G}$, let $\Phi(\cdot, y) := \langle M^c(\cdot, y), M^c(\cdot, y) \rangle$ be the compensator of the process $(M^c(\cdot, y))^2$, so that $\Phi(\cdot, y)$ is the (unique up to indistinguishability) non-decreasing continuous process such that the process $(M^c(\cdot, y))^2 - \Phi(\cdot, y)$ is a local martingale. From (3.8) it follows that 

$$
\Phi(t, y_1 + y_2, \omega) + \Phi(t, y_1 - y_2, \omega) = 2\Phi(t, y_1, \omega) + 2\Phi(t, y_2, \omega)
$$

for all $t \geq 0$ holds a.s.($P$) for all $y_1, y_2 \in \hat{G}$.

Again using the countability of $\hat{G}$, we conclude that there exists $\Omega_2 \in \mathcal{F}$ with $P(\Omega_2) = 1$ such that 

$$
\Phi(t, y_1 + y_2, \omega) + \Phi(t, y_1 - y_2, \omega) = 2\Phi(t, y_1, \omega) + 2\Phi(t, y_2, \omega)
$$

for all $t \geq 0$ holds for all $y_1, y_2 \in \hat{G}$ for all $\omega \in \Omega_2$. (3.11)

If necessary, modify the random variables $\Phi(t, y)$ for all $t \geq 0$ and all $y \in \hat{G}$ so that $\Phi(t, y, \omega) = 0$ for all $\omega \in \Omega \setminus \Omega_2$. Then (3.11) shows that $y \mapsto \Phi(t, y, \omega)$ is a quadratic form on $\hat{G}$ for each $t \geq 0$ and each $\omega \in \Omega$. Thus, $\Phi := \{\Phi(t, \cdot) : t \geq 0\}$ is a process of random continuous nonnegative quadratic forms on $\hat{G}$ such that $t \mapsto \Phi(t, y)$ is continuous and nondecreasing for each $y \in \hat{G}$. Therefore $\Phi$ has all the properties required of the second characteristic in Definition 3.5 and Theorem 3.6 is proved.

**Proposition 3.7.** If it exists, the first characteristic $\bar{B}$ is unique (up to indistinguishability) for a given choice of $g$. If it exists, the second characteristic $\Phi$ is unique (up to indistinguishability) and does not depend on the choice of $g$. The third characteristic $\nu$ exists, is unique (up to indistinguishability) and does not depend on the choice of $g$.

**Proof:** The existence and uniqueness (up to indistinguishability) of $\nu$ follow from the proof of Theorem 3.6. That $\nu$ does not depend on $g$ is obvious from its definition.

For a given choice of $g$, the special semimartingale $W(\cdot, y)$ and its canonical decomposition are unique (up to indistinguishability) for each $y \in \hat{G}$. Therefore, if $\bar{B}$ and $\bar{B}'$ are two possible candidates to be the first characteristic of $X$, then 

$$
\langle \bar{B}'(t), y \rangle = \exp\{iB(t, y)\} = \langle \bar{B}(t), y \rangle \quad \text{for all } t \geq 0
$$

holds a.s.($P$) for each $y \in \hat{G}$, where $B(t, y)$ is as in the canonical decomposition (3.5). Hence, if $Y$ is a countable dense subset of $\hat{G}$, there is a set $\Omega_1 \in \mathcal{F}$ with $P(\Omega_1) = 1$, such that $\langle \bar{B}'(t, \omega), y \rangle = \langle \bar{B}(t, \omega), y \rangle$ for all $t \geq 0$, all $y \in Y$ and all $\omega \in \Omega_1$. By continuity of $y \mapsto \langle x, y \rangle$, this implies that, when $\omega \in \Omega_1$, $\langle B'(t, \omega), y \rangle = \langle \bar{B}(t, \omega), y \rangle$ for all $y \in \hat{G}$ and all $t \geq 0$. Hence, $\bar{B}'(t, \omega) = \bar{B}(t, \omega)$ for all $t \geq 0$ whenever $\omega \in \Omega_1$. Therefore $\bar{B}$ and $\bar{B}'$ are indistinguishable.

In order to prove the statement about the second characteristic, we first prove that, for each $y \in \hat{G}$, the continuous martingale part $M^c(\cdot, y)$ of $W(\cdot, y)$ is unique and does not depend on $g$. The uniqueness of $W(\cdot, y)$ for a given choice of $g$ implies the uniqueness of $M^c(\cdot, y)$ for that choice of $g$. Therefore it is sufficient to prove that $M^c(\cdot, y)$ does not depend on $g$. Let $W'(\cdot, y)$ be the process obtained
exists $\Omega_2$. Consequently they have the same continuous martingale part $\hat{W} \in y$. Therefore, if $Y$ is a countable dense subset of $\tilde{G}$, we conclude that there exists $\Omega_2 \in \mathcal{F}$ with $P(\Omega_2) = 1$ such that $\Phi'(t, y, \omega) = \Phi(t, t, \omega)$ for all $t \geq 0$ and all $y \in Y$ whenever $\omega \in \Omega_2$. The continuity of $\Phi'(t, y, \omega)$ and $\Phi(t, t, \omega)$ with respect to $y$ then implies that $\Phi'(t, y, \omega) = \Phi(t, t, \omega)$ holds for all $t \geq 0$ and all $y \in \tilde{G}$ whenever $\omega \in \Omega_2$. Hence $\Phi$ and $\Phi'$ are indistinguishable.

**Theorem 3.8.** Let $X$ be an $\mathbb{R}$-valued semimartingale. Then all three of the semimartingale characteristics of $X$ described in Definition 3.5 exist. The second and third characteristics described in Definition 3.5 are indistinguishable from those described in Definition 1.1, if each nonnegative quadratic form $y \mapsto cy^2$ on $\mathbb{R}$ is identified with its coefficient $c$. If the truncation function $h$ and the local inner product $g$ are chosen suitably, then the first characteristics described in Definitions 1.1 and 3.5 are also indistinguishable.

**Proof.** Let $X$ be a real-valued semimartingale. The third characteristic $\nu$ is defined in the same way in the two definitions and we already know that it exists, so only the statements about the first two characteristics have to be proved. Equation (2.16) shows that, when we apply Lemma 2.1 to the sample paths of $X$, we obtain

\[W(t, y) = \left[X(t) - \sum \left\{ \Delta X(s) - k(\Delta X(s)) : 0 \leq s \leq t \right\} \right] y \tag{3.12}\]

for each $y \in \mathbb{R}$. Consider the process $B(\cdot, y)$ in the canonical decomposition (3.5). In the special situation under consideration, $B(\cdot, y)$ is indistinguishable from $B(\cdot, 1)y$. Therefore, for each $y$, $\langle B(\cdot, 1), y \rangle \equiv \exp[iB(\cdot, 1)y]$ is indistinguishable from $\exp[iB(\cdot, y)]$. Hence, $B(\cdot, 1)$ qualifies to be the first characteristic of $X$ in the sense of Definition 3.5.

Similarly, the continuous martingale part $M^c(\cdot, y)$ of $W(\cdot, y)$ is indistinguishable from $M^c(\cdot, 1)y$, and therefore the compensator $\Psi(\cdot, y)$ of its square is indistinguishable from $\Psi(\cdot, 1)y^2$ for each $y \in \mathbb{R}$. Therefore we can take $\Phi$ as the second characteristic in Definition 3.5, where $\Psi(\cdot, y) = \Psi(\cdot, 1)y^2$ for each $y$. If we replace $k$ by any continuous truncation function $h$ on $\mathbb{R}$, then the continuous martingale part of the process defined by the right hand side of (3.12) is unchanged, so $\Psi(\cdot, 1)$ is the second characteristic of $X$ in the sense of Definition 1.1.
Finally, let \( h \) be a continuous truncation function on \( \mathbb{R} \) such that \( h(-x) = -h(x) \) for all \( x \), and let \( g \) be the local inner product on \( G \times \hat{G} \) defined by \( g(x, y) = h(x)y \) for all \( x, y \in \mathbb{R} \). If \( h \) and \( g \) are used in Definitions 1.1 and 3.5, then the process \( B(\cdot, 1) \) above is the first characteristic of \( X \) in the sense of both definitions. \( \square \)

**Remark 3.9.** It is clear that Theorem 3.8 can be extended to \( \mathbb{R}^d \)-valued semimartingales for any positive integer \( d \), provided that each nonnegative quadratic form on \( \mathbb{R}^d \) is identified with its nonnegative-definite coefficient matrix.

### 4. Continuous Additive Gaussian Processes on \( G \) are Semimartingales

Suppose that the locally compact second countable abelian group \( G \) can support Gaussian distributions. This implies the existence of a non-trivial collection \( \{ \phi(t, \cdot) : t \geq 0 \} \) of continuous nonnegative quadratic forms on \( \hat{G} \) such that \( \phi(0, y) = 0 \) for all \( y \in \hat{G} \), \( t \mapsto \phi(t, y) \) is continuous for each \( y \in \hat{G} \) and \( \phi(s, y) \leq \phi(t, y) \) whenever \( 0 \leq s \leq t \). From Bingham [2] there exists on some underlying probability space \( (\Omega, \mathcal{F}, P) \) a \( G \)-valued Gaussian process \( \{ X(t) : t \geq 0 \} \) with continuous sample paths and independent increments such that

\[
E \langle X(t), y \rangle = \exp \left[-\frac{1}{2}\phi(t, y)\right] \quad \text{for all } t \geq 0, \ y \in \hat{G}.
\]  

(4.1)

In fact the existence of such a process is proved in Bingham [2] only for \( 0 \leq t \leq 1 \). But the result is easily extended to the case \( 0 \leq t < \infty \) by concatenating the sample paths of independent processes \( \{ Y_n : n = 0, 1, 2, \ldots \} \), where each \( Y_n \) is a \( G \)-valued process \( \{ Y_n(s) : 0 \leq s \leq 1 \} \) with continuous sample paths and independent increments such that

\[
E \langle Y_n(s), y \rangle = \exp \left[-\frac{1}{2}\phi(s + n, y) - \phi(n, y)\right] \quad \text{for } s \in [0, 1], \ y \in \hat{G}.
\]

The process \( X \) given by

\[
X(t) = \sum_{k=0}^{n-1} Y_k(1) + Y_n(t - n) \quad \text{for } n \leq t \leq n + 1
\]

has the required properties. We shall show that any \( G \)-valued process \( X \) with continuous sample paths and independent increments that satisfies (4.1) is a \( G \)-valued semimartingale with respect to an appropriate right continuous filtration.

Let \( \{ \mathcal{G}_t : t \geq 0 \} \) be the filtration in \( \mathcal{F} \) generated by \( X \); i.e., for each \( t \), \( \mathcal{G}_t \) is the \( \sigma \)-algebra generated by \( \{ X(s) : 0 \leq s \leq t \} \). Then \( X \) is adapted to the right-continuous filtration \( \{ \mathcal{G}_t^+ : t \geq 0 \} \), where \( \mathcal{G}_t^+ = \bigcap \{ \mathcal{G}_u : u > t \} \). For each \( y \in \hat{G} \), define the process \( N(\cdot, y) \) by

\[
N(t, y) := \langle X(t), y \rangle \exp \left[\frac{1}{2}\phi(t, y)\right].
\]

Then each \( N(\cdot, y) \) is a \( \{ \mathcal{G}_t \} \)-martingale. By the optional sampling theorem (see for example Theorem 3.22 on page 19 in Karatzas and Shreve [7])

\[
E \left[ N(t, y) \mid \mathcal{G}_{s+} \right] = N(s, y) \quad \text{a.s. for } 0 \leq s \leq t,
\]

so \( N(\cdot, y) \) is a \( \{ \mathcal{G}_t^+ \} \)-martingale. But, from the definition of \( N(t, y) \), this implies that \( \langle X(\cdot), y \rangle \) is the product of a \( \{ \mathcal{G}_t^+ \} \)-martingale and a (deterministic) finite variation process, so \( \langle X(\cdot), y \rangle \) is a \( \{ \mathcal{G}_t^+ \} \)-semimartingale. Hence \( X \) is a \( G \)-valued semimartingale with respect to the filtration \( \{ \mathcal{G}_t^+ \} \).
If we apply Lemma 2.1 to the sample paths of \( X \), we obtain, for each \( y \in \hat{G} \), a real-valued process \( W(\cdot, y) \) such that
\[
(X(t), y) = \exp \left[ i W(t, y) \right] \quad \text{for all } t \geq 0.
\]
Furthermore, \( W(\cdot, y) \) has continuous paths, independent increments and \( W(0, y) = 0 \). Therefore \( W(\cdot, y) \) is Gaussian. Because \( X(s) \) has a symmetric distribution for every \( s \) and \( g(-x, y) = -g(x, y) \) for all \( x \in G, y \in \hat{G} \), the distribution of \( W(t, y) \) is also symmetric. Therefore \( W(t, y) \) has expectation zero. Also,
\[
E \exp[i \xi W(t, y)] \bigg|_{\xi=1} = E(X(t), y) = \exp[-\frac{1}{2} \phi(t, y)],
\]
whence \( W(t, y) \) has variance \( \phi(t, y) \).

For \( \xi \in \mathbb{R} \) and \( y \in \hat{G} \) define the process \( N(\cdot, y, \xi) \) by
\[
N(t, y, \xi) := \exp \left[ i \xi W(t, y) + \frac{1}{2} \xi^2 \phi(t, y) \right].
\]
Then each \( N(\cdot, y, \xi) \) is a \( \{G_t\} \)-martingale. By the optional sampling theorem,
\[
E[N(t, y, \xi) \mid G_{\omega+}] = N(s, y, \xi) \quad \text{a.s. for } 0 \leq s \leq t
\]
and therefore
\[
E \left( \exp [i \xi W(t, y)] \mid G_{\omega+} \right) = \exp \left[ i \xi W(s, y) - \frac{1}{2} \xi^2 \{\phi(t, y) - \phi(s, y)\} \right]
\]  (4.2)
holds a.s. for \( 0 \leq s \leq t \). Generally, the exceptional set of \( P \)-measure zero in (4.2) may depend on \( \xi \), but there exist versions of the conditional expectations such that the exceptional set of \( P \)-measure zero does not depend on \( \xi \). To see this, fix a version of the conditional distribution of \( W(t, y) \) given \( G_{\omega+} \) and use this version of the conditional distribution to evaluate \( E \left( \exp [i \xi W(t, y)] \mid G_{\omega+} \right) \) for all \( \xi \in \mathbb{R} \). With these versions of the conditional expectations, both sides of (4.2) are continuous in \( \xi \). Because (4.2) holds a.s. for each \( \xi \in \mathbb{R} \), there exists \( \Omega' \in \mathcal{F} \) with \( P(\Omega') = 1 \) such that (4.2) holds for all rational \( \xi \in \mathbb{R} \) when \( \omega \in \Omega' \). The continuity in \( \xi \) of both sides of (4.2) then implies that (4.2) holds for all \( \xi \in \mathbb{R} \) when \( \omega \in \Omega' \). Hence, the conditional distribution of \( W(t, y) \) given \( G_{\omega+} \) is a.s. normal with conditional expectation \( W(s, y) \) and conditional variance \( \phi(t, y) - \phi(s, y) \). This implies that the continuous process \( W(\cdot, y) \) is a \( \{G_t\} \)-martingale and that the compensator of its square is given by \( \phi(\cdot, y) \).

It follows that \( X \) is a \( G \)-valued semimartingale with characteristics \( \vec{B} = e, \Phi = \phi \) and \( \nu = 0 \). Thus, in this example, the semimartingale characteristics exist and are the same as the canonical triple in the Lévy-Khinchine representation.

References

M. S. Bingham: School of Mathematics & Physical Sciences, University of Hull, Hull, HU6 7RX, United Kingdom

*E-mail address: m.s.bingham@hull.ac.uk*
CONVOLUTION SEMIGROUPS OF PROBABILITY MEASURES
ON GELFAND PAIRS, REVISITED

DAVID APPLEBAUM

ABSTRACT. Our goal is to find classes of convolution semigroups on Lie groups $G$ that give rise to interesting processes in symmetric spaces $G/K$. The $K$-bi-invariant convolution semigroups are a well-studied example. An appealing direction for the next step is to generalise to right $K$-invariant convolution semigroups, but recent work of Liao has shown that these are in one-to-one correspondence with $K$-bi-invariant convolution semigroups. We investigate a weaker notion of right $K$-invariance, but show that this is, in fact, the same as the usual notion. Another possible approach is to use generalised notions of negative definite functions, but this also leads to nothing new. We finally find an interesting class of convolution semigroups that are obtained by making use of the Cartan decomposition of a semisimple Lie group, and the solution of certain stochastic differential equations. Examples suggest that these are well-suited for generating random motion along geodesics in symmetric spaces.

1. Introduction

The category of Gelfand pairs is a beautiful context in which to explore probabilistic ideas. It provides an elegant mathematical formalism, and contains many important examples, not least the globally Riemannian symmetric spaces and the homogeneous trees. Until quite recently, most studies of probability measures on a Gelfand pair $(G, K)$ have focussed on the $K$-bi-invariant case. In the context that will concern us here, where the object of study is a convolution semigroup of such measures, Herbert Heyer’s paper [16] presents a masterly survey of the main developments of the theory, up to and including the early 1980s.

In fact, right $K$-invariant measures on $G$ are natural objects of study as they are in one-to-one correspondence with measures on the homogeneous space $N := G/K$. The additional assumption of left $K$-invariance certainly makes the theory extremely elegant, as it enables the use of the beautiful concept of spherical function, as introduced by Harish–Chandra; thus we may study measures in the “Fourier picture”, using the “characteristic function” given by the spherical transform. Such
an approach led to a specific Lévy–Khintchine formula, classifying infinitely divisible probability measures on non-compact symmetric spaces, in the pioneering work of Gangolli [11] (see also [21] for a more recent treatment). Another key consequence of the $K$-bi-invariance assumption is that it corresponds precisely to semigroups/Dirichlet forms that are $G$-invariant on $M$ (see Theorem 4.1 in [8]), with respect to the natural group action; so that if $N$ is a symmetric space, then the induced semigroup commutes with all isometries.

In [5], Dooley and the author developed a Lévy–Khintchine formula on non-compact semi-simple Lie groups, using a matrix-valued generalisation of Harish–Chandra’s spherical functions. In the last part of the paper, an attempt was made to project this to non-compact symmetric spaces when the convolution semigroup comprises measures that are only right $K$-invariant. Since then Liao has shown [20] that all right $K$-invariant convolution semigroups are in fact $K$-bi-invariant. In the current paper, we will ask the question – are there any natural classes of convolution semigroups, other than the $K$-bi-invariant ones, that give rise to interesting classes of Markov processes on $N$?

Since right $K$-invariance is so natural and attractive, we begin by asking whether some weaker notion of convolution may lead to any interesting conclusions. We present a candidate, but find that it once again leads to $K$-bi-invariance. Our second approach is to generalise the ideas of positive-definite and negative-definite function on the space $P$ of positive-definite spherical functions, which were first introduced, in the bi-invariant case, by Berg in [8]. The functions that Berg considered were complex-valued, but when we drop left $K$-invariance, we find that their natural generalisations must be vector-valued, and this requires us to make use of certain direct integrals of Hilbert spaces over the space $P$. Once again, however, we show that these objects lead to nothing new. Finally, in the semisimple Lie group case, we introduce a promising class of convolution semigroups, which are obtained by solving stochastic differential equations (SDEs). These SDEs are driven by vector fields that live in that part of the Cartan decomposition of the Lie algebra of $G$ which projects non-trivially to $N$. Although we haven’t developed the ideas very far herein, this concept seems more promising. In particular there are already some interesting and non-trivial examples, which involve randomising the notion of geodesic.

The organisation of the paper is as follows. Section 2 is an introduction that briefly summarises all the harmonic analysis on Gelfand pairs that we will need in the sequel. In section 3, we discuss various types of convolution semigroup, while section 4 describes the generalised notions of vector-valued positive-definite and negative-definite function. In section 5, in the Lie group/symmetric space setting, we review Hunt’s theorem and the Lévy–Khintchine formula for convolution semigroups, and section 6 puts the main results of [5] within a more general framework. Finally in section 7, we present the new class of convolution semigroups mentioned above.

**Notation.** If $G$ is a locally compact Hausdorff group then $\mathcal{B}(G)$ is its Borel $\sigma$-algebra, and $C_c(G)$ is the Banach space (with respect to the supremum norm) of all real-valued, bounded, uniformly continuous functions (with respect to the
left uniform structure)\(^1\) defined on \(G\). The closed subspace of \(C_u(G)\) comprising functions having compact support is denoted by \(C_c(G)\). If \(\mu\) is a measure on \(G\), then \(\hat{\mu}\) is the reversed measure, i.e., \(\hat{\mu}(A) = \mu(A^{-1})\), for all \(A \in B(G)\). We recall that if \(\mu_1\) and \(\mu_2\) are two finite measures on \((G, B(G))\), then their convolution \(\mu_1 \ast \mu_2\) is the unique finite measure on \((G, B(G))\) so that

\[
\int_G f(g)(\mu_1 \ast \mu_2)(dg) = \int_G \int_G f(gh)\mu_1(dg)\mu_2(dh),
\]

for all \(f \in C_c(G)\). If \(e\) is the neutral element in \(G\), then \(\delta_e\) will denote the Dirac measure at \(e\). The set \(\hat{G}\) comprises all equivalence classes (with respect to unitary conjugation) of irreducible unitary representations of \(G\), acting in some complex separable Hilbert space. If \(E\) is a real or complex Banach space, then \(B(E)\) will denote the algebra of all bounded linear operators on \(E\). If \(\mathcal{F}(G)\) is some space of functions on \(G\), and \(K\) is a closed subgroup of \(G\), we write \(\mathcal{F}_K(G)\) for the subspace comprising those functions that are right \(K\)-invariant, and we will naturally identify this subspace with the corresponding space \(\mathcal{F}(G/K)\) of functions on the homogeneous space \(G/K\) of left cosets. We choose once and for all a left-invariant Haar measure on \(G\), which is denoted by \(dg\) within integrals. Haar measure on compact subgroups is always normalised to have total mass one.

2. Gelfand Pairs and Spherical Functions

Let \((G, K)\) be a Gelfand pair, so that \(G\) is a locally compact group with neutral element \(e\), \(K\) is a compact subgroup, and the Banach algebra (with respect to convolution) \(L^1(K\backslash G/K)\) of \(K\)-bi-invariant functions is commutative. We will summarise basic facts that we will need about these structures in this section. Most of this is based on Wolf \[22\], but see also Dieudonné \[9\]. Throughout this paper, we will, where convenient, identify functions/measures/distributions on the homogeneous space \(G/K\) with right \(K\)-invariant functions/measures/distributions on \(G\). We emphasise that we do not assume that left \(K\)-invariance also holds.

If \((G, K)\) is a Gelfand pair, then Haar measure on \(G\) is unimodular. Every continuous multiplicative function from \(L^1(K\backslash G/K)\) to \(\mathbb{C}\) is of the form \(f \rightarrow \hat{f}(\omega)\), where \(\hat{f}(\omega) = \int_G f(g)\omega(g^{-1})dg\). The mapping \(\omega : G \rightarrow \mathbb{C}\) is called a (bounded) spherical function. In general a spherical function on \((G, K)\) is characterised by the property that it is a non-trivial continuous function such that for all \(g, h \in G\),

\[
\int_K \omega(gkh)dk = \omega(g)\omega(h). \tag{2.1}
\]

The set \(S(G, K)\) of all bounded spherical functions on \((G, K)\) is the maximal ideal space (or spectrum) of the algebra \(L^1(K\backslash G/K)\). It is locally compact under the weak-\(*\)-topology (and compact if \(L^1(K\backslash G/K)\) is unital). The corresponding Gelfand transform is the mapping \(f \rightarrow \hat{f}\) (usually called the spherical transform in this context). Let \(\mathcal{P} := \mathcal{P}(S, K)\) be the closed subspace of \(S(G, K)\) comprising positive definite spherical functions. For each \(\omega \in \mathcal{P}\), there exists a triple

\(^1\)In this context, uniform continuity means that given any \(\epsilon > 0\), there exists a neighbourhood \(U\) of \(e\), so that \(\sup_{x \in G} |f(g^{-1}x) - f(x)| < \epsilon\) for all \(g \in U\).
\((H_\omega, \pi_\omega, u_\omega)\), where \(H_\omega\) is a complex Hilbert space, \(\pi_\omega\) is a unitary representation of \(G\) in \(H_\omega\), and \(u_\omega \in H_\omega\) is a cyclic vector, so that for all \(g \in G\),

\[\omega(g) = \langle u_\omega, \pi_\omega(g)u_\omega \rangle.\]

Since \(\omega(e) = 1\), \(u_\omega\) is a unit vector for all \(\omega \in \mathcal{P}\); moreover the representation \(\pi_\omega\) is irreducible, and spherical in that \(\pi_\omega(k)u_\omega = u_\omega\), for all \(k \in K\). Finally, we have

\[\dim(H^K_\omega) = 1,\]

where \(H^K_\omega := \{v \in H_\omega; \pi_\omega(k)v = v \text{ for all } k \in K\}\).

There is a unique Radon measure \(\rho\) on \(\mathcal{P}\), called the Plancherel measure, such that for all bounded functions \(f \in L^1(K\backslash G/K), g \in G\),

\[f(g) = \int_\mathcal{P} \hat{f}(\omega)\omega(g)\rho(d\omega).\]

The support of \(\rho\) is the maximal ideal space \(\mathcal{P}_+ \subset \mathcal{P}\) of the C*-algebra \(C^*(G, K)\), which is the uniform closure of the range of the representation \(\psi\) in \(\mathcal{B}(L^2(K\backslash G/K))\) whose action is given by \(\psi(f)h = f * h\), for all \(f \in L^1(K\backslash G/K), h \in L^2(K\backslash G/K)\).

The Plancherel theorem states that if \(f \in L^1(K\backslash G/K) \cap L^2(K\backslash G/K)\), then \(\hat{f} \in L^2(P, \rho)\) and \(\|f\|_2 = ||\hat{f}||_2\). This isometry extends uniquely to a unitary isomorphism between \(L^2(K\backslash G/K)\) and \(L^2(P, \rho)\).

The theory described above is essential for the analysis of \(K\)-bi-invariant functions/measures/distributions on \(G\). To work with objects that are only right \(K\)-invariant we must introduce the direct integrals \(\mathcal{H}_p(G, K)\), for \(1 \leq p \leq \infty\). These spaces comprise sections \(\Psi : \mathcal{P} \to \bigcup_{\omega \in \mathcal{P}} H_\omega\) for which \(\Psi(\omega) \in H_\omega\) for all \(\omega \in \mathcal{P}\), such that for \(1 \leq p < \infty\),

\[||\Psi||_{\mathcal{H}_p} := \left(\int_\mathcal{P} ||\Psi(\omega)||_{H_\omega}^p \rho(d\omega)\right)^{\frac{1}{p}} < \infty,\]

and for \(p = \infty,||\Psi||_{\mathcal{H}_\infty} := \text{ess sup}_{\omega \in \mathcal{P}} ||\Psi(\omega)||_{H_\omega}\).

For \(1 \leq p \leq \infty\), let \(\mathcal{H}_p^0(G, K)\) be the subspace of \(\mathcal{H}_p(G, K)\) comprising sections \(\Psi\) for which \(\Psi(\omega) \in H^K_\omega\) for all \(\omega \in \mathcal{P}\). We will find it convenient in the sequel to regard \(L^p(\mathcal{P}, \rho)\) as a subspace of \(\mathcal{H}_p(G, K)\), by observing that it is precisely \(\mathcal{H}_p^0(G, K)\).

\(\mathcal{H}_p(G, K)\) is a Banach space, while \(\mathcal{H}_2(G, K)\) is a Hilbert space with inner product

\[\langle \Psi_1, \Psi_2 \rangle_{\mathcal{H}_2} = \int_\mathcal{P} \langle \Psi_1(\omega), \Psi_2(\omega) \rangle_{H_\omega} \rho(d\omega),\]

for all \(\Psi_1, \Psi_2 \in \mathcal{H}_2(G, K)\).

We introduce the Fourier cotransform for any unitary representation \(\pi\) of \(G\),

\[\pi(f) := \int_G f(g)\pi(g)dg,\]

where \(f \in L^1(G/K)\). We will need the scalar Fourier inversion formula for bounded \(f \in L^1(G/K), g \in G:\)

\[f(g) = \int_\mathcal{P} \langle \pi_\omega(f)u_\omega, \pi_\omega(g)u_\omega \rangle \rho(d\omega).\]

The vector-valued Fourier transform is the mapping \(F : L^1(G/K) \to \mathcal{H}_\infty(G, K)\) defined for each \(f \in L^1(G/K), \omega \in \mathcal{P}\) by

\[(F f)(\omega) = \pi_\omega(f)u_\omega.\]
The vector-valued Fourier inversion formula is a minor reformulation of (2.2). It states that if \( f \in L^1(G/K) \) is bounded, then \( \mathcal{F}f \in \mathcal{H}_1(G,K) \) and for all \( g \in G \):

\[
f(g) = \int_P \langle (\mathcal{F}f)(\omega), \pi_\omega(g)u_\omega \rangle \rho(d\omega).
\]

There is also a Plancherel formula within this context: if \( f \in L^1(G/K) \cap L^2(G/K) \), then \( \mathcal{F}f \in \mathcal{H}_2(G,K) \) and

\[
||\mathcal{F}f||_{\mathcal{H}_2(G,K)} = ||f||_{L^2(G/K)},
\]

and the action of \( \mathcal{F} \) extends to a unitary isomorphism between \( L^2(G/K) \) and \( \mathcal{H}_2(G,K) \).

3. Restricted Convolution Semigroups

A family \((\mu_t, t \geq 0)\) of probability measures on \((G, \mathcal{B}(G))\) is said to be a convolution semigroup if \( \mu_{s+t} = \mu_s * \mu_t \) for all \( s, t \geq 0 \). Then \( \mu_0 \) is an idempotent measure and so must be the normalised Haar measure of a compact subgroup of \( G \) (see [15], Theorem 1.2.10, p.34). A convolution semigroup is said to be continuous if vague – \( \lim_{t \to 0} \mu_t = \mu_0 \). It then follows that it is vaguely continuous on \([0, \infty)\). A continuous convolution semigroup is said to be standard if \( \mu_0 = \delta_e \). If \((\mu_t, t \geq 0)\) is standard, then \((P_t, t \geq 0)\) is a \( C_0 \)-contraction semigroup on \( C_u(G) \), where

\[
P_t f(g) = \int_G f(gh)\mu_t(dh), \text{ for all } t \geq 0, f \in C_u(G), g \in G. \tag{3.1}
\]

It is precisely the standard continuous convolution semigroups that are the laws of Lévy processes in Lie groups (see e.g. [19]).

Now let us return to Gelfand pairs \((G,K)\). Let \( \mathcal{M}_K(G) \) be the space of all right \( K \)-invariant Radon probability measures on \( G \). We say that a continuous convolution semigroup \((\mu_t, t \geq 0)\) is right \( K \)-invariant, if \( \mu_t \in \mathcal{M}_K(G) \) for all \( t \geq 0 \). In that case, \( \mu_0 \) is normalised Haar measure on \( K \), and then \( \mu_t \) is \( K \)-bi-invariant for all \( t \geq 0 \), as is shown in Proposition 2.1 of [20]. Indeed since for all \( t \geq 0, \mu_t = \mu_0 * \mu_t \), left \( K \)-invariance of \( \mu_t \) follows from that of \( \mu_0 \).

It would be desirable to be able to study families of measures on \( G \) that are right \( K \)-invariant, but not necessarily \( K \)-bi-invariant, and which capture the essential features of a convolution semigroup that we need, within a right \( K \)-invariant framework. Here is a plausible candidate. A family \((\mu_t, t \geq 0)\) of probability measures on \((G, \mathcal{B}(G))\) is said to be a right \( K \)-invariant restricted convolution semigroup if

(A1) \( \mu_t \) is right \( K \)-invariant for all \( t > 0 \),

(A2) \( \mu_0 = \delta_e \),

(A3) \( \int_G f(g)\mu_{s+t}(dg) = \int_G \int_G f(gh)\mu_s(dg)\mu_t(dh) \), for all \( f \in C_u(G/K) \),

(A4) \( \lim_{t \to 0} \int_G f(\sigma)\mu_t(d\sigma) = f(e) \), for all \( f \in C_c(G/K) \).
Notes.

(1) In (A4) we can replace $f$ with any bounded continuous right $K$-invariant function, by the argument of Theorem 1.1.9 in [15].

(2) The term “restricted convolution semigroup” is a misnomer, as (A3) appears to be too weak to define a convolution of measures in the usual sense (but see Theorem 3.1 below). Nonetheless there is a more general framework that these ideas fit into. We regard $C_u(G/K)$ as a $*$-bialgebra where the comultiplication $\Delta : C_u(G/K) \to C_u(G/K) \otimes C_u(G/K)$ is given by

$$\Delta f(g,h) = \int_K f(gkh) dk;$$

then we can interpret (A3) as a convolution of states on $C_u(G/K)$, as described in e.g. [10].

Note that $P_t$, as defined in (3.1) does not preserve the space $C_u(G/K)$. Instead we define the family of operators $T_t : C_u(G/K) \to C_u(G/K)$, for $t \geq 0$ by

$$T_t f(g) = \int_G \int_K f(gh) \mu_t(dh) dk, \text{ for all } t \geq 0, f \in C_u(G), g \in G. \quad (3.2)$$

Then by using (A1) to (A4), we can verify that $(T_t, t \geq 0)$ is a $C_0$-contraction semigroup on $C_u(G/K)$. Indeed, for $s, t \geq 0$, to verify the semigroup property, we observe that by right $K$-invariance of $\mu_s$, Fubini’s theorem, and (A3):

$$T_s(T_t f)(g) = \int_G \int_K (T_t f)(gh) \mu_s(dh) dk = \int_G \int_K \int_G \int_K f(ghk'h') \mu_t(dh') \mu_s(dh) dk' dk,$$

$$= \int_G \int_K \int_G f(ghkh') \mu_t(dh') \mu_s(dh) dk$$

$$= \int_G \int_K f(ghkhh') \mu_s(dh) dk$$

$$= \int_G f(gh) \mu_{s+t}(dh) dk

= T_{s+t} f(g).$$

Although they appear to be promising objects, as pointed out to the author by Ming Liao, restricted convolution semigroups are just continuous $K$-bi-invariant convolution semigroups, and so there is nothing new in this idea. We prove this as follows.

**Theorem 3.1.** Every right $K$-invariant restricted convolution semigroup is a continuous $K$-bi-invariant convolution semigroup.

**Proof.** Let $(\mu_t, t \geq 0)$ be a right $K$-invariant restricted convolution semigroup. For each $f \in C_u(G), g \in G$, define $f^K(g) = \int_K f(gk) dk$. Then $f^K \in C_u(G/K)$, and for all $s, t \geq 0$, using right-$K$-invariance of $\mu_{s+t}$, Fubini’s theorem, and right
K-invariance of $\mu_t$

$$\int_G f(g)\mu_{s+t}(dg) = \int_G f^K(g)\mu_{s+t}(dg)$$

$$= \int_G \int_G f^K(gh)\mu_s(dg)\mu_t(dh)$$

$$= \int_G \int_G f(gh)\mu_s(dg)\mu_t(dh).$$

By a similar argument

$$\lim_{t \to 0} \int_G f(g)\mu_t(dg) = \lim_{t \to 0} \int_G f^K(g)\mu_t(dg)$$

$$= f^K(e) = \int_K f(k)dk.$$ 

So $(\mu_t, t \geq 0)$ is a continuous right $K$ invariant convolution semigroup, with $\mu_0$ being normalised Haar measure on $K$. Hence, by Proposition 2.1 of [20], $\mu_t$ is $K$-bi-invariant for all $t > 0$. □

4. Negative Definite Functions

We define the vector-valued Fourier transform of $\mu \in \mathcal{M}_K(G)$ in the obvious way, i.e.,

$$(\mathcal{F}\mu)(\omega) = \pi_\omega(\mu)u_\omega,$$

where $\pi_\omega(\mu) = \int_G \pi_\omega(g)\mu(dg)$ for all $\omega \in \mathcal{P}$. A straightforward calculation yields:

$$(\mathcal{F}(\mu_1 * \mu_2))(\omega) = \pi_\omega(\mu_1)(\mathcal{F}\mu_2)(\omega), \quad (4.1)$$

for all $\mu_1, \mu_2 \in \mathcal{M}_K(G), \omega \in \mathcal{P}$. We also use the standard notation

$$\hat{\mu}_S(\omega) := \pi_\omega(\mu)^* = \int_G \pi_\omega(g^{-1})\mu(dg)$$

for the Fourier transform of an arbitrary bounded measure on $(G, \mathcal{B}(G))$.

Let $\mathcal{M}(\mathcal{K}\backslash G/K)$ denote the space of $K$-bi-invariant Radon probability measures on $(G, \mathcal{B}(G))$. If $\omega \in \mathcal{P}$ then the spherical transform of $\mu$ is given by

$$\hat{\mu}_S(\omega) := \int_G \omega(g^{-1})\mu(dg),$$

so that

$$\hat{\mu}_S(\omega) = \langle \hat{\mu}(\pi_\omega)u_\omega, u_\omega \rangle.$$

It is shown in Theorem 6.8 of [16] that the mapping $\mu \to \hat{\mu}_S$ is injective.

**Proposition 4.1.** The mapping $\mathcal{F}: \mathcal{M}_K(G) \to \mathcal{H}_\infty(G, K)$ is injective.

**Proof.** We follow the procedure of the proof of Lemma 2.1 in [8]. First assume that $\mu$ is absolutely continuous with respect to Haar measure, and that the Radon–Nikodym derivative $d\mu/d\nu := f \in L^1(G/K) \cap L^2(G/K)$. Then $\mathcal{F}\mu = 0$ implies that $\mathcal{F}f = 0$ and so $f = 0$ (a.e.) by (2.3). For the general case, let $(\psi_V, V \in \mathcal{V})$ be an approximate identity based on a fundamental system $\mathcal{V}$ of neighbourhoods of $e$. Now define $f_V = \psi_V * \mu$. Then for all $V \in \mathcal{V}$, we see from (4.1) that $\mathcal{F}\mu = 0$ implies that $\mathcal{F}f_V = 0$, hence $f_V = 0$ (a.e.), and it follows that $\mu = 0$, as required. □
In [8], Berg defined notions of positive- and negative-definite function that could be used to investigate $K$-bi-invariant convolution semigroups. We remind the reader of these notions. A continuous function $p : \mathcal{P} \to \mathbb{C}$ is said to be positive definite if $p = \hat{\hat{\mu}}_{S}(\omega)$ for some $\mu \in \mathcal{M}(K \backslash G/K)$, and a continuous function $q : \mathcal{P} \to \mathbb{C}$ is said to be negative definite if $q(1) = 0$ and $\exp(-tq)$ is positive definite for all $t > 0$. Berg then showed that there is a one-to-one correspondence between negative definite functions and continuous convolution semigroups in $\mathcal{M}(K \backslash G/K)$.

We extend these notions to a more general context as follows. A field $\Psi \in \mathcal{H}_{\infty}(G, K)$ is said to be generalised positive definite if there exists $\mu \in \mathcal{M}_{K}(G)$ such that $\Psi = F\mu$. A densely defined closed linear operator $Q$ on $L^{2}(\mathcal{P}, \rho) = \int_{\omega \in \mathcal{P}} H^{K}_{\omega}\rho(d\omega)$ is said to be generalised negative definite if it is diagonalisable, in that $Q = (Q(\omega), \omega \in \mathcal{P})$ with each $Q(\omega)$ acting as multiplication by a scalar in $H^{K}_{\omega}$, for $\omega \in \mathcal{P}$, and is such that

1. $Q(1) = 0$, where $Q(1)$ denotes the restriction of $Q$ to $H_{1} = H^{K}_{1}$.
2. $Q$ is the infinitesimal generator of a one-parameter contraction semigroup $(R_{t}, t \geq 0)$ acting on $L^{2}(\mathcal{P}, \rho)$.
3. For each $t \geq 0$, $R_{t}$ extends to a bounded linear operator on $\mathcal{H}_{\infty}(G, K)$, so that the mapping $\omega \mapsto R_{t}(\omega)u_{\omega}$ is positive definite.

Notes.

1. In (3), as $u_{\omega}$ is cyclic in $H_{\omega}$, it is equivalent to require

$$R_{t}(\omega)\pi_{\omega}(g)u_{\omega} = \pi_{\omega}(\mu_{t})\pi_{\omega}(g)u_{\omega},$$

for all $g \in G, t \geq 0, \omega \in \mathcal{P}$.
2. The positive definite field $(R_{t}(\omega)u_{\omega}, \omega \in \mathcal{P})$ is uniquely determined by $Q$ as the solution of a family of initial value problems in $H^{K}_{\omega}$ for $\omega \in \mathcal{P}$:

$$\frac{d\Psi(t)(\omega)}{dt} = Q(\omega)\Psi(t)(\omega),$$

with initial condition $\Psi(0)(\omega) = u_{\omega}$.

Now suppose that $\Psi$ is positive definite and that $\mu$ is $K$-bi-invariant. Then for all $\omega \in \mathcal{P}$,

$$\langle u_{\omega}, \Psi(\omega) \rangle = \langle u_{\omega}, \pi_{\omega}(\mu_{t})u_{\omega} \rangle$$

$$= \int_{G} \langle u_{\omega}, \pi_{\omega}(g)u_{\omega} \rangle \mu(dg)$$

$$= \int_{G} \omega(g)\mu(dg) = \hat{\hat{\mu}}_{S}(\omega),$$

and so $p(\omega) := \langle u_{\omega}, \Psi(\omega) \rangle$ essentially coincides with the notion of positive definite function in the bi-invariant context, as introduced by Berg in [8]. The word “essentially” is included, because Berg required his positive-definite functions to be continuous. Although we could impose continuity on our mapping $\Psi$, it is then not clear to the author how to prove the next theorem.

**Theorem 4.2.** There is a one-to-one correspondence between generalised negative definite functions on $\mathcal{P}$ and right $K$-invariant restricted convolution semigroups on $G$. 

Proof. Suppose that \((\mu_t, t \geq 0)\) is a right \(K\)-invariant restricted convolution semigroup on \(G\). Then \(R_t(\omega) = \pi_\omega(\mu_t)\) defines a one-parameter contraction \(C_0\)-semigroup (of positive real numbers) in \(H^K_\omega\). To see that \(R_t(\omega)\) preserves \(H^K_\omega\), for all \(\omega \in \mathcal{P}, t \geq 0\), note that, by Theorem 3.1, \(\mu_t\) is left \(K\)-invariant, and so for all \(k \in K\),

\[
\pi_\omega(k)R_t(\omega)u_\omega = \int_G \pi_\omega(kg)\mu_t(dg) = R_t(\omega)u_\omega.
\]

To verify the semigroup property, it is sufficient to observe that for all \(\omega \in \mathcal{P}, f_\omega \in C_u(G/K)\), where \(f_\omega(\cdot) = \langle \pi_\omega(\cdot)u_\omega, u_\omega \rangle\) and that for all \(s, t \geq 0\), by (A3),

\[
\langle \pi_\omega(\mu_s)\pi_\omega(\mu_t)u_\omega, u_\omega \rangle = \int_G \int_G \langle \pi_\omega(gh)u_\omega, u_\omega \rangle \mu_s(dg)\mu_t(dh) = \int_G \langle \pi_\omega(g)u_\omega, u_\omega \rangle \mu_{s+t}(dg) = \langle \pi_\omega(\mu_{s+t})u_\omega, u_\omega \rangle.
\]

For all \(\psi \in H_\infty(G, K)\), we define \(R_t\psi(\omega) = R_t(\omega)\psi(\omega)\). Then \(R_0 \in B(H_\infty(G, K))\) for all \(t \geq 0\), since

\[
\|R_t\psi\|_{H_\infty(G, K)} = \text{ess sup}_{\omega \in \mathcal{P}} \|R_t(\omega)\psi(\omega)\|_{H_\omega} \leq \|\psi\|_{H_\infty(G, K)}.
\]

It is easy to see that the restriction of \((R_t, t \geq 0)\) to \(L^2(\mathcal{P}, \rho)\) is in fact a contraction \(C_0\)-semigroup; indeed to verify strong continuity, we can use Lebesgue’s dominated convergence theorem to deduce that

\[
\lim_{t \to 0} \int_\mathcal{P} \|\pi_\omega(\mu_t)\Psi(\omega) - \Psi(\omega)\|^2_{H^K_\omega} \rho(d\omega) = 0,
\]

for all \(\Psi \in L^2(\mathcal{P}, \rho)\). The infinitesimal generator of \((R_t, t \geq 0)\) is then the required negative definite function. Note that since \(\pi_1(\mu_t) = 1\) for all \(t \geq 0\), it is clear that \(Q(1)u_1 = 0\).

Conversely, if \(Q\) is negative definite, it follows that there exists a family of measures \(\{\mu_t; t \geq 0\}\) in \(M_K(G)\) so that for all \(\omega \in \mathcal{P}, Q(\omega)\) is the infinitesimal generator of the semigroup \((\pi_\omega(\mu_t), t \geq 0)\) acting in \(H^K_\omega\). Hence, by (4.1) and Proposition 4.1, \((\mu_t, t \geq 0)\) is a semigroup under convolution with \(\mu_0 = \delta_e\). Indeed, for all \(s, t \geq 0, \omega \in \mathcal{P}\) we have

\[
\pi_\omega(\mu_{s+t})u_\omega = \pi_\omega(\mu_s)\pi_\omega(\mu_t)u_\omega = \pi_\omega(\mu_s \ast \mu_t)u_\omega.
\]

To show that the semigroup of measures satisfies (A3), let \(f \in C_c(G/K)\) and use scalar Fourier inversion (2.2) as follows

\[
\int_G f(g)\mu_t(dg) = \int_\mathcal{P} \int_G \langle \pi_\omega(\mu_s)u_\omega, \pi_\omega(\mu_t)u_\omega \rangle \rho(d\omega)\mu_t(dg) = \int_\mathcal{P} \int_G \langle \pi_\omega(\mu_s)u_\omega, \pi_\omega(\mu_t)u_\omega \rangle \mu_t(dg)\rho(d\omega).
\]

where the use of Fubini’s theorem to interchange integrals is justified by the fact that the sections \(\omega \to \pi_\omega(f)u_\omega \in \mathcal{H}_1(G, K)\). This latter fact also justifies the use
of Lebesgue’s dominated convergence theorem to deduce from the above that
\[
\lim_{t \to 0} \int_G f(g) \mu_t(dg) = \int_P \langle \pi_\omega(f)u_\omega, u_\omega \rangle \rho(d\omega) = f(e).
\]

The result of this theorem is negative. When combined with the conclusion of Theorem 3.1, it tells us that there is a one-to-one correspondence between generalised negative definite functions and continuous \(K\)-bi-invariant convolution semigroups, and hence between generalised negative definite functions and negative definite functions. It may be interesting for future work to investigate negative definite functions that fail to be diagonalisable.

Before we leave the topic of generalised negative definite functions, we will, for completeness, follow Berg [8], by making intrinsic characterisations of generalised positive and negative definite functions.

We say that a field \(P \in \mathcal{H}_\infty(G, K)\) is a Berg \(P\)-D function if for all \(n \in \mathbb{N}, a_1, \ldots, a_n \in \mathbb{C}, \omega_1, \ldots, \omega_n \in \mathcal{P},\)

\[
\Re \left( \sum_{i=1}^n a_i \omega_i \right) \geq 0 \text{ on } G \Rightarrow \Re \left( \sum_{i=1}^n a_i \langle u_{\omega_i}, P(\omega_i)u_{\omega_i} \rangle \right) \geq 0.
\]

A closed densely defined linear operator \(Q\) acting in \(L^2(\mathcal{P}, \rho)\) is said to be a Berg \(N\)-D function if

1. \(Q(1) = 0,\)
2. for all \(n \in \mathbb{N}, a_1, \ldots, a_n \in \mathbb{C}, \omega_1, \ldots, \omega_n \in \mathcal{P},\)

\[
\sum_{i=1}^n a_i = 0 \text{ and } \Re \left( \sum_{i=1}^n a_i \omega_i \right) \geq 0 \text{ on } G
\]

\[
\Rightarrow \Re \left( \sum_{i=1}^n a_i \langle u_{\omega_i}, Q(\omega_i)u_{\omega_i} \rangle \right) \leq 0.
\]

We generalise Theorem 5.1 in [8], where the \(K\)-bi-invariant case was explicitly considered:

**Theorem 4.3.**

1. Every generalised positive definite function on \(\mathcal{P}\) is a Berg \(P\)-D function.
2. If \(Q\) is a generalised negative definite function, then \(-Q\) is a Berg \(N\)-D function.

**Proof.**

1. Suppose that \(\Psi\) is positive definite, so that \(\psi = \mathcal{F}_\mu\) for some \(\mu \in \mathcal{M}_K(G)\). For arbitrary \(a_1, \ldots, a_n \in \mathbb{C}, \omega_1, \ldots, \omega_n \in \mathcal{P},\) we have

\[
\sum_{i=1}^n a_i \langle u_{\omega_i}, \Psi(\omega_i)u_{\omega_i} \rangle = \int_G \sum_{i=1}^n a_i \langle u_{\omega_i}, \pi_\omega(g)u_{\omega_i} \rangle \mu(dg) = \int_G \sum_{i=1}^n a_i \omega_i(g) \mu(dg),
\]

from which the required result follows easily.
(2) Now suppose that $Q$ is negative definite. Then $Q$ is the infinitesimal generator of a one-parameter contraction semigroup $(R_t, t \geq 0)$ acting on $L^2(\mathcal{P}, \rho)$, and the field $(R_t(\omega)u_\omega, \omega \in \mathcal{P})$ is positive definite. Then if \( \sum_{i=1}^n a_i = 0 \) and \( \Re (\sum_{i=1}^n a_i \omega_i) \geq 0 \) on $G$, we see that
\[
\Re \left( \sum_{i=1}^n a_i \frac{1}{t}(u_{\omega_i}, (R_t(\omega_i) - 1)u_{\omega_i}) \right) \geq 0
\]
and the result follows when we take the limit as $t \to 0$.

\[\Box\]

5. The Lévy–Khintchine Formula

In this section $G$ is a Lie group of dimension $d$ having Lie algebra $\mathfrak{g}$. Let \( \{X_1, \ldots, X_d\} \) be a basis for $\mathfrak{g}$, which we consider as acting as left-invariant vector fields on $G$. We obtain a dense subspace $\mathcal{C}_u^2(G)$ of $\mathcal{C}_u(G)$ by
\[
\mathcal{C}_u^2(G) := \{ f \in \mathcal{C}_u(G); X_i f \in \mathcal{C}_u(G) \text{ and } X_j X_k f \in \mathcal{C}_u(G) \text{ for all } 1 \leq i, j, k \leq d \},
\]
It is well-known that there exist functions $x_i \in \mathcal{C}_c(G)$ (1 \( \leq i \leq d \)) which are canonical co-ordinate functions in a co-ordinate neighbourhood of $e$, and we say that $\nu$ is a Lévy measure on $G$ if $\nu(\{e\}) = 0$ and for any co-ordinate neighbourhood $U$ of the neutral element in $G$:
\[
\int_G \left( \sum_{i=1}^d x_i(\tau)^2 \right) \nu(d\tau) < \infty \text{ and } \nu(U^c) < \infty,
\]
where $(x_1, \ldots, x_d)$ are canonical co-ordinate functions on $U$ as above.

The proof of the next celebrated theorem, goes back to the seminal work of Hunt [17]. The first monograph treatment was due to Heyer [15], and more recent treatments can be found in Liao [19], and Applebaum [4].

**Theorem 5.1** (Hunt’s theorem). Let $(\mu_t, t \geq 0)$ be a convolution semigroup of measures in $G$, with associated semigroup of operators $(P_t, t \geq 0)$ acting on $\mathcal{C}_u(G)$ in $G$ with generator $\mathcal{L}$ then
(1) $\mathcal{C}_u^2(G) \subseteq \text{Dom}(\mathcal{L})$.
(2) For each $\sigma \in G, f \in \mathcal{C}_u^2(G),
\[
\mathcal{L}f(\sigma) = \sum_{i=1}^d b^i X_i f(\sigma) + \sum_{i,j=1}^d a^{ij} X_i X_j f(\sigma)
\]
\[
+ \int_{G - \{e\}} \left( f(\sigma \tau) - f(\sigma) - \sum_{i=1}^d x^i(\tau) X_i f(\sigma) \right) \nu(d\tau),
\]
where $b = (b^1, \ldots, b^d) \in \mathbb{R}^d$, $a = (a^{ij})$ is a non-negative-definite, symmetric $d \times d$ real-valued matrix and $\nu$ is a Lévy measure on $G$.

Conversely, any linear operator with a representation as in (5.2) is the restriction to $\mathcal{C}_u^2(G)$ of the infinitesimal generator corresponding to a unique convolution semigroup of probability measures.
Now it is well-known that if \((\mu_t, t \geq 0)\) is a convolution semigroup of measures, then so is \((\tilde{\mu}_t, t \geq 0)\); from which it follows that for each \(\pi \in \hat{G}, (\tilde{\mu}_t(\pi), t \geq 0)\) is a contraction semigroup in \(H_\pi\). Let \(A_\pi\) denote the infinitesimal generator of this semigroup, and \(D_\pi\) be its domain. Following Heyer [14] pp.269–70, we may extend the domain of \(L\) to include bounded uniformly continuous functions on \(G\) that take the form \(f_{\psi,\phi}(g) = \langle \pi(g)\psi, \phi \rangle\) for all \(\psi \in D_\pi, \phi \in H_\pi, g \in G\), by observing that for all \(t \geq 0\),

\[
P_tf_{\psi,\phi}(g) = \langle \hat{\tilde{\mu}}_t(\pi)\psi, \pi(g^{-1})\phi \rangle,
\]

from which we can deduce that

\[
L_f \psi,\phi(g) = \langle A_\pi \psi, \pi(g^{-1})\phi \rangle.
\]

In the sequel, we will need the infinitesimal representation \(d_\pi\) of \(g\), corresponding to each \(\pi \in \hat{G}\), where for each \(Y \in g, -id_\pi(Y)\) is the infinitesimal generator of the strongly continuous one-parameter unitary group \((\pi(\exp(t)Y)), t \in \mathbb{R}\), where \(\exp : g \to G\) is the exponential map. Hence \(d_\pi(Y)\) is a (densely-defined) skew-adjoint linear operator acting in \(H_\pi\).

We will also need the dense set \(H_\omega^\pi\) of analytic vectors in \(H_\pi\) defined by

\[
H_\omega^\pi := \{\psi \in H_\pi; g \to \pi(g)\psi\text{ is analytic}\}.
\]

It is shown in [3] that \(H_\omega^\pi \subseteq D_\pi\).

The following Lévy–Khintchine type formula first appeared in Heyer [14], where it was established for compact Lie groups. Its extension to general Lie groups is implicit in Heyer [15]. For an alternative approach, based on operator-valued stochastic differential equations, see [3].

**Theorem 5.2.** If \((\mu_t, t \geq 0)\) is a convolution semigroup of probability measures on \(G\), then for all \(t \geq 0, \pi \in \hat{G}\),

\[
\hat{\mu}_t(\pi) = e^{tA_\pi},
\]

where for all \(\psi \in H_\omega^\pi\),

\[
A_\pi \psi = \sum_{i=1}^{d} b_i d\pi(X_i)\psi + \sum_{j,k=1}^{d} a_{jk} d\pi(X_j)d\pi(X_k)\psi
\]

\[
+ \int_G \left( \pi(\tau)\psi - \psi - \sum_{i=1}^{d} x_i(\tau)d\pi(X_i)\psi \right) \nu(d\tau),
\]

where \(b, a, \nu\) and \(x_i(1 \leq i \leq d)\) are as in Theorem 5.1.

**Proof.** This follows from (5.2) and (5.3) by the same arguments as used in the proof of Theorem 5.5.1 in [4], p.145–6. \(\square\)

### 6. Convolution Semigroups on Semisimple Lie Groups and Riemannian Symmetric Pairs

In this section we will assume that \(G\) is a Lie group and that \(K\) is a compact subgroup of \(G\). Let \(\mathfrak{g}\) denote the Lie algebra of \(K\), then it is easy to see that \(Xf = 0\) for all \(X \in \mathfrak{g}, f \in C^\infty_0(G/K)\). We write the vector space direct sum \(\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{k}^\perp\), and
we choose the basis \{X_1, \ldots, X_d\} of \mathfrak{g} so that \{X_1, \ldots, X_m\} is a basis for \mathfrak{t}^\perp, and
\{X_{m+1}, \ldots, X_d\} is a basis for \mathfrak{t}.

If \(G\) is semisimple, we have the \textit{Iwasawa decomposition} at the Lie algebra level:
\[ g = \mathfrak{t} \oplus \mathfrak{a} \oplus \mathfrak{n}, \]
where \(\mathfrak{a}\) is abelian and \(\mathfrak{n}\) is nilpotent. At the global level \(G\) is diffeomorphic to \(KAN\), where \(A\) is abelian and \(N\) is nilpotent, and we may write each \(g \in G\) as
\[ g = u(g) \exp(A(g)) n(g), \]
where \(u(g) \in K, A(g) \in \mathfrak{a}\) and \(n(g) \in N\) (see e.g. Chapter VI in [18]). Any minimal parabolic subgroup of \(G\) has a \textit{Langlands decomposition} \(MAN\) where \(M\) is the centraliser of \(A\) in \(K\). The \textit{principal series} of irreducible representations of \(G\) are obtained from finite dimensional representations of \(M\) by Mackey’s theory of induced representations. We will say more about this below.

A Gelfand pair \((G,K)\) is said to be a \textit{Riemannian symmetric pair}, if \(G\) is a connected Lie group and there exists an involutive analytic automorphism \(\sigma\) of \(G\) such that \((K \sigma)_0 \subseteq K \subseteq K \sigma\), where \(K \sigma := \{k \in K; \sigma(k) = k\}\), and \((K \sigma)_0\) is the connected component of \(e\) in \(K\). In this case, we always write \(p := \mathfrak{t}^\perp\), and note that
\[ p = \{X \in \mathfrak{g}; (d\sigma)_e(X) = -X\}. \]

We also have that \(N = G/K\) is a Riemannian symmetric space, under any \(G\)-invariant Riemannian metric on \(N\), and if \(\natural\) is the usual natural map from \(G\) to \(N\), then \((d\natural)_e : p \to T_o(X)\) is a linear isomorphism, where \(o := \natural(e)\). For details see e.g. Helgason [13], pp.209–10.

If \(G\) is semisimple, then we can find a Cartan involution \(\theta\) of \(\mathfrak{g}\), so that \((d\sigma)_e = \theta\). In this case \(K_\sigma = K\), and there is a natural Riemannian metric on \(N\), that is induced by the inner product \(B_\theta\) on \(\mathfrak{g}\), where for all \(X, Y \in \mathfrak{g}\),
\[ B_\theta(X, Y) = -B(X, \theta(Y)), \]
with \(B\) being the Killing form on \(\mathfrak{g}\) (see e.g. [18] pp.361–2).

From now on in this section, we assume that \(G\) is a noncompact, connected semisimple Lie group with finite centre, and that \(K\) is a maximal compact subgroup. Then \(G/K\) is a noncompact Riemannian symmetric space. We also assume that \(G/K\) is irreducible, i.e., that the action of \(\text{Ad}(K)\) on \(\mathfrak{g}\) is irreducible. Write \(g_{ij} = B(X_i, X_j)\), for \(i, j = 1, \ldots, m\), and define the horizontal Laplacian in \(G\) to be
\[ \Delta_H = \sum_{i,j=1}^n g_{ij}^{-1} X_i X_j, \]
where \((g_{ij}^{-1})\) is the \((i, j)\text{th}\) component of the inverse matrix to \((g_{ij})\). Then for all \(f \in C^\infty_c(N)\), we have
\[ \Delta_H(f \circ \natural) = \Delta f, \]
where \(\Delta\) is the Laplace–Beltrami operator on \(N\). We also have that for each \(\omega \in \mathcal{P}\), there exists \(c_\omega > 0\) so that
\[ \Delta_H \omega = -c_\omega \omega. \]
It is shown in [2] that if \((\mu, t \geq 0)\) is a \(K\)-bi-invariant continuous convolution semigroup, then for all \(f \in C^2_0(G)\), (5.2) reduces to

\[
\mathcal{L}f(\sigma) = a\Delta_H f(\sigma) + \int_G (f(\sigma\tau) - f(\sigma))\nu(d\tau),
\]

for all \(\sigma \in G\), where \(a \geq 0, \nu\) is a \(K\)-bi-invariant Lévy measure on \(G\), and the integral should be understood as a principal value. Then from (5.4), we obtain Gangolli’s Lévy–Khintchine formula (see also [21]), i.e., for all \(t \geq 0, \omega \in \mathcal{P}\),

\[
\frac{\omega}{\mu_t}(\omega) = e^{-t\psi_{\omega}},
\]

where

\[
\psi_{\omega} := \langle A_{\pi_\omega}u_\omega, u_\omega \rangle = -ac_{e_\omega} + \int_G (\omega(g) - 1)\nu(dg).
\]

In the remainder of this section, we will focus on more general Lévy–Khintchine formulae for standard convolution semigroups on semi-simple Lie groups.

The spherical representations of \(G\) are precisely the spherical principal series, which are obtained as follows. For each \(\lambda \in \mathfrak{a}^*\), define a representation \(\eta_\lambda\) of \(M\) on \(\mathbb{C}\) by

\[
\eta_\lambda(man) = e^{-i\lambda(\xi)},
\]

where \(m \in M, a = \exp(\xi) \in A, n \in N\). The required spherical representation \(\pi_\lambda\) acting on \(L^2(K)\) is obtained by applying the “Mackey machine” to \(\eta_\lambda\). In fact, we have for each \(g \in G, l \in K, f \in L^2(K)\),

\[
(\xi_\lambda(g)f)(l) = e^{-(i\lambda - \rho)(A(lg))}f(u(lg)),
\]

where \(\rho\) is the celebrated half-sum of positive roots (see e.g. the Appendix to [5]), and we are using the notation \(\xi_\lambda\) instead of \(\pi_{\omega_\lambda}\), for a generic element of the spherical principal series.

In this case we have \(u_\lambda := u_{\omega_\lambda} = 1\) in \(L^2(K)\) and we obtain Harish–Chandra’s beautiful formula for spherical functions:

\[
\omega_\lambda(g) = \langle u_\lambda, \pi_\lambda(g)u_\lambda \rangle = \int_K e^{(i\lambda + \rho)(A(kg))}dk,
\]

for all \(g \in G, \lambda \in \mathfrak{a}^*\). In particular, we may identify \(\mathcal{P}\) with \(\mathfrak{a}^*\).

For the general case, we explore the connection between the approach taken here, and the Lévy–Khintchine formula that was obtained in [5]. To that end, let \(\hat{K}\) be the unitary dual of \(K\), i.e., the set of all equivalence classes (up to unitary equivalence) of irreducible representations of \(K\). For each \(\pi \in \hat{K}\), let \(V_\pi\) be the finite-dimensional inner product space on which \(\pi(\cdot)\) acts, and write \(d_\pi = \dim(V_\pi)\).

For each \(\pi_1, \pi_2 \in \hat{K}, \lambda \in \mathfrak{a}^*\), define the generalised spherical function \(\Phi_{\lambda, \pi_1, \pi_2}\) by

\[
\Phi_{\lambda, \pi_1, \pi_2}(g) := \sqrt{\frac{d_{\pi_1}}{d_{\pi_2}}} \int_K e^{-(i\lambda - \rho)(A(kg))}(\pi_1(u(kg)) \otimes \overline{\pi_2}(k))dk,
\]

for all \(g \in G\), where \(\overline{\pi}\) denotes the conjugate representation associated to \(\pi\). Hence \(\Phi_{\lambda, \pi_1, \pi_2}(g)\) is a (bounded) linear operator on the space \(V_{\pi_1} \otimes \overline{V_{\pi_2}}\). The connection
with principal series representations is made apparent in Theorem 3.1 of [5], in that for all \( g \in G, u_1, v_1 \in V_{\pi_1}, u_2, v_2 \in V_{\pi_2} \),
\[
\langle \Phi_{\lambda,\pi_1,\pi_2}(g)(u_1 \otimes u_2^*), v_1 \otimes v_2^* \rangle_{V_{\pi_1} \otimes V_{\pi_2}^*} = \langle \xi_\lambda(g)f_{\pi_1}^{u_1,v_1}, f_{\pi_2}^{u_2,v_2} \rangle_{L^2(K)},
\]
where for each \( \pi \in \hat{K}, u, v \in V_\pi, k \in K, f_\pi^{u,v}(k) := \langle \pi(k)u, v \rangle \). Note that by Peter–Weyl theory, \( \mathcal{M}(K) := \text{lin. span}\{f_\pi^{u,v}(k); \pi \in \hat{K}, u, v \in V_\pi \} \) is dense in \( L^2(K) \).

If \( \mu \) is a finite measure defined on \((G, \mathcal{B}(G))\) then its generalised spherical transform is defined to be
\[
\mu_{\lambda,\pi_1,\pi_2}^{(S)} := \int_G \Phi_{\lambda,\pi_1,\pi_2}(g^{-1})\mu(dg).
\]
Then from (6.4), we easily deduce that
\[
\langle \mu_{\lambda,\pi_1,\pi_2}^{(S)}(u_1 \otimes u_2^*), v_1 \otimes v_2^* \rangle_{V_{\pi_1} \otimes V_{\pi_2}^*} = \langle \hat{\mu}\xi_\lambda f_{\pi_1}^{u_1,v_1}, f_{\pi_2}^{u_2,v_2} \rangle_{L^2(K)}. \tag{6.5}
\]

Now replace \( \mu \) by \( \mu_t \) in (6.5). In [5] a Lévy–Khinchine-type formula which extended Gangolli’s result from [11] was obtained, wherein the role of the characteristic exponent was played by
\[
\eta_{\lambda,\pi_1,\pi_2}(t) := \left. \frac{d}{dt}\mu_{\lambda,\pi_1,\pi_2}^{(S)} \right|_{t=0}.
\]
Differentiating in (6.5), we obtain
\[
\langle \eta_{\lambda,\pi_1,\pi_2}(u_1 \otimes u_2^*), v_1 \otimes v_2^* \rangle_{V_{\pi_1} \otimes V_{\pi_2}^*} = \langle A_{\xi_\lambda}f_{\pi_1}^{u_1,v_1}, f_{\pi_2}^{u_2,v_2} \rangle_{L^2(K)}, \tag{6.6}
\]
where we use the fact that \( \mathcal{M}(K) \subseteq C^\infty(K) \subseteq \text{Dom}(A_{\xi_\lambda}) \), and from here we have a direct relationship between the Lévy–Khinchine-type formula given in Theorem 5.1 of [5], and that of Theorem 5.2.

In section 6 of [5] an attempt was made to use the generalised spherical transform to obtain a Lévy–Khinchine formula for right \( K \)-invariant convolution semigroups, in the mistaken belief that there were non-trivial elements in that class that were not \( K \)-bi-invariant. The work of [20], as described in section 2 above, shows that this was erroneous.

7. A New Class of Processes on Symmetric Spaces

Let \((\Omega, \mathcal{F}, P)\) be a probability space, and \((L(t), t \geq 0)\) be a right Lévy process on \(G\) (so that it has stationary and independent left increments). Then the family of laws \((\mu_t, t \geq 0)\) is a (standard) convolution semigroup. We are interested in identifying classes of these processes so that the process \((\zeta(L(t)), t \geq 0)\) on \(N = G/K\), which is a Feller process (by Proposition 2.1 in [19], p.33), has interesting probabilistic and geometric properties. If \((\mu_t, t \geq 0)\) is \(K\)-bi-invariant, then \((\zeta(L(t)), t \geq 0)\) is a Lévy process on \(N\). Such processes were first investigated by Gangolli in [12] (see also [2, 21]), and the generic process was shown to be a Brownian motion on \(N\) interlaced with jumps having a \(K\)-bi-invariant distribution. We have seen that requiring that \((\mu_t, t \geq 0)\) is only right \(K\)-invariant gives us nothing new.
We begin with a càdlàg Lévy process \((Z(t), t \geq 0)\) taking values on \(\mathbb{R}^m\), where for each \(t \geq 0\), \(Z(t) = (Z_1(t), \ldots, Z_m(t))\), and having characteristics \((b, a, \nu)\). Assume that \(G\) is semisimple and consider the global Cartan decomposition \(G = \exp(p)K\). We induce a Lévy process \((\tilde{Z}(t), t \geq 0)\) on \(p\) by defining \(\tilde{Z}(t) = \sum_{i=1}^m Z_i(t)X_i\). As is shown in [7], Corollary to Theorem 2.4, we obtain a left Lévy process \((M(t), t \geq 0)\) on \(G\) by solving the stochastic differential equation (using the Markus canonical form \(\circ\)):

\[
dM(t) = M(t-) \circ d\tilde{Z}(t),
\]

with initial condition \(M(0) = e\) (a.s.).

The generator takes the form

\[
\mathcal{L}f(\sigma) = \sum_{i=1}^m b^i X_i f(\sigma) + \sum_{i,j=1}^m a^{ij} X_i X_j f(\sigma)
+ \int_{\mathbb{R}^m} \left[ f\left( \sigma \exp \left( \sum_{i=1}^m y^i X_i \right) \right)
- f(\sigma) - 1_{B_1}(y) \sum_{i=1}^m y^i X_i f(\sigma) \right] \nu(dy),
\]

where \(f \in C^2_b(G), \sigma \in G\). We then take \(L(t) = M(t)^{-1}\) for all \(t \geq 0\), to get the desired right Lévy process.

**Example 7.1** (Geodesics). Here the process \(Z\) has characteristics \((b, 0, 0)\). Fix \(Y = \sum_{i=1}^m b_i X_i \in p\), and consider the deterministic Lévy process \(L(t) = \exp(tY)\) for \(t \geq 0\). Then the operator \(\mathcal{L} = Y\) and \(\mathcal{L}(L(t)) = \exp(t\delta Y)\eta\), where \(\exp\) is the Riemannian exponential; i.e., \(\mathcal{L}(L(t))\) moves from \(o\) along the unique geodesic having slope \(d\delta Y \in T_o(N)\) at time zero.

**Example 7.2** (Compound Poisson Process with Geodesic Jumps). Let \((W_n, n \in \mathbb{N})\) be a sequence of independent, identically distributed random variables, taking values in \(p\), and having common law \(\eta\), and let \((N(t), t \geq 0)\) be a Poisson process of intensity 1 that is independent of all the \(W_n\)'s. Consider the Lévy process defined for \(t > 0\) by

\[
L(t) = \exp(W_{N(t)}) \exp(W_{N(t)-1}) \cdots \exp(W_1),
\]

The law of \(L(t)\) is \(\mu_t = e^{-t\delta_o} + \sum_{n=1}^\infty \frac{t^n}{n!} \eta^{\star n}\). Then

\[
L(L(t)) = \exp(d\delta(W_{N(t)})) \circ \exp(d\delta(W_{N(t)-1})) \circ \cdots \circ \exp(d\delta(W_1))\eta,
\]

describes a process which jumps along random geodesic segments. Here we slightly abuse notation so that for \(X, Y \in p\), we write \(\exp(d\delta(X)) \circ \exp(d\delta(Y))\eta\) for the geodesic that moves from time zero to time one, starting at the point \(q = \exp(d\delta(Y))\eta\), and having slope \(d\tau_g \circ d\delta(X)\), where \(g\) is the unique element of \(G\) such that \(q = \tau_g(o) := gK\).

In this case, \((P(t), t \geq 0)\) has a bounded generator,

\[
\mathcal{L}f(\sigma) = \int_g (f(\sigma \exp(Y)) - f(\sigma))d\eta(dY),
\]
for $f \in C_u(G), \sigma \in G$. The Lévy process $Z$ has characteristics $(b', 0, \eta')$. Here $\eta' := \tilde{\eta} \circ T$ and $b'_i = \int_{|y| < 1} y^i \eta'(dy)$, where $T$ is the vector space isomorphism between $\mathbb{R}^m$ and $\mathfrak{g}$, which maps each element $e_i$ of the natural basis in $\mathbb{R}^m$ to $X_i (i = 1, \ldots, m)$.

More examples can be constructed from (1) and (2) by interlacing. These extend the results of [6] (within the symmetric space context). They can also be seen as a special case of the construction in [1]. It is anticipated that the ideas in this section will be further developed in future work.

Acknowledgement. I would like to thank Ming Liao for very helpful comments, and also the referee for some useful suggestions.

References


DAVID APPLEBAUM: SCHOOL OF MATHEMATICS AND STATISTICS, UNIVERSITY OF SHEFFIELD, HICKS BUILDING, HOUNSFIELD ROAD, SHEFFIELD, S3 7RH, ENGLAND

*E-mail address*: D.Applebaum@sheffield.ac.uk
STRONG STATIONARY TIMES AND THE FUNDAMENTAL MATRIX FOR RECURRENT MARKOV CHAINS

P. J. FITZSIMMONS

ABSTRACT. We show that for a finite state space Markov chain, the occupation-time matrix up to a strong stationary time coincides with fundamental matrix of Kemeny and Snell, when each matrix is viewed as operating on functions with mean zero with respect to the stationary distribution.

1. Introduction

My aim in this short note is to point out a connection between two well-studied objects in the theory of Markov chains that appears to have gone unnoticed. I will confine my attention to discrete-time Markov chains with finite state space, but there is little doubt that the analogous results hold for Markov chains in continuous time with countable state spaces.

Throughout, \( X = (X_n) \) will be a discrete-time Markov chain with finite state space \( E = \{1, 2, \ldots, N\} \) and one-step transition matrix \( P \). We assume that \( X \) is irreducible and aperiodic. Let \( \pi \) denote the unique stationary distribution for \( X \). That is, \( \pi P = \pi \) and \( \pi \cdot 1 = 1 \). (Here \( 1 \) is an \( N \times 1 \) column of 1s. Measures on \( E \) are row vectors; function are column vectors.) As is well known, \( \lim_n P^n = \Pi \), the matrix with all rows equal to \( \pi \).

The law of \( X \) started at \( x \in E \) is \( P^x \), on the sample space \( \Omega \) of all \( E \)-valued sequences \( \omega = (\omega_n)_{n \geq 0} \). The symbol \( P^x \) will also be used for the associated expectation, and if \( \mu \) is a probability measure on \( E \) then \( P^\mu := \sum_{x \in E} \mu(x)P^x \) denotes the law (or expectation) of \( X \) under the initial distribution \( \mu \).

2. Fundamental Matrix

The fundamental matrix \( Z \) associated with \( X \) and \( P \) was introduced in [2] and generalized in [3]. We provide a bit of detail on the construction of \( Z \) for completeness.

Proposition 2.1. (a) The matrix \( I - P + \Pi \) is invertible, and \( Z := (I - P + \Pi)^{-1} \) denotes its inverse.

(b) \( \pi Z = \pi \).

(c) \( Z1 = 1 \).
Proof. (a) Let \( f : E \to \mathbb{R} \) (viewed as a column vector) be such that \((I - P + \Pi)f = 0\). Clearly \( \Pi f = c1 \), where \( c = \pi(f) := \sum_{i \in E} \pi_i f_i \). Consequently,

\[
 f - Pf + c1 = 0. \tag{2.1}
\]

Applying \( P \) on the left in (2.1) we obtain

\[
Pf - P^2f + c1 = 0, \tag{2.2}
\]

and then, adding (2.1) to (2.2),

\[
f - P^2f + 2c1 = 0. \tag{2.3}
\]

Proceeding recursively we find that

\[
f - P^nf + nc1 = 0, \tag{2.4}
\]

for \( n = 1, 2, \ldots \). Because the entries of \( P^nf \) remain bounded as \( n \to \infty \), it must be that \( c = 0 \). In particular, (2.3) now tells us that \( P^nf = f \) for all \( n \geq 1 \). But the only \( P \)-invariant functions are the constants, so

\[
f_i = \pi(f) = c = 0, \quad \forall i \in E, \tag{2.5}
\]

which means that \( f \) is the zero function. This proves that \( I - P + \Pi \) is invertible.

(b) Define \( \nu := \pi Z \). Then \( \nu(I - P + \Pi) = \pi \); that is (writing \( \tilde{c} \) for \( \sum_{i \in E} \nu_i \))

\[
\nu - \nu P + \tilde{c} \pi = \pi, \tag{2.6}
\]

and so

\[
\nu - \nu P = (1 - \tilde{c}) \pi. \tag{2.7}
\]

Multiply (2.7) on the right by \( 1 \) to see that

\[
\tilde{c} - \tilde{c} = (1 - \tilde{c}),
\]

and so \( \tilde{c} = 1 \). This yields \( \nu P = \nu \), and finally \( \nu = \pi \).

(c) The proof that \( Z1 = 1 \) follows the pattern of the proof of part (a) and is therefore omitted. \( \square \)

3. Poisson Equation

In potential theoretic terms, the fundamental matrix \( Z \) is a recurrent potential operator for \( X \), yielding solutions of the Poisson equation. More precisely, let \( f : E \to \mathbb{R} \) be given. We seek a function \( u : E \to \mathbb{R} \) such that \( u - Pu = f \). Observe that a necessary condition for this equation to have a solution is that \( \pi(f) = 0 \). Moreover, if \( u \) is a solution, then so is \( u + b1 \) for any real constant \( b \).

Given \( f : E \to \mathbb{R} \) define \( u := Zf \). Then \( u - Pu + c1 = f \) (where \( c := \pi(u) \)), and then recursively

\[
u - P^nu = \sum_{k=0}^{n-1} (P^k f - c1), \quad n \geq 1,
\]

so that

\[
u - \pi(u)1 = \sum_{k=0}^{\infty} (P^k f - c1). 
\]
As is well known, the entries of $P^k f$ converge to $\pi(f)$, at a geometric rate. It follows that $\pi(u) = c = \pi(f)$. Consequently,

$$u - \pi(u)1 = \sum_{k=0}^{\infty} (P^k f - \pi(f)1).$$

In particular, if $\pi(f) = 0$, then

$$u = \sum_{k=0}^{\infty} P^k f,$$

the series in (3.1) converging absolutely. Thus $u = f + P(\sum_{k=0}^{\infty} P^k f) = f + Pu$, so $u$ is a solution of the Poisson equation. Finally, the most general solution of the Poisson equation is $u_c := Zf + c1$, where $c = \pi(u_c)$.

4. Strong Stationary Time

Fix an initial state $x \in E$. From the work of Aldous and Diaconis [1] we know that there are (randomized) stopping times $S$, so-called strong stationary times, such that

$$P_x[X_S = i, S = k] = \pi_i \cdot P_x[S = k], \quad i \in E, k = 0, 1, 2, \ldots.$$ 

That is, under $P_x$, $X_S$ has law $\pi$ and is independent of $S$, at least on $\{S < \infty\}$. (Our chain $X$ admits such strong stationary times, and only such times will be of interest here.) Strong stationary times play a crucial role in bounding the separation distance $s_x(n)$ between $\pi$ and $P^x[X_n = \cdot]$:

$$s_x(n) := \max_i (1 - P^x[X_n = i]/\pi_i), \quad n = 0, 1, 2, \ldots.$$ 

To wit,

$$s_x(n) \leq P^x[S > n], \quad n = 0, 1, 2, \ldots,$$ 

provided $S$ is a strong stationary time (under $P^x$). See [1] (3.2); especially note that this bound is sharp in the sense that there is a strong stationary time for which equality holds in (4.1) for all $n \geq 0$. For more on these matters see [4, 5] and for the extension to continuous time see [6, 7].

Let’s now suppose that a $P^x$-strong stationary time $S_x$ has been chosen for each $x \in E$. Define $S(\omega) := S_{X_0(\omega)}(\omega), \omega \in \Omega$. Let $\mu$ be any initial distribution for $X$. Then

$$P^\mu[X_S = i, S = k] = \sum_{x \in E} P^x[X_{S_x} = i, S_x = k]\mu(x)$$

$$= \sum_{x \in E} \pi(i)P^x[S_x = k]\mu(x)$$

$$= \pi(i)P^\mu[S = k].$$

In other words, $S$ is strongly stationary for each initial distribution. Moreover,

$$P^\mu[S] = \sum_{x \in E} P^x[S_x]\mu(x) \leq \max_{x \in E} P^x[S_x] < \infty.$$
In what follows $S$ will always be such a “universal” strongly stationary time with finite expectation.

5. The Connection

Fix $S$ as at the end of section 4, and define an operator $W = W_S$ (“mean occupation measure”) by

$$Wf(x) := \mathbb{P}^x \sum_{k=0}^{S-1} f(X_k).$$

Clearly $|Wf(x)| \leq \|f\|_\infty \cdot \mathbb{P}^x[S]$ for each $x$. Moreover, $W1(x) = \mathbb{P}^x[S]$. The crucial observation is that by the simple Markov property, $W(Pf)(x) = Wf(x) - f(x) + \mathbb{P}^x[f(X_S)] = Wf(x) - f(x) + \pi(f)$. Here is our main result.

Theorem 5.1. If $\pi(f) = 0$ then $Wf(x) = Zf(x)$ for all $x \in E$.

Proof. Define $V_0 := \{f \in \mathbb{R}^E : \pi(f) = 0\}$, and note that $I - P : V_0 \to V_0$. In view of the discussion in section 3, $Z : V_0 \to V_0$ and $(I - P)Z = I$ on $V_0$. Also, by the computation preceding the statement of the theorem, $W(I - P) = I$ on $V_0$. (This is true even without the independence of $X_S$ and $S$.) Now fix $\alpha > 0$, and write

$$U^\alpha := \sum_{k=0}^\infty e^{-k\alpha} \mathbb{P}^k$$

for the $\alpha$-potential operator associated with $P$. We have, by the strong Markov property at time $S$, the independence of $X_S$ and $S$, and the fact that $X_S$ has law $\pi$,

$$U^\alpha f(x) = W^\alpha f(x) + \mathbb{P}^x[e^{-\alpha S}]\pi(U^\alpha f), \quad (5.1)$$

where

$$W^\alpha f(x) := \mathbb{P}^x \sum_{k=0}^{S-1} e^{-\alpha k} f(X_k).$$

But $\pi$ is invariant, so $\pi U^\alpha = (1 - e^{-\alpha})^{-1} \pi$, and therefore $\pi(U^\alpha f) = 0$ provided $f \in V_0$. It now follows from (5.1) that $\pi(W^\alpha f) = 0$ for each $\alpha > 0$. Sending $\alpha \downarrow 0$ we find that $\pi(Wf) = 0$ provided $\pi(f) = 0$. That is, $W : V_0 \to V_0$ as well. We have identified left and right inverses of the restriction of $I - P$ to $V_0$. It follows that this restriction is invertible and of course the left and right inverses coincide. That is, $W = Z$ on $V_0$. \hfill $\square$

Corollary 5.2. (a) With $S$ as above, but now for general $f : E \to \mathbb{R}$,

$$Wf(x) = Zf(x) + \pi(f) \cdot [\mathbb{P}^x[S] - 1], \quad \forall x \in E.$$

(b) If $R$ is a second strong stationary time, then

$$W_S f(x) - W_R f(x) = \pi(f) \cdot [\mathbb{P}^x[S] - \mathbb{P}^x[R]].$$

It may be worth noting that if $R$ and $S$ are strong stationary times, then so is their concatenation $R + S \theta R$. 

References


P. J. FITZSIMMONS: DEPARTMENT OF MATHEMATICS, UC SAN DIEGO, LA JOLLA, CA 92093-0112, USA
E-mail address: pfitzsim@ucsd.edu
URL: http://math.ucsd.edu/~pfitz
BIMODULES AND HYPERGROUPS ASSOCIATED WITH ACTIONS OF A PAIR OF GROUPS

SATOSHI KAWAKAMI, TATSUYA TSURII, AND SHIGERU YAMAGAMI

Abstract. The present paper presents conditions on the set of equivalence classes of bimodules associated with actions of a pair of finite groups on a von Neumann algebra to have the structure of a fusion rule algebra.

1. Introduction

After V. F. R. Jones [16] introduced the notion of index for subfactors, Galois theory and the theory of paragroups for inclusions of subfactors have been developed, which one initiated by A. Ocneanu’s idea [21] by applying induced representations and restrictions of representations of a pair of groups. On the other hand V. S. Sunder and N. J. Wildberger [27] investigated the constructing of fusion rule algebras associated with Dynkin diagrams appearing in the principal graphs of the inclusions of subfactors. There are many works related to bimodules and fusion rule algebras in tensor categories, for examples, V. S. Sunder [25] and [26], A. K. Vijayarajan [28], R. Schaffitzel [22] and [23], S. Yamagami [29], H. Kosaki and S. Yamagami [18] and [19], D. Evans and Y. Kawahigashi [4], M. Izumi [14] and the graph theoretical approach in [5].

Hypergroups are locally compact spaces on which the convolution and the involution of bounded measures are given analogous to the group case, where the convolution of two point measures is a probability measure with compact support (not necessarily a point measure). There exists an axiomatic approach to hypergroups initiated by C. F. Dunkl ([2], [3], 1973), R. I. Jewett ([15], 1975) and R. Spector ([24], 1975) which leads to an extensive harmonic analysis of hypergroups. For the historical background of the theory we just refer to R. I. Jewett’s fundamental paper [15] and the monograph [1] by W. R. Bloom and H. Heyer. In fact, the hypergroup was developed to be of significant applicability in probability theory where the hypergroup convolution of measure reflects a stochastic operation in the basic space of the hypergroup. Nowadays hypergroup structures are studied within various frameworks from non-commutative duality of groups to quantum groups, deformations of hypergroups [17] and bimodules.

The present work stands against a background of the above works together with our results [6], [7], [8], [9], [10], [11], [12], and [13], which are joint works with Herbert Heyer.

Received 2016-9-5; Communicated by D. Applebaum.
2010 Mathematics Subject Classification. Primary 43A62; Secondary 16D20, 20N20, 22D25.
Key words and phrases. Hypergroup, bimodule, fusion rule algebra, von Neumann algebra.
Let $G$ be a finite group and $\hat{G}$ the dual of $G$, namely the set of equivalence classes of irreducible representations of $G$. We will obtain a fusion rule algebra $\mathcal{F}(\hat{G})$ through the tensor category of $\hat{G}$ and the character hypergroup $\mathcal{K}(\hat{G})$ of $G$.

Let $\alpha$ be an action of $G$ on a von Neumann algebra $M$. We denote the fixed point algebra by $A := M^G$. In the present paper we introduce a notion of dual unitary property on the action $\alpha$ of $G$ on $M$.

In Section 3 we consider irreducible $A$-$A$ bimodules and $G$-modules in $M$. Let $\mathcal{F}(M, G, \alpha)$ be the set of equivalence classes of such irreducible $A$-$A$ bimodules and $G$-modules. Then we show that if the action $\alpha$ has dual unitary property, then $\mathcal{F}(M, G, \alpha)$ has a structure of a fusion rule algebra and $\mathcal{F}(M, G, \alpha) \cong \mathcal{F}(\hat{G})$.

By normalizing $\mathcal{F}(M, G, \alpha)$ with the dimension function we obtain a hypergroup $\mathcal{K}(M, G, \alpha)$.

In Section 4 we discuss the case $M = B(l^2(G))$ and $\alpha_g = \text{Ad}\lambda_g$ where $\lambda$ is the regular representation of $G$. We show that the action $\alpha$ has dual unitary property so that $\mathcal{F}(B(l^2(G)), G, \text{Ad}\lambda) \cong \mathcal{F}(\hat{G})$. By this fact we see that $\mathcal{F}(M, G, \alpha) \cong \mathcal{F}(\hat{G})$ for an outer action $\alpha$ of $G$ on a hyperfinite $\text{II}_1$-factor $M$.

In Section 5 we study fusion rule algebras and hypergroups associated with a pair $(G, G_0)$ where $G_0$ is a subgroup of $G$. We give the convolution on $\mathcal{F}(M^G \subset M^{G_0})$ which is the disjoint union $\mathcal{F}(M, G, \alpha)$ and $\mathcal{F}(M, G_0, \alpha)$ by inductions and restrictions of modules. Then we show that if the pair $(G, G_0)$ is admissible, then $\mathcal{F}(M^G \subset M^{G_0})$ is a fusion rule algebra. By normalizing $\mathcal{F}(M^G \subset M^{G_0})$ we obtain a hypergroup $\mathcal{K}(M^G \subset M^{G_0})$ and show that $\mathcal{K}(M^G \subset M^{G_0}) \cong \mathcal{K}(\hat{G} \cup \hat{G_0})$ where the hypergroup $\mathcal{K}(\hat{G} \cup \hat{G_0})$ is introduced in [10] and [12].

2. Preliminaries

For a finite set $K = \{c_0, c_1, \ldots, c_n\}$, we denote by $\mathbb{C}K$ the algebraic complex linear space based on $K$, namely

$$\mathbb{C}K := \left\{ \sum_{j=0}^{n} a_j c_j : a_j \in \mathbb{C} \ (j = 0, 1, 2, \ldots, n) \right\}$$

and

$$\langle \mathbb{C}K \rangle_1 := \left\{ \sum_{j=0}^{n} a_j c_j : a_j \geq 0 \ (j = 0, 1, 2, \ldots, n), \sum_{j=0}^{n} a_j = 1 \right\}.$$  

For $\mu = a_0 c_0 + a_1 c_1 + \cdots + a_n c_n \in \mathbb{C}K$, the support of $\mu$ is

$$\text{supp}(\mu) := \{ c_j \in K : a_j \neq 0 \ (j = 0, 1, 2, \cdots, n) \}.$$  

2.1. Finite hypergroups. A finite hypergroup $K = (K, \mathbb{C}K, \circ, \ast)$ consists of a finite set $K = \{c_0, c_1, \ldots, c_n\}$ together with an associative product (called convolution) $\circ$ and an involution $\ast$ in $\mathbb{C}K$ satisfying the following conditions.

(H1) The space $(\mathbb{C}K, \circ, \ast)$ is an associative $\ast$-algebra with unit $c_0$.

(H2) For $c_i, c_j \in K$, the convolution $c_i \circ c_j$ belongs to $(\mathbb{C}K)_1$.

(H3) $K^\ast = K$ i.e. $c_i^\ast \in K$ for $c_i \in K$. Moreover, $c_j = c_j^\ast$ if and only if $c_0 \in \text{supp}(c_i \circ c_j)$.  

2.2. Finite fusion rule algebras. A finite fusion rule algebra $F = (F, \circ, *)$ consists of a finite set $F = \{X_0, X_1, \ldots, X_n\}$ together with an associative product (called convolution) $\circ$ and an involution $*$ in $\mathbb{C}F$ satisfying the following conditions.

(F1) The space $(\mathbb{C}F, \circ, *)$ is an associative involutive algebra with unit $X_0$.

(F2) For $X_i, X_j \in F$, the convolution of $X_i$ and $X_j$ is given by

$$X_i \circ X_j = \sum_{k=0}^{n} a_{ij}^k X_k$$

where the $a_{ij}^k$ are non-negative integers.

(F3) $F^* = F$ i.e. $X_i^* \in F$ for $X_i \in F$. Moreover, $X_j = X_i^*$ if and only if $X_0 \in \text{supp}(X_i \circ X_j)$ and

$$X_i \circ X_j = X_0 + \sum_{k=1}^{n} a_{ij}^k X_k.$$ 

For a finite fusion rule algebra $F = \{X_0, X_1, \ldots, X_n\}$ there exists the unique dimension function $d : F \to \mathbb{R}_+$. For $X_j \in F$, put

$$c_j := \frac{1}{d(X_j)} X_j.$$ 

Then $K = \{c_0, c_1, \ldots, c_n\}$ becomes a hypergroup. We call this $K$ the hypergroup obtained by normalizing $F$, denoted $K$ by $K_d(F)$.

2.3. Hypergroup joins. Let $H = \{h_0, h_1, \ldots, h_n\}$ and $L = \{\ell_0, \ell_1, \ldots, \ell_m\}$ be finite hypergroups where $\circ_H, \circ_L$ are the convolutions of $H$ and $L$ respectively. On the set $K = H \triangledown_L \{h_0, h_1, \ldots, h_n, \ell_1, \ldots, \ell_m\}$, we define the convolution $\circ_K$ by

$$h_i \circ_K h_j := h_i \circ_H h_j,$$

$$h_i \circ_K \ell_j := \ell_j,$$

$$\ell_i \circ_K \ell_j := \ell_i \circ_L \ell_j \text{ if } \ell_j \neq \ell_i^*,$$

$$\ell_i \circ_K \ell_j := a_{ij}^0 \omega_H + \sum_{k=1}^{m} a_{ij}^k \ell_k \text{ if } \ell_j = \ell_i^*,$$

where $\omega_H$ is the normalized Haar measure of $H$ and

$$\ell_i \circ_L \ell_j = \sum_{k=0}^{m} a_{ij}^k \ell_k.$$ 

We also define the involution $*_K$ by

$$h_i^*_K := h_i^{*_H} \text{ and } \ell_i^*_K := \ell_i^{*_L}.$$ 

Then $(H \triangledown_L, \circ_K, *_K)$ becomes a finite hypergroup, which we call the hypergroup join of $H$ and $L$. 

2.4. Fusion rule algebra joins ([20], related to near-group in [14]). Let 
\( H = \{X_0, X_1, \ldots, X_n\} \) be finite fusion rule algebra and \( L = \{Y_0, Y_1\} \) the cyclic group of order two where \( \circ_H, \circ_L \) are the convolutions of \( H \) and \( L \) respectively. We assume that \( d(X_i) \) is a natural number for each \( X_i \in H \). On the set \( F = H \vee F L = \{X_0, X_1, \ldots, X_n, Y_1\} \), we define the convolution \( \circ_F \) by
\[
X_i \circ_F X_j := X_i \circ_H X_j,
X_i \circ_F Y_1 := d(X_i)Y_1,
Y_1 \circ_F Y_1 := R(H)
\]
where
\[
R(H) = \sum_{k=0}^{n} d(X_k)X_k.
\]
We also define the involution \( *_F \) by
\[
X_i^{*F} := X_i^{*H} \quad \text{and} \quad Y_1^{*F} := Y_1^{*L}.
\]
Then \( (H \vee F L, \circ_F, *_F) \) becomes a finite fusion rule algebra, which we call the fusion rule algebra join of \( H \) and \( L \).

3. Bimodules Associated with Actions of a Finite Group

Let \( \alpha \) be an action of a finite group \( G \) on a von Neumann algebra \( M \). We denote the fixed point algebra of \( M \) under the action \( \alpha \) by
\[
A := M^G := \{x \in M : \alpha_g(x) = x \text{ for all } g \in G\}.
\]
Then \( M \) is interpreted as an \( A-A \) bimodule and a \( G \)-module. We call a closed subspace \( X \) of \( M \) an \( (A,A,G) \)-module if \( X \) is an \( A-A \) bimodule and a \( G \)-module, i.e., if \( x \in X \), then \( axb \in X \) and \( \alpha_g(x) \) belong to \( X \) for \( a, b \in A \) and \( g \in G \). For two \( (A-A,G) \)-modules \( X \) and \( Y \), a linear map \( T \) from \( X \) to \( Y \) is called a \( (A-A,G) \)-module map if \( T \) satisfies
\[
T(axb) = aT(x)b \quad \text{and} \quad T(\alpha_g(x)) = \alpha_g(T(x))
\]
for \( a, b \in A \) and \( g \in G \). If there is a bijective \( (A-A,G) \)-module map from \( X \) onto \( Y \), we say that \( X \) is equivalent to \( Y \), written by \( X \cong Y \).

For an \( (A-A,G) \)-module \( X \), we denote by \( \text{End}(X) \) the space of all \( (A-A,G) \)-module maps on \( X \). The module \( X \) is called irreducible if \( \text{End}(X) = \mathbb{C} \cdot I \), where \( I \) is the identity on \( X \).

For \( x \in M \) we denote by \( V(x) \) the linear span of \( \alpha_h(x) \), \( h \in G \):
\[
V(x) := \left\{ \sum_{h \in G} c_h \alpha_h(x) : c_h \in \mathbb{C} \right\}.
\]

**Definition 3.1. (Dual unitary property)** An action \( \alpha \) of \( G \) on \( M \) has dual unitary property if there exists a unitary operator \( u(\pi) \in M \) for each \( \pi \in \hat{G} \) such that \( \alpha_g(u(\pi)) = \pi(g)u(\pi) = V(\pi) := V(u(\pi)) \).

Throughout this section we assume that the action \( \alpha \) of \( G \) on \( M \) has the dual unitary property. Let \( H(\pi) \) be the representation space of \( \pi \in \hat{G} \).
Lemma 3.2. \( \alpha_g(x) = \pi(g)x \) for all \( x \in V(\pi) \). Moreover, \( V(\pi) \cong H(\pi) \) as a \( G \)-module.

Proof. We only show that
\[
\alpha_g(\alpha_h(u(\pi))) = \pi(g)(\alpha_h(u(\pi)))
\]
for \( g, h \in G \). Indeed
\[
\begin{align*}
\alpha_g(\alpha_h(u(\pi))) &= \alpha_{gh}(u(\pi)) \\
&= \pi(gh)u(\pi) \\
&= \pi(g)\pi(h)u(\pi) \\
&= \pi(g)(\alpha_h(u(\pi))).
\end{align*}
\]

Lemma 3.3. For \( \pi \in \hat{G} \), \( AV(\pi)A = AV(\pi) = V(\pi)A \).

Proof. For \( axb \in AV(\pi)A \ (a, b \in A, x \in V(\pi)) \), since
\[
axbu(\pi)^* \in A \text{ and } u(\pi) \in V(\pi)
\]
we have
\[
axb = (abxu(\pi)^*)u(\pi) \in AV(\pi).
\]
Hence we see that \( AV(\pi)A = AV(\pi) \). The formula \( AV(\pi)A = V(\pi)A \) follows in a similar way.

For \( \pi \in \hat{G} \) we write \( X(\pi) := AV(\pi)A \).

Lemma 3.4. For \( \pi \in \hat{G} \), \( X(\pi) \) is an irreducible \( (A-A, G) \)-module.

Proof. For \( x \in V(\pi) \) and \( T \in \text{End}(X) \),
\[
T(\alpha_g(x)) = T(\pi(g)x) = \pi(g)T(x),
\]
which gives \( T\pi(g) = \pi(g)T \). Since \( \pi \) is irreducible, i.e. \( \pi(G)' = C \cdot I \) where
\[
\pi(G)' = \{S \in B(\ell^2(G)) : S\pi(G) = \pi(G)S \text{ for all } g \in G\},
\]
we see that \( T(x) = c \cdot x \ (c \in C) \) for all \( x \in V(\pi) \). For \( axb \in X(\pi) \ (a, b \in A) \),
\[
T(axb) = aT(x)b = a(c \cdot x)b = c \cdot axb,
\]
which means that \( T = c \cdot I \). Hence \( X(\pi) \) is an irreducible \( (A-A, G) \)-module.

For a finite group \( G \), write \( \hat{G} := \{\pi_0, \pi_1, \ldots, \pi_\ell\} \) where \( \pi_0 \) is the trivial representation of \( G \), and for \( \pi \in \hat{G} \), put
\[
Ch(\pi)(g) := \text{tr}(\pi(g)) \text{ and } ch(\pi) := \frac{1}{\dim \pi} Ch(\pi).
\]
Then
\[
\mathcal{F}(\hat{G}) = \{Ch(\pi) : \pi \in \hat{G}\}
\]
is a fusion rule algebra with unit \( Ch(\pi_0) \) and
\[
\mathcal{K}(\hat{G}) = \{ch(\pi) : \pi \in \hat{G}\}.
\]
is a hypergroup called the character hypergroup of $G$. We note that $\mathcal{K}_d(\mathcal{F}(\hat{G})) = \mathcal{K}(\hat{G})$ and $\mathcal{K}_d(H \vee_F L) = \mathcal{K}_d(H) \vee \mathcal{K}_d(L)$.

A finite representation $\pi$ of $G$ is decomposed as

$$\pi \cong \sum_{j=1}^{\ell} \oplus m(\pi_j)\pi_j.$$ 

Then we define $X(\pi)$ by

$$X(\pi) = \sum_{j=1}^{\ell} \oplus m(\pi_j)X(\pi_j)$$

where

$$m(\pi_j)X(\pi_j) = X(\pi_j) \oplus \cdots \oplus X(\pi_j).$$

**Proposition 3.5.** For $\pi_i, \pi_j \in \hat{G}$, $X(\pi_i)_A \otimes_A X(\pi_j) = X(\pi_i \otimes \pi_j)$.

**Proof.**

$$X(\pi_i)_A \otimes_A X(\pi_j) = AV(\pi_i) \otimes V(\pi_j)A$$

$$= AV(\pi_i \otimes \pi_j)A$$

$$= X(\pi_i \otimes \pi_j).$$

□

For an action $\alpha$ of $G$ on $M$ we denote the set of the equivalence classes of irreducible $(A-A, G)$-modules by $\mathcal{F}(M, G, \alpha)$. For $X_i$ and $X_j \in \mathcal{F}(M, G, \alpha)$, we define the convolution $X_i \ast X_j$ by

$$X_i \ast X_j := X_i A \otimes_A X_j$$

and the sum of $X_i + X_j$ by

$$X_i + X_j := X_i \oplus X_j.$$ 

**Theorem 3.6.** If an action $\alpha$ of a finite group $G$ on a von Neumann algebra $M$ has the dual unitary property, then $\mathcal{F}(M, G, \alpha)$ is an fusion rule algebra and $\mathcal{F}(M, G, \alpha) \cong \mathcal{F}(\hat{G})$.

**Proof.** This statement follows from

$$\mathcal{F}(M, G, \alpha) = \{ X(\pi) : \pi \in \hat{G} \}$$

together with Proposition 3.5. □

Let $\mathcal{K}(M, G, \alpha)$ be the hypergroup obtained by normalizing $\mathcal{F}(M, G, \alpha)$.

**Corollary 3.7.** $\mathcal{K}(M, G, \alpha) \cong \mathcal{K}(\hat{G})$. 

4. Bimodules Associated with the Regular Actions

Let $\lambda$ be the left regular representation of a finite group $G$ and $\alpha$ an action of $G$ on $M = B(\ell^2(G))$ given by $\alpha_g := \text{Ad}_\lambda$ for $g \in G$.

**Proposition 4.1.** The action $\alpha$ of $G$ on $B(\ell^2(G))$ has the dual unitary property and $\mathcal{F}(B(\ell^2(G)), G, \alpha) \cong \mathcal{F}(\hat{G})$.

**Proof.** There exists the minimal projection $p \in B(\ell^2(G))$ such that \{\alpha_h(p) : h \in G\} are mutual orthogonal projections and $\sum_{h \in G} \alpha_h(p) = 1$. We write $p_h := \alpha^{-1}(p)$. Then we see that $\alpha_g(p_h) = p_{hg^{-1}}$. We denote by $\ell^2(\hat{G})$ the space of $B(H_\pi)$-valued functions on each $\pi \in \hat{G}$.

Let $W$ be Fourier transform from $\ell^2(G)$ onto $\ell^2(\hat{G})$ given by

$$ (W \xi)(\pi) = \hat{\xi}(\pi) = \frac{1}{\sqrt{|G|}} \sum_{h \in G} \pi(h)^* \xi(h) $$

for $\xi \in \ell^2(G)$. Define

$$ q_h := W p_h W^* $$

for $h \in G$. Then $q_h$, $h \in G$ are also mutual orthogonal projections on $\ell^2(\hat{G})$ such that $\alpha_g(q_h) = q_{hg^{-1}}$ and $\sum_{h \in G} q_h = 1$.

We consider the unitary operator $\tilde{\pi}$ on $\ell^2(\hat{G})$ for $\pi \in \hat{G}$ defined by

$$ (\tilde{\pi}(g)\xi)(\tau) := \begin{cases} 
\xi(\tau) & \text{if } \tau \neq \pi, \\
\pi(g)\xi(\pi) & \text{if } \tau = \pi.
\end{cases} $$

We define a unitary operator $u(\pi)$ by

$$ u(\pi) := \sum_{h \in G} \tilde{\pi}(h)q_h. $$

Then $\alpha_g(u(\pi)) = \pi(g)u(\pi)$ on $V(\pi)$. Indeed for $u(\pi) \in V(\pi)$

$$ \alpha_g(u(\pi)) = \sum_{h \in G} \alpha_g(\tilde{\pi}(h)q_h) $n

$$ = \sum_{h \in G} \alpha_g(\tilde{\pi}(h))\alpha_g(q_h) $n

$$ = \sum_{h \in G} \lambda_g\tilde{\pi}(h)\lambda_g^*\alpha_g(q_h) $n

$$ = \sum_{h \in G} \tilde{\pi}(ghg^{-1})q_{hg^{-1}} $n

$$ = \sum_{k \in G} \tilde{\pi}(g)\tilde{\pi}(k)q_k k = hg^{-1} $n

$$ = \pi(g)\sum_{k \in G} \tilde{\pi}(k)q_k \text{ on } \ell^2(\hat{G}) $n

$$ = \pi(g)u(\pi). $$
Since $B(\ell^2(G)) \cong B(\ell^2(\hat{G}))$ through $\text{Ad}W$, then there exists a unitary operator $u(\pi) \in B(\ell^2(G))$ such that $\alpha_g(u(\pi)) = \pi(g)u(\pi)$ on $V(\pi)$. By Theorem 3.6 we see that $\mathcal{F}(B(\ell^2(G)), G, \alpha) \cong \mathcal{F}(\hat{G})$.

\[= (\text{Ad}g)(u(\pi)) \otimes \alpha_0^g(I)
= \pi_g(u(\pi)) \otimes I
= \pi_g(u(\pi) \otimes I
= \pi_g(\hat{u}(\pi)).\]

This implies that the action $\alpha = (\text{Ad} \lambda) \otimes \alpha^0$ on $B(\ell^2(G)) \otimes M_0$ has the dual unitary property, leading to the desired conclusion. \[\Box\]

**Theorem 4.2.** Let $\alpha$ be an action of a finite group $G$ on a von Neumann algebra $M$. If $(M, G, \alpha) \cong (B(\ell^2(G)) \otimes M_0, G, \text{Ad} \lambda \otimes \alpha_0)$, then $\mathcal{F}(M, G, \alpha) \cong \mathcal{F}(\hat{G})$.

**Proof.** Since by Proposition 4.1 $B(\ell^2(G))$ contains the unitary operator $u(\pi)$ for each $\pi \in \hat{G}$ such that $\alpha_g(u(\pi)) = \pi_g(u(\pi))$, we have $\hat{u}(\pi) = u(\pi) \otimes I$ is a unitary operator on $M = B(\ell^2(G)) \otimes M_0$. Then

\[
\alpha_g(\hat{u}(\pi)) = (\text{Ad} \lambda_g \otimes \alpha^0_g)(u(\pi) \otimes I)
= (\text{Ad} \lambda_g(u(\pi)) \otimes \alpha_0^g(I)
= \pi_g(u(\pi)) \otimes I
= \pi_g(u(\pi) \otimes I
= \pi_g(\hat{u}(\pi)).
\]

**Corollary 4.3.** If $\alpha$ is an outer action of $G$ on a hyperfinite II$_1$-factor $M$, then $(M, G, \alpha) \cong (B(\ell^2(G)) \otimes M_0, G, \text{Ad} \lambda \otimes \alpha_0)$ so that $\mathcal{F}(M, G, \alpha) \cong \mathcal{F}(\hat{G})$.

**Conjecture** If an outer action $\alpha$ of a finite group $G$ on a factor has Rohlin property i.e. there exists a projection $p$ such that $\alpha_h(p)$, $h \in G$ are mutually orthogonal projections and $\sum_{h \in G} \alpha_h(p) = 1$, then $\mathcal{F}(M, G, \alpha) \cong \mathcal{F}(\hat{G})$.

5. **Bimodules Associated with a Pair $(G, G_0)$ of Finite Groups**

Let $G_0$ be a subgroup of a finite group $G$ and $\alpha$ an action of $G$ on a von Neumann algebra $M$. In this section we assume that the action $\alpha$ has dual unitary property and the restriction of $\alpha$ to $G_0$ also has the dual unitary property. Then by Theorem 3.6 $\mathcal{F}(M, G, \alpha)$ and $\mathcal{F}(M, G_0, \alpha)$ are fusion rule algebras and $\mathcal{F}(M, G, \alpha) \cong \mathcal{F}(\hat{G})$ and $\mathcal{F}(M, G_0, \alpha) \cong \mathcal{F}(\hat{G}_0)$.

**Definition 5.1.** For $X(\pi) \in \mathcal{F}(M, G, \alpha)$ ($\pi \in \hat{G}$) we define the restriction of $X(\pi)$ to $G_0$ by

\[\text{res}_{G_0}^G X(\pi) := Y(\text{res}_{G_0}^G \pi),\]

and the induced module of $Y(\tau) \in \mathcal{F}(M, G_0, \alpha)$ ($\tau \in \hat{G}_0$) to $G$ by

\[\text{ind}_{G_0}^G Y(\tau) := X(\text{ind}_{G_0}^G \tau).\]

**Definition 5.2.** The convolution $*$ of $\mathcal{F}(M^G \subset M^{G_0}) = \{(X(\pi), \circ), (Y(\tau), \bullet) : \pi \in \hat{G}, \tau \in \hat{G}_0\}$ is defined by

\[(X(\pi_i), \circ) * (X(\pi_j), \circ) := (X(\pi_i) * X(\pi_j), \circ),\]
\[(X(\pi), \circ) * (Y(\tau), \bullet) := (\text{res}_{G_0}^G X(\pi) * Y(\tau), \bullet),\]
Definition 5.3. Let \((G, G_0)\) be a pair of finite groups. We call \((G, G_0)\) an admissible pair if the following conditions are satisfied:

(1) For \(\pi \in \hat{G}\) and \(\tau \in \hat{G}_0\),
\[
\text{ind}_{G_0}^G(\text{res}_{G_0}^G(\pi)) \cdot Y(\tau) = X(\pi) \cdot \text{ind}_{G_0}^G(\tau).
\]

(2) For \(\tau \in \hat{G}_0\),
\[
\text{res}_{G_0}^G(\text{ind}_{G_0}^G(\tau)) = Y(\tau) \cdot \text{res}_{G_0}^G(\text{ind}_{G_0}^G(\tau)),
\]
where \(\pi_0\) is the trivial representation of \(G_0\). 

\(\mathcal{K}(M^G \subset \hat{M}^{G_0})\) is the hypergroup obtained by normalizing \(\mathcal{F}(M^G \subset \hat{M}^{G_0})\), and the hypergroup \(\mathcal{K}(\hat{G} \cup \hat{G}_0)\) is introduced in [10].

Theorem 5.4. \(\mathcal{F}(M^G \subset \hat{M}^{G_0})\) is a fusion rule algebra if and only if \((G, G_0)\) is admissible. Moreover, \(\mathcal{F}(M^G \subset \hat{M}^{G_0}) \cong \mathcal{F}(\hat{G} \cup \hat{G}_0)\) and \(\mathcal{K}(M^G \subset \hat{M}^{G_0}) \cong \mathcal{K}(\hat{G} \cup \hat{G}_0)\).

Proof. If a finite group \(G\) together with a subgroup \(G_0\) forms an admissible pair, then the following associativity relations hold. For \(\pi_1, \pi_j, \pi_k, \pi \in \hat{G}\) and \(\tau_1, \tau_j, \tau_k, \tau \in \hat{G}_0\),

(A1) \((X(\pi_1), \o) \ast (X(\pi_j), \o) \ast (X(\pi_k), \o) = (X(\pi_1), \o) \ast ((X(\pi_j), \o) \ast (X(\pi_k), \o))\),

(A2) \(((Y(\tau), \bullet) \ast (X(\pi_1), \o)) \ast (X(\pi_j), \o) = (Y(\tau), \bullet) \ast ((X(\pi_1), \o) \ast (X(\pi_j), \o))\),

(A3) \(((Y(\tau), \bullet) \ast (Y(\tau_j), \bullet)) \ast (X(\pi), \o) = (Y(\tau), \bullet) \ast ((Y(\tau_j), \bullet) \ast (X(\pi), \o))\),

(A4) \(((Y(\tau), \bullet) \ast (Y(\tau_j), \bullet)) \ast (Y(\tau_k), \bullet) = (Y(\tau), \bullet) \ast ((Y(\tau_j), \bullet) \ast (Y(\tau_k), \bullet))\).

(A1) is clear by the fact that \(\mathcal{F}(\hat{G})\) is a fusion rule algebra.

(A2) For \(\tau \in \hat{G}_0\) and \(\pi_1, \pi_j \in \hat{G}\),
\[
((Y(\tau), \bullet) \ast (X(\pi_1), \o)) \ast (X(\pi_j), \o) = (Y(\tau) \ast \text{res}_{G_0}^G(X(\pi_1), \bullet)) \ast (X(\pi_j), \o) = (Y(\tau) \ast (\text{res}_{G_0}^G(X(\pi_1))) \ast (\text{res}_{G_0}^G(X(\pi_j))), \bullet).
\]

On the other hand,
\[
(Y(\tau), \bullet) \ast ((X(\pi_1), \o) \ast (X(\pi_j), \o)) = (Y(\tau), \bullet) \ast (X(\pi_1) \ast X(\pi_j), \o) = (Y(\tau) \ast \text{res}_{G_0}^G(X(\pi_1) \ast X(\pi_j)), \bullet) = (Y(\tau) \ast (\text{res}_{G_0}^G(X(\pi_1))) \ast (\text{res}_{G_0}^G(X(\pi_j))), \bullet).
\]

(A3) For \(\tau_1, \tau_j \in \hat{G}_0\) and \(\pi \in \hat{G}\),
\[
((Y(\tau_1), \bullet) \ast (Y(\tau_j), \bullet)) \ast (X(\pi), \o) = (\text{ind}_{G_0}^G(Y(\tau_1) \ast Y(\tau_j)), \o) \ast (X(\pi), \o)
\]

and applying the condition (1) of the definition of an admissible pair,
\[(\text{ind}_{G_0}^G(Y(\tau_i) * Y(\tau_j)) * \text{res}_{G_0}^G X(\pi)), \circ)\]
\[= (\text{ind}_{G_0}^G((Y(\tau_i) * Y(\tau_j)) * \text{res}_{G_0}^G X(\pi)), \circ)\]
\[= (\text{ind}_{G_0}^G(Y(\tau_i) * (Y(\tau_j) * \text{res}_{G_0}^G X(\pi))), \circ)\]
\[= (Y(\tau_i), \bullet) * (Y(\tau_j) * \text{res}_{G_0}^G X(\pi), \bullet)\]
\[= (Y(\tau_i), \bullet) * ((Y(\tau_j), \bullet) * (X(\pi), \circ)).\]

(A4) For \(\tau_i, \tau_j, \tau_k \in \hat{G}_0\)
\[((Y(\tau_i), \bullet) * (Y(\tau_j), \bullet)) * (Y(\tau_k), \bullet)\]
\[= (\text{ind}_{G_0}^G(Y(\tau_i) * Y(\tau_j)), \circ) * (Y(\tau_k), \bullet)\]
\[= (\text{res}_{G_0}^G(\text{ind}_{G_0}^G(Y(\tau_i) * Y(\tau_j))) * Y(\tau_k), \bullet),\]

and applying the condition (2) of the definition of an admissible pair,
\[= (\text{res}_{G_0}^G(\text{ind}_{G_0}^G Y(\tau_0)) * Y(\tau_i) * Y(\tau_j) * Y(\tau_k), \bullet)\]
This implies the associativity:
\[((Y(\tau_i), \bullet) * (Y(\tau_j), \bullet)) * (Y(\tau_k), \bullet) = (Y(\tau_i), \bullet) * ((Y(\tau_j), \bullet) * (Y(\tau_k), \bullet)).\]

The other conditions of a fusion rule algebra are easy to check.

On the other hand, if \(F(M^G \subset M^{G_0})\) is a fusion rule algebra, the associativity relations (A1), (A2), (A3) and (A4) hold. We obtain the conditions (1) and (2) of the admissible pair from the associativity relations (A3) and (A4).

6. Examples

**Example 6.1.** \(G = \mathbb{Z}_2 = \{e, g\}, G_0 = \{e\}, \hat{G} = \{\pi_0, \pi_1\}, \hat{G}_0 = \{\tau_0\}, M = M(2, \mathbb{C}).\)

The action \(\alpha\) is defined by
\[\alpha_g(x) = \lambda_g x \lambda_g^*\]
for \(x \in M\) and \(g \in G\), where \(\lambda\) is the regular representation of \(\mathbb{Z}_2\), namely
\[\lambda_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \lambda_g = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.\]
For \(x = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M(2, \mathbb{C})\), we obtain \(\alpha_g(x) = \begin{pmatrix} d & c \\ b & a \end{pmatrix}\). It is easy to see that
\[u(\pi_0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, u(\pi_1) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } u(\tau_0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.\]
Then
\[A = X(\pi_0) = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{C} \right\}.\]
By \(X(\pi_1) = A \cdot u(\pi_1)\), we obtain
\[X(\pi_1) = \left\{ \begin{pmatrix} c & -d \\ d & -c \end{pmatrix} : c, d \in \mathbb{C} \right\}.\]
It is trivial that \( Y(\tau_0) = M \). Moreover,
\[
F(M, G, \alpha) = \{ X(\pi_0), X(\pi_1) \}, \quad F(M, G_0, \alpha) = \{ Y(\tau_0) \},
\]
and
\[
res_{G_0}^G X(\pi_0) = Y(\tau_0), \quad res_{G_0}^G X(\pi_1) = Y(\tau_0) \quad \text{and} \quad \text{ind}_{G_0}^G Y(\tau_0) = X(\pi_0) + X(\pi_1) = M.
\]

Frobenius diagram of the inclusion \( G \supset G_0 \) is Dynkin diagram of type \( A_3 \):\[
\begin{array}{c}
X(\pi_0) \quad X(\pi_1) \\
\bigcirc \quad \bigcirc \\
\bigcirc \\
F(M, G, \alpha)
\end{array}
\quad Y(\tau_0) = M
\]
\[
\begin{array}{c}
F(M, G_0, \alpha)
\end{array}
\]
\[
F(M^G \subset M^{G_0}) = \{ (X(\pi_0), \circ), (X(\pi_1), \circ), (Y(\tau_0), \bullet) \}, \quad \text{where} \ (X(\pi_0), \circ) \quad \text{is unit of} \quad F(M^G \subset M^{G_0}). \quad \text{Then}
\]
\[
(X(\pi_1), \circ) * (X(\pi_1), \circ) = (X(\pi_0), \circ),
\]
\[
(X(\pi_1), \circ) * (Y(\tau_0), \bullet) = (Y(\tau_0), \bullet) * (X(\pi_1), \circ) = (Y(\tau_0), \bullet),
\]
\[
(Y(\tau_0), \bullet) * (Y(\tau_0), \bullet) = (X(\pi_0), \circ) + (X(\pi_1), \circ).
\]

We remark that \( F(M^{Z_2} \subset M) \cong \mathbb{Z}_2 \vee F \mathbb{Z}_2 \) where \( \vee F \) is a fusion rule algebra join. The above Frobenius diagram coincides with the principle graph of the inclusion of \( M^G \subset M \) by an outer action \( \alpha \) on a factor \( M \).

We have \( K(M^{Z_2} \subset M) = \{ c(\pi_0), c(\pi_1), c(\tau_0) \} \), where \( c(\pi_0) = (X(\pi_0), \circ), \)
\( c(\pi_1) = (X(\pi_1), \circ) \) and \( c(\tau_0) = \frac{1}{2}(Y(\tau_0), \bullet) \). Then \( K(M^{Z_2} \subset M) \) is a hypergroup isomorphic with \( K(\mathbb{Z}_2 \cup \{ e \}) = \mathbb{Z}_2 \vee \mathbb{Z}_2 \).

**Example 6.2.** \( G = \mathbb{Z}_3 = \{ e, g, g^2 \}, \ G_0 = \{ e \}, \ \widehat{G} = \{ \pi_0, \pi_1, \pi_2 \}, \ \widehat{G}_0 = \{ \tau_0 \}, \ M = M(3, \mathbb{C}) \).

In a similar way to the above, we obtain
\[
u(\pi_0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \nu(\pi_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix},
\]
\[
u(\pi_2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix}, \quad \nu(\tau_0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},
\]
\[
A = X(\pi_0) = \left\{ \begin{pmatrix} a_1 & a_3 & a_2 \\ a_2 & a_1 & a_3 \\ a_3 & a_2 & a_1 \end{pmatrix} : a_1, a_2, a_3 \in \mathbb{C} \right\},
\]
The Frobenius diagram of $G$ and $F$ is given by

$$X(\pi_1) = \left\{ \begin{pmatrix} b_1 & \omega b_3 & \omega^2 b_2 \\ b_2 & \omega b_1 & \omega^2 b_3 \\ b_3 & \omega b_2 & \omega^2 b_1 \end{pmatrix} : b_1, b_2, b_3 \in \mathbb{C} \right\},$$

$$X(\pi_2) = \left\{ \begin{pmatrix} c_1 & \omega^2 c_3 & \omega c_2 \\ c_2 & \omega^2 c_1 & \omega c_3 \\ c_3 & \omega^2 c_2 & \omega c_1 \end{pmatrix} : c_1, c_2, c_3 \in \mathbb{C} \right\},$$

$$Y(\tau_0) = M.$$

$${\mathcal{F}}(M, G, \alpha) = \{X(\pi_0), X(\pi_1), X(\pi_2)\}, \quad {\mathcal{F}}(M, G_0, \alpha) = \{Y(\tau_0)\}.$$

$$\text{res}_{G_0}^G X(\pi_0) = Y(\tau_0), \quad \text{res}_{G_0}^G X(\pi_1) = Y(\tau_0), \quad \text{res}_{G_0}^G X(\pi_2) = Y(\tau_0),$$

$$\text{ind}_{G_0}^G Y(\tau_0) = X(\pi_0) + X(\pi_1) + X(\pi_2),$$

and

$$M = X(\pi_0) + X(\pi_1) + X(\pi_2).$$

The Frobenius diagram of $G \supset G_0$ is a Dynkin diagram of type $D_4$:

$$\begin{array}{ccc}
X(\pi_0) & X(\pi_1) & X(\pi_2) \\
\mathcal{F}(M, G, \alpha) & & \mathcal{F}(M, G_0, \alpha) \\
\mathcal{F}(M, G, \alpha) & & \mathcal{F}(M, G_0, \alpha) \\
& & Y(\tau_0) = M
\end{array}$$

Here $\mathcal{F}(M^G \subset M^{G_0}) = \{(X(\pi_0), \circ), (X(\pi_1), \circ), (X(\pi_2), \circ), (Y(\tau_0), \bullet)\}$, where $(X(\pi_0), \circ)$ is unit of $\mathcal{F}(M^G \subset M^{G_0})$.

$$(X(\pi_1), \circ) * (X(\pi_1), \circ) = (X(\pi_2), \circ),$$

$$(X(\pi_2), \circ) * (X(\pi_2), \circ) = (X(\pi_1), \circ),$$

$$(X(\pi_1), \circ) * (X(\pi_2), \circ) = (X(\pi_0), \circ),$$

$$(X(\pi_1), \circ) * (Y(\tau_0), \bullet) = (X(\pi_2), \circ) * (Y(\tau_0), \bullet) = (Y(\tau_0), \bullet),$$

$$(Y(\tau_0), \bullet) * (Y(\tau_0), \bullet) = (X(\pi_0), \circ) + (X(\pi_1), \circ) + (X(\pi_2), \circ).$$

We remark that $\mathcal{F}(M^{Z_2} \subset M) \cong \mathbb{Z}_3 \lor \mathbb{Z}_2$ and $K(M^{Z_3} \subset M) \cong \mathbb{Z}_3 \lor \mathbb{Z}_2$.

**Example 6.3.** $G = S_3 = \mathbb{Z}_3 \rtimes \alpha \mathbb{Z}_2 = \{h_0, h_1, h_2, g, h_1 g, h_2 g\}$, where $\mathbb{Z}_3 = \{h_0, h_1, h_2\}$ and $\mathbb{Z}_2 = \{e, g\}$, $G_0 = \mathbb{Z}_2$, $\hat{G} = \{\pi_0, \pi_1, \pi_2\}$ (dim $\pi_0 = \dim \pi_1 = 1$, dim $\pi_2 = 2$), $G_0^* = \{\tau_0, \tau_1\}$ and $M = M(6, \mathbb{C}) = B(\ell^2(\hat{G}))$.

Let $\lambda$ be the regular representation of $S_3$ which is given by
BIMODULES AND HYPERGROUPS

\[
\lambda_{h_1} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & \omega & 0 & 0 & 0 & 0 \\
0 & 0 & \omega^2 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & \omega & 0 \\
0 & 0 & 0 & 0 & 0 & \omega^2
\end{pmatrix}, \quad \lambda_g = \begin{pmatrix}
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

For \( x = (x_{ij}) \in M \),

\[
\alpha_{h_1}(x) = \begin{pmatrix}
x_{11} & \omega^2 x_{12} & \omega x_{13} & x_{14} & \omega^2 x_{15} & \omega x_{16} \\
\omega x_{21} & x_{22} & \omega^2 x_{23} & \omega x_{24} & x_{25} & \omega x_{26} \\
\omega^2 x_{31} & \omega x_{32} & x_{33} & \omega^2 x_{34} & \omega x_{35} & x_{36} \\
x_{41} & \omega^2 x_{42} & \omega x_{43} & x_{44} & \omega^2 x_{45} & \omega x_{46} \\
\omega x_{51} & x_{52} & \omega^2 x_{53} & \omega x_{54} & x_{55} & \omega x_{56} \\
\omega^2 x_{61} & \omega x_{62} & x_{63} & \omega^2 x_{64} & \omega x_{65} & x_{66}
\end{pmatrix},
\]

\[
\alpha_g(x) = \begin{pmatrix}
x_{44} & x_{46} & x_{45} & x_{41} & x_{43} & x_{42} \\
x_{64} & x_{66} & x_{65} & x_{61} & x_{63} & x_{62} \\
x_{54} & x_{56} & x_{55} & x_{51} & x_{53} & x_{52} \\
x_{14} & x_{16} & x_{15} & x_{11} & x_{13} & x_{12} \\
x_{34} & x_{36} & x_{35} & x_{31} & x_{33} & x_{32} \\
x_{24} & x_{26} & x_{25} & x_{21} & x_{23} & x_{22}
\end{pmatrix}
\]

Then we see that

\[
u(\pi_0) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}, \quad \nu(\pi_1) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
\]

\[
u(\pi_2) = \begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0
\end{pmatrix},
\]

\[
u(\tau_0) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}, \quad \nu(\tau_1) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}.
\]
Moreover, it is easy to see that

\[
V(\pi_2) = \left\{ \begin{pmatrix}
0 & b & a & 0 & b & a \\
 a & 0 & b & a & 0 & b \\
 b & a & 0 & b & a & 0 \\
 0 & b & a & 0 & b & a \\
a & 0 & b & a & 0 & b \\
b & a & 0 & b & a & 0
\end{pmatrix} : a, b \in \mathbb{C} \right\},
\]

\[
A = X(\pi_0) = \left\{ \begin{pmatrix}
a_1 & 0 & 0 & a_4 & 0 & 0 \\
 0 & a_2 & 0 & 0 & a_5 & 0 \\
 0 & 0 & a_3 & 0 & 0 & a_6 \\
a_4 & 0 & 0 & a_1 & 0 & 0 \\
 0 & a_5 & 0 & 0 & a_2 & 0 \\
 0 & 0 & a_6 & 0 & 0 & a_3 \\
\end{pmatrix} : a_i \in \mathbb{C} \right\},
\]

\[
X(\pi_1) = \left\{ \begin{pmatrix}
b_1 & 0 & 0 & -b_4 & 0 & 0 \\
 0 & b_2 & 0 & 0 & -b_5 & 0 \\
 0 & 0 & b_3 & 0 & 0 & -b_6 \\
b_4 & 0 & 0 & -b_1 & 0 & 0 \\
 0 & b_5 & 0 & 0 & -b_2 & 0 \\
 0 & 0 & b_6 & 0 & 0 & -b_3 \\
\end{pmatrix} : b_i \in \mathbb{C} \right\},
\]

\[
X(\pi_2) = A \cdot V(\pi_2) = \left\{ \begin{pmatrix}
0 & c_3 & c_5 & 0 & c_9 & c_{11} \\
c_1 & 0 & c_6 & c_7 & 0 & c_{12} \\
c_2 & c_4 & 0 & c_8 & c_{10} & 0 \\
 0 & c_9 & c_{11} & 0 & c_3 & c_5 \\
 c_7 & 0 & c_{12} & c_1 & 0 & c_6 \\
c_8 & c_{10} & 0 & c_2 & c_4 & 0
\end{pmatrix} : c_i \in \mathbb{C} \right\},
\]

\[
X_1(\pi_2) = \left\{ \begin{pmatrix}
0 & d_3 & d_5 & 0 & -d_9 & -d_{11} \\
d_1 & 0 & d_6 & -d_7 & 0 & -d_{12} \\
d_2 & d_4 & 0 & -d_8 & -d_{10} & 0 \\
 0 & d_9 & d_{11} & 0 & -d_3 & -d_5 \\
d_7 & 0 & d_{12} & -d_1 & 0 & -d_6 \\
d_8 & d_{10} & 0 & -d_2 & -d_4 & 0
\end{pmatrix} : d_i \in \mathbb{C} \right\}.
\]

Here we note that \(X_1(\pi_2) \cong X(\pi_2)\) as \((A,A,G)\)-module and

\[
M = X(\pi_0) \oplus X(\pi_1) \oplus X(\pi_2) \oplus X_1(\pi_2),
\]

\[
Y(\pi_0) = \left\{ \begin{pmatrix}
e_1 & e_4 & e_7 & e_{10} & e_{16} & e_{13} \\
e_2 & e_5 & e_8 & e_{12} & e_{18} & e_{15} \\
e_3 & e_6 & e_9 & e_{11} & e_{17} & e_{14} \\
e_{10} & e_{13} & e_{16} & e_1 & e_7 & e_4 \\
e_{11} & e_{14} & e_{17} & e_3 & e_9 & e_6 \\
e_{12} & e_{15} & e_{18} & e_2 & e_8 & e_5
\end{pmatrix} : e_i \in \mathbb{C} \right\}.
\]
We see that
\[ \mathcal{F}(M, G, \alpha) = \{ X(\pi_0), X(\pi_1), X(\pi_2) \}, \quad \mathcal{F}(M, G_0, \alpha) = \{ Y(\tau_0), Y(\tau_1) \}, \]
and
\[ \text{res}^{G_0} X(\pi_0) = Y(\tau_0), \quad \text{res}^{G_0} X(\pi_1) = Y(\tau_1), \quad \text{res}^{G_0} X(\pi_2) = Y(\tau_0) + Y(\tau_1), \]
\[ \text{ind}^{G_0} Y(\tau_0) = X(\pi_0) + X(\pi_2) \quad \text{and} \quad \text{ind}^{G_0} Y(\tau_1) = X(\pi_1) + X(\pi_2). \]
We note that
\[ M = X(\pi_0) \oplus X(\pi_1) \oplus X(\pi_2) \oplus X_1(\pi_2) \cong X(\pi_0) + X(\pi_1) + 2X(\pi_2). \]
The Frobenius diagram of \( G \supset G_0 \) is a Dynkin diagram of type \( A_5 \):
\[ \mathcal{F}(M, G_0, \alpha) \]
\[ \mathcal{F}(M, G_0, \alpha) \]
\[ \mathcal{F}(M, G_0, \alpha) \]
\[ \mathcal{F}(M, G_0, \alpha) \]
Now \( \mathcal{F}(M^S \subset M^{S_2}) = \{ (X(\pi_0), \circ), (X(\pi_1), \circ), (X(\pi_2), \circ), (Y(\tau_0), \bullet), (Y(\tau_1), \bullet) \} \),
where \((X(\pi_0), \circ)\) is unit of \( \mathcal{F}(M^S \subset M^{S_2}) \). We remark that \( \mathcal{F}(M^S \subset M^{S_2}) \cong \mathcal{F}(\hat{S}_3 \cup \hat{Z}_2) \), refer to [20] and \( \mathcal{K}(M^S \subset M^{S_2}) \cong \mathcal{K}(\hat{S}_3 \cup \hat{Z}_2) \), refer to [10].

**Example 6.4.** \( G = S_3, \ G_0 = Z_2, \ G = \{ \pi_0, \pi_1, \pi_2 \}, \ G_0 = \{ \tau_0, \tau_1 \}, \ M = M(2, \mathbb{C}). \)

We define an action action \( \alpha \) of \( S_3 \) on \( M(2, \mathbb{C}) \) by
\[ \alpha_s(x) = \pi_2(s)x\pi_2(s)^* \ (s \in S_3). \]
It is easy to see that
\[ u(\pi_0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ u(\pi_1) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ u(\pi_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \]
\[ u(\tau_0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ u(\tau_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \]
\[ V(\pi_2) = \left\{ \begin{pmatrix} a & b \\ b & 0 \end{pmatrix} : a, b \in \mathbb{C} \right\}. \]
Moreover, we have
\[ A = X(\pi_0) = \left\{ \begin{pmatrix} a_1 & 0 \\ 0 & a_1 \end{pmatrix} : a_1 \in \mathbb{C} \right\}, \ X(\pi_1) = \left\{ \begin{pmatrix} a_2 & 0 \\ 0 & -a_2 \end{pmatrix} : a_2 \in \mathbb{C} \right\}, \]
Counterexample 1

\[ X(\pi_2) = A \cdot V(\pi_2) = \left\{ \begin{pmatrix} 0 & a_4 \\ a_3 & 0 \end{pmatrix} : a_3, a_4 \in \mathbb{C} \right\}, \]

\[ Y(\tau_0) = \left\{ \begin{pmatrix} b_1 & b_2 \\ b_2 & b_1 \end{pmatrix} : b_1, b_2 \in \mathbb{C} \right\}, \]
\[ Y(\tau_1) = \left\{ \begin{pmatrix} b_3 & -b_4 \\ b_4 & -b_3 \end{pmatrix} : b_3, b_4 \in \mathbb{C} \right\}, \]

\[ \mathcal{F}(M, G, \alpha) = \{ X(\pi_0), X(\pi_1), X(\pi_2) \}, \]
\[ \mathcal{F}(M, G_0, \alpha) = \{ Y(\tau_0), Y(\tau_1) \}, \]

\[ \text{res}_{G_0}^G X(\pi_0) = Y(\tau_0), \quad \text{res}_{G_0}^G X(\pi_1) = Y(\tau_1), \quad \text{res}_{G_0}^G X(\pi_2) = Y(\tau_0) + Y(\tau_1), \]

\[ \text{ind}_{G_0}^G Y(\tau_0) = X(\pi_0) + X(\pi_2), \quad \text{ind}_{G_0}^G Y(\tau_1) = X(\pi_1) + X(\pi_2) \]

and

\[ M = X(\pi_0) + X(\pi_1) + X(\pi_2). \]

The Frobenius diagram of \( G \supset G_0 \) is also a Dynkin diagram of type \( A_5 \):

\[ \begin{array}{ccc}
X(\pi_0) & X(\pi_2) & X(\pi_1) \\
\mathcal{F}(M, G, \alpha) & & \\
\mathcal{F}(M, G_0, \alpha) & & \\
\end{array} \]

**Counterexample 1** \( G = S_3, \ G = \{ \pi_0, \pi_1, \pi_2 \}, M = M(3, \mathbb{C}) \).

We define an action \( \alpha \) of \( S_3 \) on \( M(3, \mathbb{C}) \) by

\[ \alpha_s(x) := \pi(s)x\pi(s)^* \quad (x \in M) \]

where \( \pi = \pi_0 \oplus \pi_2 \). This action \( \alpha \) does not have dual unitary property and \( \mathcal{F}(M, S_3, \alpha) \) does not have a structure of fusion rule algebra.

**Counterexample 2** \( G = S_3 \times \mathbb{Z}_2, \ G_0 = \mathbb{Z}_2, \ M = M(6, \mathbb{C}), \alpha = \text{Ad}\lambda \).

Since the pair \((G, G_0)\) is not admissible, \( \mathcal{F}(M^G \subset M^{G_0}) \), does not have a structure of a fusion rule algebra by Theorem 5.4.

References


Satoshi Kawakami: Department of Mathematics, Nara University of Education, Takahatake-cho, Nara 630-8528, JAPAN
E-mail address: kawakami@nara-edu.ac.jp

Tatsuya Tsurii: Center for New Education of Science and Mathematics, Nara University of Education, Takahatake-cho, Nara 630-8528, JAPAN
E-mail address: tatsuya_tsurii@nara-edu.ac.jp

Shigeru Yamagami: Graduate School of Mathematics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, JAPAN
E-mail address: yamagami@math.nagoya-u.ac.jp
A STUDY IN LOCALLY COMPACT GROUPS
—CHABAUTY SPACE, SYLOW THEORY,
THE SCHUR-ZASSENHAUS FORMALISM,
THE PRIME GRAPH FOR NEAR ABELIAN GROUPS

WOLFGANG HERFORT, KARL H. HOFMANN, AND FRANCESCO G. RUSSO

Abstract. The class of locally compact near abelian groups is introduced and investigated as a class of metabelian groups formalizing and applying the concept of scalar multiplication. The structure of locally compact near abelian groups and its close connections to prime number theory are discussed and elucidated by graph theoretical tools. These investigations require a thorough reviewing and extension to the present circumstances of various aspects of the general theory of locally compact groups such as
– the Chabauty space of closed subgroups with its natural compact Hausdorff topology,
– a very general Sylow subgroup theory for periodic groups including their Hall systems,
– the scalar automorphisms of locally compact abelian groups,
– the theory of products of closed subgroups and their relation to semidirect products, and
– inductively monothetic groups are introduced and classified.
As applications, firstly, a complete classification is given of locally compact topologically quasihamiltonian groups, which has been initiated by F. Kümmich, and, secondly, Yu. Mukhin’s classification of locally compact topologically modular groups is retrieved and further illuminated.

1. Background

In this survey we describe what may be called a structure theory of locally compact near abelian groups. This attempt becomes clear after we describe the historical development of researching locally compact groups in a broad sense.

Keys to our understanding will be the concept of an inductively monothetic locally compact group, the Chabauty space associated canonically with a locally compact group, and a Sylow theory for closed subgroups of periodic locally compact groups reflecting typical features of the Sylow theory of finite groups. The emergence of the role of the prime numbers attached to the building blocks of...
the groups we consider points into the direction of graph theory which we shall employ rather extensively. However, as interesting as the structure theory of near abelian groups itself is, we emphasise here, that the machinery for developing and detailing it should share center stage.

1.1. Some history. The structure theory of locally compact groups has a long history. One of its roots is David Hilbert’s question of 1900 whether a locally euclidean topological group might possibly support the introduction of a differentiable parametrisation such that the group operations are in fact differentiable.

A first affirmative answer was given for compact locally euclidean groups when they were found to be real matrix groups as a consequence of the foundational work by Hermann Weyl and his student Fritz Peter in 1927 on the representation theory of compact groups. A second step was achieved when the answer was found to be “yes” for commutative locally compact groups. This emerged out of the fundamental duality theorems by Lev Semjonovich Pontryagin (1934) and Egbert van Kampen (1937). This duality forever determined the structure and harmonic analysis of locally compact abelian groups after it was widely read in the book of 1940, 1953, 1965 and 1979 by André Weil on the integration in locally compact groups [47].

A final positive answer to Hilbert’s Fifth Problem had to wait almost another two decades when, 17 years after the Second World War, the contributions of Andrew Mattei Gleason, Deane Montgomery, and Leo Zippin around 1952 provided the final affirmative answer to Hilbert’s problem. It led almost at once to the fundamental insights of Hidehiko Yamabe 1953, completing the pioneering work of Kenkichi Iwasawa (1949) providing the fundamental structure of all those locally compact groups $G$ which had a compact space $G/G_0$ of connected components: Such groups were recognised as being approximated by quotient groups $G/N$ modulo arbitrarily small compact normal subgroups $N$ in such a fashion that each $G/N$ is a Lie group, that is, one of those groups on which Hilbert had focussed in the fifth of his 23 influential problems in 1900 and which Sophus Marius Lie (1842-1899) had invented together with an ingenious algebrasiation method, long known nowadays under the name of Lie algebra theory. (S. [35], [29]) A special case arises when all $G/N$ are discrete finite groups; in this case $G$ is called profinite.

The solution of Hilbert’s Fifth Problem in the middle of last century opened up the access to the structure theory of locally compact groups to the extent they could be approximated by Lie groups, due to the rich Lie theory meanwhile developed in algebra, geometry, and functional analysis. Recently, interest in Hilbert’s Fifth Problem was rekindled in the present century under the influence of Terence Tao [46]. There are nonstandard approaches to dealing with Hilbert’s Fifth Problem by Joram Hirschfeld in [18] and recently by Lou van den Dries and Isaac Goldbring in [6].

The quest for a solution to Hilbert’s Fifth problem, at any rate, led to one major direction in the research of topological groups: in focus was the class of groups $G$ approximable by Lie group quotients $G/N$, and finally $G$ itself needed no longer to be locally compact. Such groups were called pro-Lie groups considered for the
sake of their own. (See [21], [22].) Their theory reached as far as almost connected locally compact groups go (that is, those for which $G/G_0$ is compact), including all compact and all connected locally compact groups. But not further.

Still, every locally compact group (and every pro-Lie group) $G$ has canonically and functorially attached to it a (frequently infinite dimensional) Lie algebra $\mathfrak{L}(G)$ and therefore a cardinal $\geq 0$ attached to it, namely, its topological dimension $\text{DIM} G$ (cf. e.g. [20], 9.54, p. 498ff.). Indeed, we have the following information right away:

**Proposition 1.1.** Let $G$ be a topological group which is locally compact or a pro-Lie group. Then the following statements are equivalent:

1. Every connected component of $G$ is singleton, that is, $G$ is totally disconnected.
2. $\text{DIM} G = 0$, that is, $G$ is zero-dimensional.

If $G$ is compact, then it is zero-dimensional if and only if it is profinite.

For this survey it is important to see that, in the 20th century, there was a second trend on the study of locally compact groups that is equally significant even though it is opposite to the concept of connectivity in topological groups. This trend is represented by the class of compact or locally compact zero-dimensional groups.

Such groups were encountered in field theory at an early stage. Indeed, in Galois theory the consideration of the appropriate infinite ascending family of finite Galois extensions and, finally, its union would, dually, lead to an inverse family of finite Galois groups and, in the end, to their projective limit. Thus was produced what became known as a profinite group. The Galois group of the infinite field extension, equipped with the Krull topology, is thus a profinite group. One recognised soon that profinite groups and compact totally disconnected groups were one and the same mathematical object, expressed algebraically on the one hand and topologically on the other. Comprehensive literature on this class of groups appeared much later than text books on topological groups in which connected components played a leading role. Just before the end of the 20th century totally disconnected compact groups were the protagonists of books simultaneously entitled “Profinite Groups” by John Stuart Wilson and Luis Ribes jointly with Pavel Zalesskii in 1998, while George Willis in 1994 laid the foundations of a general structure theory of totally disconnected locally compact groups if no additional algebraic information about them is available.

On the other hand, in the realm of locally compact abelian groups, the completion of the field $\mathbb{Q}$ of rational numbers with respect to any nonarchimedian valuation yields the locally compact $p$-adic fields $\mathbb{Q}_p$ as a totally disconnected counterpart of the connected field $\mathbb{R}$ of real numbers. The fields $\mathbb{Q}_p$ and their integral subrings $\mathbb{Z}_p$ were basic building blocks of ever so many totally disconnected groups, in particular the linear groups over these field and indeed all $p$-adic Lie groups which Nicolas Bourbaki judiciously included in his comprehensive treatise on Lie groups. Bourbaki’s text on Lie groups formed the culmination and certainly the endpoint of his encyclopaedic project extending over several decades.

The world of totally disconnected locally compact groups developed its own existence, methods and philosophies, partly deriving from finite group theory via
approximation through the formation of projective limits, partly through graph theory where the appropriate automorphism theory provides the fitting representation theory, and partly also through the general impact of algebraic number theory which in the text book literature is indicated by the books of HELMUT HASSE since the thirties (see e.g. [13, 12] and lastly, 1967, by the “Basic Number Theory” of ANDRÉ WEIL, see [48], whose book on the integration on locally compact groups of 1940 (with later editions through four decades) had influenced the progress of harmonic analysis of locally compact groups so much.

We pointed out that each locally compact group $G$, irrespective of any structural assumption has attached to it a (topological) Lie algebra $\mathfrak{L}(G)$ (and therefore a universal dimension). The more recent interest in zero-dimensional locally compact groups $G$ has led to a new focus on another functorially attached invariant, namely, a compact Hausdorff space $\text{SUB}(G)$ consisting of all closed subgroups of $G$ endowed with a suitable topology and now frequently called the Chabauty space of $G$. This tool is not exactly new, but has been widely utilised in applications recently. The Chabauty space is a special case of what has been called the hyperspace of a compact (or locally compact) space first introduced by VIEOTORIS (s. [7]) in topological algebra hyperspaces were used and described e.g. in [2], [23], [24], and in all recent publications where the name of Chabauty appears in the title (e.g. [8], [10], [9], and [11]).

**Notation.** We shall make use of notation coherent with the book of Hofmann and Morris, [20]. The cyclic group of order $p^n$ is denoted by $\mathbb{Z}(p^n)$ and $\mathbb{Z}(p^0)$ stands for the trivial abelian group $\{0\}$.

## 2. Introductory Definitions and Results

We now return to the second thrust of the study of locally compact groups which is concerned with the research of 0-dimensional groups.

**Definition 2.1.** A topological group $G$ is called periodic if

(i) $G$ is locally compact and totally disconnected, and
(ii) $\langle g \rangle$ is compact for all $g \in G$.

So a compact group is periodic if and only if it is profinite. A very significant portion of the locally compact groups considered here will be periodic groups. That is, we deal with totally disconnected locally compact groups in which every element is contained in a profinite subgroup. We shall say that a topological group $G$ is compactly ruled if it is the directed union of its compact open subgroups. If $G$ is a a locally compact solvable group in which every element is contained in a compact subgroup, then it is compactly ruled. The class of compactly ruled groups comprises both, the class of profinite groups and the one of locally finite groups, i.e. groups where every finite subset generates a finite subgroup only, see [30]. In many cases we assume that the periodic groups we consider are compactly ruled. These properties make them topologically special; just how close they make our groups to profinite groups remains to be seen in the course of this survey. A second significant property of the groups we study is an algebraic one: they are solvable,
indeed metabelian. Again, it is another challenge to discern just how close this makes them to abelian groups. The groups we study will be called near abelian.

In order to offer a precise definition of this class of locally compact groups we need one preliminary definition, extending a very familiar concept:

**Definition 2.2.** A topological group $G$ will be called monothetic if $G = \langle g \rangle$ for some $g \in G$, and inductively monothetic if for every finite subset $F \subseteq G$ there is an element $g \in G$ such that $\langle F \rangle = \langle g \rangle$.

We shall discuss and classify inductively monothetic locally compact groups in greater detail later; but let us observe here right away a connected example illustrating the two definitions: Indeed let $T = \mathbb{R}/\mathbb{Z}$ denote the (additively written) circle group. Then

- the 2-torus is monothetic but is not inductively monothetic, since $(\frac{1}{2}\mathbb{Z}/\mathbb{Z})^2 \subseteq T^2$ is finitely generated but is not monothetic.

Yet in the domain of totally disconnected locally compact groups

- every 0-dimensional monothetic group is inductively monothetic.

In a periodic group, each monothetic subgroup $\langle g \rangle$ is compact, equivalently, procyclic.

Now we are prepared for a definition of the class of locally compact groups whose details we shall consider here:

**Definition 2.3.** A topological group $G$ is near abelian provided it is locally compact and contains a closed abelian subgroup $A$ such that

1. $G/A$ is an abelian inductively monothetic group, and
2. every closed subgroup of $A$ is normal in $G$.

The subgroup $A$ we shall call a base for $G$.

When we eventually collect applications for this class of locally compact groups, then we shall see that for instance all locally compact groups in which two closed subgroups commute setwise form a subclass of the class of near abelian groups and that the class of all locally compact groups in which the lattice of closed subgroups is modular is likewise a subclass of the class of near abelian groups.

### 2.1. Some history of near Abelian groups.

In the world of discrete groups, near abelian groups historically appeared in a natural way when K. Iwasawa attempted the classification what is now known as quasihamiltonian and modular groups as expounded in the monograph by R. Schmidt, [44]). It was F. Kümmich (cf. [31]) who initiated in his dissertation written under the direction of Peter Plaumann and in papers developed from his thesis the study of topologically quasihamiltonian groups. These are topological groups such that $XY = YX$ is valid for any closed subgroups $X$ and $Y$ of such a group. A bit later Yu. Mukhin turned to investigating the class of locally compact topologically modular groups (cf. e.g. [37, 39]).

The properties that there be a closed normal abelian subgroup $A$ of $G$ such that $G/A$ is inductively monothetic and such that every closed subgroup of $A$ is normal in $G$ suggest themselves by the fact, proved by K. Iwasawa, that discrete quasihamiltonian groups satisfy them. In a similar vein, Mukhin, during his work
on classifying topologically modular groups, finds that these groups are all near abelian in our sense (see e.g. [38]).

An earlier article by K. H. Hofmann and F. G. Russo, was devoted to classifying compact \( p \)-groups that are topologically quasihamiltonian (cf. [25]). The major result states that such groups are at the same time topologically quasihamiltonian and near abelian with the exception of \( p = 2 \) in which case some sporadic near abelian groups are topologically quasihamiltonian while the bulk of them are not. This once again is evidence of the fact that is often quoted by number theorists and group theorists alike that 2 is the oddest of all primes.

In linear algebra, a group \( G \) of \((n + 1) \times (n + 1)\)-matrices of the form

\[
\begin{pmatrix}
  r \cdot E_n & v \\
  0 & 1
\end{pmatrix}, \quad 0 < r \in \mathbb{R}, \ v \in \mathbb{R}^n
\]

with the identity \( E_n \) of \( GL(n, \mathbb{R}) \) is a metabelian Lie group that has been called almost abelian (s. [17], p. 408, Example V.4.13). The subgroup \( A \) of all matrices with \( r = 1 \) is isomorphic to \( \mathbb{R}^n \) and every vector subspace of \( A \) is normal in \( G \) and \( G/A \cong \mathbb{R} \) is a one-dimensional Lie group which is not inductively monothetic, but we shall see that inductively monothetic groups are in some sense “rank one” group analogs.

In both cases we have a representation \( \psi : G/A \to Aut(A) \) such that \( \psi(gA)(a) = gag^{-1} \) as an essential element of structure. In the near abelian case we shall say that \( G \) is \( A \)-nontrivial if the image of \( \psi \) has more than 2 elements. Whereas in the Lie group case, the structure of an almost abelian Lie group \( G \) is comparatively simple, in the case of a group \( G \) satisfying the conditions of Definition 2.3 it is likely to be rather sophisticated as we illustrate by a result (s. [15], Theorem 7.4) in which \( C_G(A) = \ker \psi \) denotes the centraliser \( \{g \in G : (\forall a \in A) ag = ga\} \) of \( A \) in \( G \):

**Theorem 2.4.** (Structure Theorem I on Near Abelian Groups) Let \( G \) be an \( A \)-nontrivial near abelian group. Then

1. \( A \) is periodic.
2. \( G \) is totally disconnected.
3. When \( \psi(G/A) \) is compact or \( A \) is an open subgroup, then \( G \) has arbitrarily small compact open normal subgroups, that is, \( G \) is pro-discrete.
4. \( G \) itself is periodic if and only if \( G/A \) is periodic if and only if \( G/A \) is not isomorphic to a subgroup of the discrete group \( \mathbb{Q} \) of rational numbers.
5. \( C_G(A) \) is an abelian normal subgroup containing \( A \) and is maximal for this property.

This shows that for our topic, periodic locally compact groups play a significant role.

### 3. Inductively Monothetic Groups

A good understanding of near abelian groups depends on a clear insight into the concept of inductively monothetic groups. They were recently featured in [9].

We must recall the concept of a local product of a family of topological groups which in the theory of locally compact groups mediates between the idea of a
Tychonoff product of compact groups and the idea of a direct sum of a family of discrete groups; the principal applications are in the domain of abelian groups, but the concept as such has nothing to do with commutativity.

**Definition 3.1.** Let \((G_j)_{j \in J}\) be a family of locally compact groups and assume that for each \(j\) the group \(G_j\) has a normal compact open subgroup \(C_j\). Let \(P\) be the subgroup of the cartesian product of the \(G_j\) containing exactly those \(J\)-tuples \((g_j)_{j \in J}\) of elements \(g_j \in G_j\) for which the set \(\{ j \in J : g_j \notin C_j \}\) is finite. Then \(P\) contains the cartesian product \(C := \prod_{j \in J} C_j\) which is a compact topological group with respect to the Tychonoff topology. The group \(P\) has a unique group topology with respect to which \(C\) is an open subgroup. Now the local product of the family \(((G_j, C_j))_{j \in J}\) is the group \(P\) with this topology, and it is denoted by

\[ P = \prod_{j \in J} (G_j, C_j). \]

Let us note that the local product is a locally compact group with the compact open subgroup \(\prod_{i \in I} C_p\). While the full product \(\prod_{j \in J} G_j\) has its own product topology we note that in general the local product topology on \(P\) in general is properly finer than the subgroup topology. The concept of the local product was introduced and its duality theory in the commutative situation was studied by J. Braconnier in [3]. For us local products play a role most frequently with \(J\) being the set \(\pi\) of all prime numbers. This is well illustrated by the following key result on periodic locally compact abelian groups where we note, that for a locally compact abelian group \(G\) and each prime \(p\), we have a unique characteristic subgroup \(G_p\) containing all elements \(g\) for which \(\langle g \rangle\) is a profinite \(p\)-group; \(G_p\) is called the \(p\)-primary component or the \(p\)-Sylow subgroup of \(G\).

**Theorem 3.2.** (J. Braconnier) Let \(G\) be a periodic locally compact abelian group and \(C\) any compact open subgroup of \(G\). Then \(G\) is isomorphic to the local product

\[ \prod_{p \in \pi} (G_p, C_p). \]

The following remark is useful for us as a consequence of the fact that any compact \(p\)-group \(C_p\) is a \(\mathbb{Z}_p\)-module and any prime \(q \neq p\) is a unit in \(\mathbb{Z}_p\), whence \(C_p\) is divisible by \(n \in \mathbb{N}\) with \((n,p) = 1\):

**Remark 3.3.** A periodic locally compact abelian group \(G\) is divisible iff all \(p\)-Sylow subgroups \(G_p\) are divisible.

The structure of a locally compact monothetic group \(G\) is familiar to workers in the area: It is either isomorphic to the discrete group \(\mathbb{Z}\) of integers or is compact (Weil’s Lemma, s. e.g. [20], Proposition 7.43, p.348.). A compact abelian group is known if its discrete Pontryagin dual is known. A compact abelian group \(G\) is monothetic if and only if there is a morphism \(f: \mathbb{Z} \to G\) of locally compact groups with dense image, that is, if there is an injection of the discrete group \(\hat{G}\) into the character group \(\hat{T}\), that is, \(\hat{G}\) is isomorphic to a subgroup of \(\mathbb{Q}(2^{\infty}) \oplus \bigoplus_{p \in \pi} \mathbb{Z}(p^{\infty})\). (Here \(\mathbb{Z}(p^{\infty})\), as usual, is the Prüfer group \(\bigcup_{n \in \mathbb{N}} \frac{1}{p^n} \mathbb{Z}/\mathbb{Z} \subseteq T\).) Whenever \(G\) is zero-dimensional, things simplify dramatically:
Proposition 3.4. A compact zero-dimensional abelian group $G$ is monothetic if it is isomorphic to $\prod_{p \in \pi} G_p$ where the $p$-factor $G_p$ is either $\mathbb{Z}(p^m)$ for some $m \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ or $\mathbb{Z}_p$, the additive group of the ring of $p$-adic numbers.

Let us proceed to inductively monothetic locally compact groups. For periodic inductively monothetic groups it is convenient to introduce some special terminology. From Braconnier’s Theorem 3.2 we know that every periodic locally compact abelian group $G$, for any given compact open subgroup $C \subseteq G$, is (isomorphic to) the local product $\prod_{p \in \pi}(G_p, C_p)$ of its $p$-Sylow subgroups.

Definition 3.5. A topological group $G$ is called $\Pi$-procyclic, if it is a periodic locally compact abelian group and each $p$-Sylow subgroup $G_p$ is either a finite cyclic $p$-group (possibly singleton) or $\mathbb{Z}_p$, that is, $G_p$ is $p$-procyclic.

Now we can formulate the classification of inductively monothetic locally compact groups.

Theorem 3.6. (Classification Theorem of Inductively Monothetic Groups) Let $G$ be an inductively monothetic locally compact group. Then $G$ is either
(a) a $1$-dimensional compact connected abelian group, or
(b) a subgroup of the discrete group $\mathbb{Q}$, or
(c) a periodic locally compact abelian group such that $G_p$ is isomorphic to $\mathbb{Q}_p$, or $\mathbb{Z}(p^\infty)$, or $\mathbb{Z}_p$, or $\mathbb{Z}(p^{n_p})$ for some $n_p \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}$.

All inductively monothetic groups are sigma-compact, i.e. are countable unions of compact subsets.

The groups of connected type in Theorem 3.6 are monothetic; other types may or may not be monothetic. The periodic inductively monothetic groups $G$ of part (b) require special attention. First we divide the set $\pi$ of all prime numbers into disjoint sets
- $\pi_A = \{ p \in \pi : G_p \cong \mathbb{Q}_p \}$,
- $\pi_B = \{ p \in \pi : G_p \cong \mathbb{Z}(p^\infty) \}$,
- $\pi_C = \{ p \in \pi : G_p \cong \mathbb{Z}_p \}$,
- $\pi_D = \{ p \in \pi : (\exists n \in \mathbb{N}_0) G_p \cong \mathbb{Z}(p^n) \}$.

Now we fix a compact open subgroup $C$ of $G$ and identify $G$ with the local product $\prod_{p \in \pi}(G_p, C_p)$, further we define two closed characteristic subgroups as
\[
D := \prod_{p \in \pi_A \cup \pi_B}^{\text{loc}}(G_p, C_p),
\]
\[
P := \prod_{p \in \pi_C \cup \pi_D}^{\text{loc}}(G_p, C_p),
\]
and notice that $G = D \oplus P$. Both subgroups $D$ and $G$ are characteristic, and we notice that in view of Remark 3.3 $D$ is the unique largest divisible subgroup of $G$.

Theorem 3.7. (Classification of Inductively Monothetic Groups, continued) Let $G$ be a periodic inductively monothetic locally compact group. Then $G$ is the direct topological and algebraic sum $D \oplus P$ of two characteristic closed subgroups of which
D is the largest divisible subgroup of G and P is the unique largest Π-procyclic subgroup according to Definition 3.5.

We apply this information to the structure theory of a near abelian periodic group G with base A. Then G/A = D ⊕ P as in Theorem 3.7; let G_D, respectively G_P are the full inverse images for the quotient morphism G → G/A. Now G_D and G_P are closed normal subgroups such that G = G_DG_P and G_D ∩ G_P = A, and we have G_D ⊆ C_G(A) (see [15], Theorem 7.6).

**Theorem 3.8.** (Structure Theorem II on Near Abelian Groups) Let G be a periodic near abelian locally compact group with a base A such that G is A-nontrivial. Then A ⊆ G_D ⊆ C_G(A) where G_D is a normal abelian subgroup such that G/G_D ∼= G_P/A is Π-procyclic.

This portion of the basic structure theory of near abelian groups in the periodic situation will allow us to concentrate largely on the case that the factor group G/A is Π-procyclic.

### 4. Factorisation and Scaling

We begin with a definition elaborating the definition of near abelian groups.

**Definition 4.1.** Let G be a near abelian locally compact group with a base A. A closed subgroup H is called a scaling subgroup for A if

(i) H is inductively monothetic, and

(ii) G = AH.

**Example 4.2.** There exists a (discrete) abelian group G with a subgroup A which is not a direct summand and which has the following properties: A is the torsion subgroup of G of the form $A ≅ \bigoplus_{n \in \mathbb{N}} \mathbb{Z}(p^n)$ and $G/A ≅ \bigcup_{n \in \mathbb{N}} \frac{1}{2}^{n-1} p^n \mathbb{Z} ⊆ \mathbb{Q}$.

**Example 4.3.** $G ≅ \bigoplus_{n \in \mathbb{N}} \mathbb{Z}(p^n)$ and there is a subgroup A which is not a direct summand such that $G/A ≅ \mathbb{Z}(p^\infty)$.

In Example 4.2, the group G is a subgroup of $\mathbb{R}/\mathbb{Z}$ and is a construction due to D. Maier [34]. Example 4.3 is inspired by Example V in Theorem A1.32, p. 686 of [20].

These examples show that there are obstructions to a very general result asserting the existence of a scaling group for near abelian groups G with bases A.

A scaling group H, whenever it exists, is a supplement for A in G but not in general a semidirect complement. How far a supplement is from being a complement can be clarified under fairly general circumstances; we illustrate that in the following proposition.

**Proposition 4.4.** Let G be a locally compact group with a closed normal subgroup A and a closed sigma-compact subgroup H containing a compact open subgroup and satisfying $G = AH$. The inner automorphisms define a morphism $α: H → \text{Aut}(A)$ by $α(h)(a) = hah^{-1}$. Then we have the following conclusions:
(i) The semidirect product $A \rtimes_\alpha H$ is a locally compact group and the function

$$\mu : A \rtimes_\alpha H \to G, \mu(a, h) = ah,$$

is a quotient morphism with kernel $\{(h^{-1}, h) : h \in A \cap H\}$ isomorphic to $A \cap H$, mapping both $A$ and $H$ faithfully.

(ii) The factor group $G/(A \cap H)$ is a semidirect product of $A/(A \cap H)$ and $H/(A \cap H)$ and the composition

$$A \rtimes_\alpha H \to G \to G/(A \cap H)$$

is equivalent to the natural quotient morphism

$$A \rtimes_\alpha H \ightarrow \frac{A}{A \cap H} \times \frac{H}{A \cap H}$$

with kernel $(A \cap H) \times (A \cap H)$.

Notice that a scaling subgroup $H$ of a near abelian group is sigma-compact and has a compact open subgroup, so that the proposition applies in its entirety to near abelian locally compact groups. The typical “sandwich situation”

$$A \rtimes_\alpha H \to AH \to \frac{A}{A \cap H} \times \frac{H}{A \cap H}$$

is also observed in significant ways in the structure theory of compact groups (see [20], e.g. Corollary 6.75 ff.).

So one of the most pressing questions of the structure theory of near abelian locally compact groups is the following:

**Problem 1.** Under which conditions does a locally compact group $G$ with a normal subgroup $A$ such that $G/A$ is inductively monothetic contain a closed inductively monothetic subgroup $H$ such that $G = AH$?

If $G/A$ is in fact monothetic, then the answer is affirmative and easy. In more general circumstances we have the following theorems giving a partial answer to Problem 1.

**Theorem 4.5.** Let $G$ be a locally compact group with a compact normal subgroup $A$ such that $G/A$ is $\Pi$-procyclic. Then $G$ contain a $\Pi$-procyclic subgroup $H_\Pi$ such that $G = AH_\Pi$.

**Theorem 4.6.** Let $G$ be a locally compact group with a compact open normal subgroup $A$ such that $G/A$ is isomorphic to an infinite subgroup of the group $Q$. Then $G$ contains a discrete subgroup $H \cong G/A$ such that $G$ is a semidirect product $AH \cong A \rtimes H$.

It would be highly desirable to have such theorems without the hypothesis that $A$ be compact. The proofs of these theorems (see [15], Theorem 5.23ff.) make essential use of the compact Hausdorff Vietoris-Chabauty space $\mathcal{S}UB(G)$ which is attached to every locally compact group as a general invariant.

As long as this approach requires the compactness of $A$ the following theorem may be considered as fundamental for the structure theory of near abelian locally compact groups:
Theorem 4.7. Let $G$ be a locally compact near abelian group with a base $A$ such that $G$ is $A$-nontrivial and $G/A$ is $\Pi$-procyclic. Then $G$ contains a $\Pi$-procyclic scaling subgroup $H_{\Pi}$ for $A$ with $G = AH_{\Pi}$.

For a proof see [15], Theorem 11.3. The proof requires a wide spectrum of parts of our general structure theory of near abelian groups. In particular, at the root of this existence theorem is the Chabauty space $\mathcal{S}(G)$ of the group $G$ which we mentioned earlier. This theorem and Proposition 3.8 now yield the following theorem (see [15], Theorem 11.6).

Theorem 4.8. For every periodic locally compact near abelian group $G$ with base $A$ there exists a $\Pi$-procyclic closed subgroup $H_{\Pi}$ such that $G = G_{D}H_{\Pi}$ for the abelian normal subgroup $G_{D}$ with $A \subseteq G_{D} \subseteq C_{G}(A)$.

(See [15], Theorem 6.3(i,iv) and their proofs.)

In particular, Proposition 4.4 then shows us that we have

Corollary 4.9. Every periodic locally compact near abelian group $G$ is a quotient of $G_{D} \times H_{\Pi}$ modulo a subgroup isomorphic to $A \cap H_{\Pi}$.

Recall that $Z(G)$ denotes the center of a group $G$.

Corollary 4.10. For every periodic locally compact near abelian group $G$ with base $A$ we have $C_{G}(A) = G_{D}Z(G_{P})$ and $C_{G}(A) \cap H_{\Pi} \subseteq Z(G_{P})$, that is $G_{D}Z(G_{P}) \cap H_{\Pi} = Z(G_{P}) \cap H_{\Pi}$.

(See [15], Theorem 6.3(i,iv) and its proof.)

The following theorem then is rather definitive on the factorisation of a periodic near abelian locally compact group and may be considered as one of the main theorems on their structure.

Theorem 4.11. (Structure Theorem III on Periodic Near Abelian Groups) For every periodic locally compact near abelian group $G$ with a base $A$ such that $G$ is not $A$-trivial, we have

$$G = G_{D}Z(G_{P})H$$

with a $\Pi$-procyclic scaling group $H$ for $A$ in $G_{P}$, where $G_{D}Z(G_{P}) \cap H = Z(G_{P}) \cap H$.

For information as to which closed subgroups $A^{*} \subseteq AZ(G)$ containing $A$ may still be taken as base subgroups, see [15], Theorem 10.32. The role of the center $Z(G)$ in $C_{G_{P}}(A) = AZ(G_{P})$—a locally compact abelian group we know to be a local product of its $p$-primary components $A_{p}Z(G_{P})_{p}$—is still a bit mysterious; more information will be forthcoming in Theorem 7.2 below.

5. The Sylow Theory of Periodic Groups

Sylow theory, i.e., existence and conjugacy of maximal $p$-subgroups, and, more generally, of maximal $\sigma$-subgroups where $\sigma$ is a set of primes, is available for profinite groups (see [51, 43]). Several attempts have been made in order to generalize Sylow theory to noncompact and locally compact groups, see e.g. the survey from 1964 by Carin, [4], or, more recently, Platonov in [41] and Reid in [42]. Here we
shall focus on our class of compactly ruled groups. If the topology on a compactly ruled group is discrete, the group is locally finite, i.e., every finite subset generates a finite subgroup. Then our Sylow theory reduces to the one presented in the book of Kegel and Wehrfritz, see [30].

In each locally compact periodic group $G$ the concept of a $p$-group can be defined meaningfully. Indeed if $g \in G$, then $M := \langle g \rangle$ is a zero-dimensional monothetic compact group, and thus

$$M \cong \prod_{p \in \pi} M_p,$$

where for each prime $p$ the $p$-primary component $M_p$ is either $\cong \mathbb{Z}_p$ or $\mathbb{Z}(p^{n_p})$ for some $n_p = 0, 1, 2, \ldots$.

It is practical to generalize the concept of a $p$-element: For each subset $\sigma \subseteq \pi$, an element $g \in G$ is called a $\sigma$-element if $\langle g \rangle = \prod_{p \in \sigma} M_p$; if $\sigma = \{p\}$, then $g$ is called a $p$-element. The group $G$ is a $\sigma$-group, if all of its elements are $\sigma$-elements. A subgroup $S$ is called a $\sigma$-Sylow subgroup of $G$ if it is a maximal element in the set of $\sigma$-subgroups. A simple application of Zorn’s Lemma shows that every $\sigma$-element is contained in a $\sigma$-Sylow subgroup. We record:

**Lemma 5.1.** (The Closure Lemma) Let $G$ be any locally compact totally disconnected group. Then for any subset $\sigma \subseteq \pi$, then set $G_{\sigma}$ of all $\sigma$-elements of $G$ is closed in $G$.

Let us look to some traditional splitting theorems that still work in the general background of periodic locally compact groups.

**5.1. The Schur–Zassenhaus splitting.** The splitting of finite groups into products of subgroups of relatively prime orders can be generalized to the locally compact setting up to a point, as we show in the following. For locally finite groups the results to be discussed are well known, see e.g. [30]. They also relate to work of the second author, see [19, 23].

**Proposition 5.2.** Let $N$ be a closed subgroup of a locally compact periodic group $G$ and assume $N \subseteq G_{\sigma}$. Then the following conditions are equivalent:

1. $N$ is a normal Sylow subgroup.
2. $N = G_{\sigma}$.
3. $N$ is normal and $G/N$ contains no $p$-element with $p \in \sigma$.

**Definition 5.3.** Let $G$ be a locally compact periodic group and $N$ a closed subgroup. We say that $N$ satisfies the Schur-Zassenhaus Condition if and only if it satisfies the equivalent conditions of Proposition 5.2 for $\sigma = \pi(N)$.

**Theorem 5.4.** (Schur-Zassenhaus Theorem) Let $G$ be a periodic group and $N$ a closed subgroup satisfying the following two conditions:

1. $N$ satisfies the Schur-Zassenhaus Condition.
2. $G/N$ is a directed countable union of compact subgroups.

Then the following conclusions hold:

1. $N$ possesses a complement $H$ in $G$. 
(ii) Let $K$ be a closed subgroup of $G$ such that $K \cap N = \{1\}$ and assume that $G/N$ is compact. Then there is a $g \in G$ such that $gKg^{-1} = H$.

It should be remarked, that for solvable groups (such as near abelian groups) the periodic groups are always directed unions of their open compact subgroups. In such a situation condition (2) simply means that $G/N$ is sigma-compact.

The Schur-Zassenhaus configuration in the locally compact environment is delicate, since problems do arise with the product of a closed normal subgroup and a closed subgroup; such a product need not be closed, in general.

Still we do have theorems like the following:

**Theorem 5.5.** Let $N$ be a normal $\sigma$-Sylow subgroup of a locally compact periodic group $G$. Then a $(\pi \setminus \sigma)$-Sylow subgroup $H$ of $G$ exists such that $NH$ is an open and hence closed subgroup. Moreover, if $H$ is any $(\pi \setminus \sigma)$-Sylow subgroup of $G$, then $NH$ is closed in $G$ and $H$ is a complement of $N$ in $NH$, that is, $NH = N \times H$.

**5.2. Sylow subgroups commuting pairwise.** Let $p \in \pi$ denote any prime and $p' := \pi \setminus \{p\}$.

Then we have the following result:

**Lemma 5.6.** For a compactly ruled group $G$ and a prime number $p$, the following conditions are equivalent:

1. $[G_p, G_{p'}] = \{1\}$.
2. Both $G_p$ and $G_{p'}$ are subgroups, and $G = G_p \times G_{p'}$.
3. There is a unique projection $p_p : G \to G_p$ with kernel $G_{p'}$.

**Definition 5.7.** For a periodic locally compact group $G$ we write $\nu(G) = \{p \in \pi : [G_p, G_{p'}] = \{1\}\}$.

We have found the following structure theorem very useful in the context of near abelian groups generalising the well-known fact that a pronilpotent group is the cartesian product of its Sylow subgroups (cf. [43]):

**Theorem 5.8.** In a compactly ruled locally compact group $G$, the set $G_{\nu(G)}$ of $\alpha$-elements with $\alpha \cap \nu(G) = \emptyset$ is a closed normal subgroup, and all $p$-Sylow subgroups for $p \in \nu(G)$ are normal subgroups. Moreover,

$$G \cong G_{\nu(G)} \times \prod_{p \in \nu(G)}(G_p, U_p)$$

for a suitable family of compact open subgroups $U_p \subseteq G_p$ as $p$ ranges through $\nu(G)$.

**5.3. The internal structure of Sylow subgroups of near Abelian groups.**

For periodic near abelian groups, to which we can apply a Sylow theory meaningfully we assume that $G$ is a periodic near abelian locally compact group such that $G$ is nontrivial for a base $A$.

**Theorem 5.9.** Let $G$ be a periodic near abelian group and $A$ a base for which $G$ is $A$-nontrivial and which satisfies $A = C_G(A)$. Then, for every set $\sigma$ of prime numbers there is a $\sigma$-Sylow subgroup $S_\sigma$. Fix a $\sigma$-Sylow subgroup $S_\sigma$ of $G$. Then
(i) $S_\sigma \cap A$ is the $\sigma$-primary component $A_\sigma$ of $A$, equivalently, the $\sigma$-Sylow subgroup of $A$. Moreover,
(ii) $S_\sigma / A_\sigma \cong S_\sigma A / A = (G / A)_\sigma$.
(iii) If $(G / C_G(A))_\sigma \cong H / (H \cap C_G(A))$ is compact, then any two $\sigma$-Sylow subgroups of $G$ are conjugate.
(iv) $S_\sigma = C_G(A)_\sigma H_\sigma = A_\sigma Z(G)_\sigma H_\sigma$, where $H$ is as in Theorem 4.8.

(See [15], Theorem 10.1.) The case that $\sigma = \{p\}$ is an important special case. Let us note that it may happen, $S_p \subseteq A$, in which case we have $H_p = \{1\}$.

6. Scalar Automorphisms

Among the methods we are using, the specification of scalar automorphisms of a periodic locally compact abelian group is prominent. Every locally compact abelian $p$-group $A$ is a natural $\mathbb{Z}_p$-module, and in the case of a periodic locally compact abelian group

$$A = \prod_{p \in \pi} (A_p, C_p) \quad \text{(LQ)}$$

it is a natural

$$\mathbb{Z} = \prod_{p \in \pi} \mathbb{Z}_p$$

module by componentwise scalar multiplication

$$z \cdot g = (z_p)_{p \in \pi} \cdot (g_p)_{p \in \pi} = (z_p \cdot g_p)_{p \in \pi}$$

The compact ring $\mathbb{Z}$ is the profinite compactification of the ring $\mathbb{Z}$ of integers.

**Lemma 6.1.** (The Scalar Morphism Lemma) For a continuous automorphism $\alpha$ of a periodic locally compact abelian group $G$ the following conditions are equivalent:

1. $\alpha(H) \subseteq H$ for all closed subgroups $H$ of $G$.
2. $\alpha(\langle g \rangle) \subseteq \langle g \rangle$ for all $g \in G$.
3. $\alpha(g) \in \langle g \rangle$ for all $g \in G$.
4. There is an $r \in \mathbb{Z}$ such that $\alpha(a) = r \cdot a$ for all $a \in G$.

We note in passing that the first three conditions are equivalent in any locally compact group.

**Definition 6.2.** An automorphism $\alpha \in \text{Aut}(A)$ of a periodic locally compact abelian group $A$ is called a scalar automorphism. The group of all scalar automorphisms is written $\text{SAut}(A)$.

For $r \in \mathbb{Z}^\times$, the group of invertible elements of $\mathbb{Z}$, we denote the function $a \mapsto r \cdot a : A \to A$ by $\mu_r \in \text{SAut}(A)$.

**Proposition 6.3.** Let $A$ be a periodic locally compact abelian group. Then

(i) $r \mapsto \mu_r : \mathbb{Z}^\times \to \text{SAut}(A)$ is a quotient morphism of compact groups. In particular, $\text{SAut}(A)$ is a profinite group and thus does not contain any non-degenerate divisible subgroups.

(ii) The following conditions are equivalent:

(a) $\text{SAut}(A) = \{\text{id}_A, -\text{id}_A\}$,
(b) The exponent of $A$ is 2, 3, or 4.

In particular, $A$ has exponent 2 if and only if $-\text{id}_A = \text{id}_A$.

In the process of these discussions, we recover in our framework the following theorem of Mukhin [38]:

**Theorem 6.4.** Let $A$ be a locally compact abelian group written additively.

(a) If $A$ is not periodic, then $\text{SAut}(A) = \{\text{id}, -\text{id}\}$.

(b) If $A$ is periodic, then $\text{SAut}(A) = \prod_p \text{SAut}(A_p)$, where $\text{SAut}(A_p)$ may be identified with the group of units of the ring of scalars of $A_p$:

\[
\begin{cases}
\mathbb{Z}_p, & \text{if the exponent of } A_p \text{ is infinite,} \\
\mathbb{Z}_p/p^n\mathbb{Z}_p \cong \mathbb{Z}(p^n), & \text{for suitable } m \text{ otherwise.}
\end{cases}
\]

(c) In particular, $\text{SAut}(A)$ is a homomorphic image of $\tilde{\mathbb{Z}}^\times$.

(d) An automorphism $\alpha$ is in $\text{SAut}(A)$ iff there is a unit $z \in \tilde{\mathbb{Z}}^\times$ such that

\[(\forall g \in G) \alpha(g) = z \cdot g = \prod_p z_p \cdot g_p \text{ for } z = \prod_p z_p, \quad g = \prod_p g_p.\]

The significance of Mukhin’s Theorem for the structure theory of near abelian groups is visible in the very Definition 2.3 via Theorem 4.11. Indeed if $G$ is a near abelian locally compact group with a base $A$, then the inner automorphisms of $G$ induce a faithful action of the $\Pi$-procyclic factor group $G/C_G(A) \cong H/(H \cap Z(G))$ upon the base $A$. So $\text{SAut}(A)$ is a quotient of $H$ and therefore is $\Pi$-procyclic.

The structure of $G$ is largely determined by the structure of $\text{SAut}(A)$ and therefore by the group $\tilde{\mathbb{Z}}^\times$ of units of $\tilde{\mathbb{Z}}$.

### 6.1. The group of units of the profinite compactification of the ring of integers and its prime graph.

The group $\tilde{\mathbb{Z}}^\times$ is more complex than it appears at first. Its Sylow theory or primary decomposition is best understood in graph theoretical terms. The same graph theory turns out to be almost indispensable for dealing with the Sylow structure of near abelian groups in general. The graphs that we use are all subgraphs of a “universal” graph (which we also call the “master graph”) and which is used precisely to describe the Sylow theory of $\tilde{\mathbb{Z}}^\times$. We discuss it in the following.

A bipartite graph consists of two disjoint sets $U$ and $V$ and a binary relation $E \subseteq (U \cup V)^2$ such that $(u, v) \in E$ implies $u \in U$ and $v \in V$. The elements of $U \cup V$ are called vertices and the elements of $E$ are called edges. Any triple $(U, V, E)$ of this type is called a bipartite graph.

In the following we construct a special bipartite graph

$$G = (U, V, E)$$

with $U, V \subseteq \mathbb{N} \times \{0, 1\}$ as follows:

**Definition 6.5.** Define $U = \mathbb{N} \times \{1\}$, $V = \mathbb{N} \times \{0\}$. Let $n \mapsto p_n$ be the unique order preserving bijection of $\mathbb{N}$ onto the set $\pi$ of prime numbers. On $\pi$ we consider the binary relation

\[(1) \quad T = \{(p, q) \in \pi \times \pi : q = p \text{ or } p|(q-1)\}.
\]

Let $E \subseteq (\mathbb{N} \times \{0, 1\})^2$ be defined as follows

\[(2) \quad E = \{((m, 1), (n, 0)) : (p_m, q_n) \in T\}.
\]
If \( e = ((m, 1), (n, 0)) \in E \) is an edge we shall use the following notation for the prime numbers associated with \( e \):
\[
\begin{align*}
p_e &= p_m \quad & q_e &= q_n.
\end{align*}
\]

We shall call \( G = (U, V, E) \) the **prime master-graph**. In all bipartite graphs we consider in this text, the two sets \( U \) and \( V \) of vertices remain constant, while the set of edges will vary through subsets of \( E \) as defined in Definition 6.5.

### 6.2. Geometric properties of the master-graph.

The prime master-graph can be drawn and helps in forming a good intuition of the combinatorics involved.

- The set of vertices \( U \cup V \) of the master-graph is naturally contained in \( \mathbb{R}^2 = \mathbb{R} \times \mathbb{R} \), and so we can “draw” it quite naturally.

  - The elements in \( \mathbb{N} \times \{1\} \) are called the **upper vertices**, those in \( \mathbb{N} \times \{0\} \) the **lower vertices**

- The edges \( e = ((n, 1), (n, 0)) \), \( n \in \mathbb{N} \) are called **vertical**. All other edges \( e = ((m, 1), (n, 0)) \) are called **sloping**. Because of \( p_m | (q_n - 1) \) they are sloping “from left-above to right-below”. There are only vertical and sloping edges. We call vertices \( u = (m, 1) \) and \( v = (n, 0) \) **connected** iff \( e = (v, u) \in E \), i.e., if \( v \) is the upper vertex (end-point) of the edge \( e \) and \( v \) is the lower vertex (end-point) of \( e \).

Each lower vertex \( (n, 0) \) is the endpoint of one vertical and **finitely many** sloping edges. It is connected to an upper vertex \( (m, 1) \) iff \( p_m | (q_n - 1) \).

- Each upper vertex \( (m, 1) \) is connected to infinitely many lower vertices \( (n, 0) \), namely, all those for which \( p_m | (q_n - 1) \), that is, for which there is a natural number \( k \) for which \( q_n = kp_m + 1 \). Indeed, Dirichlet’s Prime Number Theorem says: Every arithmetic progression of the form \( \{ka + b : k \in \mathbb{N}\} \) with \( a \) and \( b \) relatively prime, contains infinitely many primes.

**Definition 6.6.** Let \( p \) and \( q \) be any primes, say, \( p = p_m \) and \( q = q_n \). Then
\[
E_p = \{ e : e = ((m, 1), (n, 0)) \in E \text{ such that } p | (q_n - 1) \},
\]
the set of all edges emanating downwards from the vertex \((m, 1) \in U\) will be called the cone peaking at \(p\). This cone contains infinitely many edges while

\[
\mathcal{F}_q = \{ e : e = ((m, 1), (n, 0)) \in E \text{ such that } p_m|(q - 1) \},
\]

the set of edges ending below in the vertex \((n, 0) \in V\), called the funnel pointing to \(q\), contains only finitely many edges.

All applications of prime graphs which we use in the structure theory of near abelian groups are subgraphs of this master graph.

Since for any periodic locally compact abelian group \(A\) we have a canonical surjective morphism \(\mu : \tilde{\mathbb{Z}}^\times \to \text{SAut}(A)\) we need explicit information on the primary structure – or \(p\)-Sylow structure – of \(\tilde{\mathbb{Z}}^\times\). We are now going to describe this structure in additive notation in terms of the prime master-graph \(\mathcal{G} = (U, V, E)\).

Let \(e = ((m, 1), (n, 0)) \in E\) be an edge in the master-graph.

Case 1. \(m = n\). Then we set \(S_e = \mathbb{Z}_{p_m}\).

Case 2. \(m < n\). Then \(p_m|(q_n - 1)\). Assume that the \(q_n - 1 = p_m^{k(e)} s(e)\) with \(s(e)\) relatively prime to \(p_c := p_m\). Then we set \(S_e = \mathbb{Z}(p_m^{k(e)})\).

For the following proposition we recall

\[
\tilde{\mathbb{Z}}^\times = \prod_{q \in \pi} \mathbb{Z}_q^\times.
\]

Since \(\mathbb{Z}_q^\times\) is not a \(q\)-Sylow subgroup, this is not the \(q\)-primary decomposition of \(\tilde{\mathbb{Z}}^\times\). That decomposition we describe now:

**Proposition 6.7.** (The Sylow Structure of \(\tilde{\mathbb{Z}}^\times\)) Let \(p, q \in \pi\) be primes. Then

(i) The structure of \(\mathbb{Z}_q^\times\) (in additive notation) is

\[
\prod_{e \in \mathcal{F}_q} S_e = \mathbb{Z}_q \times \prod_{e \in \mathcal{F}_q, \text{sloping}} \mathbb{Z}(p_c^{k(e)}).
\]
532 W. HERFORT, K. H. HOFMANN, AND F. G. RUSSO

(ii) The $p$-primary component or $p$-Sylow subgroup of $\tilde{\mathbb{Z}}^\times$

$$(\tilde{\mathbb{Z}}^\times)_p = \prod_{e \in \mathcal{E}_p} (\mathbb{Z}_{q_e}^\times)_{p_e}$$

is (in additive notation)

$$(\tilde{\mathbb{Z}}^\times)_p = \prod_{e \in \mathcal{E}_p} S_e = \mathbb{Z}_p \times \prod_{e \in \mathcal{E}_p, \text{sloping}} \mathbb{Z}(p^{k(e)}) .$$

6.3. The structure of the invertible scalar multiplications of an Abelian group and their prime graph. Now let $A$ be a periodic locally compact abelian group; the Sylow structure of $\text{SAut}(A)$ is now easily discussed: The quotient morphism $\mu: \tilde{\mathbb{Z}}^\times \to \text{SAut}(A)$, preserving the Sylow structures, and the structure of $\text{SAut}(A)$ described so far in Theorem 6.4 allow a precise description of the Sylow structure of $\text{SAut}(A)$.

We associate with $A$ the bipartite graph $G(A) = (U, V, E(A))$ with $U$ and $V$ as in the master-graph and with

$$E(A) = \{ e \in E : e = ((m, 1), (n, 0)) \text{ such that } \text{SAut}(A_{q_n}) \neq \{\text{id}_A\} \} ,$$

and we define

$$E_p = \{ e \in E(A) : e = ((m, 1), (n, 0)) \in e(A) \text{ such that } p \mid (q_n - 1) \},$$

the set of all edges in $G(A)$ ending at the vertex $(n, 0) \in V$ such that $\text{SAut}(A_{q_n})$ is nontrivial, and

$$F_q = \{ e \in E(A) : e = ((m, 1), (n, 0)) \in E(A) \text{ such that } p | (q - 1) \},$$

the set of all edges in $G(A)$ ending at the vertex $(n, 0) \in V$ with $q_n = q$ such that $\text{SAut}(A_q)$ is nontrivial.

We recall that for each $q$-primary component $A_q$, the ring of scalars $\text{SAut}(A_q)$ is either cyclic of order $q^r$, the exponent of $A_q$, if it is finite, and is $\cong \mathbb{Z}_q$ otherwise. Thus its $q$-primary component is

$$\cong \begin{cases} \mathbb{Z}(q^{r-1}) & \text{if the exponent of } A_q \text{ is finite} \\ \mathbb{Z}_q & \text{otherwise.} \end{cases}$$

Accordingly we define, for each edge $e = ((m, 1), (n, 0)) \in E(A)$ in the graph $G(A)$

$$S_e(A) = \begin{cases} \mathbb{Z}(q_m^{r-1}) & \text{if } m = n, \text{ and } A_q \text{ has finite exponent } q^r \\ \mathbb{Z}_{q_m} & \text{if } m = n, \text{ and } A_q \text{ has infinite exponent}, \\ \mathbb{Z}(p^{k(e)}_m) & \text{if } m < n. \end{cases}$$

Then we have, analogously to Proposition 6.7, the following theorem, complementing Proposition 6.7 and Theorem 6.4:

**Theorem 6.8.** (The Sylow Structure of $\text{SAut}(A)$) Let $A$ be a periodic locally compact abelian group and $\text{SAut}(A) = \prod_{p \in \pi} \text{SAut}(A)_p$ the $p$-primary decomposition of the profinite group $\text{SAut}(A)$. Then
(i) The $p$-primary decomposition of $\text{SAut}(A_q)$ is (additive notation assumed) 
\[ \prod_{e \in F_q}^\prime \text{SAut}(A_{q_e}) = \prod_{e \in F_q}^\prime \mathbb{S}_e(A) = \mathbb{Z}_q \times \prod_{e \in F_q}^\prime \mathbb{Z}(p_e^{k(e)}). \]

(ii) The structure of the $p$-primary component $\text{SAut}(A)_p$ of $\text{SAut}(A)$ (in additive notation) is 
\[ \prod_{e \in E_p}^\prime (\text{SAut}(A_{q_e})) = \prod_{e \in E_p}^\prime \mathbb{S}_e(A) = \mathbb{Z}_p \times \prod_{e \in E_p}^\prime \mathbb{Z}(p_e^{k(e)}). \]

This theorem illustrates the usefulness of the prime graph $\mathcal{G}(A)$ which elucidates the fine structure of $\text{SAut}(A)$. In many instances, the prime graph is equally helpful in the discussion of the Sylow structure of any periodic near abelian locally compact group $G$ (see [15], Section 10).

7. The Prime Graph of a Near Abelian Group

Staying with a periodic near abelian group $G$ which is $A$-nontrivial for a base group $A$, we investigate the interaction of the different Sylow subgroups in terms of the prime graph defined as a subgraph of the master-graph $\mathcal{G} = (U, V, E)$ of Definition 6.5 as follows:

**Definition 7.1.** Let $G$ be a periodic near abelian $A$-nontrivial locally compact group with a base group $A$ and write $G = AZ(G)H$.

A subgraph $\mathcal{G}_G = (U_G, V_G, E_G)$ of the master-graph $\mathcal{G}$ is called the prime graph of $G$ provided the following conditions are satisfied:

(i) We call $(m, 1)$ the upper $p$-vertex iff $p = p_m$ and $(n, 0)$ the lower $q$-vertex iff $q = q_n$.

(ii) An edge $e = ((m, 1), (n, 0))$ of the mastergraph is an edge in $E_G$ if and only if $[H_{pm}, A_{qm}] \neq \{1\}$. With $p = p_m$ and $q = q_n$ this edge is written $e_{pq}$ and called an edge leading from $p$ to $q$.

(iii) $(m, 1)$ is an upper vertex in $U_G$ iff $(G/C_G(A))_p \neq \{1\}$.

(iv) $(n, 0)$ is a lower vertex in $V_G$ iff $A_p \neq \{1\}$.

We have a much sharper conclusion:

**Theorem 7.2.** (Structure Theorem IV on Periodic Near Abelian Groups) Let $G$ be a periodic $A$-nontrivial near abelian group. Let $e_{pq}$ be an edge in $\mathcal{G}_G$.

Then we have the following conclusions (see [15], Theorem 10.13):

(1) If $p \neq q$, that is, $e_{pq}$ is sloping, then $p \neq 2$ and $p(q - 1)$, but above all 
(C1) for $x \in G_p \backslash C_G(A_q)$ the function $a \mapsto [x, a] : A_q \to A_q$ is an automorphism of $A_q$. In particular, $[x, A_q] = A_q$.

(C2) $A_q \cap Z(G) = \{1\}$.

(2) If $p = q$, that is, $e_{pq}$ is vertical, then there is a unit $s \in \mathbb{Z}_q^\times$ and a natural number $m \in \mathbb{N}$ such that $[x, a] = a^{m \cdot s}$ for all $a \in A_q$, that is, $[x, A_q] = A_q^m$, and $A_p \cap Z(G)$ has an exponent dividing $q^m$.

The theorem gives an impression of the circumstances in which the intersection $A_q \cap Z(G)_q$ can be nontrivial: The lower $q$-vertex has to be isolated in the prime graph in such a case.
8. Application 1: The Classification of Topologically Quasihamiltonian Groups

The following definition is due to F. Kümmich [31]:

Definition 8.1. A topological group $G$ is called topologically quasihamiltonian if $XY = YX$ holds for any pair of closed subgroups $X$ and $Y$ of $G$.

This is equivalent to saying that $XY$ is a closed subgroup whenever $X$ and $Y$ are subgroups of $G$.

With the framework provided by near abelian locally compact groups, it is possible to classify completely the class of topologically quasihamiltonian locally compact groups. The classification proceeds in two steps: In a first step we classify all locally compact topologically quasihamiltonian groups, and in a second step we classify all locally compact topologically quasihamiltonian groups in one fell swoop.

For step 1 we need a definition:

Definition 8.2. The groups $M_n$ defined by generators and relations for $n = 2, 3, \ldots$ according to

$$M_n := \langle a, b \mid b^{2^n} = 1, b^{2^n-1} = a^2, bab^{-1} = a^{-1} \rangle$$

are called generalised quaternion groups.

These groups also satisfy the relations $a^4 = 1$ and $[a, b] = a^2$ and are fully characterised by the following explicit construction:

$$M_n \cong \frac{\mathbb{Z}(4) \times \mathbb{Z}(2^n)}{\Delta},$$

where $\mathbb{Z}(2^n)$ acts on $\mathbb{Z}(4)$ by scalar multiplication with $\pm 1$ and where $\Delta$ is generated by $(s, t)$, $s = 2 + 4\mathbb{Z}$ and $t = 2^{n-1} + 2^n\mathbb{Z}$. (Cf. [25], Definition 5.8.) We note that $M_2$ is (isomorphic to) the usual group of quaternions $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ of eight elements.

Here is step 1:

Theorem 8.3. A locally compact $p$-group $G$ is topologically quasihamiltonian if and only if $G$ is near abelian with a base group $A$ and an inductively monothetic $p$-group $G/A$ and at least one of the following statements holds:

(a) $G$ is abelian.

(b) There is a $p$-pro-cyclic scaling group $H = \langle b \rangle$ such that $G = AH$ and there is a natural number $s \geq 1$, respectively, $s \geq 2$, if $p = 2$, such that $a^b = a^{1+p^s}$ for all $a$ in $A$. The group $G$ is $A$-nontrivial.

(c) $p = 2$ and $G \cong A_2 \times M_n$, where $A_2$ is an exponent 2 locally compact abelian group and $M_n$ is the generalised quaternion group of order $2^{n+1}$. In this case, $A = A_2 \times \langle a \rangle \cong \mathbb{Z}(2)^{(I_1)} \times \mathbb{Z}(2)^{I_2} \times \mathbb{Z}(4)$ with $a$ as in Definition 8.2 for suitable sets $I_1$ and $I_2$. The group $G$ is $A$-trivial.

Next step 2 (see [15], Theorem 13.9):
Theorem 8.4. Let $G$ be a locally compact periodic topologically quasihamiltonian group. Then, for each $p \in \pi(G)$, the set of $p$-elements $G_p$ is a topologically quasihamiltonian $p$-group, and there is a compact open subgroup $U_p$ in $G_p$ such that $G = G_{\nu(G)}$ is (up to isomorphism) the local product of topologically quasihamiltonian $p$-groups

$$G \cong \prod_{p \in \pi(G)}^{\text{loc}} (G_p, U_p).$$

Conversely, every group isomorphic to such a local product is a topologically quasihamiltonian group.

This theorem is proved with the aid of our Theorem 5.8: For nonperiodic abelian groups we give an algorithmic description of topologically quasihamiltonian locally compact groups in [15], Theorem 13.14. Except for $p = 2$ it turns out that topologically quasihamiltonian groups in the general locally compact domain are the same thing as near abelian groups. For the exceptional compact 2-groups that are near abelian but fail to be topologically quasihamiltonian see [25].

Theorem 8.4 can be visualised in terms of its prime graph, that all connected components are either vertical edges and its end points or are isolated vertices. If we allow ourselves the identification of the connected components of the prime graph with the subgroups they represent, we could reformulate Theorem 8.4 as follows:

Theorem 8.5. Let $G$ be a locally compact periodic topologically quasihamiltonian group. Then each connected component of the prime graph of $G$ represents a normal $p$-Sylow subgroup and $G$ is a local direct product of these subgroups.

9. Application 2: The Classification of Topologically Modular Groups

Recall that the closed subgroups of a topological group form a lattice w.r.t. inclusion “$\subseteq$” as partial order.

Definition 9.1. A topological group $G$ is called topologically modular if the lattice of closed subgroups is modular, that is, satisfies the law $X \lor (Y \cap Z) = (X \lor Y) \cap Z$ whenever $X$ is a closed subgroup of $Z$.

This is equivalent to saying that the lattice of closed subgroups does not contain a sublattice isomorphic to

(See [44], Theorem 2.1.2.) It is instructive to spend some time on an example due to Mukhin which shows that topologically modular groups can be tricky.
Example 9.2. Let $p$ be any prime and $I$ any infinite set (e.g. $I = \mathbb{N}$), set $E := \mathbb{Z}(p)$, and define define $G_j = E^2$, $C_j = \{0\} \times E$ for all $j \in I$, and set
\[ G := E^{(I)} \times E^I \cong \prod_{j \in I} (G_j, C_j), \]
where we took the discrete topology on the direct sum $E^{(I)}$ and the product topology on $E^I$. We shall identify $G$ with $\prod_{j \in I} (G_j, C_j)$ and $E^I \times E^I$ with $(E^2)^I$. The natural injection $\iota: G \to (E^2)^I = E^I \times E^I$ is continuous but is not an embedding, since it is not open onto its image.

Let $D := \{(x, x) : x \in E\} \subseteq E^2$, and
\[ \Delta = D^I = \{(x_j, x_j)_{j \in I} : x_j \in E\} \subseteq (E^2)^I \cong E^I \times E^I \]
denote the respective diagonals. Then $\Delta$ is a closed subgroup of $(E^2)^I$ and so $\iota^{-1}(\Delta) = D^{(I)}$ is a closed subgroup of $G$. We shall denote it by $Y$. This is a noteworthy and perhaps slightly unexpected fact in view of the density of $E^{(I)}$ in $E^I$. We verify as an exercise that the subgroup $Y$ is not only closed, but even discrete, since $\iota(Y)$ meets trivially every open subgroup $\{0\} \times E^K$ for a cofinite subset $K \subseteq I$.

Now the product $E^I$ is the projective limit of its finite partial products $E^F$ as $F$ ranges through the directed set $\mathcal{F}$ of finite subsets $F$ of $I$. Accordingly,
\[ G \cong E^{(I)} \times \lim_{F \in \mathcal{F}} E^F \cong \lim_{F \in \mathcal{F}} (E^{(I)} \times E^F). \]

Let $D_2 = \{(x_j)_{j \in I} : (\exists c \in E)(\forall j \in I)x_j = c\}$. Now we consider the following subgroups of $G$:
\[ X := E^{(I)} \times \{0\}, \quad Z := E^{(I)} \times D_2 \]
whence $X \subseteq Z$. Then $X \vee Y = E^{(I)} \times E^I = G$ and so $(X \vee Y) \cap Z = Z$ on the one side, while $Y \wedge Z = Y$ and so $X \vee (Y \wedge Z) = X \vee Y = G$. Hence $X \vee (Y \wedge Z) \neq (X \vee Y) \wedge Z$. Therefore $G$ is a locally compact abelian nonmodular group.

The example shows that the limit of a projective system of locally compact topologically modular group with proper bonding maps need not be a topologically modular group and that a local product of a collection of finite abelian modular groups may fail likewise to be a topologically modular group. Locally compact abelian topologically modular groups were classified by Mukhin in [37]. We now discuss the nonabelian situation.

A first step in the classification is the case of $p$-groups:

**Proposition 9.3.** Let $G$ be a compactly ruled $p$-group. Then the following statements are equivalent:

1. $G$ is a topologically modular group.
2. $G$ is a topologically quasihamiltonian group with a base group $A$ that is a topologically modular locally compact abelian group.
In contrast, however, with topologically quasihamiltonian groups, the normal $p$-Sylow subgroups are not the only building blocks of topologically modular groups. There is one additional category of building blocks which in the discrete situation were known since the pioneering work of Iwasawa in the forties of the last century, see e.g. [29].

9.1. Iwasawa $(p,q)$-factors.

**Example 9.4.** For a prime $q$ let $A$ be an additively written locally compact abelian group of exponent $q$ and be either compact or discrete. Thus, algebraically, $A$ is a vector space over the field $\text{GF}(q)$.

Now let $p$ be a prime such that $p(q-1)$. Then the multiplicative group of $\text{GF}(q)$ contains a cyclic subgroup $Z$ of order $p$. Let $C = \langle t \rangle$ be any $p$-procyclic group (that is, $C \cong \mathbb{Z}(p^k)$ for some $k \in \mathbb{N}$ or $C \cong \mathbb{Z}_p$), and let $\psi : C \to Z$ be an epimorphism. Then $C$ acts on $A$ via $r*a = \psi(r)*a$. Since $Z$ is of order $p$, the kernel of $\psi$ is an open subgroup of $C$ of index $p$.

Set $G = A \rtimes_\psi C$, the semidirect product for the action of $C$ on $A$. Then $A := A \times \{1\}$ is a base subgroup of the near abelian locally compact group $G$, and $H = \langle (0,t) \rangle = \{0\} \times C$ is a procyclic scaling $p$-subgroup.

There are many maximal $p$-subgroups of $G$, namely, each $\langle (a,t) \rangle$ for any $a \in A$, and there is one unique maximal $q$-subgroup which is normal, namely, $A$.

The simplest case arises when we take for $C$ the unique cyclic group $S_p(Z)$ of $Z$ of order $p$, in which case we have $G \cong A \times \mathbb{Z}(p)$ and the set of elements of order $p$ is $A \times (\mathbb{Z}(p) \setminus \{0\})$ and the set of $q$-elements is $A \times \{0\}$.

The class of locally compact near abelian topologically quasihamiltonian groups described in Example 9.4 is relevant enough in our classification to deserve a name:

**Definition 9.5.** A locally compact group $G$ which is isomorphic to a semidirect product $A \rtimes_\psi C$ as described in Example 9.4 will be called an Iwasawa $(p,q)$-factor. The primes $p$ and $q$ are called the primes of the factor $G$.

The prime graph $G$ of an Iwasawa $(p,q)$-factor is one sloping edge $e_{pq}$ with its endpoints.

We would like to see an abstract characterisation of a $(p,q)$-factor. For the purpose of presenting one let us formulate some terminology for an automorphic action $(h,a) \mapsto h*a : H \times A \to A$ inducing a morphism $\alpha : H \to \text{Aut}(A)$, $\alpha(h)(a) = h*a$. If $H/\ker \alpha$ is an abelian group of order $p$ for a prime number $q$, we shall say that

- the action of $H$ on $A$ is of order $p$.
- If $H$ is a subgroup of a group $G$ and $A$ is a normal subgroup of $G$, then $H$ acts on $A$ via $h*a = hah^{-1}$. If this action is of order $p$, we say that $H$ induces an action of order $p$ on $A$.

**Proposition 9.6.** Let $p$ and $q$ be primes satisfying $p|(q-1)$. A near abelian group $G$ is an Iwasawa $(p,q)$-factor if and only if it satisfies the following conditions:

(a) $A = G'$ is an abelian group of exponent $q$; it is either compact or discrete subgroup of $G$;
(b) There is a scaling group $H$ which is a procyclic $p$-group; it induces an action of order $p$ on $A$.

If these conditions are satisfied, then $G = A \rtimes H$ is a semidirect product and $Z(G) = \{h^p : h \in H\}$.

The significance of the $(p, q)$-factors for our classification is due to the following fact which requires a technical proof that is not exactly short:

**Proposition 9.7.** Let $G = AH$ be an Iwasawa $(p, q)$-factor and $A$ a topologically modular abelian group. Then $G$ is a topologically modular group.

Since a nondegenerate $(p, q)$-factor does not meet the criteria of a topologically quasihamiltonian locally compact group in Theorem 8.4, this allows us to remark a significant difference between topologically quasihamiltonian and topologically modular groups:

**Corollary 9.8.** Any nondegenerate Iwasawa $(p, q)$-factor provides a topologically modular group which is not topologically quasihamiltonian.

After a thorough discussion of compactly ruled topologically modular groups, using much of the information accumulated on near abelian groups we arrive at the following classification of periodic locally compact topologically modular groups:

**Theorem 9.9 (The Main Theorem on Topologically Modular Groups).** Let $G$ be a compactly ruled topologically modular group. Then $\pi$ is a disjoint union of a set $J$ of sets $\sigma$ of prime numbers which are either empty, or singleton sets $\sigma = \{p\}$ such that $G_\sigma$ is a normal $p$-Sylow subgroup and an Iwasawa $p$-factor, or two element sets $\{p, q\}$ such that for $p < q$ the set $G_\sigma$ is a normal $\sigma$-Sylow subgroup and an Iwasawa $(p, q)$-Factor, such that

$$G = \prod_{\sigma \in J} (G_\sigma, C_\sigma)$$

for a family of compact open subgroups $C_\sigma \subseteq G_\sigma$. In particular, $G$ is a periodic near abelian locally compact group.

Conversely, every near abelian locally compact $G$ of this form is a topologically modular locally compact group.

We notice that, in the prime graph of $G$, the Sylow $p$-subgroups $G_p$ constitute the connected components of either isolated vertices or vertical edges with its endpoints, while the Sylow subgroups $G_{\{p, q\}}$ which are Iwasawa $(p, q)$-components are connected components consisting of sloping edges with their endpoints. Moreover, every prime graph having such connected components can be realised as the prime graph of a periodic locally compact topologically modular group.

**Acknowledgment.** We gratefully acknowledge Sidney A. Morris for his reading our text and helping us to make the introduction to a complicated network of mathematics more lucid. The second author looks back with deep appreciation on at least 45 years of mathematical contacts about locally compact groups with Herbert Heyer, to whom this survey is dedicated. The referee for this article contributed additional references and helped us to clean up some typographical flaws. We are grateful for such a careful reading of our text.
References


41. Platonov, V. P.: Periodic and compact subgroups of topological groups, Sibirsk. Mat. Z. 7 (1966), 854–877. [Russian]


POSITIVE DEFINITENESS ON SPHERES AND
HYPERBOLIC SPACES

WALTER R. BLOOM* AND N. J. WILDBERGER

Abstract. We consider two different concepts of positive definiteness, the metric version due to Schoenberg, and the often used algebraic version for a hypergroup. The two notions are the same for the unit sphere in euclidean space and the associated hypergroup of spherical random walks, but in general the metric concept is stronger. We determine explicit convolution structures on spheres and classical hyperbolic spaces geometrically and investigate large dimensional limits.

1. Introduction

Associated with the sphere \( S^n \) and hyperbolic space \( \mathbb{H}^n_+ \) are convolution structures, called commutative hypergroups, on the intervals \([0, \pi]\) and \([0, \infty]\) respectively. The characters of these hypergroups are Gegenbauer (ultraspherical) polynomials in the case of the sphere, and conical (associated Legendre) functions in the case of hyperbolic space. These characters are essentially the same as the spherical functions of these symmetric spaces when viewed as homogeneous spaces. There is also a related notion of positive definite function, and by a form of Bochner’s theorem any such hypergroup positive definite function is a suitable non-negative linear combination (possibly in the integral sense) of characters (see for example [6]).

In [9] Schoenberg introduced a quite different notion of positive definiteness for a metric space \( M \). He described these functions for the finite-dimensional spheres \( S^n \) and also the infinite-dimensional sphere \( S^\infty \), and showed how relations between them lead to representations of positive definite functions in terms of Gegenbauer polynomials and powers of \( \cos x \) respectively. This notion was also investigated in [1] and [5].

In this paper we would like to reconcile the two notions of positive definiteness for the sphere and extend these ideas to hyperbolic spaces. The next section introduces the various notions of positive definiteness. Section 4 describes the hypergroup associated with the sphere \( S^n \) and its characters. Our approach is geometric and does not rely on any symmetric space theory. In particular, no detailed knowledge of the isometry groups is required.

Received 2016-9-2; Communicated by D. Applebaum.

2010 Mathematics Subject Classification. Primary 43A62; Secondary 43A35.

Key words and phrases. Convolution, hypergroup, positive definite, Bochner, Schoenberg, sphere, hyperbolic space.

* Part of the work for this paper was carried out while the first author was visiting the University of New South Wales.
Sections 5 and 6 treat the classical hyperbolic spaces $\mathbb{H}^n_+$ in a similar fashion. The discussion of the characters, or spherical functions, involves the distinction between the principal series and supplementary series.

In Section 7 we give a direct proof to show that for either of these families of spaces, Schoenberg positive definiteness implies hypergroup positive definiteness. For the sphere, results of Schoenberg together with known descriptions of the spherical functions also give the converse.

In Section 9 the corresponding result for the non-compact homogeneous space $SL(2, \mathbb{C})/SU(2)$ (the Naimark hypergroup) is considered, but it isn’t known whether the notions are equivalent. One problem is that the Plancherel measure on the dual of this double coset hypergroup has support only on the principal series characters.

Schoenberg’s work leads us to ask if there is a hypergroup structure on the infinite sphere $\mathbb{S}^\infty$ with characters generating its positive definite functions. We show that in fact the relevant structure is a semigroup, and that this semigroup is the natural infinite limit of the hypergroups associated with the finite spheres $\mathbb{S}^n$ as $n$ approaches $\infty$. The corresponding limiting structure for the classical hyperbolic spaces $\mathbb{H}^n_+$ is also investigated.

Our investigations focus on the important cases of spheres and hyperbolic spaces, which are rank one symmetric spaces. The study of spherical functions on higher rank symmetric spaces, called Gelfand pairs, along with their resulting asymptotics has been studied by others, for example one can consider compact Lie groups themselves as Gelfand pairs under conjugation and their asymptotics as the dimension goes to infinity. The main works here are [7] for compact symmetric spaces of type A, and [8] for those of types B and C, and following a somewhat different approach, [10] in the unitary case. For a comprehensive overview see [4].

2. Positive Definite Functions

A complex-valued function $f$ on a group $G$ is \textit{positive definite} if

$$\sum_{i,j=1}^{n} f(x_i x_j^{-1}) \xi_i \xi_j \geq 0$$

for all choices of $x_i \in G$, $\xi_i \in \mathbb{C}$ and $n \in \mathbb{N}$. Let $P(G)$ denote the set of all continuous positive definite functions on $G$.

If $G = \mathbb{R}$, then Bochner’s theorem establishes that any positive definite function is the Fourier-Stieltjes transform of a bounded non-decreasing function $F$, that is

$$f(x) = \int_{-\infty}^{\infty} e^{ix\xi} F(d\xi).$$

More generally, if $G$ is a locally compact abelian group, then the Weil-Povzner-Raikov theorem states that there is a bounded positive measure $\mu$ on $\widehat{G}$ such that $f$ is the Fourier-Stieltjes transform of $\mu$.

We now consider the corresponding notion on a commutative hypergroup $K$; see [2] for details of hypergroups. These structures are defined through a convolution structure on their measure algebra $M(K)$; there is no actual group multiplication.
However, the convolution product

\[ f(x * y) := \int f \, d(\epsilon_x * \epsilon_y) \]

is defined for any measurable function \( f \) for which the integral exists. Note that \( \epsilon_x * \epsilon_y \) as the convolution of two point masses is a probability measure, but rarely has a single point support, so that \( x * y \) has no meaning on its own. Nevertheless we can use this to develop a theory of positive definiteness. A complex-valued function \( f \) on a hypergroup \( K \) is positive definite if

\[ \sum_{i,j=1}^{n} f(x_i * x_j^-) \xi_i \overline{\xi_j} \geq 0 \]

for all choices of \( x_i \in K, \xi_i \in \mathbb{C} \) and \( n \in \mathbb{N} \). Let \( P(K) \) denote the set of all continuous positive definite functions on \( K \). Unlike the group case, functions in \( P(K) \) are not necessarily bounded. Nevertheless, they satisfy the properties

1. \( f(e) \geq 0 \),
2. \( f(x_i * x_j^-) \geq 0 \) for all \( x \in K \),
3. \( f(x^-) = \overline{f(x)} \) for all \( x \in K \),
4. \( f(e) = \|f\|_\infty \) whenever \( f \) is bounded.

It is easily seen that \( P(K) \) is closed under

1. linear combinations with non-negative coefficients,
2. pointwise convergence to a continuous limit.

Schoenberg introduced another notion of positive definiteness in the case of a metric space \( M \) with distance function \( d(x,y) \). A continuous real-valued even function \( g \) defined on the interval \([−d(x,y),d(x,y)]\) is positive definite if

\[ \sum_{i,j=1}^{n} g(d(x_i,x_j)) \xi_i \overline{\xi_j} \geq 0 \]

for all \( x_i \in M, \xi_i \in \mathbb{C} \) and \( n \in \mathbb{N} \). Schoenberg showed that the set \( \Psi(M) \) of positive definite functions on \( M \) is closed under

1. linear combinations with non-negative coefficients,
2. pointwise convergence to a continuous limit,
3. pointwise multiplication.

It should be emphasised that the third property is very strong, one that is not in general enjoyed by hypergroups.

If \( M \subset N \) then \( \Psi(M) \supset \Psi(N) \). This observation led Schoenberg to relations between Gegenbauer (ultraspherical functions) \( P_n^{(\alpha)}(\cos t) \), generated for \( \alpha > 0 \) by the expansion

\[ (1 - 2r \cos t + r^2)^{-\alpha} = \sum_{n=0}^{\infty} r^n P_n^{(\alpha)}(\cos t), \quad (2.1) \]

and for \( \alpha = 0 \) by

\[ P_n^{(0)}(\cos t) = \cos nt = T_n(\cos t), \]
where $T_n$ denotes the Tchebychev polynomial of the first kind of degree $n$. Let $\mathbb{S}^m$ denote the unit sphere in $(m + 1)$-space and $\mathbb{S}^\infty$ the unit sphere in Hilbert space. Since we may assume that

$$\mathbb{S}^1 \subset \mathbb{S}^2 \subset \cdots \subset \mathbb{S}^m \subset \cdots \subset \mathbb{S}^\infty$$

it follows that

$$\mathcal{P}(\mathbb{S}^1) \supset \mathcal{P}(\mathbb{S}^2) \supset \cdots \supset \mathcal{P}(\mathbb{S}^m) \supset \cdots \supset \mathcal{P}(\mathbb{S}^\infty)$$

and in fact $\mathcal{P}(\mathbb{S}^\infty)$ is exactly the intersection of all the sets $\mathcal{P}(\mathbb{S}^m)$, $m = 1, 2, \cdots$.

Schoenberg showed that each function $g \in \mathcal{P}(\mathbb{S}^m)$ can be represented as

$$g(t) = \sum_{n=0}^{\infty} a_n W_n^m(\cos t),$$

where $a_n \geq 0$ satisfies $\sum_{n=0}^{\infty} a_n < \infty$ and $P_n$ is a normalised Legendre polynomial of degree $n$. More generally, each function $g \in \mathcal{P}(\mathbb{S}^m)$ has a similar representation as

$$g(t) = \sum_{n=0}^{\infty} a_n W_n^m(\cos t)$$

where the $W_n^\alpha$ are normalisations of the Gegenbauer polynomials via

$$W_n^\alpha(x) = P_n^\alpha(x) / P_n^\alpha(1).$$

The functions $g$ in $\mathcal{P}(\mathbb{S}^\infty)$ have a similar representation as

$$g(t) = \sum_{n=0}^{\infty} a_n (\cos t)^n,$$

where $a_n \geq 0$ and $\sum_{n=0}^{\infty} a_n < \infty$.

3. Positive Definite Functions and Homogeneous Spaces

Let $G$ be a locally compact group and $H$ a compact subgroup with normalised Haar measure $\omega_H$. (For the sphere $\mathbb{S}^n$ and hyperbolic space $\mathbb{H}^n_+$, $G, H$ will be chosen so that $G/H \cong [0, \pi], \mathbb{R}_+$ respectively.) Consider the following sequence of mappings and an associated correspondence between functions:

$$G \xrightarrow{\pi} G/H \xrightarrow{\pi'} G//H$$

$$f^\circ (g) = f^\circ (gH) = f(HgH) \quad \quad (3.1)$$

We note that $f^\circ$ is constant on double $H$-cosets.

**Theorem 3.1.** If $f^\circ$ is positive definite on the group $G$, then $f$ is positive definite on the hypergroup $G//H$.

**Proof.** Assume $f^\circ$ to be positive definite on the group $G$ and recall that it is $H$–bi-invariant. Then we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j f^\circ(g_i g_j^{-1}) \geq 0$$
for all choices of $g_i \in G, c_i \in \mathbb{C}$ and $n \in \mathbb{N}$. Now for $k_1, k_2, \ldots, k_n \in H$ we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j f^\flat \left( (k_i g_i k_i^{-1}) (k_j g_j k_j^{-1})^{-1} \right) \geq 0.$$ 

Integrating this expression $n$ times over $H$ gives

$$0 \leq \int_H \int_H \cdots \int_H \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j f^\flat \left( (k_i g_i k_i^{-1}) (k_j g_j k_j^{-1})^{-1} \right) \omega_H (dk_1) \omega_H (dk_2) \ldots \omega_H (dk_n)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \int_H f^\flat (g_i k_i^{-1}) \omega_H (dk)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \int_H \varepsilon_{g_i} * f^\flat * \varepsilon_{g_j} (k) \omega_H (dk)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \int_H \varepsilon_{g_i} (Hg_i H) * Hg_j^{-1} H$$

which shows that $f$ is hypergroup positive definite. Note that in the first equality we have used the property that $f^\flat$ is constant on $H$--double cosets. \qed

4. Convolution on the Sphere $\mathbb{S}^n$

We follow the ideas in [1] to develop the hypergroup structure of the unit $n$--sphere $\mathbb{S}^n$ in $\mathbb{R}^{n+1}$. Let $G = SO(n+1)$ denote the group of rotations of $\mathbb{R}^{n+1}$ and view $H = SO(n)$ as a subgroup of $SO(n+1)$. We may assume that $G$ left acts transitively on the $n$--sphere $\mathbb{S}^n$, with the stabiliser subgroup of the north pole $s_0 = (0,0,\ldots,0,1)$ being $H$. Thus $\mathbb{S}^n \cong G/H$, the space $\{gH \}$ of left cosets of $H$ in $G$. The orbits of the subgroup $H$ on $\mathbb{S}^n$ are the meridian hypercircles $C_\varphi$ parallel to $\mathbb{R}^n \subset \mathbb{R}^{n+1}$, indexed by the angle $\varphi$ from $0$ to $\pi$. At the endpoints $0$ and $\pi$ these hypercircles reduce to the north and south poles $s_0$ and $-s_0$ respectively. The orbits of $H$ on $G/H$ form the double coset space $G/H$, which can be identified with the set of double cosets $\{HgH \}$ of $H$ in $G$. Referring to (3.1) we see that $f^\flat$ is a function on $\mathbb{S}^n$ constant on meridian circles centered on the $(0,0,\ldots,0,1)$-axis, and $f^\flat$ is a function on $G$ constant on $H$--double cosets.

The space $K = G/H$ is a natural commutative hypergroup in one of the following equivalent ways. Given two functions $f$ and $g$ on $K$ define their convolution by

$$(f * g)^\flat = f^\flat * g^\flat$$

involving the usual convolution of functions on the group $G$. Alternatively we may define

$$(f * g) (x) = \int_K f (x * y) g(y) \omega_K (dy),$$
where

\[ f(x \ast y) = \int_K f d(\epsilon_x \ast \epsilon_y) \]

is defined in terms of the hypergroup product \( \epsilon_x \ast \epsilon_y \) of the point measures at \( x \) and \( y \). In other words the convolution of functions is determined by knowledge of the products of pairs of point measures, which are by the definition of a hypergroup necessarily probability measures on \( K \). This gives a clear probabilistic interpretation to convolution in the case of a double coset hypergroup \( K = G//H \).

Let us now consider how to determine the hypergroup structure geometrically for the \( n \)-sphere \( S^n \cong SO(n + 1)/SO(n) \).

We begin at the north pole \( s_0 = (0, 0, \ldots, 0, 1) \), take a random step of distance \( x \) to a point \( P \) and then another random step from \( P \) of distance \( y \) to a point \( Q \). In this case random means that if we consider the meridian which is the \((n - 1)\) sphere \( T = \{s \in S^n : d(P, s) = y\} \), then \( T \) is a homogeneous space for the group \( SO(n) \) acting as rotations which fix the point \( P \), and we are taking the probability distribution for \( Q \) to be uniform on \( T \) with respect to this group. To determine the probability density function \( g_{x,y}(r) \) for the distance \( r \) from \( s_0 \) to \( Q \) we analyze the spherical triangle \( s_0PQ \).

Then distance means angular distance on the surface of the unit sphere. The portion of the sphere \( T \) for which the angle \( \angle s_0PQ \) lies in \([\theta, \theta + d\theta]\) has measure

\[
\frac{1}{c_n} \sin^{n-2} \theta d\theta,
\]

where

\[
c_n = \int_0^\pi \sin^{n-2} \theta d\theta = \frac{\Gamma\left(\frac{n-1}{2}\right) \sqrt{\pi}}{\Gamma\left(\frac{n}{2}\right)}. \tag{4.1}
\]

If one of \( x, y \) is 0 or \( \pi \), then the resulting probability is concentrated at one point. Assume then that neither \( x \) nor \( y \) is 0 or \( \pi \). In the spherical triangle \( s_0PQ \) let \( \theta \) denote the angle \( \angle s_0PQ \). Then the spherical cosine law asserts that

\[
\cos r = \cos x \cos y + \sin x \sin y \cos \theta, \tag{4.2}
\]

and since \( \sin x, \sin y \geq 0 \) we have

\[
\cos x \cos y - \sin x \sin y \leq \cos r \leq \cos x \cos y + \sin x \sin y.
\]

This is just

\[
\cos(x + y) \leq \cos r \leq \cos(x - y). \tag{4.3}
\]

Consider \( x \) and \( y \) to be fixed and take infinitesimals in (4.2) to obtain

\[
\sin r \, dr = \sin x \sin y \sin \theta \, d\theta. \tag{4.4}
\]
Then (4.2) gives
\[ \sin^2 \theta = 1 - \left( \frac{\cos r - \cos x \cos y}{\sin x \sin y} \right)^2 \]
\[ = \frac{1 - \cos^2 y - \cos^2 x - \cos^2 r + 2 \cos r \cos x \cos y}{\sin x \sin y}^2 \]
\[ = \frac{(\cos (x - y) - \cos r) (\cos r - \cos (x + y))}{\sin x \sin y}^2, \]
and so
\[ \sin \theta = \frac{[(\cos (x - y) - \cos r) (\cos r - \cos (x + y))]^{\frac{1}{2}}}{\sin x \sin y}. \quad (4.5) \]
Now
\[ g_{x,y}^{(n)} (r) \ dr = \frac{1}{c_n} \sin^{n-2} \theta \ d\theta, \quad (4.6) \]
and appealing to (4.4) and (4.5) we have
\[ g_{x,y}^{(n)} (r) = \frac{1}{c_n} \frac{\sin r \sin^{n-2} \theta}{\sin x \sin y \ sin \theta} \]
\[ = \frac{\sin r}{c_n \ sin x \ sin y} \left( \frac{[(\cos (x - y) - \cos r) (\cos r - \cos (x + y))]^{\frac{1}{2}}}{\sin x \sin y} \right)^{n-3} \]
\[ = \frac{\sin r}{c_n} \frac{[\sin x \sin y]^{n-2}}{[\sin x \sin y]^{n-2}}, \]
valid for \( r \) satisfying (4.3). It follows that \( K = [0, 1] \) becomes a hypergroup with
\[ f (x * y) = \int_0^1 f (r) g_{x,y}^{(n)} (r) \ dr, \]
or equivalently, using (4.6) and appealing to (4.2),
\[ f (x * y) = \frac{1}{\pi} \int_0^\pi f \left( \cos^{-1} (cos x \cos y + sin x \sin y \cos \theta) \right) \ d\theta \quad (4.7) \]
We identify \( K^{(n)} = SO(n + 1) / S(n) \) with the interval \( [0, \pi] \) of values of the angle \( \varphi \) parametrising the meridians of \( S^n \). The characters of the associated hypergroups \( K^{(n)} \cong [0, \pi] \) will be those functions \( \chi \) satisfying
\[ \chi (x) \chi (y) = \int_0^\pi \chi (r) g_{x,y}^{(n)} (r) \ dr, \]
or equivalently, using (4.7),
\[ \chi (x) \chi (y) = \frac{1}{c_n} \int_0^\pi \chi \left( \cos^{-1} (cos x \cos y + sin x \sin y \cos \theta) \right) \sin^{n-2} \theta \ d\theta \quad (4.8) \]
for all \( x, y \in K^{(n)} \). The solutions of (4.8) are given by
\[
\psi^n_k (r) = \frac{k! \Gamma (n-1)}{\Gamma (n-1+k)} P^l_{n-1} (\cos r)
\]
for \( k = 0, 1, 2, \cdots \), where \( P^l_k \) is the Gegenbauer polynomial of order \( l, k \) defined as in (2.1). The \( c_n \) are chosen to ensure that \( \psi^n_k (0) = 1 \).

For \( n = 2 \) the Gegenbauer polynomial reduces to the Legendre polynomial \( P^l_k \) and for \( n = 3 \) it reduces to the Tchebychev polynomial \( P^1_k \) of the second kind. The latter has the explicit formula
\[
U_k (\cos r) = \frac{\sin (k+1)r}{\sin r}
\]
(see [11]).

It follows that the set of positive definite functions on the hypergroup \( K^{(n)} \) agrees with the set of positive definite functions in the sense of Schoenberg ([9]) for the sphere \( S^n \) with the (natural) spherical distance \( d(n) \).

5. Convolution on the Hyperbolic Plane

The probabilistic development of the spherical hypergroup of the sphere \( S^n \) given above has a direct analogue for the classical hyperbolic spaces \( \mathbb{H}^n_+ \). Our treatment largely avoids details about the structure of the isometry group \( SO (n, 1) \) of \( \mathbb{H}^n_+ \) which is a significant advantage over more traditional approaches. In fact just as in the spherical case the algebraic structure contained in the associated ‘hyperbolic hypergroups’ is thus a direct consequence of the geometry of the hyperbolic spaces.

We begin in three-dimensional space with the Lorentzian inner product
\[
\langle v, v' \rangle = \langle (x, y, z), (x', y', z') \rangle = -xx' - yy' + zz'
\]
and define the classical hyperbolic plane \( \mathbb{H}^2_+ \) to be the sheet of the hyperboloid
\[
\{ v : \langle v, v \rangle = 1 \}
\]
through the point \( O = (0, 0, 1) \). This is a restricted, classical form of hyperbolic geometry concentrating only on the interior of the light cone, which is not as general as the universal hyperbolic geometry introduced in ([12]), which rather looks at the entire three-dimensional space. The classical metric on \( \mathbb{H}^2_+ \) is given by the hyperbolic distance \( d(P, Q) \) defined by
\[
\cosh d(P, Q) = \langle P, Q \rangle.
\]
As a homogeneous space
\[
\mathbb{H}^2 \cong SO (2, 1)/SO (2) \cong [0, \infty),
\]
with \( T = SO (2) \) regarded as those isometries fixing the point \( O \). The orbits of \( T \) on \( \mathbb{H}^2 \) are circles centered at \( O \) indexed by their hyperbolic distance from \( O \), which can take any value in \( K = [0, \infty) \). We may describe the convolution of the circles either indirectly, by appealing to the group convolution of \( T \) bi-invariant measures in \( G = SO (2, 1) \), or directly by analyzing the geometry of random walks.
on $\mathbb{H}^2$. As in the case of the sphere, it is this latter approach that we prefer on account of its more immediate and elementary nature.

To convolve the circles of radius $x$ and $y$ around $O$, note first that if one or both of $x, y$ is 0, then it acts as the identity, so in what follows we assume both $x$ and $y$ are non-zero. Choose a point $P$ on the circle of radius $x$ around $O$, and then choose randomly a point $Q$ on the circle of radius $y$ around $P$. Here again, randomly means with respect to the invariant probability measure on such a circle coming from the circle of isometries fixing $P$. The notion of angle between two directions in the hyperbolic plane is the same as the Euclidean one and is preserved by isometries. The hyperbolic triangle $OPQ$ with $\theta$ representing the angle $\angle OPQ$ and the length $r$ of the side $OQ$ satisfy the hyperbolic cosine law

$$\cosh r = \cosh x \cosh y - \sinh x \sinh y \cos \theta.$$  \hfill (5.1)

It follows that $r$ satisfies the inequalities

$$\cosh (x - y) \leq \cosh r \leq \cosh (x + y).$$  \hfill (5.2)

With $x$ and $y$ fixed, taking differentials of (5.1) gives

$$\sinh r \, dr = \sinh x \sinh y \sin \theta \, d\theta.$$  \hfill (5.3)

As in the case of the sphere, the measure of that portion of the circle around $P$ corresponding to $[\theta, \theta + d\theta]$ is $\frac{1}{\pi} \sinh \theta \, d\theta$ so that the probability density for the distance $r$ is

$$m_{x,y}(r) = \frac{\sinh r}{\pi \sinh x \sinh y \sin \theta}.$$  \hfill (5.4)

To obtain an expression involving only $x, y, r$, we have from (5.1)

$$\sin^2 \theta = 1 - \left( \frac{\cosh x \cosh y - \cosh r}{\sinh x \sinh y} \right)^2$$

$$= 1 - \frac{\cosh^2 x - \cosh^2 y - \cosh^2 r + 2 \cosh r \cosh x \cosh y}{(\sinh x \sinh y)^2}$$

$$= \frac{(\cosh r - \cosh (x - y)) (\cosh (x + y) - \cosh r)}{(\sinh x \sinh y)^2}.$$  

Take a square root to get

$$\sin \theta = \frac{[(\cosh r - \cosh (x - y)) (\cosh (x + y) - \cosh r)]^{\frac{1}{2}}}{\sinh x \sinh y}.$$  \hfill (5.4)

Thus

$$m_{x,y}(r) = \frac{\sinh r}{\pi [(\cosh r - \cosh (x - y)) (\cosh (x + y) - \cosh r)]^{\frac{1}{2}}}$$

which is valid for all $r$ satisfying (5.2). It follows that $K = [0, \infty)$ becomes a hypergroup with

$$f(x * y) = \int_0^\infty f(r) \, m_{x,y}(r) \, dr$$
or equivalently, using $m_{x,y}(r) \, dr = d\theta/\pi$ and appealing to (5.1),

$$f(x*y) = \frac{1}{\pi} \int_0^\pi f(\cosh^{-1}(\cosh x \cosh y - \sinh x \sinh y \cos \theta)) \, d\theta.$$  

This corresponds to [2], p.236 with $a = 1$. Haar measure for this hypergroup is

$$\omega(dx) = (\sinh^2 x) \lambda_{\mathbb{R}_+} (dx).$$

A character on this (non-compact) hypergroup is defined to be a continuous function $\chi$ satisfying

$$\chi(x) \chi(y) = \chi(x*y) = \frac{1}{\pi} \int_0^\pi \chi(\cosh^{-1}(\cosh x \cosh y - \sinh x \sinh y \cos \theta)) \, d\theta = \frac{1}{\pi} \int_0^\pi \chi(\cosh^{-1}(\cosh x \cosh y + \sinh x \sinh y \cos \theta)) \, d\theta,$$

where the second equality arises on replacing $\theta$ by $\pi - \theta$. It can be shown that the bounded characters are the so-called conical functions

$$\chi_{-\frac{1}{2}+i\kappa} (r) = p_{-\frac{1}{2}+i\kappa} (\cosh r) = \frac{1}{2\pi} \int_0^{2\pi} (\cosh r + \sinh r \cos \theta)^{-\frac{1}{2}+i\kappa} \, d\theta,$$

where either $\kappa \in \mathbb{R}_+$ (the principal series) or $\kappa \in i [0, \frac{1}{2}]$ (the supplementary series). The identity character is the element of the supplementary series, where $\kappa = \frac{1}{2}$. It is well known that the Plancherel measure for this hypergroup is supported on the principal series characters. Note the symmetry $p_{-\frac{1}{2}+i\kappa} = p_{-\frac{1}{2}-i\kappa}$.

Conical functions are examples of associated Legendre functions, where the lower parameter is $-\frac{1}{2} + i\kappa$ for some $\kappa \geq 0$. In this case the upper parameter is 0 and in MAPLE these are given by LegendreP\((-\frac{1}{2} + i\kappa, 0, \cosh x)\).

6. Higher-dimensional Hyperbolic Hypergroups

The discussion of the previous section generalises to the case of hyperbolic (real) $n$-space $\mathbb{H}_+^n$. In $(n+1)$-dimensional space with inner product

$$\langle v, v' \rangle = \langle (x_1, \cdots, x_n, x_{n+1}), (x_1, \cdots, x_n, x_{n+1}) \rangle = -x_1x'_1 - \cdots - x_nx'_n + x_{n+1}x'_{n+1}$$

and hyperbolic distance $d(v, v')$ defined by

$$\cosh d(v, v') = \langle v, v' \rangle,$$

define $\mathbb{H}_+^n$ to be the sheet of the hyperboloid

$$\{ v : \langle v, v \rangle = 1 \}$$

passing through the point $O = (0, \cdots, 0, 1)$. As a homogeneous space

$$\mathbb{H}_+^n \cong SO(n, 1)/SO(n)$$

with the stabiliser subgroup $SO(n)$ of the point $O$ acting as (ordinary) rotations about the $x_{n+1}$ axis. The non-trivial orbits of $SO(n)$ on $\mathbb{H}_+^n$ are $n$-spheres which
we call zonal meridians. The meridian consisting of the set of points in $\mathbb{H}^n_+$ with hyperbolic distance from $O$ some fixed $d \geq 0$ may be described as the intersection of $\mathbb{H}^n_+$ with the hyperplane given by $x_{n+1} = \cosh^{-1} d$ or $\{ v : \langle v, e_{n+1} \rangle = \cosh^{-1} d \}$. Since any such meridian is determined by the positive number $d$, the orbit space of all zonal meridians is

$$K \cong SO(n,1)/SO(n) \cong [0, \infty).$$

Since the group $SO(n,1)$ acts on $\mathbb{H}^n_+$ as isometries, any point $P$ on $\mathbb{H}^n_+$ may be obtained as $gP$ for some $g \in SO(n,1)$ (actually we may take $g$ to be in the connected component $SO(n,1)_+$ of the identity). Then the set of points with hyperbolic distance $d$ from a general point $P$ will be called the meridian with centre $P$ and radius $d$, and is obtained by intersecting $\mathbb{H}^n_+$ with the $SO(n,1)$-translate of the hyperplane $\{ v : \langle v, ge_{n+1} \rangle = \cosh^{-1} d \}$. Note that $ge_{n+1}$ is also on $\mathbb{H}^n_+$. The general meridian is topologically an $n$-sphere and in particular carries a unique probability measure that is invariant under the subgroup of isometries of $SO(n,1)$ fixing its centre.

Geodesics through $O$ are precisely the intersections of $\mathbb{H}^n_+$ with planes through the centre $(0, \cdots, 0, 0)$. It follows that general geodesics are obtained by intersecting $\mathbb{H}^n_+$ with general planes passing through the centre.

Fix $x, y \in \mathbb{R}_+$ both non-zero (otherwise the product is trivial) and suppose we choose a point $P$ randomly on the meridian of centre $O$ and radius $x$ and then choose a point $Q$ randomly on the meridian of centre $P$ and radius $y$. To determine the probability density function $m_{x,y}^{(n)}(r)$ for the hyperbolic distance $r$ from $O$ to $Q$ we proceed exactly as in the case of $\mathbb{H}^2$ and analyze the hyperbolic triangle $OPQ$ as above. The only difference is that the portion of the meridian sphere $T$ for which the angle $\angle OPQ$ lies in $[\theta, \theta + d\theta]$ has measure $\frac{1}{c_n} \sin^n \theta d\theta$, where $c_n$ is as before.

For $n \geq 2$, appealing to (5.3) and (5.4) we obtain

$$m_{x,y}^{(n)}(r) = \frac{1}{c_n} \sin r \sin^{n-2} \theta$$

$$= \frac{\sinh r}{c_n \sinh x \sinh y} \left[ \frac{\left( \cosh r - \cosh (x - y) \right) \left( \cosh (x + y) - \cosh r \right)}{(\sinh x \sinh y)^2} \right]^\frac{n-3}{2}$$

$$= \frac{\sinh r}{c_n} \frac{\left( \cosh r - \cosh (x - y) \right) \left( \cosh (x + y) - \cosh r \right)^{\frac{n-3}{2}}}{[\sinh x \sinh y]^{n-2}}$$

valid for all $r$ satisfying (5.2). As before, $K = [0, \infty)$ becomes a hypergroup with

$$f(x \ast y) = \int_0^\infty f(r) m_{x,y}^{(n)}(r) \, dr,$$
or equivalently, using \( m_{x,y}^{(n)} (r) \) \( dr = \frac{1}{c_n} \sin^{n-2} \theta d\theta \) and appealing to (5.1) with the obvious change of variable,

\[
f(x * y) = \frac{1}{c_n} \int_0^\pi f \left( \cosh^{-1} \left( \cosh x \cosh y + \sinh x \sinh \cosh \theta \right) \right) \sin^{n-2} \theta d\theta.
\]

(6.1)

This hypergroup is sometimes called a hyperbolic hypergroup (see [2], p. 237 and [13]). The characters for this convolution are given by

\[
\psi_{\lambda}^{(n-1)/2}(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right)} \int_0^\pi (\cosh x + \cos t \sinh x)^{i\kappa - \frac{n-1}{2}} \sin^{n-2} t dt,
\]

where \( \kappa = \sqrt{\lambda - \frac{(n-1)^2}{2}} \). Those for the hyperbolic plane are obtained when \( n = 2 \).

7. Schoenberg Positive Definite Implies Hypergroup Positive Definite

In this section \( M \) will denote either the sphere \( \mathbb{S}^n \) or the hyperbolic space \( \mathbb{H}^n_+ \), each considered as a metric space with distance function \( d \). Then as a homogeneous space \( M \cong G/H \), where \( G \) is the group of isometries of \( M \) and \( H \) is the compact subgroup fixing a distinguished point \( O \). The double coset hypergroup \( K = G/H \) is then either the finite interval \([0, \pi]\) in the case of \( \mathbb{S}^n \) or the infinite interval \([0, \infty]\) in the case of \( \mathbb{H}^n_+ \), where an \( H \) orbit on \( M \) (or a double coset) is identified with its distance from \( O \). Let \( dh \) denote the normalised invariant (Haar) measure on \( H \) so that

\[
\int_H dh = 1.
\]

Lemma 7.1. Fix points \( m, m' \) on \( M \) and let \( x = d(O,m), y = d(O,m') \) so that \( x, y \in K \). Then for any function \( f \) on \( K \)

\[
\int_H f(d(m, hm')) dh = f(x * y).
\]

Proof. For fixed \( m \) and \( m' \) consider the quantity \( d(m, hm') \) as \( h \) varies over the subgroup \( H \). To go from \( m \) to \( hm' \) we can take a step of size \( x \) to \( O \) and then a step of distance \( y \) to \( hm' \). This means the average of the values of \( f(d(m, hm')) \) over \( H \) is equal to \( f(x * y) \) by the definition of hypergroup convolution. \( \square \)

Theorem 7.2. If a function \( f \) on \( K \) is positive definite in the metric sense of Schoenberg, then it is a positive definite function on the hypergroup \( K = G/H \).

Proof. Suppose that \( f \in \mathcal{P}(K) \), which means that

\[
\sum_{i,j=1}^n f(d(m_i, m_j)) \xi_i \overline{\xi_j} \geq 0
\]

for any positive integer \( n \), any points \( m_i \) on \( M \) and any complex numbers \( \xi_i \), where \( i = 1, 2, \ldots, n \). Averaging each of \( m_1, \ldots, m_n \) along its respective orbit by \( H \) gives

\[
\int_H \cdots \int_H \sum_{i,j=1}^n f(d(h_i m_i, h_j m_j)) \xi_i \overline{\xi_j} dh_1 \cdots dh_n \geq 0
\]
or
\[ \sum_{i,j=1}^{n} \left( \int_{H} \int_{H} f( (h_{i}m_{i}, h_{j}m_{j}) ) \, dh_{i}dh_{j} \right) \xi_{i}\xi_{j} \geq 0. \]

If we consider just one term in this double sum, say the one involving \( m_{i} \) and \( m_{j} \), then by \( H \)-invariance of the spherical distance \( d \) and Lemma 7.1
\[
\int_{H} \int_{H} f( (h_{i}m_{i}, h_{j}m_{j}) ) \, dh_{i}dh_{j} = \int_{H} f( (m_{i}, hm_{j}) ) \, dh = f( x_{i} \ast x_{j} ),
\]
where \( x_{i} = d( O, m_{i} ) \) is interpreted as an element of the hypergroup \( K \). This shows that
\[ \sum_{i,j=1}^{n} f( x_{i} \ast x_{j} ) \xi_{i}\xi_{j} \geq 0 \]
for all choices of \( x_{i} \in K, \xi_{i} \in \mathbb{C} \) and \( n \in \mathbb{N} \).

\[ \Box \]

Remark 7.3. The proof immediately extends to rank-one symmetric spaces.

8. Hyperbolic Hypergroups

The hyperbolic hypergroup \(( \mathbb{R}_{+}, \ast \) convolution is given by (with the notation in [2], 3.5.65, p.237)
\[
\epsilon_{x} \ast \epsilon_{y} = \frac{\Gamma( b + 1 )}{\sqrt{\pi} \Gamma( b + \frac{1}{2} )} \int_{0}^{\pi} \epsilon_{\text{arccosh}( \cosh x \cosh y + \cos t \sinh x \sinh y )} \sin^{2b} t \, dt,
\]
where \( b = \rho/2 \) and \( \rho > 0 \) so that \( A(x) = \sinh^{2b} t = \sinh^{4b} t \). In [13] the convolution (taking into account the different notation) is exactly the same but with \( \rho > -1 \). However in the body of this paper (p.217), (8.1) takes the form
\[
\epsilon_{x} \ast \epsilon_{y} = \frac{\Gamma( \beta + \frac{1}{2} )}{\sqrt{\pi} \Gamma( \beta )} \int_{-1}^{1} \epsilon_{\text{arccosh}( \cosh x \cosh y + u \sinh x \sinh y )} \left( 1 - u^{2} \right)^{\beta - 1} \, du,
\]
which is easily obtained from (8.1) with the substitution \( u = \cos t \) and replacing \( b \) by \( \beta - \frac{1}{2} \), in which case \( b > -\frac{1}{2} \) (that is \( \rho > -1 \)) is equivalent to \( \beta > 0 \).

We now consider explicit representations for these characters. In [13] the function
\[
\psi^{\beta}_{\lambda}( x ) = \frac{\Gamma( \beta + \frac{1}{2} )}{\sqrt{\pi} \Gamma( \beta )} \int_{-1}^{1} (\cosh x + u \sinh x)^{\sqrt{\beta^{2} - \lambda}\beta - \beta ( 1 - u^{2} )^{\beta - 1} \, du
\]
satisfies \( \psi^{\beta}_{\lambda}( 0 ) = \| \psi^{\beta}_{\lambda} \|_{\infty} = 1 \) when \( 4\beta^{2} \text{Re} \lambda \geq (\text{Im} \lambda)^{2} \) (in which case \( \psi^{\beta}_{\lambda} \) is bounded), \( \psi^{\beta}_{\lambda} \) is real-valued when \( \lambda \geq 0 \) (in which case \( \psi^{\beta}_{\lambda} \) is a character) and \( \psi^{\beta}_{\lambda} \) is non-negative when \( \lambda \in [0, \beta^{2}] \). With the change of variable \( u = \cos t \) we obtain
\[
\psi^{\beta}_{\lambda}( x ) = \frac{\Gamma( \beta + \frac{1}{2} )}{\sqrt{\pi} \Gamma( \beta )} \int_{0}^{\pi} (\cosh x + \cos t \sinh x)^{\sqrt{\beta^{2} - \lambda}\beta - \beta \sin^{2\beta - 1} t} \, dt.
\]
(8.2)
The integrand in (8.2) is just that given for the toroidal function

\[ P^{\mu}_{\nu - \frac{1}{2}}(\cosh x) = \frac{\Gamma(\nu + \mu + \frac{1}{2}) \sinh^\mu x}{\Gamma(\nu - \mu + \frac{1}{2}) 2^\mu \sqrt{\pi} \Gamma(\mu + \frac{1}{2})} \int_0^\pi \frac{\sin^{2\mu} t}{(\cosh x + \cos t \sinh x)^{\nu + \mu + \frac{1}{2}}} dt \]

in [3], Section 3.13, p.173.

Now the above parametrisation of the characters of \((\mathbb{R}_+, \ast)\) in [13] differs from that in [2]. In [13] we have \(\mathbb{R}_+ \cong \mathbb{R}_+\), where the supplementary series corresponds to \([0, \beta^2]\) and the principal series \([\beta^2, \infty)\), whereas in [2] the dual space is given by

\[ \mathbb{R}_+ \cong \mathbb{R}_+ \cup i[0, \rho], \]

where the supplementary series corresponds to \(i[0, \rho]\) and the principal series \([0, \infty)\). The correspondence is easily described as

<table>
<thead>
<tr>
<th>Zeuner</th>
<th>Correspondence</th>
<th>Supplementary</th>
<th>Principal</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\psi^{\rho}_{\lambda^2 + \rho^2})</td>
<td>(\uparrow)</td>
<td>([0, \rho^2])</td>
<td>([\rho^2, \infty))</td>
<td>(\uparrow)</td>
</tr>
<tr>
<td>Bloom/Heyer</td>
<td>(\phi^\rho_{\lambda})</td>
<td>(i[0, \rho])</td>
<td>([0, \infty))</td>
<td>(\phi^\rho_{i\rho})</td>
</tr>
</tbody>
</table>

The difference is that the characters in [2] are solutions to the Sturm-Liouville equation

\[ L^\lambda_{\phi^\rho} = (\lambda^2 + \rho^2) \phi^\rho_{\lambda}, \quad \phi^\rho_{\lambda}(0) = 1, \quad (\phi^\rho_{\lambda})'(0) = 0 \] (8.3)

whereas those in [13] solve the equation

\[ L^\beta_{\psi^\lambda} = \lambda \psi^\lambda_{\beta}, \quad \psi^\lambda_{\beta}(0) = 1, \quad (\psi^\lambda_{\beta})'(0) = 0. \] (8.4)

In both cases the differential operator is given by

\[ L^\lambda_{\phi^\rho} f = -f'' - \frac{A^\rho}{A^\lambda} f' = -f'' - 2\alpha \frac{\cosh}{\sinh} f'. \]

9. Positive Definiteness on the Naimark Hypergroup

In the case \(\rho = 1\) (so that \(b = \frac{1}{2}\) and \(A(x) = \sinh^2 x\)) our hyperbolic hypergroup is the well-known Naimark hypergroup, which has convolution

\[ \epsilon x \ast \epsilon y = \frac{1}{2 \sinh x \sinh y} \int_{|x-y|}^{x+y} \epsilon t \sinh t dt. \] (9.1)

Indeed, with the substitution

\[ \cosh u = \cosh x \cosh y + \sinh x \sinh y \cos t \]

we have

\[ \sinh u \frac{du}{dt} = -\sinh x \sinh y \sin t, \]

and \(t = 0\) gives

\[ \cosh u = \cosh x \cosh y + \sinh x \sinh y = \cosh(x + y), \]

so that \(u = x + y\), and \(t = \pi\) gives

\[ \cosh u = \cosh x \cosh y - \sinh x \sinh y = \cosh(x - y), \]
so that \( u = |x - y| \). Thus the integral (6.1) becomes (with \( n = 3 \))

\[
\varepsilon_x \ast \varepsilon_y = \frac{1}{2} \int_{x+y}^{x-y} \varepsilon_u \sinh u (-\sinh x \sinh y)^{-1} \, du
\]

\[
= \frac{1}{2 \sinh x \sinh y} \int_{|x-y|}^{x+y} \varepsilon_u \sinh u \, du.
\]

The characters of the Naimark hypergroup are given by

\[
\phi^1_\lambda (x) = \begin{cases} 
\sin \lambda x / \lambda \sinh x, & \lambda \neq 0, \\
\frac{x}{\sinh x}, & \lambda = 0.
\end{cases}
\]

Note that with \( \lambda = \beta = 1 \) we have

\[
\psi^1_1 (x) = \frac{x}{\sinh x} = \phi^1_0 (x).
\]

Now it should be observed that for \( \lambda \in (0, \infty) \) the characters oscillate, whereas for \( \lambda = i \xi \) where \( 0 < \xi \leq 1 \) we have

\[
\phi^1_{i \xi} (x) = \sin \frac{i \xi x}{i \xi \sinh x} = \frac{\sinh \xi x}{\xi \sinh x} \geq 0.
\]

In both cases (except when \( \xi = 1 \)) the characters belong to \( C_0 (\mathbb{R}_+) \). All characters are automatically positive definite.

Now Zeuner ([13], Proposition 6.1) has shown that the product of two characters on the Naimark hypergroup is also positive definite. Since continuous positive definite functions are themselves positive mixtures of characters it follows that the product of two continuous positive definite functions on the Naimark hypergroup is also positive definite.

This result indicates that the positive definite functions on the Naimark hypergroup satisfy the stronger multiplicative property (3) of Schoenberg positive definiteness, and indeed it is to be expected that the two notions of positive definiteness would coincide for this hyperbolic hypergroup, but this remains open.

### 10. Convolution Measures in Higher Dimensions

Our formulae for the explicit hypergroup structures on spheres and classical hyperbolic spaces are worth looking at separately for the interesting dependence on the dimension that they exhibit. For example, the formula that describes the measures appearing in the hypergroup structure on the \( n \) dimensional sphere

\[
g^{(n)}_{x,y} (r) = \frac{\sin r \left[ (\cos (x - y) - \cos r) (\cos r - \cos (x + y)) \right]^{n-3}}{c_n \left[ \sin x \sin y \right]^{n-2}},
\]

where \( c_n \) is given by (4.1), has quite a different character for different values of \( n \). This connects with the well-known geometrical understanding that low-dimensional spheres have a lot of surface area around the poles, while higher-dimensional spheres have a lot of surface area around the equator. We consider two examples, with both distributions centring on \( \cos^{-1} (\cos 1 \cos 2) \simeq 1.7976 \). The first is \( g^{(5)}_{1,2} (r) \) and the second is \( g^{(1500)}_{1,2} (r) \). Both are probability measures obtained
by convolving circles of radii 1 and 2 on respectively a 5-dimensional sphere and 1500-dimensional sphere. The distribution in the latter case concentrates much more on the value \(\cos^{-1}(\cos 1 \cos 2)\).

The corresponding convolutions in hyperbolic space exhibit a similar phenomenon. The measure appearing in the hypergroup structure on the \(n\)-dimensional classical hyperbolic space is given by

\[
m_{x,y}^{(n)}(r) = \frac{\sinh r \left[(\cosh r - \cosh (x - y)) (\cosh (x + y) - \cosh r)\right]^{\frac{n-3}{2}}}{[\sinh x \sinh y]^{n-2}}.
\]

The probability measures \(m_{1,2}^{(5)}(r)\) and \(m_{1,2}^{(1500)}(r)\) concentrate on

\[
\cosh^{-1}(\cosh 1 \cosh 2) \simeq 2.4444.
\]

They are obtained by convolving circles of radii 1 and 2 on respectively a 5-dimensional classical hyperbolic space and 1500-dimensional hyperbolic space.

It is natural to ask what happens in the limit with these measures as the dimension \(n\) goes to infinity.

### 11. Infinite Limits

We first consider the convolution of spheres in \(\mathbb{R}^n\).

**Proposition 11.1.** For \(x, y > 0\)

\[
\lim_{n \to \infty} 2r \frac{c_n}{\Gamma(n/2)} \frac{\left[(r^2 - (x-y)^2) \left( (x+y)^2 - r^2 \right) \right]^{\frac{n-3}{2}}}{(2xy)^{n-2}} \chi_{\{x,y\}}(r) = \varepsilon \sqrt{x^2 + y^2}, \tag{11.1}
\]

where the limit is taken distributionally.

**Proof.** First observe that \( |x-y| \leq \sqrt{x^2 + y^2} \leq x + y \) for all \(x, y \geq 0\). We consider

\[
\int_{|x-y|}^{x+y} \frac{2r}{c_n} \frac{\left[(r^2 - (x-y)^2) \left( (x+y)^2 - r^2 \right) \right]^{\frac{n-3}{2}}}{(2xy)^{n-2}} dr
\]

\[
= \frac{1}{c_n (2xy)^{n-2}} \int_{(x-y)^2}^{(x+y)^2} \frac{u - (x-y)^2}{(x+y)^2 - u} \left( (x+y)^2 - u \right)^{\frac{n-3}{2}} du
\]

\[
= \frac{\Gamma \left( \frac{n}{2} \right)}{\Gamma \left( \frac{n-1}{2} \right)} \frac{\Gamma \left( \frac{n-1}{2} \right)^2 (2xy)^{n-2}}{\sqrt{\pi} \Gamma(n-1)} = 1,
\]

the latter step using the Legendre duplication formula, so that the left-hand side of (11.1) is a limit of probability density functions on \(\mathbb{R}_+\). We show that as a pointwise limit this vanishes for \(r \neq \sqrt{x^2 + y^2}\) and the result will then follow.

For \(r \neq \sqrt{x^2 + y^2}\) we first consider

\[
|x-y| < r < \sqrt{x^2 + y^2}.
\]
Then $r^2 = x^2 + y^2 - \gamma$ for some $\gamma \in (0, 2xy)$ and

$$2r \left[ \left( r^2 - (x - y)^2 \right) \left( (x + y)^2 - r^2 \right) \right]^{\frac{n-2}{2}}$$

$$= \frac{2\sqrt{x^2 + y^2 - \gamma}}{c_n} \left[ \left( x^2 + y^2 - \gamma - (r_1 - r_2)^2 \right) \left( (x + y)^2 - (x^2 + y^2 - \gamma) \right) \right]^{\frac{n-2}{2}}$$

$$= \frac{2\sqrt{x^2 + y^2 - \gamma}}{c_n} \left[ (2xy - \gamma)(2xy + \gamma) \right]^{\frac{n-2}{2}}$$

$$= \frac{2\sqrt{x^2 + y^2 - \gamma}}{c_n} \left[ 4x^2y^2 - \gamma^2 \right]^{\frac{n-3}{2}}$$

$$= \frac{\sqrt{x^2 + y^2 - \gamma}}{xy\sqrt{\pi}} \left[ 1 - \frac{\gamma^2}{4x^2y^2} \right]^{\frac{n-3}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

$$(11.2)$$

$$\sim \frac{\sqrt{x^2 + y^2 - \gamma}}{xy\sqrt{\pi}} \left[ 1 - \frac{\gamma^2}{4x^2y^2} \right]^{\frac{n-3}{2}} \left( \frac{n}{2} \right)^{\frac{1}{2}}$$

$$(11.3)$$

$$\to 0 \text{ as } n \to \infty,$$

where for (11.2) we refer to (4.1), and for (11.3) see in [3], 1.18(5). A similar argument holds for $\sqrt{x^2 + y^2} < r < x + y$. 

The corresponding result, with an analogous proof, for the convolution of hypercircles in $\mathbb{S}^n$ is the following.

**Proposition 11.2.** The limit

$$\lim_{n \to \infty} \frac{\sin r \left[ (\cos (x - y) - \cos r) \cos r - \cos (x + y) \right]^{\frac{n-3}{2}}}{c_n} \chi_{|x-y|,x+y}(r)$$

$$= \varepsilon_{\cos^{-1}(\cos x \cos y)}$$

exists distributionally.

Similarly, for the convolution of hypercircles in $\mathbb{H}^n$ we can provide an analogous computation to show the following.
Proposition 11.3. The limit

\[
\lim_{n \to \infty} \frac{\sinh r \left[ (\cosh (x - y) - \cosh r) (\cosh r - \cosh (x + y)) \right]^{\frac{n-3}{2}}}{(\sinh x \sinh y)^n} \chi_{[|x-y|, x+y]}(r) = \varepsilon \cosh^{-1} (\cosh x \cosh y)
\]

exists distributionally.

12. Schoenberg Sets of Positive Definite Functions

We can relate the above calculations to a semigroup structure in the limiting cases. Let \( P_n = P_n(S^n) \) denote the set of functions on \([0, \pi]\) that are of the form

\[
f(r) = \sum_{k=0}^\infty a_k \psi_k^n (r) \quad \text{where} \quad a_k \geq 0, \quad \sum_{k=0}^\infty a_k < \infty.
\]

(In [1] it is observed that the definition of these sets is valid for arbitrary strictly positive \( n \) using the definition of the Gegenbauer polynomials given above.) While this definition does not extend to the case \( n = \infty \), it is true that \( \psi_k^n (r) \to (\cos r)^k \) as \( n \to \infty \). This leads us to define \( P_\infty \) to be the set of functions on \([0, \pi]\) that are of the form

\[
f(r) = \sum_{k=0}^\infty a_k (\cos r)^k \quad \text{where} \quad a_k \geq 0, \quad \sum_{k=0}^\infty a_k < \infty,
\]

which is consistent with Schoenberg's result that \( P_\infty \) so defined consists exactly of the positive definite functions on \([0, \pi]\) for the metric space \((S^n, d)\).

From the general considerations of Schoenberg it follows that \( P_n \) is a semigroup for all values \( n = 1, 2, \cdots, \infty \) and that

\[ P_1 \supset P_2 \supset P_3 \cdots \supset P_\infty. \]

From what we have observed for any positive integer \( n \), \( P_n \) consists of exactly all the positive definite functions on the hypergroup

\[ K^{(n)} = SO (n+1) / SO (n) \cong [0, \pi]. \]

It is then reasonable to ask if there is a hypergroup \( K^{(\infty)} \) associated with meridian ‘circles’ on the infinite-dimensional sphere \( S^\infty \) and whether \( P_\infty \) is then the set of positive definite functions on \( K^{(\infty)} \).

The problem with the definition of the hypergroup \( K^{(\infty)} \) is that although \( S^{(\infty)} \) is a homogeneous space for an infinite-dimensional Lie group there is no invariant probability measure on it that will readily allow a convenient notion of convolution. Nevertheless there is in a precise sense a limiting object of the hypergroups \( K^{(n)} \) as \( n \to \infty \), which we see is not a hypergroup but rather a semigroup, and its positive definite functions are then exactly \( P_\infty \). The semigroup multiplication is given by

\[ x \cdot y := \cos^{-1} (\cos x \cos y), \]

which is a consequence of Proposition 11.2.

Conjecture 12.1. It would be expected that a similar observation would be valid for \( P_n(H^n) \) for which a hyperbolic version of the Schoenberg results would be needed. At this stage the question is still open.
References

WALTER R. BLOOM: SCHOOL OF ENGINEERING AND INFORMATION TECHNOLOGY, MURDOCH UNIVERSITY, PERTH, WESTERN AUSTRALIA 6150, AUSTRALIA
E-mail address: w.bloom@murdoch.edu.au
URL: http://www.murdoch.edu.au/

N. J. WILDBERGER: SCHOOL OF MATHEMATICS, UNIVERSITY OF NEW SOUTH WALES, SYDNEY, NSW 2052, AUSTRALIA
E-mail address: n.wildberger@unsw.edu.au
URL: http://web.maths.unsw.edu.au/~norman/
GENERALIZED COMMUTATIVE ASSOCIATION SCHEMES, HYPERGROUPS, AND POSITIVE PRODUCT FORMULAS

MICHAEL VOIT

Abstract. It is well known that finite commutative association schemes in the sense of the monograph of Bannai and Ito lead to finite commutative hypergroups with positive dual convolutions and even dual hypergroup structures. In this paper we present several discrete generalizations of association schemes which also lead to associated hypergroups. We show that discrete commutative hypergroups associated with such generalized association schemes admit dual positive convolutions at least on the support of the Plancherel measure. We hope that examples for this theory will lead to the existence of new dual positive product formulas in near future.

1. Motivation

The following setting appears quite often in the theory of Gelfand pairs, spherical functions, and associated special functions:

Let \((G_n, H_n)_{n \in \mathbb{N}}\) be a sequence of Gelfand pairs, i.e., locally compact groups \(G_n\) with compact subgroups \(H_n\) such that the Banach algebras \(M_b(G_n||H_n)\) of all bounded, signed \(H_n\)-biinvariant Borel measures on \(G_n\) are commutative. Assume also that the double coset spaces \(G_n//H_n := \{ H_n g H_n : g \in G_n \}\) with the quotient topology are homeomorphic with some fixed locally compact space \(D\). Then the space \(M_b(D)\) carries the canonical double coset hypergroup convolutions \(*_n\).

A non-trivial \(H_n\)-biinvariant continuous function \(\varphi_n \in C(G_n)\) is called spherical if the product formula

\[
\varphi_n(g)\varphi_n(h) = \int_{H_n} \varphi_n(ghk) \, d\omega_{H_n}(k) \quad (g, h \in G_n)
\]

(1.1)

holds with the normalized Haar measure \(\omega_{H_n}\) of \(H_n\). Via \(G_n//H_n \cong D\), we may identify the spherical functions of \((G_n, H_n)\) with the nontrivial continuous functions on \(D\), which are multiplicative w.r.t. \(*_n\).

For all relevant examples of such series \((G_n, H_n)_{n}\), the spherical functions are parameterized by some spectral parameter set \(\chi(D)\) independent on \(n\), and the associated functions \(\varphi_n : \chi(D) \times D \to \mathbb{C}\) can be embedded into a family of special functions which depend analytically on \(n\) in some domain \(A \subset \mathbb{C}\), where these functions are spherical for some integers \(n\). In many cases, we can determine

Received 2016-8-31; Communicated by D. Applebaum.

2010 Mathematics Subject Classification. Primary 43A62; Secondary 05E30, 33C54, 33C67, 20N20, 43A90.

Key words and phrases. Association schemes, Gelfand pairs, hypergroups, Hecke pairs, spherical functions, positive product formulas, dual convolution, distance-transitive graphs.
these special functions and obtain concrete versions of the product formula (1.1), in which $n$ appears as a parameter. Based on Carleson’s theorem, a principle of analytic continuation (see e.g. [32], p.186), it is often easy to extend the positive product formula for $\varphi_n$ in the group cases to a continuous range of parameters. Usually, this extension leads to a continuous family of commutative hypergroups.

Classical examples of such positive product formulas are the well-known product formulas of Gegenbauer for the normalized ultraspherical polynomials

$$F_k^{(\alpha,\alpha)}(x) = 2F_1(2\alpha + k + 1, -k, \alpha + 1; (1 - x)/2) \quad (k \in \mathbb{N}_0) \quad (1.2)$$
on D = [-1,1] \text{ with } \chi(D) = \mathbb{N}_0 \text{ and for the modified Bessel functions}$$

$$\Lambda_\alpha(x,y) := j_\alpha(xy) \quad \text{with} \quad j_\alpha(z) := {}_0F_1(\alpha + 1; -z^2/4) \quad (y \in \mathbb{C}) \quad (1.3)$$
on $D = [0, \infty]$ with $\chi(D) = \mathbb{C}$; see e.g. the survey [1]. In both cases, this works for $\alpha \in [-1/2, \infty]$ where $\alpha = (n-1)/2$ corresponds to the Gelfand pair $(G_n,H_n)$ with $G_n = SO(n+2)$, $H_n = SO(n+1)$ and $G_n = SO(n+1) \times \mathbb{R}^{n+1}$, $H_n = SO(n+1)$ respectively. The continuous ultraspherical product formula can be extended to Jacobi polynomials [17] which generalizes the product formulas for the spherical functions of the projective spaces over $\mathbb{F} = \mathbb{R}, \mathbb{C}$ and the quaternions $\mathbb{H}$. Further prominent semisimple, rank one examples are the Gelfand pairs associated with the hyperbolic spaces over $\mathbb{F} = \mathbb{R}, \mathbb{C}, \mathbb{H}$ with the groups

$$\begin{align*}
\mathbb{F} = \mathbb{R} : & \quad G = SO_0(1,k), \quad K = SO(k) \\
\mathbb{F} = \mathbb{C} : & \quad G = SU(1,k), \quad K = SU(1) \times SU(k)) \\
\mathbb{F} = \mathbb{H} : & \quad G = Sp(1,k), \quad K = Sp(1) \times Sp(k).
\end{align*}$$

Here, $D = [0, \infty]$ with $\chi(D) = \mathbb{C}$ where the spherical functions are Jacobi functions [23]. Besides these classical rank-one examples, reductive examples as well as several examples of higher rank were studied; see e.g. [19] and references there for disk polynomials as well as [24], [28], [29], [30], [31], [38], [39] in the higher rank case, and [18] for hypergeometric functions associated with root systems. Moreover, series of discrete examples were studied in the setting of trees, graphs, buildings, association schemes, Hecke pairs, and other discrete structures; see e.g. [3], [4], [9], [8], [12], [25], [27], [37] where sometimes the connection to hypergroups is suppressed. This discrete setting forms the main topic of this paper.

Before going into details, we return to the general setting. Besides positive product formulas for $\varphi_n(\lambda,.)$ on $D$ originating from (1.1), there exist dual product formulas for the functions $\varphi_n(.,x) (x \in D)$ on suitable subsets of $\chi(D)$ for the group cases. To explain this, consider the closed set $P_n(D) \subset \chi(D)$ of spectral parameters $\lambda$ for which $\varphi_n(\lambda,.) \in C(D)$ corresponds to a positive definite function on $G_n$. For $\lambda_1, \lambda_2 \in P_n(D)$, then $\varphi_n(\lambda_1,.) \cdot \varphi_n(\lambda_2,.)$ is also positive definite on $G_n$, which implies that $\varphi_n(\lambda_1,.) \cdot \varphi_n(\lambda_2,.)$ is positive definite on the double coset hypergroup $D$. Therefore, by Bochner’s theorem for commutative hypergroups (see [20]), there exists a unique probability measure $\mu_{n,\lambda_1,\lambda_2}$ on $P_n(D)$ with the dual product formula

$$\varphi_n(\lambda_1,x) \cdot \varphi_n(\lambda_2,x) = \int_{P_n(D)} \varphi_n(\lambda,x) \, d\mu_{n,\lambda_1,\lambda_2}(\lambda) \quad \text{for all} \quad x \in D. \quad (1.4)$$
For related results of harmonic analysis for Gelfand pairs and commutative hypergroups we refer to [7], [11], [13], [20]. The dual product formula (1.4) has an interpretation in terms of group representations of $G_n$ of class 1, and for several classes of examples there exist explicit formulas for $\mu_{n,\lambda_1,\lambda_2}$ which again can be extended to a continuous parameter range $n$ where usually the positivity remains available. This is for instance trivial for $G_n = SO(n+1) \ltimes \mathbb{R}^{n+1}$, $H_n = SO(n+1)$, with $P_n(D) = [0, \infty]$, where, due to the symmetry of $\Lambda_\alpha$ in $x, y$ in (1.3), the dual product formula agrees with the original one for $\alpha \in [-1/2, \infty]$. Moreover, for $G_n = SO(n+2)$, $H_n = SO(n+1)$, it is well known that $P_n(D) = \chi(D) = \mathbb{N}_0$, and that the dual formula (1.4) corresponds to the well-known positive product linearization for the ultraspherical polynomials for $\alpha \in ]-1/2, \infty[$ which is part of a famous explicit positive product linearization formula for Jacobi polynomials [16], [1]. On the other hand, for hyperbolic spaces, the positive dual product formula is more involved. Here $P_n(D)$ depends on $n$ (see [14], [15]) with $[0, \infty] \subset P_n(D) \subset [0, \infty) \cup i \cdot [0, \infty] \subset \mathbb{C}$. Moreover, the known explicit formulas for the Lebesgue densities of the measures $\mu_{n,\lambda_1,\lambda_2}$ via triple integrals for $\lambda_1, \lambda_2 \in [0, \infty]$ e.g. in [23] are quite involved.

This picture is typical for many Gelfand pairs. We only mention the Gelfand pairs and orthogonal polynomials associated with homogeneous trees and infinite distance transitive graphs with $D = \mathbb{N}_0$ and $\chi(D) = \mathbb{C}$ in [27], [37], where dual product formulas are computed by brute force. For examples of higher rank, the existence of positive dual product formulas seems to be open except for group cases and simple self dual examples like the Bessel examples above. Such self dual examples appear e.g. as orbit spaces when compact subgroups of $U(n)$ act on $\mathbb{R}^n$ which leads e.g. to examples associated with matrix Bessel functions in [29].

In summary, there exist many continuous families of commutative hypergroup structures with explicit convolutions, for which the multiplicative functions are known special functions, and where the existence of dual product formulas is unknown except for the group parameters. The intention of this paper is to start some systematic research beyond the group cases by using more general algebraic structures behind commutative hypergroup structures.

Commutative association schemes as in [4], [3] might form such algebraic objects where these schemes are defined as finite sets $X \neq \emptyset$ with some partition $D$ of $X \times X$ with certain intersection properties; see Section 3 for details. The most prominent examples appear as homogeneous spaces $X = G/H$ for subgroups $H$ of finite groups $G$. It is known that all finite association schemes lead to hypergroup structures on $D$ which are commutative if and only if so are the association scheme. In the group case $X = G/H$, this hypergroup is just the double coset hypergroup $G//K$. There exist commutative association schemes which do not appear as homogeneous spaces $X = G/H$ (see [4]) such that the class of commutative hypergroups associated with association schemes extends the class of commutative double coset hypergroups. Moreover, by Section 2.10 of [4], these commutative hypergroups admit positive dual product formulas. Therefore, finite commutative association schemes form a tool to establish dual positive product formulas for some finite commutative hypergroups beyond double coset hypergroups.
The aim of this paper is to show that certain generalizations of commutative association schemes also lead to commutative hypergroups with dual positive product formulas. We first study possibly infinite, commutative association schemes where the theory of [4] can be extended canonically. This obvious extension is very rigid as the use of partitions leads to a very few examples only. For a further extension, we observe that all association schemes admit \( \{0,1\} \)-valued adjacency matrices labeled by \( D \) which are stochastic after some renormalization. We translate the axioms of association schemes into a system of axioms for these matrices. As now the integrality conditions vanish, we obtain more examples. We show that also in this generalized case associated commutative hypergroups exist and that, under some restrictions, dual positive product formulas exist; see Section 5.

We expect that our approach may be extended to non-discrete spaces \( X \) where families of Markov kernels labeled by some locally compact space \( D \) instead of stochastic matrices are used. We shall study this non-discrete generalization in a forthcoming paper. It covers the group cases with \( X = G/H, D = G//H \) for Gelfand pairs \( (G,H) \) as well two known continuous series of examples with \( X = \mathbb{R}^2, D = [0,\infty] \) and \( X = S^2 \) the 2-sphere in \( \mathbb{R}^3, D = [-1,1] \) due to Kingman [22] and Bingham [6]. The associated commutative hypergroups on \( D \) with dual positive product formulas will be just the hypergroups associated with Bessel functions and ultraspherical polynomials mentioned above. Even if for these examples the dual positive product formulas are well known, these examples might be a hint that generalized continuous commutative association schemes might form a powerful tool to derive the existence of dual positive product formulas. We hope that this approach can be applied to certain Heckman-Opdam Jacobi polynomials of type BC and hypergeometric functions of type BC (see [18]), which generalize the spherical functions of compact and noncompact Grassmannians and for which continuous families of commutative hypergroup structures exist by [28] and [30].

This paper is organized as follows. In Section 2 we recapitulate some facts about commutative hypergroups. Section 3 is then devoted to possibly infinite association schemes and associated hypergroups where the definition remains very close to the classical one in [4]. We there in particular study the relations to the double coset hypergroups \( G//H \) for compact open subgroups \( H \) of \( G \) and for Hecke pairs \( (G,H) \). In Section 4 we prove that for all commutative hypergroups associated with such association schemes there exist positive dual convolutions and dual product formulas at least on the support of the Plancherel measure. In Section 5 we propose a discrete generalization of association schemes without integrality conditions. We show that under some conditions, many results of Sections 3 and 4 remain valid including the existence of dual product formulas.

2. Hypergroups

Hypergroups form an extension of locally compact groups. For this, remember that the group multiplication on a locally compact group \( G \) leads to the convolution \( \delta_x \ast \delta_y = \delta_{xy} \) (\( x,y \in G \)) of point measures. Bilinear, weakly continuous extension of this convolution then leads to a Banach-\( * \)-algebra structure on the Banach space \( M_b(G) \) of all signed bounded regular Borel measures with the total variation norm \( \|\cdot\|_{TV} \). In the case of hypergroups we only require a convolution \( \ast \) for measures.
which admits most properties of a group convolution. We here recapitulate some well-known facts; for details see [11], [20], [7].

**Definition 2.1.** A hypergroup $\langle D, \ast \rangle$ is a locally compact Hausdorff space $D$ with a weakly continuous, associative, bilinear convolution $\ast$ on the Banach space $M_b(D)$ of all bounded regular Borel measures with the following properties:

1. For all $x, y \in D$, $\delta_x \ast \delta_y$ is a compactly supported probability measure on $D$ such that the support $\text{supp}(\delta_x \ast \delta_y)$ depends continuously on $x, y$ w.r.t. the so-called Michael topology on the space of all compacta in $X$ (see [20] for details).
2. There exists a neutral element $e \in D$ with $\delta_x \ast \delta_e = \delta_e \ast \delta_x = \delta_x$ for $x \in D$.
3. There exists a continuous involution $x \mapsto \bar{x}$ on $X$ such that for all $x, y \in D$, $e \in \text{supp}(\delta_x \ast \delta_y)$ holds if and only if $y = \bar{x}$.
4. If for $\mu \in M_b(D)$, $\mu^-$ denotes the image of $\mu$ under the involution, then $(\delta_x \ast \delta_y)^- = \delta_{\bar{y}} \ast \delta_{\bar{x}}$ for all $x, y \in D$.

A hypergroup is called **commutative** if the convolution $\ast$ is commutative. It is called **symmetric** if the involution is the identity.

**Remark 2.2.**

1. The identity $e$ and the involution $\bar{\cdot}$ above are unique.
2. Each symmetric hypergroup is commutative.
3. For each hypergroup $\langle D, \ast \rangle$, $\langle M_b(D), \ast \rangle$ is a Banach-$\ast$-algebra with the involution $\mu \mapsto \mu^*$ with $\mu^*(A) := \mu(\bar{A})$ for Borel sets $A \subset D$.
4. For a second countable locally compact space $D$, the Michael topology agrees with the well-known Hausdorff topology; see [24].

The most prominent examples are double coset hypergroups $G//H := \{HgH : g \in G\}$ for compact subgroups $H$ of locally compact groups $G$:

**Example 2.3.** Let $H$ be a compact subgroup of a locally compact group $G$ with identity $e$ and with the unique normalized Haar measure $\omega_H \in M^1(H) \subset M^1(G)$, i.e. $\omega_H$ is a probability measure. Then the space

$$M_b(G||H) := \{\mu \in M_b(G) : \mu = \omega_H \ast \mu \ast \omega_H\}$$

of all $H$-biinvariant measures in $M_b(G)$ is a Banach-$\ast$-subalgebra of $M_b(G)$. With the quotient topology, $G//H$ is a locally compact space, and the canonical projection $p_G//H : G \to G//H$ is continuous, proper and open. Now consider the push forward (or image-measure mapping) $\tilde{p}_G//H : M_b(G) \to M_b(G//H)$ with $\tilde{p}_G//H(\mu)(A) = \mu(p_{G//H}^{-1}(A))$ for $\mu \in M_b(G)$ and Borel sets $A \subset G//H$. It is easy to see that $\tilde{p}_G//H$ is an isometric isomorphism between the Banach spaces $M_b(G||H)$ and $M_b(G//H)$ w.r.t. the total variation norms, and that the transfer of the convolution on $M_b(G||H)$ to $M_b(G//H)$ leads to a hypergroup $G//H, \ast$) with identity $HeH$ and involution $HgH \mapsto Hg^{-1}H$. For details see [20].

Let us consider some typical discrete double coset hypergroups $G//H$:

**Example 2.4.** Let $\Gamma$ be the vertex set of a locally finite, connected undirected graph with the graph metric $d : \Gamma \times \Gamma \to N_0 := \{0, 1, \ldots\}$. A bijective mapping $g : \Gamma \to \Gamma$ is called an automorphism of $\Gamma$ if $d(g(a), g(b)) = d(a, b)$ for all $a, b \in \Gamma$. 

ASSOCIATION SCHEMES AND HYPERGROUPS 565
Clearly, the set $\text{Aut}(\Gamma)$ of all automorphisms is a topological group w.r.t. the topology of pointwise convergence, i.e., we regard $\text{Aut}(\Gamma)$ as subspace of $\Gamma^\Gamma$ equipped with the product topology. Assume now that $\text{Aut}(\Gamma)$ acts transitively on $\Gamma$. It is well-known and easy to see that $\text{Aut}(\Gamma)$ is a totally disconnected locally compact group which contains the stabilizer subgroup $H_x \subset G$ of any $x \in \Gamma$ as a compact open subgroup. $\Gamma$ can be identified with $\text{Aut}(\Gamma)/H_x$, and the discrete orbit space $\Gamma/H_x := \{H_x(y) : y \in \Gamma\}$ with the discrete double coset space $\text{Aut}(\Gamma)/H_x$.

We also consider another kind of double coset hypergroups:

**Example 2.5.** Let $G$ be a discrete group with some subgroup $H$. $(G,H)$ is called a Hecke pair if the so-called Hecke condition holds, i.e., if each double coset $HgH$ $(g \in G)$ decomposes into an at most finite number of right cosets $g_1H,...,g_{\text{ind}(HgH)}H$ where $\text{ind}(HgH) \in \mathbb{N}$ is called the (right-)index of $HgH$. Hecke pairs are studied e.g. in [25], [26] where left coset are taken.

For Hecke pairs, $G//H$ carries a discrete hypergroup structure due to [26]. To describe the associated convolution, take $a,b \in G$ and consider the disjoint decompositions $HaH = \bigcup_{i=1}^{n_a} a_iH, HbH = \bigcup_{j=1}^{n_b} b_jH$. If we put $\mu(HcH) := |\{(i,j) : a_i b_j H = c H\}| \in \mathbb{N}_0$, then $\mu(HcH)$ is independent of the representative $c$ of $HcH$, and

$$\delta_{HaH} \ast \delta_{HbH} := \sum_{HcH \in G//H} \frac{\mu(HcH) \cdot \text{ind}(HcH)}{\text{ind}(HaH) \cdot \text{ind}(HbH)} \delta_{HcH}$$

$(a,b \in G)$ generates a hypergroup structure on $G//H$.

The notion of Haar measures on hypergroups is similar to groups:

**Definition 2.6.** Let $(D, \ast)$ be a hypergroup, $x, y \in D$, and $f \in C_c(D)$ a continuous function with compact support. We write $x f(y) := f(x \ast y) := \int_K f d(\delta_x \ast \delta_y)$ and $f_x(y) := f(y \ast x)$ where, by the hypergroup axioms, $f_x, x f \in C_c(D)$ holds.

A non-trivial positive Radon measure $\omega \in M^+(D)$ is called a left or right Haar measure if

$$\int_D x f \, d\omega = \int_D f \, d\omega \quad \text{or} \quad \int_D f_x \, d\omega = \int_D f \, d\omega$$

$(f \in C_c(D), x \in D)$ respectively. $\omega$ is called a Haar measure if it is a left and right Haar measure. If $(D, \ast)$ admits a Haar measure, then it is called unimodular.

The uniqueness of left and right Haar measures and their existence for particular classes are known for a long time by Dunkl, Jewett, and Spector; see [7] for details. The general existence was settled only recently by Chapovsky [10]:

**Theorem 2.7.** Each hypergroup admits a left and a right Haar measure. Both are unique up to normalization.

**Examples 2.8.** (1) Let $(D, \ast)$ be a discrete hypergroup. Then, by [20], left and right Haar measures are given by

$$\omega_l(\{x\}) = \frac{1}{(\delta_x \ast \delta_x)(\{x\})}, \quad \omega_r(\{x\}) = \frac{1}{(\delta_x \ast \delta_x)(\{x\})},$$

$(x \in D)$. 

Moreover (1) The functions on group with Haar measure \( \omega \) we collect some essential well-known results:

(2) A function \( f \) for \( \alpha \in \mathbb{C} \) and those which vanish at infinity respectively. Hence

\[ (\alpha(x \ast y)) = \alpha(x) \cdot \alpha(y) \]

for all \( x, y \in D \).

(4) Let \( \alpha \in D \) consider the Fourier transforms are given by

\[ f(x) = \int_D f(\alpha) \alpha(x) d\omega(x), \quad \hat{\mu}(\alpha) = \int_D \alpha(x) d\mu(x) \quad (\alpha \in \hat{D}) \]

with \( \hat{f} \in C_0(\hat{D}), \hat{\mu} \in C_0(\hat{D}) \) and \( \|\hat{f}\|_\infty \leq \|f\|_1, \|\hat{\mu}\|_\infty \leq \|\mu\|_{TV}. \)

(3) There exists a unique positive measure \( \pi \in M^+(\hat{D}) \) such that the Fourier transform \( \hat{\cdot} : C_0(\hat{D}) \cap L^2(\hat{D}) \to C_0(\hat{D}) \cap L^2(\hat{D}, \pi) \) is an isometry. \( \pi \) is called the Plancherel measure on \( \hat{D} \).

Notice that, different from l.c.a. groups, the support \( S := \text{supp} \pi \) may be a proper closed subset of \( \hat{D} \). Quite often, we even have \( 1 \not\in S \).

(5) If \( f \in C_0(D) \) is called positive definite on the hypergroup \( D \) for all \( n \in \mathbb{N}, x_1, \ldots, x_n \in D \) and \( c_1, \ldots, c_n \in \mathbb{C}, \sum_{k,l=1}^n c_k \bar{c}_l \cdot f(x_k \ast \bar{x}_l) \geq 0 \). Obviously, all characters \( \alpha \in \hat{D} \) are positive definite.

We collect some essential well-known results:

**Facts 2.10.** (1) (Theorem of Bochner, [20]) A function \( f \in C_0(D) \) is positive definite if and only if \( f = \hat{\mu} \) for some \( \mu \in M_0^+(\hat{D}) \). In this case, \( \mu \) is a probability measure if and only if \( \hat{\mu}(e) = 1 \).

(2) For \( f, g \in L^2(D) \), the convolution product \( f \ast g(x) := \int f(x \ast y) g(y) d\omega(y) \)

(\( x \in D \)) satisfies \( f \ast g \in C_0(D) \). Moreover, for \( f \in L^2(D) \), \( f^*(x) = \overline{f(x)} \)

satisfies \( f^* \in L^2(D) \), and \( f \ast f^* \in C_0(D) \) is positive definite; see [20], [7].

(3) A function \( f \in C_0(D) \) is the inverse Fourier transform \( \hat{\mu} \) for some \( \mu \in M_0^+(\hat{D}) \) with \( \text{supp} \mu \subset S \) if and only if \( f \) is the compact-uniform limit of positive definite functions of the form \( h \ast h^*, h \in C_c(D) \); see [35].

(4) Let \( \alpha \in \hat{D} \). Then \( \alpha \in S \) if and only if \( \alpha \) is the compact-uniform limit of positive definite functions of the form \( h \ast h^*, h \in C_c(D) \); see [35].
(5) There exists precisely one positive character $\alpha_0 \in S$ by [33], [7].
(6) If $\mu \in M^1(\hat{D})$ satisfies $\hat{\mu} \geq 0$ on $\hat{D}$, then its support $\text{supp } \mu$ contains at least one positive character; see [36].

In contrast to l.c.a. groups, products of positive definite functions on $D$ are not necessarily positive definite; see e.g. Section 9.1C of [20] for an example with $|D| = 3$. However, sometimes this positive definiteness of products is available.

If for all $\alpha, \beta \in \hat{D}$ (or a subset of $\hat{D}$ like $S$) the products $\alpha \beta$ are positive definite, then by Bochner’s theorem 2.10(1), there are probability measures $\delta_\alpha \hat{\ast} \delta_\beta \in M^1(\hat{D})$ with $(\delta_\alpha \hat{\ast} \delta_\beta)^\vee = \alpha \beta$, i.e., we obtain dual positive product formulas as claimed in Section 1. Under additional conditions, $(\hat{D}, \hat{\ast})$ then carries a dual hypergroup structure with $1$ as identity and complex conjugation as involution. This for instance holds for all compact commutative double coset hypergroups $G//H$ by [11]. For non-compact Gelfand pairs $(G, H)$ we have dual positive convolutions on $S$; see [20], [36]. These convolutions usually do not generate a dual hypergroup structure, and sometimes $\alpha \beta$ is not positive definite on $D$ for some $\alpha, \beta \in \hat{D}$; see Theorem 4.9 for an example.

3. Discrete Association Schemes

In this section we extend the classical notion of finite association schemes in a natural way. As references for finite association schemes we recommend [4], [3].

**Definition 3.1.** Let $X, D$ be nonempty, at most countable sets and $(R_i)_{i \in D}$ a disjoint partition of $X \times X$ with $R_e \neq \emptyset$ for $e \in D$ and the following properties:

1. There exists $e \in D$ with $R_e = \{(x, x) : x \in X\}$.
2. There exists an involution $i \mapsto \bar{i}$ on $D$ such that for $i \in D$, $R_{\bar{i}} = \{(y, x) : (x, y) \in R_i\}$.
3. For all $i, j, k \in D$ and $(x, y) \in R_k$, the number
   $$p^k_{i,j} := |\{z \in X : (x, z) \in R_i \text{ and } (z, y) \in R_j\}|$$
   is finite and independent of $(x, y) \in R_k$.

Then $\Lambda := (X, D, (R_i)_{i \in D})$ is called an association scheme with intersection numbers $(p^k_{i,j})_{i,j,k \in D}$ and identity $e$.

An association scheme is called commutative if $p^k_{i,j} = p^k_{j,i}$ for all $i, j, k \in D$. It is called symmetric (or hermitian) if the involution on $D$ is the identity. Moreover, it is called finite, if so are $X$ and $D$.

Finite association schemes above are obviously precisely association schemes in the sense of the monographs [4].

Association schemes have the following interpretation: We regard $X \times X$ as set of all directed paths from points in $X$ to points in $X$. The set of paths is labeled by $D$ which might be colors, lengths, or difficulties of paths. Then $R_e$ is the set of trivial paths, and $R_i$ is the set of all reversed paths in $R_i$. Axiom (3) is some kind of a symmetry condition and the central part of the definition.

Association schemes may be described with the aid of adjacency matrices:
Definition 3.2. The adjacency matrices \( A_i \in \mathbb{R}^{X \times X} \ (i \in D) \) of an association scheme \((X, D, (R_i)_{i \in D})\) are given by
\[
(A_i)_{x,y} := \begin{cases} 
1 & \text{if } (x, y) \in R_i \\
0 & \text{otherwise}
\end{cases} \quad (i \in D, x, y \in X).
\]
The adjacency matrices have the following obvious properties:

1. \( A_e \) is the identity matrix \( I_X \).
2. \( \sum_{i \in D} A_i \) is the matrix \( J_X \) whose entries are all equal to 1.
3. \( A_i^2 = A_i \) for \( i \in D \).
4. For all \( i \in D \) and all rows and columns of \( A_i \), all entries are equal to zero except for finitely many cases (take \( k = e, j = i \) in 3.1(3)).
5. For \( i, j \in D \), \( A_i A_j = \sum_{k \in D} P_{i,j}^k A_k \).
6. An association scheme \( \Lambda \) is commutative if and only if \( A_i A_j = A_j A_i \) for all \( i, j \in D \).
7. An association scheme \( \Lambda \) is symmetric if and only if all \( A_i \) are symmetric.

In particular, each symmetric association scheme is commutative.

Definition 3.3. Let \((X, D, (R_i)_{i \in D})\) be an association scheme. The valency of \( R_i \) or \( i \in D \) is defined as \( \omega_i := p_{i,i}^e \). Obviously, the \( \omega_i \) satisfy
\[
\omega_i = |\{ z \in X : (x, z) \in R_i \}| \in \mathbb{N} \tag{3.1}
\]
further, \( \omega_e = 1 \), and
\[
|X| = \sum_{i \in K} \omega_i \in \mathbb{N} \cup \{ \infty \}. \tag{3.2}
\]

Remark 3.4. For \( i \in D \), the renormalized matrices \( S_i := \frac{1}{\omega_i} A_i \in \mathbb{R}^{X \times X} \) are stochastic, i.e., all rows sum are equal to 1 by (3.1). Notice that also all column sums are finite by 3.2(4). Moreover, by 3.2(5), the stochastic matrices \( S_i \) satisfy
\[
S_i S_j = \sum_{k \in D} \frac{\omega_i}{\omega_j} p_{i,j}^k S_k \quad \text{for} \quad i, j \in D. \tag{3.3}
\]
Since both sides of (3.3) are stochastic, absolute convergence leads to
\[
\sum_{k \in D} \frac{\omega_i}{\omega_j} p_{i,j}^k = 1 \quad \text{for} \quad i, j \in D. \tag{3.4}
\]
The formulas (3.3) and (3.4) are the starting point in Section 5 for the extension of association schemes to a continuous setting.

We now collect some relations about the intersection numbers and valencies:

Lemma 3.5. For all \( i, j, k, l, m \in D \):

1. \( p_{e,i}^l = p_{i,e}^l = \delta_{i,j} \) and \( p_{i,j}^m = \omega_i \delta_{i,j} \).
2. \( p_{i,j}^l = p_{j,i}^l \).
3. \( \sum_{l \in K} \omega_l \bar{p}_{i,j}^l = \omega_i \) for all \( l \in D \), and, in particular, for all \( l, i \in D \), \( p_{i,j}^l \neq 0 \) holds for finitely many \( j \in K \) only.
4. \( \omega_i \bar{p}_{i,j}^l = \omega_i \bar{p}_{i,j}^l \) and \( \omega_j \bar{p}_{i,j}^l = \omega_j \bar{p}_{i,j}^l \).
5. \( \sum_{l \in K} \omega_l \bar{p}_{i,j}^l = \omega_i \omega_j \) and \( \sum_{l \in K} \omega_l \bar{p}_{i,j}^l = \omega_i \omega_j \).
6. \( \sum_{l \in K} \bar{p}_{i,j}^l \cdot \bar{p}_{k,l}^m = \sum_{l \in K} \bar{p}_{i,j}^l \cdot \bar{p}_{k,l}^m \).
7. If \( \bar{p}_{i,j}^k > 0 \), then \( \frac{\omega_k}{\omega_i} = \frac{\omega_i}{\omega_j} = \frac{\omega_j}{\omega_i} \).

ASSOCIATION SCHEMES AND HYPERGROUPS 569
Proof. Part (1) is obvious, and part (2) follows from
\[
\sum_i p_{i,j}^T A_i = A_j^T A_i = (A_i A_j)^T = \sum_i p_{i,j}^T A_i = \sum_i p_{i,j} A_i
\]
and a comparison of coefficients. Part (3) is a consequence of
\[
\omega_i J_X = A_i J_X = \sum_j A_i A_j = \sum_j \sum_i p_{i,j} A_i = \sum_i \left(\sum_j p_{i,j}\right) A_i.
\]
For the proof of the first statement of (4), notice that by part (1), for \(x \in X\)
\[
\omega_l \cdot p_{i,j} = \sum_k p_{i,j}^k (A_k A_l)_{x,x} = (A_i A_j A_l)_{x,x} = \sum_k p_{j,i}^k (A_i A_k)_{x,x} = \omega_i \cdot p_{j,i}^k = \omega_i \cdot p_{i,j}^k;
\]
the second statement of part (4) follows in a similar way. Moreover, the statements in (5) are consequences of
\[
\omega_1 \cdot \omega_j J_X = A_i A_j J_X = \sum_i p_{i,j} A_i J_X = \sum_i p_{i,j} \cdot \omega_l J_X
\]
and
\[
\omega_1 \cdot \omega_j J_X = J_X A_i A_j = \sum_i p_{i,j} J_X A_i = \sum_i p_{i,j} \cdot \omega_l J_X.
\]
Part (6) follows from \((A_i A_j) A_k = A_i (A_j A_k)\) and comparison of coefficients in the expansions. Finally, parts (4) and (2) imply
\[
\frac{\omega_k}{\omega_1 \cdot \omega_j} p_{i,j}^k = \frac{p_{i,j}^k}{\omega_j} = \frac{p_{j,i}^k}{\omega_i} = \frac{\omega_k}{\omega_i \omega_j} p_{i,j}^k,
\]
which proves (7).
\[\square\]

Typical examples of finite or infinite association schemes are connected with homogeneous spaces \(G/H\) in one of the following two ways:

Example 3.6. Let \(G\) be a second countable, locally compact group with a compact open subgroup \(H\) and with neutral element \(e\). Then the quotient \(X := G/H\) as well as the double coset space \(D := G//H\) are at most countable, discrete spaces w.r.t. the quotient topology. Consider the partition \((R_i)_{i \in D}\) of \(X \times X\) with
\[
R_{HgH} := \{ (xH,yH) \in X \times X : HgH = Hx^{-1}yH \}.
\]
Then \((X,D,(R_i)_{i \in D})\) forms an association scheme with neutral element \(HeH = H\) and involution \(H^{-1}H\). In fact, the axioms (1), (2) in 3.1 are obvious. For axiom (3) consider \(g,h,x,y,\tilde{x},\tilde{y} \in G\) with \((xH,yH)\) and \((\tilde{x}H,\tilde{y}H)\) in the same partition \(R_{Hx^{-1}yH}\), i.e., with \(H\tilde{x}^{-1}\tilde{y}H = Hx^{-1}yH\). Thus, there exist \(h_1, h_2 \in H\) and \(w \in G\) with \(\tilde{x}^{-1} = h_1 x^{-1} w^{-1}\) and \(\tilde{y} = wyh_2\). Therefore, for any \(zH \in X, H\tilde{x}^{-1}zH = Hx^{-1}w^{-1}zH\) and \(Hz^{-1}yH = H(w^{-1}z)^{-1}H\), which means that \(zH \mapsto w^{-1}zH\) establishes a bijective mapping between \(\{ zH : Hx^{-1}zH = HgH, Hx^{-1}yH = HhH \}\) and \(\{ zH : H\tilde{x}^{-1}yH = HgH, Hz^{-1}yH = HhH \}\).

This shows that the intersection number \(p_{HgH,HhH}^{Hx^{-1}yH}\) is independent of the choice of \(x, y\). Moreover, due to compactness, each double coset \(HgH\) decomposes into finitely many cosets \(xH\) which shows that \(\{ zH : Hx^{-1}zH = HgH \}\) and thus the intersection number \(p_{HgH,HhH}^{Hx^{-1}yH}\) is finite as claimed.

It follows from the definition of \(R_{HgH}\) and Eq. (3.1) that for \(g \in G\) the valency \(\omega_{HgH}\) of the double coset \(HgH \in G//H\) is given by the finite number of different cosets \(xH \in G/H\) contained in \(HgH\).
The preceding example also works for Hecke pairs:

**Example 3.7.** Let \((G, H)\) be a Hecke pair, i.e., each double coset \(HgH\) decomposes into finitely many cosets \(xH\). Then the same partition as in Example 3.6 leads to an association scheme \((G/H, G//H, (R_j)_{j \in D})\). We skip the proof. We remark that again the valency \(\omega_{HgH}\) of a double coset \(HgH \in G//H\) is \(\text{ind}(HgH)\), i.e., the number of cosets \(xH \in G/H\) contained in \(HgH\).

Association schemes lead to discrete hypergroups. The associated convolution algebras are just the Bose-Mesner algebras for finite association schemes in [4].

**Proposition 3.8.** Let \(\Lambda := (X, D, (R_i)_{i \in D})\) be an association scheme with intersection numbers \(p^k_{i,j}\) and valencies \(\omega_i\). Then the product \(*\) with

\[
\delta_i * \delta_j := \sum_{k \in D} \frac{\omega_k}{\omega_i \omega_j} \cdot p^k_{i,j} \delta_k
\]

can be extended uniquely to an associative, bilinear, \(\| \cdot \|_{TV}\)-continuous mapping on \(M_b(D)\). \((D, *)\) is a discrete hypergroup with the left and right Haar measure

\[
\Omega_l := \sum_{i \in D} \omega_i \delta_i \quad \text{and} \quad \Omega_r := \Omega_l^* := \sum_{i \in D} \omega_i \delta_i\quad (3.5)
\]

respectively. This hypergroup is commutative or symmetric if and only if so is \(\Lambda\).

**Proof.** By Lemma 3.5(5), \(\delta_i * \delta_j\) is a probability measure with finite support for \(i, j \in D\). Thus, \(*\) can be extended uniquely to a bilinear continuous mapping on \(M_b(D)\). The associativity of \(*\) follows from Lemma 3.5(6) for point measures, and, in the general case, by the unique bilinear continuous extension. The remaining hypergroup axioms now follow from Lemma 3.5. Clearly, \(*\) is commutative if and only if so is \(\Lambda\). The same holds for symmetry, as the involution on a hypergroup and on an association scheme are unique.

Finally, by Example 2.8, a left Haar measure is given by

\[
\Omega_l(\{i\}) = \frac{1}{p^i_{i,i}} \omega_i \omega_i \omega_i = \omega_i \quad (i \in D).
\]

The same argument works for right Haar measures. \(\square\)

**Example 3.9.** Let \(G\) be a second countable, locally compact group with a compact open subgroup \(H\). Consider the associated quotient association scheme \((X = G/H, D = G//H, (R_j)_{j \in D})\) as in Example 3.6. Then the associated hypergroup \((D = G//H, *)\) according to Proposition 3.8 is the double coset hypergroup in the sense of Section 2.

For the proof notice that both hypergroups live on the discrete space \(D\). We must show that for all \(x, y, g \in G\) the products \((\delta_{HxH} * \delta_{HyH})(\{HgH\})\) are equal.

Let us compute this first for the double coset hypergroup \(G//H\): By Example 3.6, \(HxH\) decomposes into \(\omega_{HxH}\) many disjoint cosets \(x_1 H, \ldots, x_{\omega_{HxH}} H\). This shows that \(HxH\) also decomposes into \(\omega_{Hx^{-1}H}\) many disjoint cosets of the form \(Hx_1^{-1}H, \ldots, Hx_{\omega_{Hx^{-1}H}}^{-1}H\). Moreover, \(HyH\) decomposes into \(\omega_{HyH}\) many disjoint
cosets \( y_1H, \ldots, y_\omega H \). Using the normalized Haar measure \( \omega_H \) of \( H \) we obtain

\[
((\omega_H * \delta_x * \omega_H) * (\omega_H * \delta_y * \omega_H))(HgH) = \frac{1}{\omega_{H_x^{-1}H} \cdot \omega_{HyH}} \sum_{k=1}^{\omega_{H_x^{-1}H}} \sum_{l=1}^{\omega_{H_yH}} (\omega_H * \delta_{x^{-1}} * \delta_{y_l} * \omega_H)(HgH).
\]

Therefore, by the definition of the double coset convolution in Section 2,

\[
(\delta_{HxH} * \delta_{HyH})(\{HgH\}) = \frac{|\{(k, l) : Hx^{-1}y_lH = HgH\}|}{\omega_{H_x^{-1}H} \cdot \omega_{HyH}}.
\]

We next check that the hypergroup associated with the association scheme in Example 3.6 leads to the same result. For this we employ the beginning of the proof of 3.5(4) and observe that

\[
\omega_{H_gH} \cdot p_{HxH,HyH}^{HgH} = (A_{HxH} A_{HyH} A_{H^{-1}gH})_{eH,eH}
\]

for the trivial coset \( eH \in X = G/H \). This shows with the notations above that

\[
\omega_{HgH} \cdot p_{HxH,HyH}^{HgH} = |\{(k, l) : Hx^{-1}y_lH = HgH\}|.
\]

The convolution in Proposition 3.8 now again leads to (3.7) as claimed.

The same result can be obtained for Hecke pairs. We omit the obvious proof.

**Example 3.10.** Let \((G, H)\) be a Hecke pair. Consider the associated quotient association scheme \((X = G/H, D = G//H, (R_j)_{j \in D})\) as in Example 3.7. Then the associated hypergroup \((D = G//H, \ast)\) according to Proposition 3.8 is the double coset hypergroup in the sense of Section 2.

We next discuss a property of association schemes which is valid in the finite case, but not necessarily in infinite cases.

**Definition 3.11.** An association scheme \( \Lambda \) with valencies \( \omega_i \) is called **unimodular** if \( \omega_i = \omega_{\bar{i}} \) for all \( i \in D \).

**Lemma 3.12.**

1. If an association scheme is commutative or finite, then it is unimodular.
2. An association scheme \((X, D, (R_i)_{i \in D})\) is unimodular if and only if the associated discrete hypergroup \((D, \ast)\) is unimodular.
3. If \((X, D, (R_i)_{i \in D})\) is unimodular, then \( S_i^T = S_i \) \( (i \in D) \).

**Proof.** The commutative case in (1) is trivial. Now let the set \( X \) of an association scheme finite. Then, for \( h \in D \), the adjacency matrices \( A_h \) and \( \bar{A}_h \) have \(|X| \cdot \omega_h \) and \(|X| \cdot \omega_{\bar{h}} \) times the entry 1 respectively. \( A_h = \bar{A}_h^T \) now implies \( \omega_h = \omega_{\bar{h}} \).

Part (2) is clear by Eq. (3.5) in Proposition 3.8; (3) is also clear. \( \square \)

**Remark 3.13.**

1. There exist non-unimodular association schemes in the group context 3.6. In fact, in [21], examples of totally disconnected groups \( G \) with compact open subgroups \( H \) are given where the associated double coset hypergroups \((G//H, \ast)\) are not unimodular. Therefore, by Lemma 3.12(2), the corresponding association schemes are not unimodular. For examples in the context of Hecke pairs see [25].
Remark 3.14. Let \( \Lambda \) be an unimodular association scheme with the associated hypergroup \((D, \ast)\) as in Proposition 3.8. Then, by (3.3) and 3.12(3), the (finite) linear span of the matrices \( S_i \) \((i \in D)\) forms an \(*\)-algebra which is via \( S_i \to \delta_i \) \((i \in D)\) isomorphic with the \(*\)-algebra \((M_f(D), \ast)\) of all signed measures on \( D \) with finite support and the convolution from Proposition 3.8.

We briefly study concepts of automorphisms of association schemes.

Definition 3.15. 

(1) Let \( \Lambda = (X, D = G, (R_i)_{j \in D}) \) and \( \tilde{\Lambda} = (\tilde{X}, \tilde{D}, (\tilde{R}_j)_{j \in \tilde{D}}) \) be association schemes. A pair of \((\varphi, \psi)\) of bijective mappings \( \varphi : X \to \tilde{X} \), \( \psi : D \to \tilde{D} \) is called an isomorphism from \( \Lambda \) onto \( \tilde{\Lambda} \) if for all \( x, y \in X \) and \( i \in D \) with \((x, y) \in R_i \), \( (\varphi(x), \varphi(y)) \in \tilde{R}_{\psi(i)} \).

(2) If \( \Lambda = \tilde{\Lambda} \), then an isomorphism is called an automorphism.

(3) If \((\varphi, \psi)\) is an automorphism with \( \psi \) as identity, then \( \varphi \) is called a strong automorphism. The set \( \text{Saut}(\Lambda) \) of all strong automorphisms of \( \Lambda \) is obviously a group.

Remark 3.16. Let \( \Lambda = (X, D, (R_j)_{j \in D}) \) be an association scheme. If we regard \( \text{Saut}(\Lambda) \) as subspace of the space \( X^X \) of all maps from \( X \) to \( X \) with the product topology, then \( \text{Saut}(\Lambda) \) obviously is a topological group. For \( x_0 \in X \) we consider the stabilizer subgroup \( \text{Stab}_{x_0} := \{ \varphi \in \text{Saut}(\Lambda) : \varphi(x_0) = x_0 \} \) which is obviously open in \( \text{Saut}(\Lambda) \). \( \text{Stab}_{x_0} \) is also compact. In fact, for each \( g \in \text{Stab}_{x_0} \) and \( y \in X \) with \((x, y) \in R_i \) with \( i \in D \), we have \((x, g(y)) = (g(x), g(y)) \in R_i \). The axioms of an association scheme show that \( \{g(y) : g \in \text{Stab}_{x_0}\} \) is finite \( y \in X \). This shows that \( \text{Stab}_{x_0} \) is compact. In particular, \( \text{Stab}_{x_0} \) is a compact open neighborhood of the identity of \( \text{Saut}(\Lambda) \) which shows that \( \text{Saut}(\Lambda) \) is locally compact.

This argument is completely analog to Example 2.4 for graphs.

Example 3.17. Let \( G \) be a locally compact group with compact open subgroup \( H \), or let \((G, H)\) be a Hecke pair. Consider the associated association scheme \( \Lambda = (X = G/H, D = G\!/\!H, (R_j)_{j \in D}) \) as in 3.6 or 3.7. Then for each \( g \in G \), the mapping \( T_g : G/H \to G/H, xH \mapsto gxH \) defines a strong automorphism of \( \Lambda \). This obviously leads to a group homomorphism \( T : G \to \text{Saut}(\Lambda) \).

Remark 3.18. Let \((\varphi, \psi)\) be an automorphism of an association scheme \( \Lambda = (X, D, (R_j)_{j \in D}) \). Then \( \psi : D \to D \) is an automorphism of the associated hypergroup on \( D \) according to Proposition 3.8. This follows immediately from the definition of the convolution in 3.8 and the definitions of \( p^i_j \) and \( \omega_i \).

The constructions of association schemes are parallel above for groups \( G \) with compact open subgroups \( H \) and for Hecke pairs \((G, H)\). This is not an accident.

If \( \tilde{G} \) is a locally compact group with compact open subgroup \( \tilde{H} \), then w.r.t. the discrete topology, \((\tilde{G}, \tilde{H})\) is also a Hecke pair, and the association scheme \((\tilde{X} = \tilde{G}/\tilde{H}, \tilde{D} = \tilde{G}\!/\!\tilde{H}, (\tilde{R}_j)_{j \in \tilde{D}})\) does not depend on the topologies. In this way, the approach via Hecke pairs seems to be the more general one. On the other hand, the following theorem shows that both approaches are equivalent:
Theorem 3.19. Let \((G, H)\) be a Hecke pair. Then there is a totally disconnected locally compact group \(\tilde{G}\) with an compact open subgroup \(H\) such that the associated association schemes \((G/H, D = G//H, (R_j)_{j\in D})\) and \((\tilde{G}/H, D = \tilde{G}//H, (\tilde{R}_j)_{j\in D})\) are isomorphic. The associated double coset hypergroups are also isomorphic.

Proof. Let \((G, H)\) be a Hecke pair and \(\Lambda := (G/H, D = G//H, (R_j)_{j\in D})\) the associated association scheme. Consider the homomorphism \(T : G \to \text{Saut}(\Lambda)\) according to 3.17, where \(G\) acts transitively on \(X = G/H\). Hence, the totally disconnected locally compact group \(\tilde{G} := \text{Saut}(\Lambda)\) of 3.16 acts transitively on \(X = G/H\). We define \(\tilde{H} := \text{Stab}_H\) as a compact open subgroup of \(\tilde{G}\).

We next consider the normal subgroup \(N := \bigcap_{x \in G} xHx^{-1} \leq H\) of \(G\). Then obviously \((G/N)/(H/N) \cong G/H\), \((G/N)/(H/N) \cong G/H\), and \((G/N, H/N)\) forms a Hecke pair for which the associated association scheme is isomorphic with \(\Lambda\). Using this division by \(N\), we may assume from now on that \(N = \{e\}\).

If this is the case, we obtain that the homomorphism \(G \to \tilde{G}, g \mapsto T_g\) from 3.17 with \(T_g(xH) = gxH\) is injective. In fact, if for some \(g \in G\) and all \(x \in G\) we have \(T_g(xH) = gxH = xH\), it follows that \(g \in N = \{e\}\) as claimed. We thus may assume that \(G\) is a subgroup of \(\tilde{G}\). We then readily obtain \(H = G \cap \tilde{H}\).

We now consider the closures \(\tilde{G}, \tilde{H}\) of \(G, H\) in \(\tilde{G}\). Then \(\tilde{G}\) is a locally compact, totally disconnected topological group with \(\tilde{H} \subset \tilde{G}\) as compact subgroup. Moreover \(\tilde{H} = G \cap \tilde{H}\) yields \(H = G \cap \tilde{H}\) which implies that \(H\) is in fact a compact open subgroup of \(\tilde{G}\). Moreover, as \(\tilde{G}\) acts transitively on \(G/H\) with \(H\) as stabilizer subgroup of \(eH \in G/H\), we may identify \(\tilde{G}/\tilde{H}\) with \(G/H\). To make this more explicit we claim that

for all \(g \in \tilde{G}\) there exists \(g \in G\) with \(g\tilde{H} = g\tilde{H}\). \((3.8)\)

In fact, each \(\tilde{g} \in \tilde{G}\) is the limit of some \(g_\alpha \in G\). As \(\tilde{g}\tilde{H}\) is an open neighborhood of \(\tilde{g} \in \tilde{G}\), we obtain \(g_\alpha \in \tilde{g}\tilde{H}\) for large indices which proves \((3.8)\). Moreover, as for \(g_1, g_2 \in G\) the relation \(g_1\tilde{H} = g_2\tilde{H}\) implies \(g_1^{-1}g_2 \in H \cap G = H\), we obtain with \((3.8)\) that the mapping

\[\varphi : G/K \to \tilde{G}/\tilde{H}, \quad gH \mapsto g\tilde{H}\]

is a well defined bijective mapping. Moreover, it is clear that the orbits of the action of \(H\) and \(\tilde{H}\) on \(G/H\) are equal. This shows that also the mapping

\[\psi : G//K \to \tilde{G}//\tilde{H}, \quad HgH \mapsto Hg\tilde{H}\]

is well defined and bijective. If we compare the pair \((\varphi, \psi)\) with the construction of the associations schemes in Examples 3.6 and 3.7, we see that the schemes \(\Lambda\) and \((X = G/H, D = G//H, (\tilde{R}_j)_{j\in D})\) are isomorphic via \((\varphi, \psi)\) as claimed.

Finally, the statement about the associated double coset hypergroups is clear by Example 3.9. \(\square\)

We finally present the following trivial commutativity criterion:

Lemma 3.20. If an association scheme \(\Lambda = (X, D, (R_i)_{i\in D})\) admits an automorphism \((\varphi, \psi)\) with \(\psi(i) = i\) for \(i \in D\), then \(\Lambda\) is commutative.

Proof. For \(i, j, k \in D\), \(p^k_{i,j} = p^k_{j,i} = p^{\psi(k)}_{\psi(i),\psi(j)} = p^k_{j,i}\), where the last equality follows easily from the definition of an automorphism and the definition of \(p^k_{j,i}\). \(\square\)
Remark 3.21. Lemma 3.20 is a natural extension of the following well-known criterion for Gelfand pairs:
Let $H$ be a compact subgroup of a locally compact group $G$ such that there is a continuous automorphism $T$ of $G$ with $x^{-1} \in HT(x)H$ for $x \in G$. Then $(G, H)$ is a Gelfand pair, i.e., $G//H$ is a commutative hypergroup.
If $H$ is compact and open in $G$, then we obtain this criterion from Lemma 3.20 when applied to association scheme $\Lambda = (X, D, H)$ with the automorphism $(\varphi, \psi)$ with $\varphi(gH) = T(g)H$ and $\psi(HgH) = HT(g)H$.

4. Commutative Association Schemes and Dual Product Formulas
In this section let $\Lambda = (X, D, (R_i)_{i \in D})$ be a commutative association scheme and $(D, \ast)$ the associated commutative discrete hypergroup as in Proposition 3.8.

By Remark 3.14, the linear span $A(X)$ of the matrices $S_i$ $(i \in D)$ is a commutative $\ast$-algebra which is isomorphic with the $\ast$-algebra $(M_f(D), \ast)$ via
$$\mu \in M_f(D) \iff S_{\mu} := \sum_{i \in D} \mu(i)S_i \in A(X).$$
Moreover, as $(f \ast g)\Omega = f\Omega \ast g\Omega$ and $f^\ast \Omega = (f\Omega)^\ast$ for all $f, g \in C_c(D)$ and the Haar measure $\Omega$ of $(D, \ast)$ by [20], the $\ast$-algebra $A(X)$ is also isomorphic with the commutative $\ast$-algebra $(C_c(D), \ast)$ via
$$f \in C_c(D) \iff S_f := \sum_{i \in D} f(i)\omega_iS_i = \sum_{i \in D} f(i)A_i \in A(X).$$
By this construction, $S_f = S_{f\Omega}$ for $f \in C_c(D)$. We notice that even for $f \in C_b(D)$, the matrix $S_f := \sum_{i \in D} f(i)\omega_iS_i$ is well-defined, and that products of $S_f$ with matrices with only finitely many nonzero entries in all rows and columns are well-defined. In particular, we may multiply $S_f$ on both sides with any $S_{\mu}, \mu \in M_f(D)$.

For each function $f \in C_b(D)$ we define the function $F_f : X \times X \to \mathbb{C}$ with
$$F_f(x, y) = f(i) \text{ for the unique } i \in D \text{ with } (x, y) \in R_i.$$ Then $F_f$ is constant on all partitions $R_i$, and by our construction,
$$F_f(x, y))_{x,y\in X} = S_f.$$ In particular, $f(e)$ is equal to the diagonal entries of $S_f$ (where these are equal).

The following definition is standard for kernels.

Definition 4.1. A function $F : X \times X \to \mathbb{C}$ is called positive definite on $X \times X$ if for all $n \in \mathbb{N}$, $x_1, \ldots, x_n \in X$ and $c_1, \ldots, c_n \in \mathbb{C}$,
$$\sum_{k,l=1}^{n} c_k\bar{c_l} \cdot F(x_k, x_l) \geq 0,$$ i.e., the matrices $(F(x_k, x_l))_{k,l}$ are positive semidefinite. This in particular implies that these matrices are hermitian, i.e.,
$$F(x, y) = \overline{F(y, x)} \text{ for all } x, y \in X.$$

As the pointwise products of positive semidefinite matrices are again positive semidefinite (see e.g. Lemma 3.2 of [5]), we have:
Lemma 4.2. If $F, G : X \times X \to \mathbb{C}$ are positive definite, then the pointwise product $F \cdot G : X \times X \to \mathbb{C}$ is also positive definite.

Moreover, for $f \in C_b(D)$, the positive definiteness of $f$ on $D$ is related to that of $F_f$ on $X \times X$. We begin with the following observation:

Lemma 4.3. Let $f \in C_b(D)$. If $F_f$ is positive definite, then $f$ is positive definite on the hypergroup $(D, \ast)$.

Proof. For $f$ define the reflected function $f^-(x) := f(\bar{x})$ for $x \in D$. Let $n \in \mathbb{N}$, $x_1, \ldots, x_n \in D$ and $c_1, \ldots, c_n \in \mathbb{C}$. Let $\mu := \sum_{k=1}^{n} c_k \delta_{x_k} \in M_f(D)$ and observe that by the definition of the convolution of a function with a measure and by commutativity,

$$ P := \sum_{k,l=1}^{n} c_k \bar{c}_l \cdot f^-(x_k \ast \bar{x}_l) = \int_D f^-(\bar{\mu} \ast \mu^\ast) = (\mu \ast \mu^\ast \ast f)(\bar{e}) = (\mu \ast f \ast \mu^\ast)(\bar{e}). $$

where by the considerations above, this is equal to the diagonal entries of

$$ S_{\mu \ast f \ast \mu^\ast} = S_\mu \cdot S_f \cdot S_{\mu^\ast}^T = S_\mu \cdot F_f(x,y)_{x,y \in X \ast S_{\mu^\ast}^T}. $$

This matrix product exists and is positive semidefinite, as so is $F_f(x,y)_{x,y \in X \ast S_{\mu^\ast}^T}$ by our assumption. As the diagonal entries of a positive semidefinite matrix are nonnegative, we obtain $P \geq 0$. Hence, $f^-$ and thus $f$ is positive definite.

Here is a partially reverse statement.

Lemma 4.4. Let $f \in C_c(D)$. Then, by 2.10(2), $f \ast f^\ast$ is positive definite on $D$, and $F_{f \ast f^\ast}$ is positive definite.

Proof. By our considerations above, $(F_{f \ast f^\ast}(x,y))_{x,y \in X} = S_f \cdot S_f^T$. This proves that $F_{f \ast f^\ast}$ is positive definite.

We now combine Lemma 4.4 with Fact 2.10(4) and the trivial observation that pointwise limits of positive definite functions are positive definite. This shows:

Corollary 4.5. Let $\alpha \in S \subset \hat{D}$ be a character in the support of the Plancherel measure. Then $F_\alpha : X \times X \to \mathbb{C}$ is positive definite.

Lemmas 4.2 and 4.3 now lead to the following central result of this paper:

Theorem 4.6. Let $(D, \ast)$ be a commutative discrete hypergroup which is associated with some association scheme $\Lambda = (X, D, (R_i)_{i \in D})$. Let $\alpha, \beta \in S \subset \hat{D}$ be characters in the support of the Plancherel measure. Then $\alpha \cdot \beta$ is positive definite on $D$, and there exists a unique probability measure $\delta_\alpha \ast \delta_\beta \in M^1(D)$ with $(\delta_\alpha \ast \delta_\beta)^\vee = \alpha \cdot \beta$. The support of this measure is contained in $S$.

Furthermore, for all $\alpha \in S$, the unique positive character $\alpha_0$ in $S$ according to 2.10(5) is contained in the support of $\delta_\alpha \ast \delta_\alpha$.

Proof. 4.5, 4.2, and 4.3 show that $\alpha \cdot \beta$ is positive definite. Bochner’s theorem 2.10(1) now leads to the probability measure $\delta_\alpha \ast \delta_\beta \in M^1(D)$. The assertions about the supports of $\delta_\alpha \ast \delta_\beta$ and $\delta_\alpha \ast \delta_\alpha$ follow from Theorem 2.1(4) of [36].

\[\square\]
For finite association schemes we have a stronger result. It is shown in Section 2.10 of [4] in another, but finally equivalent way:

**Theorem 4.7.** Let \((D, \ast)\) be a finite commutative hypergroup which is associated with some association scheme \(\Lambda = (X, D, (R_i)_{i \in I})\). Then \((\hat{D}, \ast)\) is a hypergroup.

**Proof.** For finite hypergroups we have \(S = \hat{D}\), and the unique positive character in \(S\) is the identity \(1\). Therefore, if we take \(1\) as identity and complex conjugation as involution, we see that almost all hypergroup properties of \((D, \ast)\) follow from Theorem 4.6. We only have to check that for \(\alpha \neq \beta \in \hat{D}\), \(1\) is not contained in the support of \(\delta_\alpha \ast \delta_\beta\). For this we recapitulate that for all \(\gamma, \rho \in \hat{D}\), \(\hat{\gamma} \rho d\Omega = ||\gamma||^2 \delta_{\gamma, \rho}\) with the Kronecker-\(\delta\). Therefore, with 12.16 of [20],

\[
(\delta_\alpha \ast \delta_\beta)(\{1\}) = \int_{\hat{D}} \mathbf{1}_\{1\} d(\delta_\alpha \ast \delta_\beta) = \int_{\hat{D}} \hat{1} d(\delta_\alpha \ast \delta_\beta) = \\
= \int_{\hat{D}} \mathbf{1} (\delta_\alpha \ast \delta_\beta)^\vee d\Omega = \int_{\hat{D}} \alpha \beta d\Omega = ||\gamma||^2 \delta_{\alpha, \beta} = 0.
\]

This completes the proof. \(\square\)

We present some examples related to Gelfand pairs which show that the infinite case is more involved than in Theorem 4.7; for details see [37].

**Example 4.8.** Let \(a, b \geq 2\) be integers. Let \(C_b\) the complete undirected graph graph with \(b\) vertices, i.e., all vertices of \(C_b\) are connected. We now consider the infinite graph \(\Gamma := \Gamma(a, b)\) where precisely \(a\) copies of the graph \(C_b\) are tacked together at each vertex in a tree-like way, i.e., there are no other cycles in \(\Gamma\) than those in a copy of \(C_b\). For \(b = 2\), \(\Gamma\) is the homogeneous tree of valency \(a\). We denote the distance function on \(\Gamma\) by \(d\).

It is clear that the group \(G := \text{Aut}(\Gamma)\) of all graph automorphisms acts on \(\Gamma\) in a distance-transitive way, i.e., for all \(v_1, v_2, v_3, v_4 \in \Gamma\) with \(d(v_1, v_3) = d(v_3, v_4)\) there exists \(g \in \Gamma\) with \(g(v_1) = v_3\) and \(g(v_2) = v_4\). \(\text{Aut}(\Gamma)\) is a totally disconnected, locally compact group w.r.t. the topology of pointwise convergence, and the stabilizer subgroup \(H \subset G\) of any fixed vertex \(e \in \Gamma\) is compact and open. We identify \(G/H\) with \(\Gamma\), and \(G//H\) with \(\mathbb{N}_0\) by distance transitivity. We now study the association scheme \(\Lambda = (\Gamma \simeq G/H, \mathbb{N}_0 = G//H, (R_i)_{i \in \mathbb{N}_0})\) and the double coset hypergroup \((\mathbb{N}_0 \simeq G//H, \ast)\). As in the case of finite distance-transitive graphs in [4], \(\Lambda\) and \((\mathbb{N}_0, \ast)\) are commutative and associated with a sequence of orthogonal polynomials in the Askey scheme [2].

More precisely, by [34], the hypergroup convolution is given by

\[
\delta_m \ast \delta_n = \sum_{k = |m - n|}^{m+n} g_{m,n,k} \delta_k \in M^1(\mathbb{N}_0) \quad (m, n \in \mathbb{N}_0) \quad (4.1)
\]

with

\[
g_{m,n,m+n} = \frac{a-1}{a} > 0, \quad g_{m,n,|m-n|} = \frac{1}{a(a-1)^{m \wedge n-1}(b-1)^{m \wedge n}} > 0, \quad g_{m,n,|m-n|+2k+1} = \frac{b-2}{a(a-1)^{m \wedge n-k-1}(b-1)^{m \wedge n-k}} \geq 0 \quad (k = 0, \ldots, m \wedge n - 1),
\]
By Eq. (4.4), we have the dual space $[2)$. By the orthogonality relations in [2], the normalized orthogonality measure $P_{1,n+1} = \frac{a-1}{a}$, $g_{n,1,n} = \frac{b-2}{a(b-1)}$, $g_{n,1,n-1} = \frac{1}{a(b-1)}$.

We define a sequence of orthogonal polynomials $(P_{n}^{(a,b)})_{n \geq 0}$ by

$$P_{0}^{(a,b)} := 1, \quad P_{1}^{(a,b)}(x) := \frac{2}{a} \sqrt{\frac{a-1}{b-1}} : x + \frac{b-2}{a(b-1)}.$$ 

and the three-term-recurrence relation

$$P_{n}^{(a,b)} = \frac{1}{a(b-1)} P_{n-1}^{(a,b)} + \frac{b-2}{a(b-1)} P_{n}^{(a,b)} + \frac{a-1}{a} P_{n+1}^{(a,b)} \quad (n \geq 1).$$

Then,

$$P_{n}^{(a,b)} = \sum_{k=m-n}^{m+n} g_{m,n,k} P_{k}^{(a,b)} \quad (m, n \geq 0).$$

We also notice that the formulas above are correct for all $a, b \in [2, \infty]$, and that Eq. (4.1) then still defines commutative hypergroups $(\mathbb{N}_{0}, \ast)$.

We discuss some properties of the $P_{n}^{(a,b)}$ from [34], [37]. Eq. (4.2) yields

$$P_{n}^{(a,b)} \left( \frac{z + z^{-1}}{2} \right) = c(z) z^{n} + \frac{c(z^{-1}) z^{-n}}{((a-1)(b-1))^{n/2}} \quad \text{for } z \in \mathbb{C} \setminus \{0, \pm 1\}$$

with

$$c(z) := \frac{(a-1)z - z^{1+2(b-2)(a-1)}{(a-1)(b-1)}^{-1/2}}{a(z^2 - 1)}.$$  

(4.5)

We define

$$s_{0} := s_{0}^{(a,b)} := \frac{2 - a - b}{2 \sqrt{(a-1)(b-1)}}, \quad s_{1} := s_{1}^{(a,b)} := \frac{ab - a - b + 2}{2 \sqrt{(a-1)(b-1)}}.$$  

(4.6)

Then

$$P_{n}^{(a,b)}(s_{1}) = 1, \quad P_{n}^{(a,b)}(s_{0}) = (1-b)^{-n} \quad (n \geq 0).$$

(4.7)

It is shown in [37] that the $P_{n}^{(a,b)}$ fit into the Askey-Wilson scheme (pp. 26–28 of [2]). By the orthogonality relations in [2], the normalized orthogonality measure $\rho = \rho^{(a,b)} \in M^{1}(\mathbb{R})$ is

$$d\rho^{(a,b)}(x) = w^{(a,b)}(x)dx \quad \text{for } a \geq b \geq 2$$

(4.8)

and

$$d\rho^{(a,b)}(x) = w^{(a,b)}(x)dx \quad \text{for } b > a \geq 2$$

(4.9)

with

$$w^{(a,b)}(x) := \frac{a}{2\pi} \cdot \frac{(1-x^{2})^{1/2}}{(s_{1} - x)(x - s_{0})}.$$  

For $a, b \in \mathbb{R}$ with $a, b \geq 2$, the numbers $s_{0}, s_{1}$ satisfy

$$-s_{1} \leq s_{0} \leq -1 < 1 \leq s_{1}.$$  

(4.10)

By Eq. (4.4), we have the dual space

$$\mathcal{D} \simeq \{ x \in \mathbb{R} : (P_{n}^{(a,b)}(x))_{n \geq 0} \text{ is bounded } \} = [-s_{1}, s_{1}].$$  

(4.11)
Moreover, for all $x,y \in \mathbb{R}$, then the following statements are equivalent for $x \neq y$:

1. $x \in [s_0^{(a,b)}, s_1^{(a,b)}]$.
2. The mapping $\Gamma \times \Gamma \to \mathbb{R}$, $(v_1, v_2) \mapsto P^{(a,b)}_{d(v_1, v_2)}(x)$ is positive definite;
3. The mapping $g \mapsto P^{(a,b)}_{d(gH,e)}(x)$ is positive definite on $G$.

Moreover, for all $x, y \in [s_0^{(a,b)}, s_1^{(a,b)}]$ there exists a unique probability measure $\mu_{x,y} \in M^1([-s_1^{(a,b)}, s_1^{(a,b)}])$ with

$$P_n^{(a,b)}(x) \cdot P_n^{(a,b)}(y) = \int_{-s_1^{(a,b)}}^{s_1^{(a,b)}} P_n^{(a,b)}(z) \, d\mu_{x,y}(z) \quad \text{for all } n \in \mathbb{N}_0. \quad (4.10)$$

Furthermore, for certain integers $b \geq 2$ there exist $x, y \in [-s_1^{(a,b)}, s_0^{(a,b)}]$ for which no probability measure $\mu_{x,y} \in M^1(\mathbb{R})$ exists which satisfies (4.10).

**Remark 4.10.** The preceding examples show that for Gelfand pairs $(G, H)$ with $H$ a compact open subgroup, characters $\alpha \in (G//H)^\wedge$ of the double coset hypergroup may not correspond to a positive definite function on $G$ and not to a positive definite kernel on $G/H$. On the other hand, as in the examples in parts (2) and (3) of Theorem 4.9, these positive definiteness conditions are equivalent in general. This well-known fact follows immediately from the definitions of positive definite functions on $G$ and positive definite kernel on $G/H$.

### 5. Generalized Discrete Association Schemes

In this section we propose a generalization of the definition of association schemes from Section 3 via stochastic matrices. We focus on the commutative case and have applications in mind to dual product formulas as in Theorem 4.6.

One might expect at a first glance that for each discrete commutative hypergroup $(D, \ast)$ there is a natural kind of a generalized association scheme with $X = D$ and transition matrices $S_i$ with $(S_i)_{j,k} := (\delta_i \ast \delta_j)(\{k\})$ for $i, j, k \in D$. We then would obtain $F_j(x, y) = f(x \ast y)$ for $x, y \in D$ and $f \in C(D)$ with the notations in Section 4. These matrices $S_i$ and the functions $F_j$ have much in common with the construction above. However, one central point is different in the view of dual product formulas: In Section 4 we used the fact that the $R_i$ form partitions which led to the central identity $F_{1g \ast} = F_{F_g \ast}$ for functions $f, g$ on $D$. This is obviously usually not correct for $F_j(x, y) = f(x \ast y)$. In particular, such a simple approach would lead to a contradiction with Example 9.1C of Jewett [20].

Having this problem in mind, we propose a definition which skips the integrality condition and which keeps the partition condition. The integrality conditions will be replaced by the existence of a positive invariant measure on $X$ which replaces the counting measure on $X$ in Sections 3 and 4.
Definition 5.1. Let $X, D$ be nonempty, at most countable sets and $(R_i)_{i \in D}$ a disjoint partition of $X \times X$ with $R_i \neq \emptyset$ for $i \in D$. Let $\tilde{S}_i \in \mathbb{R}^{X \times X}$ for $i \in D$ be stochastic matrices. Assume that:

(1) For all $i, j, k \in D$ and $(x, y) \in R_k$, the number
$$p^k_{i,j} := |\{ z \in X : (x, z) \in R_i \text{ and } (z, y) \in R_j \}|$$
is finite and independent of $(x, y) \in R_k$.

(2) For all $i \in D$ and $x, y \in X$, $\tilde{S}_i(x, y) > 0$ if and only if $(x, y) \in R_i$.

(3) For all $i, j, k \in D$ there exist (necessarily nonnegative) numbers $\tilde{p}^k_{i,j}$ with
$$\tilde{S}_i \tilde{S}_j = \sum_{k \in D} \tilde{p}^k_{i,j} \tilde{S}_k.$$

(4) There exists an identity $e \in D$ with $\tilde{S}_e = I_X$ as identity matrix.

(5) There exists a positive measure $\pi \in M^+(X)$ with $\text{supp } \pi = X$ and an involution $i \mapsto \bar{i}$ on $D$ such that for all $i \in D$, $x, y \in X$,
$$\pi(\{y\}) \tilde{S}_i(y, x) = \pi(\{x\}) \tilde{S}_i(x, y).$$

Then $\Lambda := (X, D, (R_i)_{i \in D}, (\tilde{S}_i)_{i \in D})$ is called a generalized association scheme. $\Lambda$ is called commutative if $\tilde{S}_i \tilde{S}_j = \tilde{S}_j \tilde{S}_i$ for all $i, j \in D$. It is called symmetric if the involution is the identity. $\Lambda$ is called finite, if so are $X$ and $D$.

Example 5.2. If $\Lambda = (X, D, (R_i)_{i \in D})$ is an unimodular association scheme as in Section 3, and if we take the stochastic matrices $(S_i)_{i \in D}$ as in Remark 3.4, then $(X, D, (R_i)_{i \in D}, (S_i)_{i \in D})$ is a generalized association scheme. In fact, axioms (1)-(4) are clear, and for axiom (5) we take the involution of $\Lambda$ and $\pi$ as the counting measure on $X$. Lemma 3.12(3) and unimodularity then imply axiom (5).

Lemma 5.3. Let $(X, D, (R_i)_{i \in D}, (\tilde{S}_i)_{i \in D})$ be a generalized association scheme. Then:

(1) The triplet $(X, D, (R_i)_{i \in D})$ is an association scheme.

(2) The positive measure $\pi \in M^+(X)$ from axiom (5) is invariant, i.e., for all $y \in X$ and $i \in D$,
$$\sum_{x \in X} \pi(\{x\}) \tilde{S}_i(x, y) = \pi(\{y\}).$$

(3) The deformed intersection numbers $\tilde{p}^k_{i,j}$ satisfy $\sum_{k \in D} \tilde{p}^k_{i,j} = 1$ and
$$\tilde{p}^k_{i,j} > 0 \iff p^k_{i,j} > 0 \quad (i, j, k \in D).$$

(4) For all $p \in [1, \infty]$ and $i \in D$, the transition operator
$$T_{\tilde{S}_i} f(x) := \sum_{y \in X} \tilde{S}_i(x, y) f(y) \quad (x \in X)$$
is a continuous linear operator on $L^p(X, \pi)$ with $\|T_{\tilde{S}_i}\| \leq 1$. Moreover, for $p = 2$, the $T_{\tilde{S}_i}$ satisfy the adjoint relation $T_{\tilde{S}_i}^* = T_{\tilde{S}_i}^\ast$.

Proof. Part (1) is clear; we only remark that the axiom regarding the involution on $D$ follows from axioms (2) and (3) above.

For part (2) we use axiom (5) which implies for $i \in D$, $y \in X$ that
$$\sum_{x \in X} \pi(\{x\}) S_i(x, y) = \pi(\{y\}) \sum_{x \in X} S_i(y, x) = \pi(\{y\}).$$
Part (3) and the adjoint relation in (4) are obvious. Moreover, for $p \geq 1$, the invariance of $\pi$ w.r.t. $S_i$ and Hölder’s inequality easily imply $\|T_{\tilde{S}_i}\| \leq 1$. □
In summary, we may regard generalized association schemes as deformations of given classical association schemes with deformed intersection numbers $\tilde{p}_{i,j}^k$.

Generalized association schemes also lead to discrete hypergroups:

**Proposition 5.4.** Let $\Lambda := (X,D,(R_i)_{i\in D},(\tilde{S}_i)_{i\in D})$ be a generalized association scheme with deformed intersection numbers $\tilde{p}_{i,j}^k$. Then the product $\hat{*}$ with

$$\delta_i \hat{*} \delta_j := \sum_{k \in D} \tilde{p}_{i,j}^k \delta_k$$

can be extended uniquely to an associative, bilinear, scheme with deformed intersection numbers $\tilde{\delta}_{i,j}$. Moreover, by Lemma 5.3(3), $\hat{*}$ is associative. Clearly, $\hat{*}$ is commutative if and only if so is $\Lambda$.

**Proof.** As the product of matrices is associative, the convolution $\hat{*}$ is associative. Moreover, by Lemma 5.3(3), $\delta_i \hat{*} \delta_j$ is a probability measure for $i,j \in D$ whose support is the same as for the hypergroup convolution of the association scheme $(X,D,(R_i)_{i\in D})$ in Proposition 3.8. This readily shows that $(D,\hat{*})$ is a hypergroup.

Clearly, $\hat{*}$ is commutative or symmetric if and only if so is $\Lambda$. Finally, the statement about the Haar measures follows from Example 2.8(1). □

We now extend the approach of Section 4 to dual positive product formulas for discrete commutative hypergroups $(D,\hat{*})$ which are associated with generalized commutative association schemes $\Lambda := (X,D,(R_i)_{i\in D},(\tilde{S}_i)_{i\in D})$.

For this we study the linear span $A(X)$ of the matrices $\tilde{S}_i$ ($i\in D$). If we identify the $\tilde{S}_i$ with the transition operators $T_{\tilde{S}_i}$, we may regard $A(X)$ as a commutative $*$-subalgebra of the C*-algebra $\mathcal{B}(L^2(X,\pi))$ of all bounded linear operators on $L^2(X,\pi)$ by Lemma 5.3(4). As in Section 4, $A(X)$ is isomorphic with the $*$-algebra $(M_f(D),\hat{*})$ of all measures with finite support via

$$\mu \in M_f(D) \longleftrightarrow \tilde{S}_\mu := \sum_{i\in D} \mu(i) \tilde{S}_i \in A(X).$$

Moreover, using the Haar measure $\Omega$ of $(D,\hat{*})$, $A(X)$ is also isomorphic with the commutative $*$-algebra $(C_c(D),\hat{*})$ via

$$f \in C_c(D) \longleftrightarrow \tilde{S}_f := \sum_{i\in D} f(i) \omega_i \tilde{S}_i \in A(X).$$

As in Section 4, we have $\tilde{S}_f = \tilde{S}_{f0}$ for $f \in C_c(D)$.

We even may define the matrices $\tilde{S}_f := \sum_{i\in D} f(i) \omega_i \tilde{S}_i$ for $f \in C_b(D)$, and we may form matrix products of $\tilde{S}_f$ with matrices with only finitely many nonzero entries in all rows and columns. In particular, we may multiply $\tilde{S}_f$ on both sides with any $\tilde{S}_\mu, \mu \in M_f(D)$. We need the following notion of positive definiteness:

**Definition 5.5.** Let $A \in \mathbb{C}^{X \times X}$ be a matrix. Then for all $g_1, g_2 \in C_f(X) := \{g : X \to \mathbb{C} \text{ with finite support}\}$, we may form

$$\langle Ag_1,g_2 \rangle_\pi := \sum_{x \in X} (Ag_1)(x) \cdot \overline{(g_2)(x)} \cdot \pi(\{x\}) \in \mathbb{C}.$$
For each \( f \in C_b(D) \) we define the function \( F_f : X \times X \to \mathbb{C} \) with \( F_f(x, y) = f(i) \) for the unique \( i \in D \) with \((x, y) = R_i \) as in Section 4. Then different from Section 4, we usually have \( (F_f(x, y))_{x, y \in X} \neq \hat{S}_f \). We thus have to state some results from Section 4 for the matrices \( \hat{S}_f \) instead of \( (F_f(x, y))_{x, y \in X} \).

Lemmas 4.3 and 4.4 now read as follows:

**Lemma 5.6.** Let \( f \in C_b(D) \). If \( \hat{S}_f \) is positive definite, then \( f \) is positive definite on the hypergroup \((D, \hat{\ast})\).

**Proof.** For \( f \) define \( f^-(x) := f(x) \) for \( x \in D \). Let \( n \in \mathbb{N}, x_1, \ldots, x_n \in D \) and \( c_1, \ldots, c_n \in \mathbb{C} \). Let \( \mu := \sum_{k=1}^n c_k \delta_{x_k} \in M_f(D) \) and observe that as in the proof of Lemma 4.3, \( P := \sum_{k,l=1}^n c_k \delta_k \ast f^-(x_k \ast x_l) = (\mu \ast f \ast \mu^*)(e) \) where this is equal to the diagonal entries of \( \hat{S}_\mu \ast f \ast \mu^* = \hat{S}_\mu \cdot \hat{S}_f \cdot \hat{S}_\mu^* \).

This matrix product exists and is \( \pi \)-positive semidefinite. Therefore, by 5.5, the diagonal entries of this matrix are nonnegative and thus \( P \geq 0 \). Hence, \( f^- \) and thus \( f \) is positive definite. \( \square \)

**Lemma 5.7.** Let \( f \in C_c(D) \). Then, by 2.10(2), \( f \ast f^* \) is positive definite on \((D, \hat{\ast})\), and \( S_{f \ast f^*} \) is \( \pi \)-positive definite.

**Proof.** By our considerations above, \( S_{f \ast f^*} = \hat{S}_f \hat{S}_f^* \). This proves the claim. \( \square \)

We now combine Lemma 5.7 with 2.10(4) and conclude as in Corollary 4.5:

**Corollary 5.8.** Let \( \alpha \in (D, \hat{\ast})^\wedge \) be a character in the support of the Plancherel measure. Then \( \hat{S}_\alpha \) is \( \pi \)-positive definite.

We now want to combine Corollary 5.8 with Lemmas 4.2 and 5.6 to derive an extension of Theorem 4.6. Here, however we would need \( (F_f(x, y))_{x, y \in X} = \hat{S}_f \). We can overcome this problem with some additional condition which relates the characters of \((D, \hat{\ast})\) with the characters of the commutative hypergroup \((D, \ast)\) which is associated with the association scheme \((X, D, (R_i)_{i \in D})\):

**Definition 5.9.** Let \( \Lambda := (X, D, (R_i)_{i \in D}, (\hat{S}_i)_{i \in D}) \) be a generalized commutative association scheme with the associated hypergroups \((D, \hat{\ast})\) and \((D, \ast)\) as above. We say that \( \alpha \in (D, \hat{\ast})^\wedge \) has the positive connection property if \( \alpha \) is positive definite on \((D, \ast)\), and if the associated kernel \( F_\alpha : X \times X \to \mathbb{C} \) is positive semidefinite.

We say that \( \Lambda \) has the positive connection property, if all \( \alpha \in (D, \hat{\ast})^\wedge \) have the positive connection property.

To understand this condition, consider a finite generalized commutative association scheme \( \Lambda \). If a character \( \alpha \in (D, \hat{\ast})^\wedge \) can be written as a nonnegative linear combination of the characters of \((D, \ast)\), then \( \alpha \) is positive definite on \((D, \ast)\), and by Corollary 4.5, the associated kernel \( F_\alpha \) is positive semidefinite.

In this way, the positive connection property of \( \Lambda \) roughly means that all \( \alpha \in (D, \hat{\ast})^\wedge \) admit nonnegative integral representations w.r.t. the characters of \((D, \ast)\). Such positive integral representations are well known for many families of special functions and often easier to prove than positive product formulas.

We now use the positive connection property and obtain the following extension of Theorem 4.6:
Theorem 5.10. Let \((D, \ast)\) be a commutative discrete hypergroup which corresponds to the generalized association scheme \(\Lambda := (X, D, (R_i)_{i \in D}, (\tilde{S}_i)_{i \in D})\). Let \(\alpha, \beta \in (D, \ast)^N\) be characters such that \(\alpha \ast \beta \in \mathcal{M}(D, \ast)^N\) with \((\delta_\alpha \ast \delta_\beta)^\vee = \alpha \ast \beta\). The support of this measure is contained in \(S\).

Proof. By Corollary 5.8, \(\tilde{S}_\alpha\) is \(\pi\)-positive definite, i.e., \((\pi\{x\})(\tilde{S}_\alpha)_{x,y} \in X\) is a positive semidefinite matrix. Furthermore, as \(\beta\) has the positive connection property, \((F_\beta(x, y))_{x,y \in X}\) is also positive semidefinite. We conclude from Lemma 4.2 that the pointwise product \((\pi\{x\})(\tilde{S}_\alpha)_{x,y} \in X\) is also positive semidefinite. On the other hand, for all \(x, y \in X\) and \(i \in D\) with \((x, y) \in R_i\), we have

\[
(\tilde{S}_\alpha)_{x,y} F_\beta(x, y) = \beta(i) \alpha(i) \omega_i(\tilde{S}_i)_{x,y} = (\tilde{S}_{\alpha \beta})_{x,y}.
\]

Therefore, \(\tilde{S}_{\alpha \beta}\) is \(\pi\)-positive definite, and, by Lemma 5.6, \(\alpha \beta\) is positive definite on \((D, \ast)\). As in the proof of Theorem 4.6, Bochner’s theorem 2.10(1) together with Theorem 2.1(4) of [36] lead to the theorem. \(\square\)

We remark that Theorem 4.7 can be also established for finite generalized commutative association schemes similar to Theorem 5.10.

We complete the paper with an example.

Example 5.11. Consider the group \((\mathbb{Z}, +)\), on which the group \(H := \mathbb{Z}_2 = \{\pm 1\}\) acts multiplicatively as group of automorphisms. Consider the semidirect product \(G := \mathbb{Z} \rtimes \mathbb{Z}_2\) with \(H := \mathbb{Z}_2\) as finite subgroup. We then consider \(X := G/H = \mathbb{Z}\) and \(D := G/H = \mathbb{N}_0\) (with the canonical identifications) as well as the corresponding commutative association scheme \((\mathbb{Z}, \mathbb{N}_0, (R_k)_{k \in \mathbb{N}_0})\) with the associated double coset hypergroup \((\mathbb{N}_0, \ast)\) and the associated transition matrices

\[
S_0 = I, \quad S_k(x, y) = \frac{1}{2} \delta_{k,|x-y|} \quad (k \in \mathbb{N}, x, y \in \mathbb{Z})
\]

with the Kronecker-\(\delta\).

Now fix some parameter \(r > 0\) and put \(p := e^r / (e^r + e^{-r}) \in ]0, 1[\). Define the deformed stochastic matrices

\[
\tilde{S}_0 = I, \quad \tilde{S}_k(x, y) = \frac{1}{p^k + (1-p)^k} \left( p^k \delta_{k,y-x} + (1-p)^k \delta_{k,x-y} \right)
\]

for \(k \in \mathbb{N}_0, x, y \in \mathbb{Z}\). It can be easily checked that for \(k \in \mathbb{N}\),

\[
\tilde{S}_k \tilde{S}_l = \frac{p^{k+l} + (1-p)^{k+l}}{p^k + (1-p)^k} \tilde{S}_{k+l} + \left( \frac{p^{k+l} + (1-p)^{k+l}}{p^k + (1-p)^k} \right) \tilde{S}_{k-1}
= \frac{\cosh((k+1)r)}{2 \cosh(kr) \cosh(r)} \tilde{S}_{k+1} + \frac{\cosh((k-1)r)}{2 \cosh(kr) \cosh(r)} \tilde{S}_{k-1}.
\]

(5.1)

Induction yields that for \(k, l \in \mathbb{N}_0\),

\[
\tilde{S}_k \tilde{S}_l = \frac{\cosh((k+l)r)}{2 \cosh(kr) \cosh(lr)} \tilde{S}_{k+l} + \frac{\cosh((k-l)r)}{2 \cosh(kr) \cosh(lr)} \tilde{S}_{k-l}.
\]

(5.2)

We thus obtain the axioms (1)-(4) of 5.1. Moreover, with the measure \(\pi\{x\} := \left( \frac{p}{1-p} \right)^x = e^{2rx} (x \in \mathbb{Z})\) and the identity as involution on \(\mathbb{N}_0\) we also obtain axiom
5.1(5). We conclude from (5.2) that the associated hypergroup \((\mathbb{N}_0, \ast)\) is the so-called discrete \(\text{cosh}\)-hypergroup; see [40] and 3.4.7 and 3.5.72 of [7]. The characters of \((\mathbb{N}_0, \ast)\) are given by 
\[
\alpha_\lambda(n) := \frac{\cos(\lambda n)}{\cosh(r n)} \quad (n \in \mathbb{N}_0, \lambda \in [0, \pi] \cup i \cdot [0, r] \cup \{\pi + iz : z \in [0, r]\})
\]
where in this parameterization, \(\alpha_\lambda\) is in the support \(S\) of the Plancherel measure precisely for \(\lambda \in [0, \pi]\). Using 
\[
\frac{\cos(\lambda n)}{\cosh(r n)} = \frac{1}{2} \int_{-\infty}^{\infty} \cos(t n) \cosh((t + \lambda/r)\pi/2) \, dt \quad \text{for} \quad \lambda \in \mathbb{C}, |\Im \lambda| < r
\]
(see (1) in [40] and references there) and degenerated formulas for \(\lambda = ir, \pi + ir\), we see readily that each character of \((\mathbb{N}_0, \ast)\) has a positive integral representation w.r.t. characters of \((\mathbb{N}_0, \ast)\). As for the hypergroup \((\mathbb{N}_0, \ast)\), the support of the Plancherel measure is equal to \((\mathbb{N}_0, \ast)^\wedge\), we conclude that the generalized association scheme \((\mathbb{Z}, \mathbb{N}_0, (R_k)_{k \in \mathbb{N}_0}, (S_k)_{k \in \mathbb{N}_0})\) has the positive connection property. We thus may apply Theorem 5.10. The associated dual convolution can be determined explicitly similar to [40].

Notice that the hypergroups \((\mathbb{N}_0, \ast)\) and \((\mathbb{N}_0, \overset{\sim}{\ast})\) are related by a deformation of the convolution via some positive character as described in [33].

References


Michael Voit: Fakultät Mathematik, Technische Universität Dortmund, Vogelpothsweg 87, D-44221 Dortmund, Germany
E-mail address: michael.voit@math.tu-dortmund.de
CONDITIONS FOR STATIONARITY AND ERGODICITY OF TWO-FACTOR AFFINE DIFFUSIONS

BEÁTA BOLYOG AND GYULA PAP

Abstract. Sufficient conditions are presented for the existence of a unique stationary distribution and exponential ergodicity of two-factor affine diffusion processes.

1. Introduction

We consider general 2-dimensional two-factor affine diffusion processes
\[
\begin{align*}
\text{d}Y_t &= (a - bY_t) \text{d}t + \sigma_1 \sqrt{Y_t} \text{d}W_t, \\
\text{d}X_t &= (\alpha - \beta Y_t - \gamma X_t) \text{d}t + \sigma_2 \sqrt{Y_t}(\varrho \text{d}W_t + \sqrt{1 - \varrho^2} \text{d}B_t) + \sigma_3 \text{d}L_t,
\end{align*}
\]
for \( t \in [0, \infty) \), where \( a \in [0, \infty) \), \( b, \alpha, \beta, \gamma \in \mathbb{R} \), \( \sigma_1, \sigma_2, \sigma_3 \in [0, \infty) \), \( \varrho \in [-1, 1] \) and \((W_t, B_t, L_t)_{t \in [0, \infty)}\) is a 3-dimensional standard Wiener process. Affine processes are joint generalizations of continuous state branching processes and Ornstein–Uhlenbeck type processes, and they have applications in financial mathematics, see, e.g., in Duffie et al. [7]. The aim of the present paper is to extend the results of Barczy et al. [1] for the processes given in (1.1), where the case of \( \beta = 0, \ \varrho = 0, \ \sigma_1 = 1, \ \sigma_2 = 1, \ \sigma_3 = 0 \) is covered. We give sufficient conditions for the existence of a unique stationary distribution and exponential ergodicity, see Theorems 3.1 and 4.1, respectively. These results can be used in a forthcoming paper for studying parameter estimation for this model. An important observation is that it is enough to prove the results for the special case of \( \varrho = 0 \), since there is a non-singular linear transform of a 2-dimensional affine diffusion process which is a special 2-dimensional affine diffusion process with \( \varrho = 0 \), see Proposition 2.5. Otherwise, the method of the proofs are the same as in Barczy et al. [1].

2. The Affine Two-factor Diffusion Model

Let \( \mathbb{N}, \mathbb{Z}_+, \mathbb{R}, \mathbb{R}_+, \mathbb{R}_{++}, \mathbb{R}_-, \mathbb{R}_{--} \) and \( \mathbb{C} \) denote the sets of positive integers, non-negative integers, real numbers, non-negative real numbers, positive real numbers, non-positive real numbers, negative real numbers and complex numbers, respectively. For \( x, y \in \mathbb{R} \), we will use the notations \( x \wedge y := \min(x, y) \) and \( x \vee y := \max(x, y) \). By \( C^2(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R}) \), we denote the set of twice continuously differentiable real-valued functions on \( \mathbb{R}_+ \times \mathbb{R} \) with compact support. We will
denote the convergence in distribution and equality in distribution by \( \overset{D}{\longrightarrow} \) and \( \overset{=}{\longrightarrow} \), respectively.

We start with the definition of a two-factor affine process.

**Definition 2.1.** A time-homogeneous Markov process \((Y_t, X_t)_{t \in \mathbb{R}_+}\) with state space \(\mathbb{R}_+ \times \mathbb{R}\) is called a **two-factor affine process** if its (conditional) characteristic function takes the form

\[
\mathbb{E}[e^{i(u_1Y_t + u_2X_t)} | (Y_0, X_0) = (y_0, x_0)] = \exp\{y_0G_1(t, u_1, u_2) + x_0G_2(t, u_1, u_2) + H(t, u_1, u_2)\}, \quad (u_1, u_2) \in \mathbb{R}^2,
\]

for \((y_0, x_0) \in \mathbb{R}_+ \times \mathbb{R}, \ t \in \mathbb{R}_+\), where \(G_1(t, u_1, u_2), G_2(t, u_1, u_2), H(t, u_1, u_2) \in \mathbb{C}\).

Let \((\Omega, \mathcal{F}, (F_t)_{t \in \mathbb{R}_+}, \mathbb{P})\) be a filtered probability space satisfying the usual conditions, i.e., \((\Omega, \mathcal{F}, \mathbb{P})\) is complete, the filtration \((F_t)_{t \in \mathbb{R}_+}\) is right-continuous and \(F_0\) contains all the \(\mathbb{P}\)-null sets in \(\mathcal{F}\). Let \((W_t, B_t, L_t)_{t \in [0, \infty)}\) be a 3-dimensional standard \((\mathbb{F})\)-Wiener process.

The next proposition is about the existence and uniqueness of a strong solution of the SDE (1.1).

**Proposition 2.2.** Let \((\eta_0, \xi_0)\) be a random vector independent of the process \((W_t, B_t, L_t)_{t \in \mathbb{R}_+}\) satisfying \(\mathbb{P}(\eta_0 \in \mathbb{R}_+) = 1\). Then for all \(a \in \mathbb{R}_+, \ b, \alpha, \beta, \gamma \in \mathbb{R}, \ \sigma_1, \sigma_2, \sigma_3 \in \mathbb{R}_+, \ \vartheta \in [-1, 1]\), there is a (pathwise) unique strong solution \((Y_t, X_t)_{t \in \mathbb{R}_+}\) of the SDE (1.1) such that \(\mathbb{P}((Y_0, X_0) = (\eta_0, \xi_0)) = 1\) and \(\mathbb{P}(Y_t \in \mathbb{R}_+ \text{ for all } t \in \mathbb{R}_+) = 1\). Further, for all \(s, t \in \mathbb{R}_+\) with \(s \leq t\), we have

\[
Y_t = e^{-b(t-s)}Y_s + a \int_s^t e^{-b(t-u)} du + \sigma_1 \int_s^t e^{-b(t-u)} \sqrt{Y_u} dW_u \quad (2.1)
\]

and

\[
X_t = e^{-\gamma(t-s)}X_s + \int_s^t e^{-\gamma(t-u)}(\alpha - \beta Y_u) du + \sigma_2 \int_s^t e^{-\gamma(t-u)} \sqrt{Y_u} (\vartheta dW_u + \sqrt{1-\vartheta^2} dB_u) + \sigma_3 \int_s^t e^{-\gamma(t-u)} dL_u \quad (2.2)
\]

Moreover, \((Y_t, X_t)_{t \in \mathbb{R}_+}\) is a two-factor affine process with infinitesimal generator

\[
(A_{(Y,X)}f)(y, x) = (a - by)f_1'(y, x) + (\alpha - \beta y - \gamma x)f_2'(y, x)
+ \frac{1}{2} y\sigma_1^2 f_1''(y, x) + 2\vartheta \sigma_1 \sigma_2 f_1''(y, x) + \sigma_2^2 f_2''(y, x) + \sigma_3^2 f_3''(y, x),
\]

where \((y, x) \in \mathbb{R}_+ \times \mathbb{R}, \ f \in C_c^2(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R}), \text{ and } f_i', i \in \{1, 2\}, \text{ and } f_{i,j}'', i, j \in \{1, 2\}, \text{ denote the first and second order partial derivatives of } f \text{ with respect to its } i\text{-th and } i\text{-th and } j\text{-th variables.}

Conversely, every two-factor affine diffusion process is a (pathwise) unique strong solution of a SDE (1.1) with suitable parameters \(a \in \mathbb{R}_+, \ b, \alpha, \beta, \gamma \in \mathbb{R}, \ \sigma_1, \sigma_2, \sigma_3 \in \mathbb{R}_+\) and \(\vartheta \in [-1, 1]\).
Proof. Equation (1.1) is a special case of the equation (6.6) in Dawson and Li [6], and Theorem 6.2 in Dawson and Li [6] implies that for any initial value \((\eta_0, \xi_0)\) with \(\mathbb{P}(\{\eta_0, \xi_0\} \in \mathbb{R}_+ \times \mathbb{R}) = 1\) and \(\mathbb{E}(\eta_0) < \infty, \mathbb{E}(\xi_0) < \infty\), there exists a pathwise unique non-negative strong solution satisfying \(\mathbb{P}((Y_0, X_0) = (\eta_0, \xi_0)) = 1\) and \(\mathbb{P}(Y_t \in \mathbb{R}_+ \text{ for all } t \in \mathbb{R}_+) = 1\).

Applications of the Itô’s formula to the processes \((e^{bt}Y_t)_{t \in \mathbb{R}_+}\) and \((e^{\gamma t}X_t)_{t \in \mathbb{R}_+}\) give formulas (2.1) and (2.2), respectively.

The form of the infinitesimal generator (2.3) readily follows by (6.5) in Dawson and Li [6]. Further, Theorem 6.2 in Dawson and Li [6] also implies that \(Y\) is a continuous state and continuous time branching process with infinitesimal generator given in the Proposition.

The converse follows from Theorems 6.1 and 6.2 in Dawson and Li [6]. \(\Box\)

Next we present a result about the first moment of \((Y_t, X_t)_{t \in \mathbb{R}_+}\) together with its asymptotic behavior as \(t \to \infty\). Note that the formula for \(\mathbb{E}(Y_t)\), \(t \in \mathbb{R}_+\), is well known.

**Proposition 2.3.** Let \((Y_t, X_t)_{t \in \mathbb{R}_+}\) be the unique strong solution of the SDE (1.1) satisfying \(\mathbb{P}(Y_0 \in \mathbb{R}_+) = 1\) and \(\mathbb{E}(Y_0) < \infty, \mathbb{E}(X_0) < \infty\). Then

\[
\begin{bmatrix}
\mathbb{E}(Y_t) \\
\mathbb{E}(X_t)
\end{bmatrix} = \begin{bmatrix}
e^{-bt} - \beta e^{-\gamma t} \int_0^t e^{(\gamma-b)u} du & 0 \\
0 & e^{-bt} - \beta e^{-\gamma t} \int_0^t e^{(\gamma-b)u} du
\end{bmatrix} \begin{bmatrix}
\mathbb{E}(Y_0) \\
\mathbb{E}(X_0)
\end{bmatrix}
+ \begin{bmatrix}
-\beta e^{-\gamma t} \int_0^t e^{(\gamma-b)u} du \\
-\beta e^{-\gamma t} \int_0^t e^{(\gamma-b)u} du
\end{bmatrix} \begin{bmatrix}
0 \\
0
\end{bmatrix} \begin{bmatrix}
\alpha \\
\alpha
\end{bmatrix},
\]

Consequently, as \(t \to \infty\), if \(b \in \mathbb{R}_++\), then \(\mathbb{E}(Y_t) = \frac{\alpha}{\gamma} + O(e^{-bt})\) and

\[
\mathbb{E}(X_t) = \begin{cases}
\frac{\alpha}{\gamma} - \frac{\alpha \beta}{\gamma^2} + O(e^{-(\beta+\gamma)t}), & \gamma \in \mathbb{R}_+,
\alpha - \frac{\alpha \beta}{\gamma} + O(1), & \gamma = 0,
\left(\beta \text{ } \mathbb{E}(Y_0) + \mathbb{E}(X_0) - \frac{\alpha}{\gamma} + \frac{\alpha \beta}{\gamma^2} - \frac{\alpha^2}{b(\gamma-b)}\right)e^{-\gamma t} + O(1), & \gamma \in \mathbb{R}_-;
\end{cases}
\]

if \(b = 0\), then \(\mathbb{E}(Y_t) = at + O(1)\), and

\[
\mathbb{E}(X_t) = \begin{cases}
-\frac{\alpha}{\gamma} + O(1), & \gamma \in \mathbb{R}_+,
-\frac{\alpha}{\gamma} + O(1), & \gamma = 0,
\left(\frac{\beta}{\gamma} \mathbb{E}(Y_0) + \mathbb{E}(X_0) - \frac{\alpha}{\gamma} + \frac{\alpha \beta}{\gamma^2} - \frac{\alpha^2}{b(\gamma-b)}\right)e^{-\gamma t} + O(t), & \gamma \in \mathbb{R}_-;
\end{cases}
\]

if \(b \in \mathbb{R}_-\), then \(\mathbb{E}(Y_t) = (\mathbb{E}(Y_0) - \frac{\alpha}{\gamma})e^{-bt} + O(1)\), and

\[
\mathbb{E}(X_t) = \begin{cases}
\left(\beta \text{ } \mathbb{E}(Y_0) + \mathbb{E}(X_0) - \frac{\alpha}{\gamma} + \frac{\alpha \beta}{\gamma^2} - \frac{\alpha^2}{b(\gamma-b)}\right)e^{-\gamma t} + O(1), & \gamma \in \mathbb{R}_+,
\left(\frac{\beta}{\gamma} \mathbb{E}(Y_0) + \mathbb{E}(X_0) - \frac{\alpha}{\gamma} + \frac{\alpha \beta}{\gamma^2} - \frac{\alpha^2}{b(\gamma-b)}\right)e^{-\gamma t} + O(t), & \gamma = 0,
\left(\beta \text{ } \mathbb{E}(Y_0) + \mathbb{E}(X_0) - \frac{\alpha}{\gamma} + \frac{\alpha \beta}{\gamma^2} - \frac{\alpha^2}{b(\gamma-b)}\right)e^{-\gamma t} + O(e^{-bt}), & \gamma \in (b,0),
\left(\frac{\beta}{\gamma} \mathbb{E}(Y_0) + \mathbb{E}(X_0) - \frac{\alpha}{\gamma} + \frac{\alpha \beta}{\gamma^2} - \frac{\alpha^2}{b(\gamma-b)}\right)e^{-\gamma t} + O(e^{-bt}), & \gamma \in (-\infty,b).
\end{cases}
\]
Proof. It is sufficient to prove the statement in the case when \((Y_0, X_0) = (y_0, x_0)\) with an arbitrary \((y_0, x_0) \in \mathbb{R}_+ \times \mathbb{R}\), since then the statement of the proposition follows by the law of total expectation.

The formula for \(E(Y_t), t \in \mathbb{R}_+\), can be found, e.g., in Cox et al. [5, Equation (19)] or Jeanblanc et al. [9, Theorem 6.3.3.1]. Next we observe that
\[
\left( \int_0^t e^{-\gamma(t-u)} \sqrt{Y_u} d(\varrho W_u + \sqrt{1-\varrho^2} B_u) \right)_{t \in \mathbb{R}_+}
\]
(2.4)
is a square integrable martingale, since
\[
E \left[ \left( \int_0^t e^{-\gamma(t-u)} \sqrt{Y_u} d(\varrho W_u + \sqrt{1-\varrho^2} B_u) \right)^2 \right] = \int_0^t e^{-2\gamma(t-u)} E(Y_u) \, du < \infty,
\]
where the finiteness of the integral follows from
\[
E(Y_s) = e^{-b s} y_0 + a \int_0^s e^{-b u} \, du, \quad s \in \mathbb{R}_+,
\]
see, e.g., Cox et al. [5, Equation (19)], Jeanblanc et al. [9, Theorem 6.3.3.1] or Proposition 3.2 in Barczy et al. [1]. In a similar way,
\[
\left( \int_0^t e^{-\gamma(t-u)} dL_u \right)_{t \in \mathbb{R}_+}
\]
(2.5)
is a square integrable martingale, since
\[
E \left[ \left( \int_0^t e^{-\gamma(t-u)} dL_u \right)^2 \right] = \int_0^t e^{-2\gamma(t-u)} \, du < \infty.
\]
Taking expectations of both sides of the equation (2.2) and using the martingale property of the processes in (2.4) and (2.5), we have
\[
E(X_t) = e^{-\gamma t} x_0 + \int_0^t e^{-\gamma(t-u)} (\alpha - \beta E(Y_u)) \, du
\]
\[
= e^{-\gamma t} x_0 + \alpha \int_0^t e^{-\gamma(t-u)} \, du - \beta \int_0^t e^{-\gamma(t-u)} \left( e^{-b u} y_0 + a \int_0^u e^{-b v} \, dv \right) \, du
\]
\[
= e^{-\gamma t} x_0 - \beta y_0 e^{-\gamma t} \int_0^t e^{(\gamma-b) u} \, du + a \int_0^t e^{-\gamma u} \, du - \beta a e^{-\gamma t} \int_0^t e^{\gamma u} \left( \int_0^u e^{-b v} \, dv \right) \, du,
\]
t \in \mathbb{R}_+.

The asymptotic behavior of \(E(Y_t)\) as \(t \to \infty\) does not depend on \(\gamma\), which can be derived from
\[
E(Y_t) = e^{-b t} y_0 + a \int_0^t e^{-b u} \, du = \begin{cases} \frac{y_0}{b} + (y_0 - \frac{a}{b}) e^{-bt}, & b \neq 0, \\ y_0 + b t, & b = 0. \end{cases}
\]
The asymptotic behavior of $E(X_t)$ as $t \to \infty$ does depend on $b$ and $\gamma$ as well. We have
\[
\int_0^t e^{-\gamma v} dv = \begin{cases} \frac{1-e^{-\gamma t}}{\gamma}, & \gamma \neq 0, \\ t, & \gamma = 0, \end{cases} \quad e^{-\gamma t} \int_0^t e^{(\gamma-b)u} du = \begin{cases} \frac{e^{bt} - e^{-\gamma t}}{\gamma-b}, & b \neq \gamma, \\ \frac{e^{-bt}}{\gamma-b}, & b = \gamma, \end{cases}
\]
\[
eq \begin{cases} \frac{1-e^{-\gamma t}}{b\gamma} - \frac{e^{-bt} - e^{-\gamma t}}{(\gamma-b)b}, & b \neq 0, \gamma \neq 0, b \neq \gamma, \\ \frac{\gamma}{b} - \frac{1-e^{-bt}}{b^2}, & b \neq 0, \gamma = 0, \\ \frac{\gamma}{\gamma^2} - \frac{1-e^{-\gamma t}}{\gamma^2}, & b = 0, \gamma \neq 0, \\ \frac{1}{b^2}, & b = 0, \gamma = 0. \end{cases}
\]
Consequently, if $b \neq 0$, $\gamma \neq 0$ and $b \neq \gamma$, then
\[
E(X_t) = e^{-\gamma t}x_0 - \beta y_0 \frac{e^{-bt} - e^{-\gamma t}}{\gamma-b} + \frac{\alpha}{\gamma}(1-e^{-\gamma t}) \\
- \frac{a\beta}{b\gamma}(1-e^{-\gamma t}) + \frac{a\beta}{(\gamma-b)b}(e^{-bt} - e^{-\gamma t}) \\
= \frac{\alpha}{\gamma} \frac{a\beta}{b\gamma} + \left( -\frac{\beta}{\gamma-b}y_0 + \frac{a\beta}{(\gamma-b)b} \right) e^{-bt} \\
+ \left( \frac{\beta}{\gamma-b}y_0 + x_0 - \frac{\alpha}{\gamma} + \frac{a\beta}{b\gamma} - \frac{a\beta}{(\gamma-b)b} \right) e^{-\gamma t}.
\]
Moreover, if $b \neq 0$, $\gamma \neq 0$ and $b = \gamma$, then
\[
E(X_t) = e^{-\gamma t}x_0 - \beta y_0 te^{-\gamma t} + \frac{\alpha}{\gamma}(1-e^{-\gamma t}) - \frac{a\beta}{b\gamma}(1-e^{-\gamma t}) + \frac{a\beta}{b}te^{-\gamma t} \\
= \frac{\alpha}{\gamma} \frac{a\beta}{b\gamma} + \left( x_0 - \frac{\alpha}{\gamma} + \frac{a\beta}{b\gamma} \right) e^{-\gamma t} + \left( -\beta y_0 + \frac{a\beta}{b} \right) e^{-\gamma t}.
\]
Further, if $b \neq 0$ and $\gamma = 0$, then
\[
E(X_t) = e^{-\gamma t}x_0 - \beta y_0 \frac{1-e^{-bt}}{b} + \alpha t - \frac{a\beta}{b\gamma}t + \frac{a\beta}{b^2}(1-e^{-bt}) \\
= -\frac{\beta}{b}y_0 + \frac{a\beta}{b^2} + \left( \alpha - \frac{a\beta}{b} \right)t + \left( \frac{\beta}{b}y_0 + x_0 - \frac{a\beta}{b^2} \right) e^{-bt}.
\]
In a similar way, if $b = 0$ and $\gamma \neq 0$, then
\[
E(X_t) = e^{-\gamma t}x_0 - \beta y_0 \frac{1-e^{-\gamma t}}{\gamma} + \frac{\alpha}{\gamma}(1-e^{-\gamma t}) - \frac{a\beta}{\gamma^2} \left( \frac{t}{\gamma} - \frac{1-e^{-\gamma t}}{\gamma^2} \right) \\
= -\frac{\beta}{\gamma}y_0 + \frac{\alpha}{\gamma^2} + \frac{a\beta}{\gamma^2} - \frac{a\beta}{\gamma}t + \left( \frac{\beta}{\gamma}y_0 + x_0 - \frac{\alpha}{\gamma} + \frac{a\beta}{\gamma^2} \right) e^{-\gamma t},
\]
and if $b = 0$ and $\gamma = 0$, then

$$E(X_t) = x_0 - \beta y_0 t + \alpha t - a\beta t^2. $$

The asymptotic behavior of $E(X_t)$ as $t \to \infty$ can be derived from the above formulas. \hfill \Box

Based on the asymptotic behaviour of the expectations $(E(Y_t), E(Z_t))$ as $t \to \infty$, we introduce a classification of affine diffusion processes given by the SDE (1.1).

**Definition 2.4.** Let $(Y_t, X_t)_{t \in \mathbb{R}_+}$ be the unique strong solution of the SDE (1.1) satisfying $P(Y_0 \in \mathbb{R}_+) = 1$. We call $(Y_t, X_t)_{t \in \mathbb{R}_+}$ *subcritical*, *critical* or *supercritical* if $b \wedge \gamma \in \mathbb{R}_{++}$, $b \wedge \gamma = 0$ or $b \wedge \gamma \in \mathbb{R}_{--}$, respectively.

The next proposition describes a non-singular linear transform of a 2-dimensional affine diffusion process which is a special 2-dimensional affine diffusion process with $\varrho = 0$.

**Proposition 2.5.** Let us consider the 2-dimensional affine diffusion model (1.1) with $a \in \mathbb{R}_+$, $b, \alpha, \beta, \gamma \in \mathbb{R}$, $\sigma_1, \sigma_2, \sigma_3 \in \mathbb{R}_+$, $\varrho \in [-1, 1]$, and with a random initial value $(\eta_0, \zeta_0)$ independent of $(W_t, B_t, L_t)_{t \in \mathbb{R}_+}$ satisfying $P(\eta_0 \in \mathbb{R}_+) = 1$. Put

$$c := \begin{cases} 0, & \text{if } \sigma_1 = 0, \\ \frac{\sigma_2 \varrho}{\varrho^2}, & \text{if } \sigma_1 > 0, \end{cases} \quad Z_t := X_t - cY_t, \quad t \in \mathbb{R}_+. \tag{2.6}$$

Then the process $(Y_t, Z_t)_{t \in \mathbb{R}_+}$ is a regular affine process with infinitesimal generator

$$(A(Y, Z)) f(y, z) = (a - by) f_1'(y, z) + (A - By - \gamma z) f_2'(y, z) + \frac{1}{2} y \left[ \sigma_1^2 f_1''(y, z) + \Sigma_2^2 f_2''(y, z) \right] + \frac{1}{2} \Sigma_3^2 f_2''(y, z), \tag{2.7}$$

for $(y, z) \in \mathbb{R}_+ \times \mathbb{R}$ and $f \in C^2_b(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R})$, where

$$A := \alpha - ca, \quad B := \beta - c(b - \gamma), \quad \Sigma_2 := \begin{cases} \sigma_2, & \text{if } \sigma_1 = 0, \\ \sigma_2 \sqrt{1 - \varrho^2}, & \text{if } \sigma_1 > 0. \end{cases}$$

**Proof.** If $\sigma_1 = 0$, then the statement follows from Proposition 2.2.

If $\sigma_1 > 0$, then, by Itô’s formula, $(Y_t, Z_t)_{t \in \mathbb{R}_+}$ is the unique strong solution of the SDE

$$\begin{aligned} &dY_t = (a - By_t) dt + \sigma_1 \sqrt{Y_t} dW_t, \\
&dZ_t = (A - BY_t - \gamma Z_t) dt + \Sigma_2 \sqrt{Y_t} dB_t + \sigma_3 dL_t. \tag{2.8} \end{aligned}$$

with random initial value $(\eta_0, \zeta_0 - \varrho \eta_0)$. By Proposition 2.2, $(Y_t, Z_t)_{t \in \mathbb{R}_+}$ is a regular affine process with infinitesimal generator (2.7). \hfill \Box
3. Stationarity

The following result states the existence of a unique stationary distribution of the affine diffusion process given by the SDE (1.1). Let \( \mathbb{C}_- := \{ z \in \mathbb{C} : \text{Re}(z) \leq 0 \} \).

**Theorem 3.1.** Let us consider the 2-dimensional affine diffusion model (1.1) with \( a \in \mathbb{R}_+, \ b \in \mathbb{R}_+, \ \alpha, \beta \in \mathbb{R}, \ \gamma \in \mathbb{R}_++, \ \sigma_1, \sigma_2, \sigma_3 \in \mathbb{R}_+, \ \varrho \in [-1,1], \) and with a random initial value \((\eta_0, \zeta_0)\) independent of \((W_t, B_t, L_t)_{t \in \mathbb{R}_+}\) satisfying \( \mathbb{P}(\eta_0 \in \mathbb{R}_+) = 1 \). Then

(i) \((Y_t, X_t) \xrightarrow{d} (Y_\infty, X_\infty)\) as \( t \to \infty \), and we have

\[
\mathbb{E}(e^{u_1 Y_\infty + i \lambda_2 X_\infty}) = \exp \left\{ a \int_0^\infty K_s(u_1, \lambda_2) \, ds + i \frac{\alpha}{\gamma} \lambda_2 - \frac{\sigma_2^2}{4 \gamma} \lambda_2^2 \right\} \tag{3.1}
\]

for \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}, \) where \( K_t(u_1, \lambda_2) \), \( t \in \mathbb{R}_+ \), is the unique solution of the (deterministic) differential equation

\[
\begin{aligned}
\frac{\partial K_t}{\partial t}(u_1, \lambda_2) &= -bK_t(u_1, \lambda_2) - i\beta e^{-\gamma t} \lambda_2 + \frac{1}{2} \sigma_2^2 \kappa_t(u_1, \lambda_2)^2 \\
&\quad + i \sigma_1 \sigma_2 e^{-\gamma t} \lambda_2 \kappa_t(u_1, \lambda_2) - \frac{1}{4} \sigma_2^2 e^{-2\gamma t} \lambda_2^2,
\end{aligned} \tag{3.2}
\]

(ii) supposing that the random initial value \((\eta_0, \zeta_0)\) has the same distribution as \((Y_\infty, X_\infty)\) given in part (i), \((Y_t, X_t)_{t \in \mathbb{R}_+}\) is strictly stationary.

**Proof.** First we check that it is enough to prove the statement (i) for the special affine diffusion process \((Y_t, Z_t)_{t \in \mathbb{R}_+}\) given in Proposition 2.5. Hence we suppose that (i) holds for \((Y_t, Z_t)_{t \in \mathbb{R}_+}\), and we check that then (i) holds for \((Y_t, X_t)_{t \in \mathbb{R}_+}\) as well.

If \( \sigma_1 = 0 \) then \((Y_t, X_t)_{t \in \mathbb{R}_+} = (Y_t, Z_t)_{t \in \mathbb{R}_+}\), hence (i) trivially holds for \((Y_t, X_t)_{t \in \mathbb{R}_+}\) as well.

If \( \sigma_1 > 0 \) then \((Y_t, Z_t) \xrightarrow{d} (Y_\infty, Z_\infty)\) as \( t \to \infty \), and we have

\[
\mathbb{E}(e^{u_1 Y_\infty + i \lambda_2 Z_\infty}) = \exp \left\{ a \int_0^\infty K_s(u_1, \lambda_2) \, ds + i \frac{\alpha}{\gamma} \lambda_2 - \frac{\sigma_2^2}{4 \gamma} \lambda_2^2 \right\}
\]

for \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}, \) where \( K_t(u_1, \lambda_2) \), \( t \in \mathbb{R}_+ \), is the unique solution of the differential equation

\[
\begin{aligned}
\frac{\partial K_t}{\partial t}(u_1, \lambda_2) &= -bK_t(u_1, \lambda_2) - i\beta e^{-\gamma t} \lambda_2 \\
&\quad + \frac{1}{2} \sigma_1^2 K_t(u_1, \lambda_2)^2 + \frac{1}{2} \Sigma_2^2 e^{-2\gamma t} \lambda_2^2,
\end{aligned} \tag{3.3}
\]

By the continuous mapping theorem, we obtain

\[
(Y_t, X_t) = \left( Y_t, Z_t + \frac{\sigma_2 \varrho}{\sigma_1} Y_t \right) \xrightarrow{d} (Y_\infty, Z_\infty + \frac{\sigma_2 \varrho}{\sigma_1} Y_\infty) =: (Y_\infty, X_\infty) \quad \text{as} \quad t \to \infty.
\]
Moreover, for each \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}\),

\[
E(e^{u_1 Y_\infty + i \lambda_2 X_\infty}) = E\left(\exp\left(u_1 Y_\infty + i \lambda_2 \left(Z_\infty + \frac{\sigma_2 \theta}{\sigma_1} Y_\infty\right)\right)\right)
\]

\[
= E\left(\exp\left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2\right) Y_\infty + i \lambda_2 Z_\infty\right)
\]

\[
= \exp\left\{a \int_0^\infty K_s \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right) ds + i \frac{A}{\gamma} \lambda_2 - \frac{\sigma_3^2}{4 \gamma} \lambda_2^2\right\}
\]

\[
= \exp\left\{a \int_0^\infty K_s \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right) ds + i \frac{\alpha}{\gamma} \lambda_2 - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2 - \frac{\sigma_3^2}{4 \gamma} \lambda_2^2\right\}
\]

\[
= \exp\left\{a \int_0^\infty \kappa_s(u_1, \lambda_2) ds + i \frac{\alpha}{\gamma} \lambda_2 - \frac{\sigma_3^2}{4 \gamma} \lambda_2^2\right\},
\]

where

\[
\kappa_t(u_1, \lambda_2) := K_t \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right) - i \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2
\]

for \(t \in \mathbb{R}_+\) and \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}\). Using that \(K_t(u_1, \lambda_2), \ t \in \mathbb{R}_+\), satisfies the differential equation (3.3), we get

\[
\frac{\partial \kappa_t}{\partial t}(u_1, \lambda_2) = \frac{\partial K_t}{\partial t} \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right) + i \gamma \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2
\]

\[
= -b K_t \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right) - i B e^{-\gamma t} \lambda_2 + \frac{1}{2} \sigma_1^2 K_t \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right)^2
\]

\[
- \frac{1}{2} \sigma_2^2 e^{-2 \gamma t} \lambda_2^2 + i \gamma \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2
\]

\[
= -b \left(\kappa_t(u_1, \lambda_2) + i \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2\right) - i \left(\beta - \frac{\sigma_2 \theta}{\sigma_1} (b - \gamma)\right) e^{-\gamma t} \lambda_2
\]

\[
+ \frac{1}{2} \sigma_1^2 \left(\kappa_t(u_1, \lambda_2) + i \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2\right)^2
- \frac{1}{2} \sigma_2^2 (1 - \theta^2) e^{-2 \gamma t} \lambda_2^2 + i \gamma \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2
\]

\[
= -b \kappa_t(u_1, \lambda_2) - i \beta e^{-\gamma t} \lambda_2 + \frac{1}{2} \sigma_1^2 \kappa_t(u_1, \lambda_2)^2
+ i \theta \sigma_1 \sigma_2 e^{-\gamma t} \lambda_2 \kappa_t(u_1, \lambda_2) - \frac{1}{2} \sigma_2^2 e^{-2 \gamma t} \lambda_2^2,
\]

and

\[
\kappa_0(u_1, \lambda_2) = K_0 \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right) - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2 = \left(u_1 + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2\right) - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2 = u_1,
\]

hence \(\kappa_t(u_1, \lambda_2), \ t \in \mathbb{R}_+\), is a solution of the differential equation (3.2). In a similar way, if \(\kappa_t(u_1, \lambda_2), \ t \in \mathbb{R}_+\), satisfies the differential equation (3.2), then

\[
K_t(u_1, \lambda_2) := \kappa_t \left(u_1 - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2\right) + i \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2, \quad t \in \mathbb{R}_+,
\]
is a solution of the differential equation (3.3) for each \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}\), since
\[
\frac{\partial K_t}{\partial t}(u_1, \lambda_2) = \frac{\partial \kappa_t}{\partial t} \left( u_1 - \frac{i \sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2 \right) - i \gamma \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2
\]
\[
= -b \kappa_t \left( u_1 - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2 \right) - i \beta e^{-\gamma t} \lambda_2 + \frac{1}{2} \gamma_1^2 \kappa_t \left( u_1 - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2 \right)^2
\]
\[
+ i \frac{\sigma_2 \sigma_1}{\sigma_1} e^{-\gamma t} \frac{\kappa_t}{\lambda_2} \left( u_1 - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2 \right) - \frac{1}{2} \frac{\sigma_2^2}{\sigma_1^2} e^{-2\gamma t} \lambda_2^2 - i \gamma \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2
\]
\[
= -b \left( K_t(u_1, \lambda_2) - \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2 \right) - i \beta e^{-\gamma t} \lambda_2 + \frac{1}{2} \gamma_1^2 \left( K_t(u_1, \lambda_2) - \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2 \right)^2
\]
\[
+ i \frac{\sigma_2 \sigma_1}{\sigma_1} e^{-\gamma t} \frac{\kappa_t}{\lambda_2} \left( K_t(u_1, \lambda_2) - \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2 \right) - \frac{1}{2} \frac{\sigma_2^2}{\sigma_1^2} e^{-2\gamma t} \lambda_2^2 - i \gamma \frac{\sigma_2 \theta}{\sigma_1} e^{-\gamma t} \lambda_2
\]
\[
= -b K_t(u_1, \lambda_2) - i B e^{-\gamma t} \lambda_2 + \frac{1}{2} \gamma_1^2 K_t(u_1, \lambda_2)^2 - \frac{1}{2} \frac{\sigma_2^2}{\sigma_1^2} e^{-2\gamma t} \lambda_2^2,
\]
and
\[
K_0(u_1, \lambda_2) = \kappa_0 \left( u_1 - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2 \right) + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2 = \left( u_1 - i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2, \lambda_2 \right) + i \frac{\sigma_2 \theta}{\sigma_1} \lambda_2 = u_1.
\]
Consequently, \( \kappa_t(u_1, \lambda_2), \ t \in \mathbb{R}_+ \), is the unique solution of the differential equation (3.2).

(i): We prove this part for the special linear transform described in Proposition 2.5 in three steps.

**Step 1.** By Theorem 6.1 in Dawson and Li [6] and Proposition 2.2, we have
\[
E \left( e^{u(Y_t, Z_t)} \mid (Y_0, Z_0) = (y_0, z_0) \right) = e^{(iy_0, z_0), \psi_t(u)} + \phi_t(u)
\]
(3.4)
for \( u \in \mathbb{C}_- \times (i \mathbb{R}), \ (y_0, z_0) \in \mathbb{R}_+ \times \mathbb{R}, \ t \in \mathbb{R}_+, \) for all \( u = (u_1, u_2) \in \mathbb{C}_- \times (i \mathbb{R}), \) we have \( \psi_t(u) = (\psi_t^{(1)}(u), e^{-\gamma t} u_2), \ t \in \mathbb{R}_+, \) where \( \psi_t^{(1)}(u), \ t \in \mathbb{R}_+, \) is a solution of the Riccati equation
\[
\begin{cases}
\frac{\partial \psi_t^{(1)}}{\partial t}(u) = R(\psi_t^{(1)}(u), e^{-\gamma t} u_2), & t \in \mathbb{R}_+,
\psi_0^{(1)}(u) = u_1,
\end{cases}
\]
(3.5)
the function \( \mathbb{R}_+ \times (\mathbb{C}_- \times (i \mathbb{R})) \ni (t, u) \mapsto \psi_t^{(1)}(u) \) is continuous, and
\[
\phi_t(u) = \int_0^t F(\psi_s^{(1)}(u), e^{-\gamma s} u_2) \, ds, \quad t \in \mathbb{R}_+,
\]
where the (complex valued) functions \( F \) and \( R \) are given by
\[
F(u) = au_1 + i u_2 + \frac{1}{2} \sigma_2^2 u_2^2, \quad R(u) = -bu_1 - \sigma_1^2 u_2^2 + \frac{1}{2} \Sigma^2 u_2^2 + \frac{1}{2} \Sigma^2 u_2^2
\]
for \( u = (u_1, u_2) \in \mathbb{C}_- \times (i \mathbb{R}). \) Note that for every \( u = (u_1, u_2) \in \mathbb{C}_- \times (i \mathbb{R}) \) and \( t \in \mathbb{R}_+, \) we have \( \psi_t^{(1)}(u) \in \mathbb{C}_- \) and \( \phi_t(u) \in \mathbb{C}_-. \) Indeed,
\[
|E(\psi_t^{(1)}(u, Y_t, Z_t)) \mid (Y_0, Z_0) = (y_0, z_0))| \leq E(|e^{u(Y_t, Z_t)}| \mid (Y_0, Z_0) = (y_0, z_0)) \leq 1,
\]
since \(|e^{(u, (Y_t, Z_t))}| = e^{\text{Re}(u, (Y_t, Z_t))} = e^{Y_t \text{Re}(u_1)} \leq 1\) by \(Y_t \geq 0\) and \(\text{Re}(u_1) \leq 0\). Consequently,
\[
|e^{((y_0, z_0), \psi_t(u)) + \phi_t(u)}| = e^{\text{Re}((y_0, z_0), \psi_t(u)) + \text{Re}(\phi_t(u))} = e^{y_0 \text{Re}(\psi_t^{(1)}(u)) + \text{Re}(\phi_t(u))} \leq 1,
\]
hence \(y_0 \text{Re}(\psi_t^{(1)}(u)) + \text{Re}(\phi_t(u)) \leq 0\). Putting \(y_0 = 0\), we obtain \(\text{Re}(\phi_t(u)) \leq 0\), thus \(\phi_t(u) \in \mathbb{C}_-\). Further, for each \(y_0 > 0\), we have \(\text{Re}(\psi_t^{(1)}(u)) \leq -\text{Re}(\phi_t(u))/y_0\), thus letting \(y_0 \to \infty\), we obtain \(\text{Re}(\psi_t^{(1)}(u)) \leq 0\), and hence \(\psi_t^{(1)}(u) \in \mathbb{C}_-\).

Moreover, for all \(t \in \mathbb{R}_+\) and \(u = (u_1, u_2) \in \mathbb{C}_- \times (i\mathbb{R})\), we have
\[
\phi_t(u) = e^{\int_0^t (a\psi_s^{(1)}(u) + Ae^{-\gamma s}u_2 + \frac{1}{2}\sigma_s^2(e^{-\gamma s}u_2)^2) ds} = e^{\int_0^t \psi_s^{(1)}(u) ds + A\int_0^t u_2 \frac{1-e^{-\gamma t}}{\gamma} + \frac{1}{2}\sigma_s^2u_2^2 \frac{1-e^{-\gamma t}}{2\gamma}}.
\]

In fact, we have
\[
\mathbb{E}(e^{u_1 Y_t + i\lambda_2 Z_t} | (Y_0, Z_0) = (y_0, z_0)) = \exp \left\{ y_0 K_t(u_1, \lambda_2) + iz_0 e^{-\gamma t} \lambda_2 + g_t(u_1, \lambda_2) \right\},
\]
for \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}\) and \((y_0, z_0) \in \mathbb{R}_+ \times \mathbb{R}\), where
\[
g_t(u_1, \lambda_2) := a \int_0^t K_s(u_1, \lambda_2) ds + iA\lambda_2 \int_0^t \frac{1-e^{-\gamma t}}{\gamma} - \frac{1}{2}\sigma_s^2\lambda_2^2 \frac{1-e^{-\gamma t}}{2\gamma},
\]
and \(K_t(u_1, \lambda_2), t \in \mathbb{R}_+,\) is the unique solution of the differential equation (3.3). Indeed, by (3.4) with \(u_2 = i\lambda_2\), we have
\[
\mathbb{E}(e^{u_1 Y_t + i\lambda_2 Z_t} | (Y_0, Z_0) = (y_0, z_0)) = \exp \left\{ y_0 \psi_t^{(1)}(u_1, i\lambda_2) + iz_0 e^{-\gamma t} \lambda_2 + a \int_0^t \psi_s^{(1)}(u_1, i\lambda_2) ds + iA\lambda_2 \int_0^t \frac{1-e^{-\gamma t}}{\gamma} - \frac{1}{2}\sigma_s^2\lambda_2^2 \frac{1-e^{-\gamma t}}{2\gamma} \right\},
\]
for \((y_0, z_0) \in \mathbb{R}_+ \times \mathbb{R}\), where
\[
\left\{ \frac{\partial \psi_t^{(1)}}{\partial \lambda_2}(u_1, i\lambda_2) = -b\psi_t^{(1)}(u_1, i\lambda_2) - B(e^{-\gamma t}i\lambda_2)
\right\} + \frac{\sigma_t^2}{2}[\psi_t^{(1)}(u_1, i\lambda_2)]^2 + \frac{\Sigma_t^2}{2}(e^{-\gamma t}i\lambda_2)^2,
\]
and hence, for the function \(K_t(u_1, \lambda_2) := \psi_t^{(1)}(u_1, i\lambda_2), t \in \mathbb{R}_+,\) we obtain the differential equation (3.3). Recall that \(K_t(u_1, \lambda_2) \in \mathbb{C}_-\) for all \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}\). The uniqueness of the solution of the differential equation (Cauchy problem) (3.3) follows by general results of Duffie et al. [7, Propositions 6.1, 6.4 and Lemma 9.2].

Step 2. We show that there exists \(C_2 \in \mathbb{R}_{++}\) (depending on the parameters \(b\) and \(\gamma\)), and for each \((u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}\), there exists \(C_1(u_1, \lambda_2) \in \mathbb{R}_+\) (depending on the parameters \(b, B, \gamma, \sigma_1\) and \(\Sigma_2\)), such that
\[
|K_t(u_1, \lambda_2)| \leq C_1(u_1, \lambda_2)e^{-C_2 t}, \quad t \in \mathbb{R}_+.
\]
Let us introduce the functions \( v_t(u_1, \lambda_2), \ t \in \mathbb{R}_+, \) and \( w_t(u_1, \lambda_2), \ t \in \mathbb{R}_+, \) by

\[
v_t(u_1, \lambda_2) := -\text{Re}(K_t(u_1, \lambda_2)), \quad w_t(u_1, \lambda_2) := \text{Im}(K_t(u_1, \lambda_2)), \quad t \in \mathbb{R}_+.
\]

We observe that, as a consequence of (3.3), the function \( (v_t(u_1, \lambda_2), w_t(u_1, \lambda_2)), \ t \in \mathbb{R}_+, \) is the unique solution of the system of the Riccati equations

\[
\begin{align*}
\frac{\partial v_t(u_1, \lambda_2)}{\partial t} &= -bv_t(u_1, \lambda_2) - \frac{1}{2}\sigma_1^2(v_t(u_1, \lambda_2)^2 - w_t(u_1, \lambda_2)^2) + \frac{1}{2}\Sigma_t^2 e^{-2\gamma t}\lambda_2^2, \\
\frac{\partial w_t(u_1, \lambda_2)}{\partial t} &= -bw_t(u_1, \lambda_2) - Be^{-\gamma t}\lambda_2 - \sigma_1^2 v_t(u_1, \lambda_2)w_t(u_1, \lambda_2), \\
v_0(u_1, \lambda_2) &= -\text{Re}(u_1), \\
w_0(u_1, \lambda_2) &= \text{Im}(u_1).
\end{align*}
\]

(3.9)

Note that \( K_t(u_1, \lambda_2) \in \mathbb{C}_- \) implies \( v_t(u_1, \lambda_2) \geq 0 \) for all \( (u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R} \). Clearly, the function \( w_t(u_1, \lambda_2), \ t \in \mathbb{R}_+, \) is the unique solution of the inhomogeneous linear differential equation

\[
\begin{align*}
\frac{\partial w_t(u_1, \lambda_2)}{\partial t} &= -f_t(u_1, \lambda_2)w_t(u_1, \lambda_2) - B\lambda_2 e^{-\gamma t}, \quad t \in \mathbb{R}_+, \\
w_0(u_1, \lambda_2) &= \text{Im}(u_1),
\end{align*}
\]

(3.10)

with \( f_t(u_1, \lambda_2) := b + \sigma_1^2 v_t(u_1, \lambda_2), \ t \in \mathbb{R}_+ \). The general solution of the homogeneous linear differential equation

\[
\frac{\partial \tilde{w}_t(u_1, \lambda_2)}{\partial t} = -f_t(u_1, \lambda_2)\tilde{w}_t(u_1, \lambda_2), \quad t \in \mathbb{R}_+,
\]

takes the form

\[
\tilde{w}_t(u_1, \lambda_2) = C e^{-\int_0^t f_s(u_1, \lambda_2) \, dz}, \quad t \in \mathbb{R}_+,
\]

where \( C \in \mathbb{R} \). By variation of constants, the function

\[
\mathbb{R}_+ \ni t \mapsto -B\lambda_2 e^{-\int_0^t f_s(u_1, \lambda_2) \, dz} \int_0^t e^{-\gamma s + \int_0^s f_z(u_1, \lambda_2) \, dz} \, ds, \quad t \in \mathbb{R}_+,
\]

is a particular solution of the inhomogeneous linear differential equation (3.10). Hence a general solution of the inhomogeneous linear differential equation takes the form

\[
w_t(u_1, \lambda_2) = C e^{-\int_0^t f_s(u_1, \lambda_2) \, dz} - B\lambda e^{-\int_0^t f_s(u_1, \lambda_2) \, dz} \int_0^t e^{-\gamma s + \int_0^s f_z(u_1, \lambda_2) \, dz} \, ds
\]

for \( t \in \mathbb{R}_+ \). Taking into account of the initial value \( w_0(u_1, \lambda_2) = \text{Im}(u_1) \), we obtain \( C = \text{Im}(u_1) \). Consequently,

\[
|w_t(u_1, \lambda_2)| \leq |\text{Im}(u_1)| e^{-\int_0^t f_s(u_1, \lambda_2) \, dz} + |B\lambda_2| \int_0^t e^{-\gamma s + \int_0^s f_z(u_1, \lambda_2) \, dz} \, ds
\]

for \( t \in \mathbb{R}_+ \). Applying \( f_t(u_1, \lambda_2) \geq b > 0, \ t \in \mathbb{R}_+, \) we get \( e^{-\int_0^t f_s(u_1, \lambda_2) \, dz} \leq e^{-bt}, \ t \in \mathbb{R}_+, \) and

\[
\int_0^t e^{-\gamma s - \int_0^s f_z(u_1, \lambda_2) \, dz} \, ds \leq \int_0^t e^{-\gamma s - (t-s)b} \, ds = \begin{cases} \frac{e^{-\gamma t} - e^{-bt}}{b - \gamma} & b \neq \gamma, \\ tc^{-bt} \leq 2e^{-bt/2} & b = \gamma, \end{cases}
\]
\[ te^{-bt} \leq e^{-bt/2} \sup_{t \in \mathbb{R}_+} te^{-bt/2}, \text{ where } \sup_{t \in \mathbb{R}_+} te^{-bt/2} = 2e^{-1/b}. \] Summarizing, we have
\[ |w_1(u_1, \lambda_2)| \leq C_3(u_1, \lambda_2)e^{-C_2t}, \quad t \in \mathbb{R}_+, \] with
\[ C_3(u_1, \lambda_2) := |\text{Im}(u_1)| + |B\lambda_2| \left( \frac{1}{|b - \gamma|} \mathbb{I}_{\{b \neq \gamma\}} + \frac{2}{eb} \mathbb{I}_{\{b = \gamma\}} \right) \in \mathbb{R}_+, \]
\[ C_2 := \min\{\gamma, b/2\} \in \mathbb{R}_+. \]

Using (3.9) and (3.11), we obtain
\[
\begin{cases}
\frac{\partial u}{\partial t}(u_1, \lambda_2) \leq -bv_1(u_1, \lambda_2) + C_4(u_1, \lambda_2)e^{-C_2t}, & t \in \mathbb{R}_+, \\
v_0(u_1, \lambda_2) = -\text{Re}(u_1),
\end{cases}
\]
with \( C_4(u_1, \lambda_2) := (\sigma_1^2 C_3(u_1, \lambda_2)^2 + \Sigma_2^2 \lambda_2^2)/2 \in \mathbb{R}_+. \) By the help of a version of the comparison theorem (see, e.g., Volkmann [17]), we can derive the inequality \( v_1(t, u_1, \lambda_2) \leq \tilde{v}_1(u_1, \lambda_2) \) for all \( t \in \mathbb{R}_+ \) and \( (u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R} \), where \( \tilde{v}_1(u_1, \lambda_2) \), \( t \in \mathbb{R}_+, \) is the unique solution of the inhomogeneous linear differential equation
\[
\begin{cases}
\frac{\partial \tilde{u}}{\partial t}(u_1, \lambda_2) \leq -b\tilde{v}_1(u_1, \lambda_2) + C_4(u_1, \lambda_2)e^{-C_2t}, & t \in \mathbb{R}_+, \\
\tilde{v}_0(u_1, \lambda_2) = -\text{Re}(u_1).
\end{cases}
\]
This differential equation has the same form as (3.10), hence the solution takes the form
\[ \tilde{v}_1(u_1, \lambda_2) = -\text{Re}(u_1)e^{-bt} + C_4(u_1, \lambda_2)e^{-bt} \int_0^t e^{-C_2s + bs} ds, \quad t \in \mathbb{R}_+. \]
We have \( b - C_2 \geq b/2 > 0 \) and \( b > b/2 \geq C_2 \), thus
\[ 0 \leq v_1(u_1, \lambda_2) \leq \tilde{v}_1(u_1, \lambda_2) = -\text{Re}(u_1)e^{-bt} + C_4(u_1, \lambda_2) \frac{e^{-C_2t} - e^{-bt}}{b - C_2} \leq C_5(u_1, \lambda_2)e^{-C_2t}, \quad t \in \mathbb{R}_+, \]
with \( C_5(u_1, \lambda_2) := -\text{Re}(u_1) + 2C_4(u_1, \lambda_2)/b \in \mathbb{R}_+. \) Using (3.11), we conclude
\[ |K_1(u_1, \lambda_2)| = \sqrt{v_1(u_1, \lambda_2)^2 + w_1(u_1, \lambda_2)^2} \leq C_1(u_1, \lambda_2)e^{-C_2t}, \quad t \in \mathbb{R}_+, \]
with \( C_1(u_1, \lambda_2) := C_5(u_1, \lambda_2) + C_3(u_1, \lambda_2) \in \mathbb{R}_+ \), and we obtain (3.8).

**Step 3.** For each \( (u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R} \), the function \( h_{u_1, \lambda_2} : \mathbb{R}^2 \to \mathbb{C}, \) given by \( h_{u_1, \lambda_2}(y, z) := e^{u_1y + i\lambda_2 z}, \quad (y, z) \in \mathbb{R}^2, \) is bounded and continuous, since \( |e^{u_1y + i\lambda_2 z}| = e^{\text{Re}(u_1)y} \leq 1. \) Hence, by the continuity theorem, by (3.6) and by the portmanteau theorem, to prove (i), it is enough to check that for all \( (u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R} \) and \( (y_0, z_0) \in \mathbb{R}_+ \times \mathbb{R}, \)
\[
\lim_{t \to \infty} [y_0K_1(u_1, \lambda_2) + iz_0e^{-t\lambda_2} + g_1(u_1, \lambda_2)] = a \int_0^\infty K_u(u_1, \lambda_2) ds + \frac{A}{7}\lambda_2 - \frac{\sigma_1^2}{4\gamma}\lambda_2^2 =: g_\infty(u_1, \lambda_2), \tag{3.12}
\]
and that the function \( C_\times \mathbb{R} \ni (u_1, \lambda_2) \mapsto g_\infty(u_1, \lambda_2) \) is continuous. Indeed, using (3.6) and the independence of \((\eta_0, \zeta_0)\) and \((W_t, B_t, L_t)_{t \in \mathbb{R}_+}\), the law of total expectation yields that

\[
\mathbb{E}(e^{u_1 Y_t + i \lambda_2 Z_t}) = \int_0^\infty \int_-^\infty \mathbb{E}(e^{u_1 Y_t + i \lambda_2 Z_t} \mid (Y_0, Z_0) = (y_0, z_0)) \mathbb{P}_{(Y_0, Z_0)}(dy_0, dz_0)
\]

\[
= \int_0^\infty \int_-^\infty \exp\left\{y_0 K_t(u_1, \lambda_2) + i z_0 e^{-\gamma t} \lambda_2 + g_t(u_1, \lambda_2)\right\} \mathbb{P}_{(Y_0, Z_0)}(dy_0, dz_0)
\]

for all \((u_1, \lambda_2) \in C_\times \mathbb{R}\), where \(\mathbb{P}_{(Y_0, Z_0)}\) denotes the distribution of \((Y_0, Z_0)\) on \(\mathbb{R}_+ \times \mathbb{R}\), and hence (3.12) and the dominated convergence theorem implies that

\[
\lim_{t \to \infty} \mathbb{E}(e^{u_1 Y_t + i \lambda_2 Z_t}) = \int_0^\infty \int_-^\infty e^{g_\infty(u_1, \lambda_2)} \mathbb{P}_{(Y_0, Z_0)}(dy_0, dz_0) = e^{g_\infty(u_1, \lambda_2)}
\]

for all \((u_1, \lambda_2) \in C_\times \mathbb{R}\). Then, using the continuity of the function \( C_\times \mathbb{R} \ni (u_1, \lambda_2) \mapsto g_\infty(u_1, \lambda_2) \) (which will be checked later on), the continuity theorem implies \((Y_t, X_t) \xrightarrow{D} (Y_\infty, X_\infty)\) as \(t \to \infty\), and then, applying the portmanteau theorem for the functions \(h_{u_1, \lambda_2}, (u_1, \lambda_2) \in C_\times \mathbb{R}\), yields (i).

Next we turn to prove (3.12). By (3.8) and \(\gamma > 0\), we have

\[
\lim_{t \to \infty} |y_0 K_t(u_1, \lambda_2) + i z_0 e^{-\gamma t} \lambda_2| = 0.
\]

Recall that

\[
g_t(u_1, \lambda_2) = a \int_0^t K_s(u_1, \lambda_2) ds + i \lambda_2 \frac{1 - e^{-\gamma t}}{\gamma} - \frac{1}{2} \sigma^2 \lambda_2^2 \frac{1 - e^{-2\gamma t}}{2\gamma}.
\]

Since \(\gamma > 0\), we have \(\lim_{t \to \infty} \frac{1 - e^{-\gamma t}}{\gamma} = \frac{1}{\gamma}\) and \(\lim_{t \to \infty} \frac{1 - e^{-2\gamma t}}{2\gamma} = \frac{1}{2\gamma}\), and by the dominated convergence theorem, we get

\[
\lim_{t \to \infty} \int_0^t K_s(u_1, \lambda_2) ds = \int_0^\infty K_s(u_1, \lambda_2) ds.
\]

Indeed, by (3.8), \(|K_s(u_1, \lambda_2) 1_{[0,t]}(s)| \leq |K_s(u_1, \lambda_2)|\) for all \(t \in \mathbb{R}_+\) and \(s \in [0,t]\), and

\[
\int_0^\infty |K_s(u_1, \lambda_2)| ds \leq C_1(u_1, \lambda_2) \int_0^\infty e^{-C_2 s} ds \leq \frac{C_1(u_1, \lambda_2)}{C_2} < \infty.
\]

The continuity of the function \( C_\times \mathbb{R} \ni (u_1, \lambda_2) \mapsto g_\infty(u_1, \lambda_2) \) can be checked as follows. It will follow if we prove that for all \(s \in \mathbb{R}_+\), the function \(K_s\) is continuous. Namely, if \((u_1^{(n)}, \lambda_2^{(n)})_n \in \mathbb{N}\) is a sequence in \( C_\times \mathbb{R}\), such that \(\lim_{n \to \infty} (u_1^{(n)}, \lambda_2^{(n)}) = (u_1, \lambda_2)\), where \((u_1, \lambda_2) \in C_\times \mathbb{R}\), then \(\lim_{n \to \infty} K_s(u_1^{(n)}, \lambda_2^{(n)}) = K_s(u_1, \lambda_2)\) for all \(s \in \mathbb{R}_+\), and, by (3.8),

\[
|K_s(u_1^{(n)}, \lambda_2^{(n)})| \leq C_1(u_1^{(n)}, \lambda_2^{(n)}) e^{-C_2 s}, \quad n \in \mathbb{N}, \quad s \in \mathbb{R}_+.
\]

Since the sequence \((u_1^{(n)}, \lambda_2^{(n)})_n \in \mathbb{N}\) is bounded (since it is convergent), the dominated convergence theorem implies

\[
\lim_{n \to \infty} \int_0^\infty K_s(u_1^{(n)}, \lambda_2^{(n)}) ds = \int_0^\infty K_s(u_1, \lambda_2) ds,
\]
which shows the continuity of $g_{\infty}$. Finally, the continuity of the function $K_s$ follows from the continuity of the function $\psi_s^{(1)}$.

(ii): First we check that the one-dimensional distributions of $(Y_t, X_t)_{t \in \mathbb{R}^+}$ are translation invariant and have the same distribution as $(Y_\infty, X_\infty)$ has. Clearly, it is enough to prove that the one-dimensional distributions of $(Y_t, Z_t)_{t \in \mathbb{R}^+}$ are translation invariant and have the same distribution as $(Y_\infty, Z_\infty)$ has. Using (3.1), (3.6), the tower rule and the independence of $(Y_0, Z_0)$ and $(W, B, L)$, it is enough to check that for all $t \in \mathbb{R}^+$ and $(u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}$, 

$$
E \left( \exp \left\{ K_t(u_1, \lambda_2) Y_\infty + ie^{-\gamma t} \lambda_2 Z_\infty + g_t(u_1, \lambda_2) \right\} \right) = \exp \left\{ a \int_0^\infty K_s(u_1, \lambda_2) \, ds + i \frac{A}{\gamma} \lambda_2 - \frac{\sigma^2}{4 \gamma^2} \lambda_2^2 \right\}.
$$

By (3.1), (3.7) and using the fact that $K_t(u_1, \lambda_2) \in \mathbb{C}_-$ for all $t \in \mathbb{R}^+$ and $(u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}$ (see Step 1 of the proof of part (i)), we have 

$$
E \left( \exp \left\{ K_t(u_1, \lambda_2) Y_\infty + ie^{-\gamma t} \lambda_2 Z_\infty + g_t(u_1, \lambda_2) \right\} \right) = \exp \left\{ a \int_0^\infty K_s(K_t(u_1, \lambda_2), e^{-\gamma t} \lambda_2) \, ds + i \frac{A}{\gamma} e^{-\gamma t} \lambda_2 - \frac{\sigma^2}{4 \gamma^2} e^{-2\gamma t} \lambda_2^2 + g_t(u_1, \lambda_2) \right\} = \exp \left\{ a \left( \int_0^\infty K_s(K_t(u_1, \lambda_2), e^{-\gamma t} \lambda_2) \, ds + \int_0^t K_s(u_1, \lambda_2) \, ds \right) + i \frac{A}{\gamma} \lambda_2 - \frac{\sigma^2}{4 \gamma^2} \lambda_2^2 \right\}.
$$

Hence it remains to check that 

$$
\int_0^\infty K_s(u_1, \lambda_2) \, ds = \int_0^\infty K_s(K_t(u_1, \lambda_2), e^{-\gamma t} \lambda_2) \, ds + \int_0^t K_s(u_1, \lambda_2) \, ds, \quad t \in \mathbb{R}^+,
$$

for all $(u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}$, i.e., 

$$
\int_0^t K_s(u_1, \lambda_2) \, ds = \int_0^\infty K_s(K_t(u_1, \lambda_2), e^{-\gamma t} \lambda_2) \, ds, \quad t \in \mathbb{R}^+.
$$

For this it is enough to check that 

$$
K_s(K_t(u_1, \lambda_2), e^{-\gamma t} \lambda_2) = K_{s+t}(u_1, \lambda_2), \quad s, t \in \mathbb{R}^+,
$$

for all $(u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}$, or equivalently, 

$$
K_t(K_s(u_1, \lambda_2), e^{-\gamma s} \lambda_2) = K_{t+s}(u_1, \lambda_2), \quad s, t \in \mathbb{R}^+,
$$

(3.13)

for all $(u_1, \lambda_2) \in \mathbb{C}_- \times \mathbb{R}$. By (3.3), we have 

$$
\frac{\partial K_{s+t}}{\partial t}(u_1, \lambda_2) = -b K_{s+t}(u_1, \lambda_2) - i Be^{-\gamma(s+t)} \lambda_2 + i \frac{1}{2} \sigma^2 K_{s+t}(u_1, \lambda_2)^2 - \frac{1}{2} \Sigma^2 e^{-2\gamma(s+t)} \lambda_2^2
$$

for $t \in \mathbb{R}^+$ with initial condition $K_{s+0}(u_1, \lambda_2) = K_s(u_1, \lambda_2)$. Note also that, again by (3.3), 

$$
\frac{\partial K_t}{\partial t}(K_s(u_1, \lambda_2), e^{-\gamma s} \lambda_2) = -b K_t(K_s(u_1, \lambda_2), e^{-\gamma s} \lambda_2) - i Be^{-\gamma t}(e^{-\gamma s} \lambda_2) + i \frac{1}{2} \sigma^2 K_t(K_s(u_1, \lambda_2), e^{-\gamma s} \lambda_2)^2 + \frac{1}{2} \Sigma^2 e^{-2\gamma t}(e^{-\gamma s} \lambda_2)^2
$$
for \( t \in \mathbb{R}_+ \) with initial condition \( K_0(K_s(u_1, \lambda_2), e^{-s \lambda_2}) = K_s(u_1, \lambda_2) \). Hence, for all \( s \in \mathbb{R}_+ \), the left and right sides of (3.13), as functions of \( t \in \mathbb{R}_+ \), satisfy the differential equation (3.3) with \( e^{-s \lambda_2} \) instead of \( \lambda_2 \) and with the initial value \( K_s(u_1, \lambda_2) \). Since (3.3) has a unique solution for all non-negative initial values, we obtain (3.13).

Finally, the strict stationarity (translation invariance of the finite dimensional distributions) of \((Y_t, X_t)_{t \in \mathbb{R}_+}\) follows by the chain’s rule for conditional expectations using also that it is a time homogeneous Markov process.

4. Exponential Ergodicity

In the subcritical case, the following result states the exponential ergodicity for the process \((Y_t, X_t)_{t \in \mathbb{R}_+}\). As a consequence, according to the discussion after Proposition 2.5 in Bhattacharya [3], one also obtains a strong law of large numbers (4.3) for \((Y_t, X_t)_{t \in \mathbb{R}_+}\).

Theorem 4.1. Let us consider the 2-dimensional affine diffusion model (1.1) with \( a, b \in \mathbb{R}_+, \alpha, \beta \in \mathbb{R}_+, \gamma \in \mathbb{R}_+, \sigma_1 \in \mathbb{R}_+, \sigma_2, \sigma_3 \in \mathbb{R}_+ \) and \( g \in [-1,1] \) with a random initial value \((y_0, \zeta_0)\) independent of \((W_t, B_t, L_t)_{t \in \mathbb{R}_+}\) satisfying \(\mathbb{P}(y_0 \in \mathbb{R}_+) = 1\). Suppose that \((1-g^2)\sigma_2^2 + \sigma_3^2 > 0\). Then the process \((Y_t, X_t)_{t \in \mathbb{R}_+}\) is exponentially ergodic, namely, there exist \(\delta \in \mathbb{R}_+, B \in \mathbb{R}_+ \) and \(\kappa \in \mathbb{R}_+\), such that

\[
\sup_{|g| \leq V+1} \left| \mathbb{E}(g(Y_t, X_t) | (Y_0, X_0) = (y_0, x_0)) - \mathbb{E}(g(Y_\infty, X_\infty)) \right| \leq B(V(y_0, x_0)+1)e^{-\delta t} \tag{4.1}
\]

for all \( t \in \mathbb{R}_+ \) and \((y_0, x_0) \in \mathbb{R}_+ \times \mathbb{R}\), where the supremum is running for Borel measurable functions \(g : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}\),

\[
V(y, x) := y^2 + \kappa x^2, \quad (y, x) \in \mathbb{R}_+ \times \mathbb{R}, \tag{4.2}
\]

and the distribution of \((Y_\infty, X_\infty)\) is given by (3.1) and (3.2). Moreover, for all Borel measurable functions \(f : \mathbb{R}^2 \to \mathbb{R}\) with \(\mathbb{E}(|f(Y_\infty, X_\infty)|) < \infty\), we have

\[
\mathbb{P}\left( \lim_{T \to \infty} \frac{1}{T} \int_0^T f(Y_s, X_s) \, ds = \mathbb{E}(f(Y_\infty, X_\infty)) \right) = 1. \tag{4.3}
\]

Proof. First we check that it is enough to prove (4.1) for the special affine diffusion process \((Y_t, Z_t)_{t \in \mathbb{R}_+}\) given in Proposition 2.5. Hence we suppose that (4.1) holds for \((Y_t, Z_t)_{t \in \mathbb{R}_+}\) with \(\delta \in \mathbb{R}_+, B \in \mathbb{R}_+ \) and \(\kappa \in \mathbb{R}_+\), and we check that then (4.1) holds for \((Y_t, X_t)_{t \in \mathbb{R}_+}\) with \(\delta \in \mathbb{R}_+\), with some appropriate \(B \in \mathbb{R}_+\), and with \(\kappa \in \mathbb{R}_+\) as well. Let \(g : \mathbb{R}^2 \to \mathbb{R}\) be a Borel measurable function with \(|g(y, x)| \leq V(y, x) + 1 = y^2 + \kappa x^2 + 1\), \((y, x) \in \mathbb{R}_+ \times \mathbb{R}\). Then \(g(Y_t, X_t) = g(Y_t, Z_t + cY_t) = h(Y_t, Z_t), \quad t \in \mathbb{R}_+\), where \(h : \mathbb{R}^2 \to \mathbb{R}\) is given by \(h(y, z) := g(y, z + cy), \quad (y, z) \in \mathbb{R}^2\). Clearly, \(h\) is also a Borel measurable function. Moreover, (i) of Theorem 3.1 implies \((Y_t, X_t) \xrightarrow{D} (Y_\infty, X_\infty)\) as \(t \to \infty\). Again by (i) of Theorem 3.1, we obtain \((Y_t, X_t) = (Y_t, Z_t + cY_t) \xrightarrow{D} (Y_\infty, Z_\infty + cY_\infty)\) as \(t \to \infty\), consequently, \((Y_\infty, X_\infty) \xrightarrow{D} (Y_\infty, Z_\infty + cY_\infty)\), and hence, \(\mathbb{E}(g(Y_\infty, X_\infty)) = \mathbb{E}(h(Y_\infty, Z_\infty)) = \mathbb{E}(g(Y_\infty, Z_\infty)) + \mathbb{E}(c g(Y_\infty, Z_\infty)) \xrightarrow{D} \mathbb{E}(g(Y_\infty, Z_\infty)) = \mathbb{E}(h(Y_\infty, Z_\infty)) = \mathbb{E}(g(Y_\infty, X_\infty))\).
\( \mathbb{E}(g(Y_\infty, Z_\infty + cY_\infty)) = \mathbb{E}(h(Y_\infty, Z_\infty)) \). We conclude

\[
\left| \mathbb{E} \left( g(Y_t, X_t) \big| (Y_0, X_0) = (y_0, x_0) \right) - \mathbb{E}(g(Y_\infty, X_\infty)) \right|
\]

\[
= \left| \mathbb{E} \left( g(Y_t, Z_t + cY_t) \big| (Y_0, Z_0 + cY_0) = (y_0, x_0) \right) - \mathbb{E}(g(Y_\infty, X_\infty)) \right|
\]

\[
= \left| \mathbb{E} \left( h(Y_t, Z_t) \big| (Y_0, Z_0) = (y_0, z_0) \right) - \mathbb{E}(h(Y_\infty, Z_\infty)) \right|
\]

with \( z_0 := x_0 - cy_0 \). By the assumption, (4.1) holds for \((Y_t, Z_t)_{t \in \mathbb{R}_+}\) and Borel measurable functions \( \bar{g} : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R} \) with \( |\bar{g}(y, z)| \leq V(y, z) + 1 = y^2 + \kappa z^2 + 1 \), \((y, z) \in \mathbb{R}_+ \times \mathbb{R}\). We have

\[
|g(y, z)| = |g(y, z + cy)| \leq y^2 + \kappa(z + cy)^2 + 1 \leq y^2 + 2\kappa(z^2 + c^2y^2) + 1
\]

\[
= (1 + 2\kappa c^2)y^2 + 2\kappa z^2 + 1 \leq C(y^2 + \kappa z^2 + 1)
\]

with \( C := \max\{1 + 2\kappa c^2, 2\} \), hence we can apply (4.1) for \((Y_t, Z_t)_{t \in \mathbb{R}_+}\) and the Borel measurable function \( \frac{1}{C}h(y, z) \), \((y, z) \in \mathbb{R}_+ \times \mathbb{R}\). We obtain

\[
\left| \mathbb{E} \left( \frac{1}{C}h(Y_t, Z_t) \big| (Y_0, Z_0) = (y_0, z_0) \right) - \mathbb{E}\left( \frac{1}{C}h(Y_\infty, Z_\infty) \right) \right| \leq B(y_0^2 + \kappa z_0^2 + 1)e^{-\delta t},
\]

and hence

\[
\left| \mathbb{E} \left( g(Y_t, X_t) \big| (Y_0, X_0) = (y_0, x_0) \right) - \mathbb{E}(g(Y_\infty, X_\infty)) \right|
\]

\[
\leq BC(y_0^2 + \kappa(x_0 - cy_0)^2 + 1)e^{-\delta t} \leq BC(y_0^2 + 2\kappa(x_0^2 + c^2y_0^2) + 1)e^{-\delta t}
\]

\[
= BC((1 + 2\kappa c^2)y_0^2 + 2\kappa x_0^2 + 1)e^{-\delta t} \leq BC^2(y_0^2 + \kappa x_0^2 + 1)e^{-\delta t},
\]

thus (4.1) holds for \((Y_t, X_t)_{t \in \mathbb{R}_+}\) with \( \delta \in \mathbb{R}_{++} \), \( \tilde{B} := BC^2 \in \mathbb{R}_{++} \) and with \( \kappa \in \mathbb{R}_{++} \).

Next we prove (4.3) for the special affine diffusion process \((Y_t, Z_t)_{t \in \mathbb{R}_+}\) given in Proposition 2.5. We use the notations of Meyn and Tweedie [12], [13]. Using Theorem 6.1 (so called Foster-Lyapunov criteria) in Meyn and Tweedie [13], it is enough to check that

(a) \((Y_t, Z_t)_{t \geq 0}\) is a right process (defined on page 38 in Sharpe [15]);
(b) all compact sets are petite for some skeleton chain (skeleton chains and petite sets are defined on pages 491, 500 in Meyn and Tweedie [12], and page 550 in Meyn and Tweedie [11], respectively);
(c) there exist \( c \in \mathbb{R}_{++} \) and \( d \in \mathbb{R} \) such that the inequality

\[
(A_nV)(y, z) \leq -cV(y, z) + d,
\]

holds for all \( n \in \mathbb{N} \), where \( O_n := \{(y, z) \in \mathbb{R}_+ \times \mathbb{R} : ||(y, z)|| < n\} \) for each \( n \in \mathbb{N} \), and \( A_n \) denotes the extended generator of the process \((Y_t^{(n)}, Z_t^{(n)})_{t \in \mathbb{R}_+}\) given by

\[
(Y_t^{(n)}, Z_t^{(n)}) := \begin{cases} (Y_t, Z_t), & \text{for } t < T_n, \\ (0, n), & \text{for } t \geq T_n, \end{cases}
\]

where the stopping time \( T_n \) is defined by \( T_n := \inf\{t \in \mathbb{R}_+ : (Y_t, Z_t) \in (\mathbb{R}_+ \times \mathbb{R}) \setminus O_n\} \). (Here we note that instead of \((0, n)\) we could have chosen any fixed state in \((\mathbb{R}_+ \times \mathbb{R}) \setminus O_n\), and we could also have defined
\[(Y_t^{(n)},Z_t^{(n)})_{t \in \mathbb{R}_+} \text{ as the stopped process } (Y_{t \wedge T_n},Z_{t \wedge T_n})_{t \in \mathbb{R}_+}, \text{ see Meyn and Tweedie [13, page 521].}\]

To prove (a), it is enough to show that the process \((Y_t,Z_t)_{t \in \mathbb{R}_+}\) is a (weak) Feller process (see Meyn and Tweedie [12, Section 3.1]), strong Markov process with continuous sample paths, see, e.g., Meyn and Tweedie [12, page 498]. According to Proposition 8.2 (or Theorem 2.7) in Duffie et al. [7], the process \((Y_t,Z_t)_{t \in \mathbb{R}_+}\) is a Feller Markov process. Since \((Y_t,Z_t)_{t \in \mathbb{R}_+}\) has continuous sample paths almost surely (especially, it is càdlàg), it is automatically a strong Markov process, see, e.g., Theorem 1 on page 56 in Chung [4].

To prove (b), in view of Proposition 6.2.8 in Meyn and Tweedey [14], it is sufficient to show that the skeleton chain \((Y_n,Z_n)_{n \in \mathbb{Z}_+}\) is irreducible with respect to the Lebesgue measure on \(\mathbb{R}_+ \times \mathbb{R}\) (see, e.g., Meyn and Tweedey [13, page 520]), and admits the Feller property. The skeleton chain \((Y_n,Z_n)_{n \in \mathbb{Z}_+}\) admits the Feller property, since the process \((Y_t,Z_t)_{t \in \mathbb{R}_+}\) is a Feller process. In order to check irreducibility of the skeleton chain \((Y_n,Z_n)_{n \in \mathbb{Z}_+}\) with respect to the Lebesgue measure on \(\mathbb{R}_+ \times \mathbb{R}\), it is enough to prove that the conditional distribution of \((Y_1,Z_1)\) given \((Y_0,Z_0)\) is absolutely continuous (with respect to the Lebesgue measure on \(\mathbb{R}_+ \times \mathbb{R}\)) with a conditional density function \(f_{(Y_1,Z_1) | (Y_0,Z_0)} : \mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}_+\) such that \(f_{(Y_1,Z_1) | (Y_0,Z_0)}(y,z \mid y_0,z_0) > 0\) for all \((y,z) \in \mathbb{R}_+ \times \mathbb{R}\). Indeed, the Lebesgue measure on \(\mathbb{R}_+ \times \mathbb{R}\) is \(\sigma\)-finite, and if \(B\) is a Borel set in \(\mathbb{R}_+ \times \mathbb{R}\) with positive Lebesgue measure, then

\[
\mathbb{E} \left( \sum_{n=0}^{\infty} \mathbb{1}_B(Y_n,Z_n) \bigg| (Y_0,Z_0) = (y_0,z_0) \right) \geq \mathbb{P}( (Y_1,Z_1) \in B \big| (Y_0,Z_0) = (y_0,z_0) )
\]

\[
= \int_B f_{(Y_1,Z_1) | (Y_0,Z_0)}(y,z \mid y_0,z_0) \, dy \, dz > 0
\]

for all \((y_0,z_0) \in \mathbb{R}_+ \times \mathbb{R}\). The existence of \(f_{(Y_1,Z_1) | (Y_0,Z_0)}\) with the required property can be checked as follows. By Theorem 2.2, we have

\[
Y_1 = e^{-b} \left( y_0 + a \int_0^1 e^{bu} \, du + \sigma_1 \int_0^1 e^{bu} \sqrt{Y_u} \, dW_u \right),
\]

\[
Z_1 = e^{-\gamma} \left( z_0 + \int_0^1 e^{\gamma u} (A - BY_u) \, du + \Sigma_2 \int_0^1 e^{\gamma u} \sqrt{Y_u} \, dB_u + \sigma_3 \int_0^1 e^{\gamma u} \, dL_u \right),
\]

provided that \((Y_0,Z_0) = (y_0,z_0)\), \((y_0,z_0) \in \mathbb{R}_+ \times \mathbb{R}\). Recall that a two-dimensional random vector \(\zeta\) is absolutely continuous if and only if \(V\zeta + v\) is absolutely continuous for all invertible matrices \(V \in \mathbb{R}^{2 \times 2}\) and for all vectors \(v \in \mathbb{R}^2\), and if the density function of \(\zeta\) is positive on a set \(S \subset \mathbb{R}^2\), then the density function of \(V\zeta + v\) is positive on the set \(VS + v\). Hence it is enough to check that the random vector

\[
\left( \sigma_1 \int_0^1 e^{bu} \sqrt{Y_u} \, dW_u, I \right)
\]

with

\[
I := -B \int_0^1 e^{\gamma u} Y_u \, du + \Sigma_2 \int_0^1 e^{\gamma u} \sqrt{Y_u} \, dB_u + \sigma_3 \int_0^1 e^{\gamma u} \, dL_u
\]

(4.4)
is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^2 \) having a
density function being strictly positive on the set
\[
\left\{ y \in \mathbb{R} : y > -y_0 - a \int_0^1 e^{bu} \, du \right\} \times \mathbb{R}.
\]
For all \( y \leq -y_0 - a \int_0^1 e^{bu} \, du \) and \( z \in \mathbb{R} \), we have
\[
\mathbb{P} \left( \sigma_1 \int_0^1 e^{bu} \sqrt{Y_u} \, dW_u < y, I < z \right) = \mathbb{P} \left( e^b Y_1 - y_0 - a \int_0^1 e^{bu} \, du < y, I < z \right)
\leq \mathbb{P}(Y_1 < 0) = 0,
\]
since \( \mathbb{P}(Y_1 \geq 0) = 1 \). Note that the conditional distribution of \( I \) given \( (Y_t)_{t \in [0,1]} \) is a normal distribution with mean \(-B \int_0^1 e^{\gamma u} Y_u \, du\) and with variance
\[
\gamma^2 := \Sigma_2^2 \int_0^1 e^{2\gamma u} Y_u \, du + \sigma_3^2 \int_0^1 e^{2u} \, du
\]
due to the fact that \((Y_t)_{t \in [0,1]}\) and \((B_t, L_t)_{t \in \mathbb{R}^+}\) are independent. Indeed, \((Y_t)_{t \in \mathbb{R}^+}\) is adapted to the augmented filtration corresponding to \( \eta_0 \) and \((W_t)_{t \in \mathbb{R}^+}\) (see, e.g., Karatzas and Shreve [10, page 285]), and using the independence of the standard Wiener processes \( W \) and \( B \), and Problem 2.7.3 in Karatzas and Shreve [10], one can argue that this augmented filtration is independent of the filtration generated by \( B \). Here we call the attention that the condition \((1 - \theta^2)\sigma_2^2 + \sigma_3^2 > 0\) implies \( \mathbb{P}(\gamma^2 \in \mathbb{R}^+) = 1 \). Indeed, \( \Sigma_2^2 = (1 - \theta^2)\sigma_2^2 \), and the assumption \( a > 0 \) yields \( \mathbb{P}(\int_0^1 e^{2\gamma u} Y_u \, du \in \mathbb{R}^+) = 1 \). Hence, using again the independence of the standard Wiener processes \( W \), \( B \) and \( L \), we get for all \( y > -y_0 - a \int_0^1 e^{bu} \, du \) and \( z \in \mathbb{R} \), by the law of total expectation,
\[
\mathbb{P} \left( \sigma_1 \int_0^1 e^{bu} \sqrt{Y_u} \, dW_u < y, I < z \right)
= \mathbb{P} \left( e^b Y_1 - y_0 - a \int_0^1 e^{bu} \, du < y, I < z \right)
= \mathbb{E} \left( \mathbb{P} \left( Y_1 < e^{-b} \left( y + y_0 + a \int_0^1 e^{bu} \, du \right) \right. \left. , I < z \right| (Y_t)_{t \in [0,1]} \right)
= \mathbb{E} \left( \mathbb{E} \left( \mathbb{1}_{\{Y_1 < e^{-b} \left( y + y_0 + a \int_0^1 e^{bu} \, du \right) \}} \, \mathbb{1}_{\{I < z\}} \right| (Y_t)_{t \in [0,1]} \right)
= \mathbb{E} \left( \mathbb{E} \left( \mathbb{1}_{\{Y_1 < e^{-b} \left( y + y_0 + a \int_0^1 e^{bu} \, du \right) \}} \right) \mathbb{P}(I < z \mid (Y_t)_{t \in [0,1]}) \right)
= \mathbb{E} \left( \left( \int_{\mathbb{R}^+} \frac{1}{\sqrt{2\pi T^2}} \exp \left\{ \frac{-(w + B \int_0^1 e^{\gamma u} Y_u \, du)^2}{2T^2} \right\} \, dw \right) \right)
= \int_{\mathbb{R}^+} \frac{1}{\sqrt{2\pi T^2}} \exp \left\{ \frac{-(w + B \int_0^1 e^{\gamma u} Y_u \, du)^2}{2T^2} \right\} \, dw
\times f_{Y_1 \mid Y_0}(v \mid y_0) \, dv \, dw
is known that for each \( y \in C_Tweedie \). We also note that, by Duffie et al. [7, Theorem 2.7], for functions \( y, z \) have \( \eta \) such that \( f_Y |_{y_0}(u | y_0) > 0 \) for Lebesgue a.e. \( u \in \mathbb{R}_{++} \), see, e.g., Cox et al. [5, Equation (18)], Jeanblanc et al. [9, Proposition 6.3.2.1] or Ben Alaya and Kebaier [2, the proof of Proposition 2] in case of \( y_0 \in \mathbb{R}_{++} \), and Ikeda and Watanabe [8, page 222] in case of \( y_0 = 0 \).

In what follows we will make use of the following simple observation: if \( \xi \) and \( \eta \) are random variables such that \( P(\xi \in \mathbb{R}_{++}) = 1 \), \( E(\xi) < \infty \), \( P(\eta \in \mathbb{R}_{++}) = 1 \), and \( \eta \) is absolutely continuous with a density function \( f_\eta \) having the property \( f_\eta(x) > 0 \) Lebesgue a.e. \( x \in \mathbb{R}_{++} \), then \( E(\xi | \eta = y) > 0 \) Lebesgue a.e. \( y \in \mathbb{R}_{++} \). For a proof, see, e.g., the proof of Theorem 4.1 in the extended arXiv version of Barczy et al. [1].

Now we turn back to the proof that the random vector (4.4) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^2 \) with a density function being strictly positive on the set \( \{ y \in \mathbb{R} : y > -y_0 - a \int_0^1 e^{bu} du \} \times \mathbb{R} \). Using that \( f_Y |_{y_0}(e^{-b}(v + y_0 + a \int_0^1 e^{bu} du) | y_0) > 0 \) for all \( v > -y_0 - a \int_0^1 e^{bu} du \), there exists a measurable function \( g : \mathbb{R}^2 \to \mathbb{R}_+ \) such that \( g(v, w) > 0 \) for \( v > -y_0 - a \int_0^1 e^{bu} du \), \( w \in \mathbb{R} \), and

\[
P \left( \sigma_1 \int_0^1 e^{bu} \sqrt{Y_u} dW_u < y, I < z \right) = \begin{cases} \int_y^\infty g(v, w) dv dw & \text{if } y > -y_0 - a \int_0^1 e^{bu} du, \, z \in \mathbb{R}, \\ 0 & \text{if } y \leq -y_0 - a \int_0^1 e^{bu} du, \, z \in \mathbb{R}. \end{cases}
\]

Consequently, the random vector (4.4) is absolutely continuous with density function \( g \) having the desired property.

To prove (c), first we note that, since the sample paths of \( (Y_t, Z_t)_{t \in \mathbb{R}_+} \) are almost surely continuous, for each \( n \in \mathbb{N} \), the extended generator has the form

\[
(A_n f)(y, z) = (a - by) f'_1(y, z) + (A - By - \gamma z) f'_2(y, z) + \frac{\gamma z}{2} f''_{1,1}(y, z) + \frac{\sigma^2}{2} f''_{2,2}(y, z)
\]

for all \( (y, z) \in O_n \) and \( f \in C^2(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R}) \), see, e.g., page 538 in Meyn and Tweedie [13]. We also note that, by Duffie et al. [7, Theorem 2.7], for functions \( f \in C^2(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R}) \), \( A_n f = A f \) on \( O_n \), where \( A \) denotes the (non-extended) generator of the process \( (Y_t, Z_t)_{t \in \mathbb{R}_+} \). For the function \( V \) defined in (4.2), we have \( V \in C^2(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R}) \) and

\[
V'_1(y, z) = 2y, \quad V'_2(y, z) = 2\kappa z, \quad V''_{1,1}(y, z) = 2, \quad V''_{2,2}(y, z) = 2\kappa
\]
Thus for \((y, z) \in \mathbb{R}_+ \times \mathbb{R}\), and hence for all \(n \in \mathbb{N}\) and \(c \in \mathbb{R}_{++}\),

\[
(A_n V)(y, z) + c V(y, z) = 2(a - by)y + 2\kappa(A - By - \gamma z)z + y(\sigma_1^2 + \kappa \Sigma_2^2) + \kappa \sigma_3^2 + cy^2 + cz^2
\]

\[
= -(2b - c)y^2 - 2\kappa Byz - \kappa(2\gamma - c)z^2 + c_1 y + 2\kappa A z + \kappa \sigma_3^2
\]

for all \((y, z) \in O_n\) with \(c_1 := 2a + \sigma_1^2 + \kappa \Sigma_2^2\).

Thus

\[
(A_n V)(y, z) + c V(y, z) = -c_2 \left( y + \frac{\kappa B z}{c_2} \right)^2 - c_3 z^2 + c_1 \left( y + \frac{\kappa B z}{c_2} \right) + c_4 z + \kappa \sigma_3^2
\]

for all \((y, z) \in O_n\) with \(c_2 := 2b - c, \quad c_3 := \kappa \left( 2\gamma - c - \frac{\kappa B^2}{c_2} \right), \quad c_4 := 2\kappa A - c_1 \frac{\kappa B}{c_2}\),

whenever \(c_2 \neq 0\). Consequently,

\[
(A_n V)(y, z) + c V(y, z) = -c_2 \left( y + \frac{\kappa B z}{c_2} - \frac{c_1}{2c_2} \right)^2 - c_3 \left( z - \frac{c_4}{2c_3} \right)^2 + d
\]

for all \((y, z) \in O_n\) with \(d := \frac{c_1^2}{4c_2} + \frac{c_3^2}{4c_4} + \kappa \sigma_3^2\),

whenever \(c_2 \neq 0\) and \(c_3 \neq 0\). Let us choose

\[
c \in (0, 2 \min\{b, \gamma\}), \quad \kappa \in \left(0, \frac{(2\gamma - c)(2b - c)}{B^2}\right).
\]

Then \(c_2 > 0\) and \(c_3 > 0\), hence

\[
(A_n V)(y, z) \leq -c V(y, z) + d, \quad (y, z) \in O_n, \quad n \in \mathbb{N},
\]

and the proof is complete. \(\square\)

5. Moments of the Stationary Distribution

**Theorem 5.1.** Let us consider the 2-dimensional affine diffusion model (1.1) with \(a \in \mathbb{R}_+, \quad b \in \mathbb{R}_{++}, \quad \alpha, \beta \in \mathbb{R}, \quad \gamma \in \mathbb{R}_{++}, \quad \sigma_1, \sigma_2, \sigma_3 \in \mathbb{R}_+, \quad \rho \in [-1, 1]\), and the random vector \((Y_\infty, X_\infty)\) given by Theorem 3.1. Then all the (mixed) moments of \((Y_\infty, X_\infty)\) of any order are finite, i.e., we have \(\mathbb{E}(Y_\infty^n | X_\infty^p) < \infty\) for all \(n, p \in \mathbb{Z}_+,\) and the recursion

\[
\mathbb{E}(Y_\infty^n X_\infty^p) = \frac{1}{nb + p\gamma} \left[ \left( na + \frac{1}{2} n(n - 1) \sigma_1^2 \right) \mathbb{E}(Y_\infty^{n-1} X_\infty^p) - p\beta \mathbb{E}(Y_\infty^{n+1} X_\infty^{p-1}) \right.
\]

\[
+ p(\alpha + n \sigma_1 \sigma_2) \mathbb{E}(Y_\infty^n X_\infty^{p-1})
\]

\[
+ \frac{1}{2} p(p - 1) \sigma_2^2 \mathbb{E}(Y_\infty^{n+1} X_\infty^{p-2})
\]

\[
+ \frac{1}{2} p(p - 1) \sigma_3^2 \mathbb{E}(Y_\infty^n X_\infty^{p-2}) \right],
\]
holds for all \( n, p \in \mathbb{Z}_+ \) with \( n + p \geq 1 \), where \( \mathbb{E}(Y^k \mathcal{X}\ell) := 0 \) for \( k, \ell \in \mathbb{Z} \) with \( k < 0 \) or \( \ell < 0 \). Especially,

\[
\begin{align*}
\mathbb{E}(Y_\infty) & = \frac{a}{b}, \quad \mathbb{E}(Y^2_\infty) = \frac{a(2a + \sigma_1^2)}{2b}, \quad \mathbb{E}(Y^3_\infty) = \frac{a(a + \sigma_1^2)(2a + \sigma_1^2)}{2b^3}, \\
\mathbb{E}(X_\infty) & = \frac{ba - a\beta}{b^\gamma}, \quad \mathbb{E}(Y_\infty X_\infty) = a \mathbb{E}(X_\infty) - \beta \mathbb{E}(Y^2_\infty) + (\alpha + \varphi \sigma_1 \sigma_2) \mathbb{E}(Y_\infty), \\
\mathbb{E}(X^2_\infty) & = \frac{-2\beta \mathbb{E}(Y_\infty X_\infty) + 2\alpha \mathbb{E}(X_\infty) + \sigma_2^2 \mathbb{E}(Y_\infty) + \sigma_3^2}{2\gamma}, \\
\mathbb{E}(Y^2_\infty X_\infty) & = \frac{(2a + \sigma_1^2) \mathbb{E}(Y_\infty X_\infty) - \beta \mathbb{E}(Y^3_\infty) + (\alpha + \varphi \sigma_1 \sigma_2) \mathbb{E}(Y^2_\infty)}{2b + \gamma}, \\
\mathbb{E}(Y_\infty X^2_\infty) & = \frac{a \mathbb{E}(X^2_\infty) - 2\beta \mathbb{E}(Y^2_\infty X_\infty) + 2(\alpha + \varphi \sigma_1 \sigma_2) \mathbb{E}(Y_\infty X_\infty)}{b + 2\gamma} + \frac{\sigma_2^2 \mathbb{E}(Y^2_\infty) + \sigma_3^2 \mathbb{E}(Y_\infty)}{b + 2\gamma}.
\end{align*}
\]

If \( \sigma_1 > 0 \) then the Laplace transform of \( Y_\infty \) takes the form

\[
\mathbb{E}(e^{-\lambda Y_\infty}) = \left(1 + \frac{\sigma_2^2}{2b} \lambda \right)^{-2a/\sigma_1^2}, \quad \lambda \in \mathbb{R}_+, \tag{5.1}
\]

i.e., \( Y_\infty \) has gamma distribution with parameters \( 2a/\sigma_1^2 \) and \( 2b/\sigma_1^2 \), hence

\[
\mathbb{E}(Y^\kappa_\infty) = \frac{\Gamma \left( \frac{2a}{\sigma_1^2} + \kappa \right)}{(2b)^{\kappa} \Gamma \left( \frac{2a}{\sigma_1^2} \right)}, \quad \kappa \in \left( -\frac{2a}{\sigma_1^2}, \infty \right).
\]

If \( \sigma_1 > 0 \) and \( (1 - \varphi^2)\sigma_2^2 + \sigma_3^2 > 0 \) then the distribution of \( (Y_\infty, X_\infty) \) is absolutely continuous.

**Proof.** We may and do suppose that all the mixed moments of \( (Y_0, X_0) \) are finite and \( \mathbb{P}(Y_0 > 0) = 1 \), since, due to Theorem 3.1, the distribution of \( (Y_\infty, X_\infty) \) does not depend on the initial value of the model. First we show that

\[
\int_0^t \mathbb{E}(Y^n_u X^{2p}_u) \, du < \infty \quad \text{for all } t \in \mathbb{R}_+ \text{ and } n, p \in \mathbb{Z}_+. \tag{5.2}
\]

One can easily check that it is enough to prove (5.2) for the special affine diffusion process \( (Y_t, Z_t)_{t \in \mathbb{R}_+} \) given in Proposition 2.5. Indeed, then

\[
\int_0^t \mathbb{E}(Y^n_u X^{2p}_u) \, du = \int_0^t \mathbb{E}(Y^n_u (Z_u + cY_u)^{2p}) \, du \\
= \sum_{k=0}^{2p} \binom{2p}{k} c^{2p-k} \int_0^t \mathbb{E}(Y^{n+2p-k} Z^{2k}_u) \, du < \infty.
\]
Applying (2.5) and the power means inequality \((a + b + c + d)^2 \leq 4^{2p-1}(a^{2p} + b^{2p} + c^{2p} + d^{2p})\), \(a, b, c, d \in \mathbb{R}\), we obtain

\[
\int_0^t \mathbb{E}(Y_u^n Z_0^{2p}) \, du \leq 4^{2p-1} \int_0^t \mathbb{E} \left[ Y_u^n \left( e^{-2p\gamma u} Z_0^{2p} + \left( \int_0^u e^{-\gamma(u-v)} (A - By_v) \, dv \right)^2 \right) \right. \\
\left. + \left( \int_0^u e^{-\gamma(u-v)} \sqrt{Y_v} \, dB_v \right)^2 \right] \, du \\
\int_0^u e^{-\gamma(u-v)} \sqrt{Y_v} \, dB_v \right)^{2p} \right] \, du 
\]

for all \(t \in \mathbb{R}_+\) and \(n, p \in \mathbb{Z}_+\). Since for all \(u \in [0, t]\), the distribution of \(\int_0^u e^{-\gamma(u-v)} \, dB_v\) is a normal distribution with mean 0 and with variance \(\int_0^u e^{-2\gamma(u-v)} \, dv\), and the conditional distribution of \(\int_0^u e^{-\gamma(u-v)} \sqrt{Y_v} \, dB_v\) with respect to the \(\sigma\)-algebra generated by \((Y_s)_{s \in [0, t]}\) is a normal distribution with mean 0 and with variance \(\int_0^u e^{-2\gamma(u-v)} \, dv\), to prove (5.2), it is enough to show that, for all \(t \in \mathbb{R}_+\) and \(n, p \in \mathbb{Z}_+\),

\[
\int_0^t \mathbb{E}(e^{-2p\gamma u} Y_u^n Z_0^{2p}) \, du < \infty, \quad \int_0^t \mathbb{E}(Y_u^n) \, du < \infty, \\
\int_0^t \mathbb{E} \left[ Y_u^n \left( \int_0^u e^{-\gamma(u-v)} Y_v \, dv \right)^{2p} \right] \, du < \infty, \\
\int_0^t \mathbb{E} \left[ Y_u^n \left( \int_0^u e^{-2\gamma(u-v)} Y_v \, dv \right)^p \right] \, du < \infty, 
\]

which can be checked by standard arguments, see, e.g., in the arXiv version of the proof of Theorem 4.2 in Barczy et al. [1].

For all \(n, p \in \mathbb{Z}_+\), using the independence of \(W, B\) and \(L\), by Itô’s formula, we have

\[
d(Y_t^n X_t^p) = nY_t^{n-1} X_t^p [(a - bY_t) \, dt + \sigma_1 \sqrt{Y_t} \, dW_t] \\
\quad + pY_t^n X_t^{p-1} [(\alpha - \beta Y_t - \gamma X_t) \, dt + \sigma_2 \sqrt{Y_t} (\rho \, dW_t + \sqrt{1 - \rho^2} \, dB_t) + \sigma_3 \, dL_t] \\
\quad + \frac{1}{2} \sigma_1^2 \sigma_2 (n - 1)Y_t^{n-2} X_t^p \, dt + \frac{1}{2} \rho (p - 1) Y_t^n X_t^{p-2} (\sigma_2 Y_t + \sigma_3^2) \, dt \\
\quad + npY_t^{n-1} X_t^{p-1} \rho \sigma_1 \sigma_2 Y_t \, dt 
\]

for \(t \in \mathbb{R}_+\). Writing the SDE above in an integrated form and taking the expectation of both sides, we have

\[
\mathbb{E}(Y_t^n X_t^p) - \mathbb{E}(Y_0^n X_0^p) = \int_0^t \left[ na \mathbb{E}(Y_u^{n-1} X_u^p) - nb \mathbb{E}(Y_u^n X_u^p) + p\alpha \mathbb{E}(Y_u^n X_u^{p-1}) \\
- p\beta \mathbb{E}(Y_u^{n+1} X_u^{p-1}) - p\gamma \mathbb{E}(Y_u^n X_u^p) \\
+ \frac{1}{2} \sigma_1^2 \sigma_2 (n - 1) \mathbb{E}(Y_u^{n-1} X_u^p) + \frac{1}{2} \rho (p - 1) \mathbb{E}(Y_u^{n+1} X_u^{p-2}) \\
+ \frac{1}{2} \sigma_2^2 \rho (p - 1) \mathbb{E}(Y_u^n X_u^{p-2}) + \rho \sigma_1 \sigma_2 np \mathbb{E}(Y_u^n X_u^{p-3}) \right] \, du 
\]
for all $t \in \mathbb{R}_+$, where we used that

$$
\left( \int_0^t Y_u^{-1/2} X_u^p \, dW_u \right)_{t \in \mathbb{R}_+}, \quad \left( \int_0^t Y_u^{-1/2} X_u^{p-1} \, dW_u \right)_{t \in \mathbb{R}_+},
$$

$$
\left( \int_0^t Y_u^{1/2} X_u^p \, dB_u \right)_{t \in \mathbb{R}_+}, \quad \left( \int_0^t Y_u^{1/2} X_u^{p-1} \, dB_u \right)_{t \in \mathbb{R}_+},
$$

are continuous square integrable martingales due to (5.2), see, e.g., Ikeda and Watanabe [8, page 55]. Introduce the functions $f_{n,p}(t) := \mathbb{E}(Y_t^n X_t^p)$, $t \in \mathbb{R}_+$, for $n,p \in \mathbb{Z}_+$. Then we have

$$
f'_{n,p}(t) = -(nb + p\gamma)f_{n,p}(t) - p\beta f_{n+1,p-1}(t) + \left( na + \frac{1}{2} \sigma_1^2 n(n-1) \right) f_{n-1,p}(t)
$$

$$
+ p(\alpha + \varphi_1 \sigma_2 n)f_{n,p-1}(t) + \frac{1}{2} \sigma_2^2 p(p-1) f_{n+1,p-2}(t) + \frac{1}{2} \sigma_2^2 p(p-1) f_{n,p-2}(t)
$$

for $t \in \mathbb{R}_+$, where $f_{k,\ell}(t) := 0$ if $k,\ell \in \mathbb{Z}$ with $k < 0$ or $\ell < 0$. Hence for all $M \in \mathbb{N}$, the functions $f_{n,p}$, $n,p \in \mathbb{Z}_+$ with $n + p \leq M$ satisfy a homogeneous linear system of differential equations with constant coefficients. The eigenvalues of the coefficient matrix of the above mentioned system of differential equations are $-(kb + \ell \gamma)$, $k,\ell \in \mathbb{Z}_+$ with $k + \ell \leq M$ and 0. Thus, for all $n,p \in \mathbb{Z}_+$, the function $f_{n,p}$ is a linear combination of the functions $e^{-(kb+\ell \gamma)t}$, $t \in \mathbb{R}_+$, $k,\ell \in \mathbb{Z}_+$ with $k + \ell \leq n + p$, and the constant function. Consequently, for all $f_{n,p}$, the function $f_{n,p}$ is bounded and the limit $\lim_{t \to \infty} f_{n,p}(t)$ exists and finite. By the moment convergence theorem (see, e.g., Stroock [16, Lemma 2.2.1]), $\lim_{t \to \infty} f_{n,p}(t) = \lim_{t \to \infty} \mathbb{E}(Y_t^n X_t^p) = \mathbb{E}(Y_\infty^n X_\infty^p)$, $n,p \in \mathbb{Z}_+$. Indeed, by Theorem 3.1 and the continuous mapping theorem, $Y_t^n X_t^p \xrightarrow{D} Y_\infty^n X_\infty^p$ as $t \to \infty$, and the family $\{Y_t^n X_t^p : t \in \mathbb{R}_+\}$ is uniformly integrable. This latter fact follows from the boundedness of the function $f_{2n,2p}$, see, e.g., Stroock [16, condition (2.2.5)]. Hence we conclude that all the mixed moments of $(Y_\infty, X_\infty)$ are finite.

Next, we calculate these mixed moments. We may and do suppose that the initial value $(Y_0, X_0)$ is independent of $(W_t, B_t, L_t)_{t \in \mathbb{R}_+}$, and its distribution is the same as that of $(Y_\infty, X_\infty)$, since, due to Theorem 3.1, the distribution of $(Y_\infty, X_\infty)$ does not depend on the initial value of the model. Then, by Theorem 3.1, the process $(Y_t, X_t)_{t \in \mathbb{R}_+}$ is strictly stationary, and hence, $f_{n,p}(t) = \mathbb{E}(Y_\infty^n X_\infty^p)$ for all $t \in \mathbb{R}_+$ and $n,p \in \mathbb{Z}_+$. The above system of differential equations for the functions $f_{n,p}$, $n,p \in \mathbb{Z}_+$, yields the recursion for $\mathbb{E}(Y_\infty^n X_\infty^p)$, $n,p \in \mathbb{Z}_+$. By this recursion, one can calculate the moments listed in the theorem.

The fact that, in case of $\sigma_1 > 0$, the random variable $Y_\infty$ has gamma distribution with parameters $2a/\sigma_1^2$ and $2b/\sigma_2^2$ follows by Cox et al. [5, Equation (20)].

Finally, we prove that the distribution of $(Y_\infty, X_\infty)$ is absolutely continuous whenever $\sigma_1 > 0$ and $(1 - \varrho^2)\sigma_2^2 + \sigma_3^2 > 0$. Let us consider a 2-dimensional affine diffusion model (1.1) with random initial value $(Y_0, X_0)$ independent of $(W_t, B_t, L_t)_{t \in [0,\infty)}$ having the same distribution as that of $(Y_\infty, X_\infty)$. Then, by part (ii) of Theorem 3.1, the process $(Y_t, X_t)_{t \in [0,\infty)}$ is strictly stationary. Hence
it is enough to prove that the distribution of \((Y_1, X_1)\) is absolutely continuous. According to the proof of part (b) in the proof of Theorem 3.1, the conditional distribution of \((Y_1, X_1)\) given \((Y_0, X_0)\) is absolutely continuous. This clearly implies that the (unconditional) distribution of \((Y_1, X_1)\) is absolutely continuous, and hence, the distribution of \((Y_\infty, X_\infty)\) is absolutely continuous. □

References


BÉÁTA BOLYOG: BOLYAI INSTITUTE, UNIVERSITY OF SZEGERD, ARADI VÉRTANÚK TERE 1, H–6720 SZEGERD, HUNGARY

E-mail address: bbeata@math.u-szeged.hu

GYULA PAP: BOLYAI INSTITUTE, UNIVERSITY OF SZEGERD, ARADI VÉRTANÚK TERE 1, H–6720 SZEGERD, HUNGARY

E-mail address: papgy@math.u-szeged.hu

URL: http://www.math.u-szeged.hu/~papgy/index_en.html