Notes on Topics in Classical Analysis, Math 502, Spring 2017

Eric A. Carlen[1]
Rutgers University

April 13, 2017

0.1 Markov kernels

0.1 DEFINITION. Let $(\Omega, \mathcal{M}, \mu)$ be a measure space. A *Markov kernel* on $(\Omega, \mathcal{M}, \mu)$ is a non-negative function K on $(\Omega \times \Omega, \mathcal{M} \otimes \mathcal{M})$ such that for each $x \in \Omega$,

$$\int_\Omega K(x,y)\mathrm{d}\mu(y) = 1\ . \tag{0.1}$$

That is, for each x, $K(x,y)\mathrm{d}\mu(y)$ is a probability measure on $(\Omega, \mathcal{M})$. A Markov kernel K is *doubly stochastic* in case both

$$\int_\Omega K(x,y)\mathrm{d}\mu(y) = 1 \quad \text{and} \quad \int_\Omega K(x,y)\mathrm{d}\mu(x) = 1 \tag{0.2}$$

for all x, y. A Markov kernel K is *symmetric* in case $K(x,y) = K(y,x)$ for all $x, y \in \Omega$. Every symmetric Markov kernel is doubly stochastic.

In probability theory, Markov kernels arise in the description of Markov jump process with a *state space* Ω. When a jump from state x occurs, the probability of jumping from x into a measurable set $E \subset \Omega$ is given by $\int_E K(x,y)\mathrm{d}\mu(y)$. Markov kernels play a prominent role in analysis as well; the results in the rest of this section give a first indication of why this is the case.

0.2 THEOREM. *Let K be a doubly stochastic Markov kernel on $(\Omega, \mathcal{M}, \mu)$. For each $p \in [1, \infty]$, define a linear operator P_K on $L^p(\Omega, \mathcal{M}, \mu) \cap L^\infty(\Omega, \mathcal{M}, \mu)$ by*

$$P_K f(x) = \int_\Omega K(x,y) f(y)\mathrm{d}\mu(y) = 1\ . \tag{0.3}$$

Then $\|P_K f\|_p \le \|f\|_p$, so that P_K extends by continuity to a contraction on $L^p(\Omega, \mathcal{M}, \mu)$.

Proof. Suppose f is bounded and measurable. Then for each x, f is integrable with respect to $K(x,y)\mathrm{d}\mu(y)$, and

$$\left|\int_\Omega K(x,y)f(y)\mathrm{d}\mu(y)\right| \le \int_\Omega K(x,y)|f(y)|\mathrm{d}\mu(y) \le \|f(y)\|_\infty \int_\Omega K(x,y)\mathrm{d}\mu(y) = \|f(y)\|_\infty\ .$$

Now suppose that for $p \in [1, \infty)$, $f \in L^p(\Omega, \mathcal{M}, \mu)$. Since $t \mapsto |t|^p$ is convex on $\mathbb{R}$, Jensen's inequality implies

$$\left|\int_\Omega K(x,y)f(y)\mathrm{d}\mu(y)\right|^p \le \left(\int_\Omega K(x,y)|f(y)|\mathrm{d}\mu(y)\right)^p \le \int_\Omega K(x,y)|f(y)|^p\mathrm{d}y\ .$$

Then by the Fubini-Toninelli Theorem,

$$\int_\Omega \left| \int_\Omega K(x,y)f(y)\mathrm{d}\mu(y) \right|^p \mathrm{d}\mu(x) \le \int_\Omega \left(\int_\Omega K(x,y)\mathrm{d}\mu(x) \right) |f(y)|^p \mathrm{d}\mu(y) = \|f\|_p^p \ .$$

□

Let K be a symmetric Markov kernel on $(\Omega, \mathcal{M}, \mu)$, and let P_K also denote the corresponding contraction $\mathcal{H} := L^2(\Omega, \mathcal{M}, \mu)$. It is easy to see that P_K is self-adjoint.

Convolution operators provide an important class of examples. Let ρ be a non-negative Borel function on $\mathbb{R}$ such that $\int_{\mathbb{R}} \rho \mathrm{d}x = 1$. In this case, we say that ρ is a probability density on $(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Given such a function ρ, define $K(x,y) = \rho(x-y)$. Then it is evident that $K(x,y)$ is a doubly stochastic Markov kernel. Therefore, for each $p \in [1,\infty]$ we may define an operator T_ρ acting on $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ by

$$T_\rho f(x) = \int_{\mathbb{R}} \rho(x-y)f(y)\mathrm{d}y =: \rho * f(x) \ . \tag{0.4}$$

Then $\rho * f$ is called the convolution of ρ and f, and T_ρ is the operator of convolution by ρ. As an immediate consequence of Theorem 0.2 we have that for any probability density ρ, and $p \in [1,\infty]$ and and $f \in L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$,

$$\|\rho * f\|_p \le \|f\|_p \ . \tag{0.5}$$

If ρ is any probability density, then for all $\lambda > 0$,

$$\rho_\lambda(x) = \lambda^{-1}\rho(x/\lambda) \tag{0.6}$$

is also a probability density. If ρ has support in a compact interval $[-L, L]$, ρ_λ has support in a the interval $[-\lambda L, \lambda L]$. Let f be a continuous compactly supported function on $\mathbb{R}$. Then for all $\epsilon > 0$, there is a $\delta > 0$ so that $|x-y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon$. Then for λ such that $\lambda L < \delta$,

$$|\rho_\lambda * f(x) - f(x)| = \left| \int_{\mathbb{R}} \rho_\lambda(x-y)(f(y) - f(x))\mathrm{d}y \right| \le \int_{\mathbb{R}} \rho_\lambda(x-y)|f(y) - f(x)|\mathrm{d}y \le \epsilon$$

since $\rho_\lambda(x-y) = 0$ if $|y-x| \ge \delta$. Also note that if f is supported in the interval $[-R, R]$, then $\rho_\lambda * f$ is supported in the interval $[-\lambda L - R, \lambda L + R]$ since $\rho_\lambda(y-x)f(y) = 0$ if $|x| > R + \lambda L$. It follows that $|\rho_\lambda * f - f|$ is bounded by ϵ and is supported in $[-R-1, R+1]$ for all λ such that $\lambda L < \min\{\delta, 1\}$. For such λ, $\|\rho_\lambda f - f\|_p \le \epsilon(2R+2)^{1/p}$. We have proved an important special case of the following theorem:

0.3 THEOREM. *For any probability density ρ on $\mathbb{R}$, let ρ_λ be defined for $\lambda > 0$ by (0.6). Then for all $p \in [1,\infty)$, and all $f \in L^p(\mathbb{R}, \mathcal{M}, \mathrm{d}x)$,*

$$\lim_{\lambda to 0} \|\rho_\lambda * f - f\|_p = 0 \ .$$

Proof. Define $\eta_L = \int_{[-L,L]^c} \rho(x)\mathrm{d}x$, and note that $\lim_{L\to\infty} \eta_L = 0$. For all L large enough that $\eta_L < 1$, Define

$$\rho^{(L)}(x) := (1-\eta_L)^{-1} 1_{[-L,L]}(x)\rho(x) \quad \text{and} \quad r^{(L)}(x) := \eta_L^{-1} 1_{[-L,L]^c}(x)\rho(x) \ .$$

Then evidently both $\rho^{(L)}$ and $r^{(L)}$ are probability densities $\rho = (1-\eta_L)\rho^{(L)} + \eta_L r^{(L)}$ is a convex combination of them. Hence for all $f \in L^p(\mathbb{R}, \mathcal{M}, \mathrm{d}x)$, Likewise, for all $\lambda > 0$, $\rho_\lambda = (1-\eta_L)\rho_\lambda^{(L)} + \eta_L r_\lambda^{(L)}$, and then for all $p \in [1,\infty)$, and all $f \in L^p(\mathbb{R}, \mathcal{M}, \mathrm{d}x)$,

$$\begin{aligned} \|\rho_\lambda * f - f\|_p &= \|(1-\eta_L)(\rho_\lambda^{(L)} * f - f) + \eta_L(r_\lambda^{(L)} * f - f)\|_p \\ &\leq (1-\eta_L)\|\rho_\lambda^{(L)} * f - f\|_p + \eta_L\|r_\lambda^{(L)} * f - f\|_p \\ &\leq (1-\eta_L)\|\rho_\lambda^{(L)} * f - f\|_p + \eta_L 2\|f\|_p \ , \end{aligned}$$

where we used (0.5) in the last step,

Now fix $\epsilon > 0$, and choose L sufficiently large that $\eta_L 2\|f\|_p < \epsilon/3$. Since $C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$, we may choose $g \in C_c(\mathbb{R})$ such that $\|f - g\|_p < \epsilon/3$. Then using Minkowski's inequality and (0.5),

$$\|\rho_\lambda^{(L)} * f - f\|_p \leq \|\rho_\lambda^{(L)} * (f-g)\|_p + \|\rho_\lambda^{(L)} * g - g\|_p + \|f - g\|_p \leq \|\rho_\lambda^{(L)} * g - g\|_p + 2\epsilon/3 \ .$$

Altogether, we have that

$$\|\rho_\lambda * f - f\|_p \leq \|\rho_\lambda^{(L)} * g - g\|_p + \epsilon \ ,$$

and since ρ_L has support in $[-L, L]$ and since $g \in C_c(\mathbb{R})$, by what we have explained just before the statement of the theorem, $\lim_{\lambda\to 0} \|\rho_\lambda^{(L)} * g - g\|_p = 0$. Since $\epsilon > 0$ is arbitrary, the proof is complete. □

0.2 The basic facts about convolution

Throughout this section, L^p denotes $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Let $f, g \in L^1 \cap L^\infty$. Then the integral

$$g * f(x) := \int_{\mathbb{R}} g(x-y) f(y) \mathrm{d}y \tag{0.7}$$

converges for all x, and we have the point-wise inequality

$$|g * f(x)| \leq |g| * |f|(x) \tag{0.8}$$

If $g \neq 0$, so that $\|g\|_1 \neq 0$, $\rho := \|g\|_1^{-1}|g|$ is a probability density. Therefore, by (0.5), for all $p \in [1, \infty]$,

$$\|g * f\|_p \leq \||g| * |f|\|_p = \|g\|_1 \|\rho * |f|\|_p \leq \|g\|_1 \|f\|_p \ . \tag{0.9}$$

In particular, $g * f \in L^1 \cap L^\infty$. Thus, convolution is a product on $L^1 \cap L^\infty$, making it an algebra. By the Fubini-Toninelli Theorem, and the change of variables $w = y - x$,

$$\int_{\mathbb{R}} f * g \mathrm{d}x = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g(x-y) f(y) \mathrm{d}x \right) \mathrm{d}y = \left(\int_{\mathbb{R}} g(w) \mathrm{d}w \right) \left(\int_{\mathbb{R}} f(y) \mathrm{d}y \right) \ , \tag{0.10}$$

so that when f and g are non-negative, $\|f * g\|_1 = \|f\|_1 \|g\|_1$.

Making the change of variables $y = x - w$,

$$g * f(x) = \int_{\mathbb{R}} g(x-y) f(y) \mathrm{d}y = \int_{\mathbb{R}} f(x-w) g(w) \mathrm{d}w = f * g(x) \tag{0.11}$$

Thus the convolution product on $L^1 \cap L^\infty$ is commutative. It is also associative: Let $f, g, h \in L^1 \cap L^\infty$. We have already seen that $f * g \in L^1 \cap L^\infty$. Then making the change of variables $y = u - w$, and applying the Fubini-Toninelli Theorem,

$$\begin{aligned} (f * g) * h(x) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x - w - y) g(y) \mathrm{d}y \right) h(w) \mathrm{d}w \\ &= \int_{\mathbb{R}} f(x-u) \left(\int_{\mathbb{R}} g(u-w) h(w) \mathrm{d}w \right) \mathrm{d}u = f * (g * h)(x) \ . \end{aligned} \tag{0.12}$$

There is an important complement to (0.9): Since $L^1 \cap L^\infty \subset L^p$ for all $p \in [1, \infty]$, for $p \in (1, \infty)$, $g \in L^{p/(p-1)}$ and $f \in L^p$, and then by Hölder's inequality, for all x,

$$|g * f(x)| \leq \int_{\mathbb{R}} |g(x-y)||f(y)| \mathrm{d}y \leq \|g\|_{p/(p-1)} \|f\|_p \ .$$

Therefore,

$$\|g * f\|_\infty \leq \|g\|_{p/(p-1)} \|f\|_p \ . \tag{0.13}$$

For all $\lambda > 0$ and all $f \in L^1$, define

$$f_\lambda(x) = \lambda^{-1} f(x/\lambda) \tag{0.14}$$

which reduces to (0.6) when f is a probability density. A simple change of variables shows that $\|f_\lambda\|_1 = \|f\|_1$ for all $\lambda > 0$. A simple computation shows that for all $f, g \in L^1$,

$$g_\lambda * f_\lambda(x) = \lambda^{-2} \int_{\mathbb{R}} g(x/\lambda - y/\lambda) f(y/\lambda) \mathrm{d}y = \lambda^{-1} f * g(x/\lambda) = (g * f)_\lambda(x) \ . \tag{0.15}$$

Another simple computation shows that for $f \in L^1 \cap L^\infty$, $p \in [1, \infty)$ and $\lambda > 0$, with f_λ defined in (0.14),

$$\|f_\lambda\|_p = \lambda^{1/p - 1} \|f\|_p \ . \tag{0.16}$$

Now suppose that for some $p, q, r \in [1, \infty)$, there is a finite constant C such that for all $f, g \in L^1 \cap L^\infty$,

$$\|g * f\|_r \leq C \|g\|_q \|f\|_p \ . \tag{0.17}$$

Then replacing f and g by f_λ and g_λ, and using (0.15),

$$\|g * f\|_r = \lambda^{1-1/r} \|(g * f)_\lambda\|_r = \lambda^{1-1/r} \|g_\lambda * f_\lambda\|_r \leq C \|g_\lambda\|_q \|f_\lambda\|_p = \lambda^{1/q + 1/p - 1/r - 1} C \|g\|_q \|f\|_p \ .$$

Since the left side is independent of λ, the right had side must be independent of λ also, since if $1/q + 1/p - 1/r - 1 > 0$, taking $\lambda \to 0$, we would conclude that $\|g * f\|_r = 0$, and thus that $g * f = 0$ almost everywhere. If $1/q + 1/p - 1/r - 1 < 0$, taking $\lambda \to \infty$ brings us to the same conclusion. But for non-negative f, g, $\|g * f\|_1 = \|g\|_1 \|f\|_1$, and so $g * f = 0$ almost everywhere implies that $f = g = 0$.

Therefore, there is no finite constant C for which the inequality (0.17) can hold in general unless

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1 \ . \tag{0.18}$$

It turns out that when (0.18) is satisfied, then (0.17) is also valid in general, with a constant C no greater than 1. This is Young's inequality for convolution. We have already seen several important cases, namely the case in which $r = p, q = 1$, which is (0.9) and the case $r = \infty$,$q = p/(p-1)$, which is (0.13).

There are several ways to prove the remaining cases of Young's inequality with the constant $C = 1$. The inequality actually holds in these cases with an optimal constant C that is strictly less than 1, We shall prove this sharp form of Young's inequality, which is due to Beckner and Brascamo and Lieb later in the Chapter. The main tool used in the proof that we give is the heat semigroup, our next topic. For now, we close this section with the following lemma that summarizes our main conclusions.

0.4 LEMMA. *The convolution product on $L^1 \cap L^\infty$ is commutative and associative, and the inequalities (0.9) and (0.13) for all $f, g \in L^1 \cap L^\infty$. Moreover, since $L^1 \cap L^\infty$ is dense in L^r for all $1 < r < \infty$, the map $(g, f) \mapsto g * f$ extends by continuity to a map from $L^\times L^p$ to L^p for which (0.9) is valid, and to a map from $L^{p/(p-1)} \times L^p$ to L^∞ for which (0.13) is valid.*

0.3 The Heat semigroup

The Gaussian probability density $\gamma(x)$ associated to the *normal distribution* of probability theory is given by

$$\gamma(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \ .$$

For $t > 0$, define $\gamma_t(x) = (2\pi t)^{-1/2} e^{-x^2/2t}$. This is the Gaussian probability density with mean 0 and variance t. That is,

$$\int_{\mathbb{R}} \gamma_t(x)\mathrm{d}x = 1 \ , \qquad \int_{\mathbb{R}} x\gamma_t(x)\mathrm{d}x = 0 \quad \text{and} \quad \int_{\mathbb{R}} x^2\gamma_t(x)\mathrm{d}x = t \ . \tag{0.19}$$

Note that the definition $\gamma_t(x) = t^{-1/2}\gamma(x/\sqrt{t})$ duffers slightly from (0.6) due to the square roots on the right. In the present context, this scaling will turn out to be more natural. The following lemma explains why.

0.5 LEMMA. *For all $s, t > 0$,*

$$\gamma_t * \gamma_s = \gamma_{t+s} \ . \tag{0.20}$$

Proof. Note that $\gamma_t(x-y)\gamma_s(y) = \dfrac{1}{2\pi\sqrt{st}} e^{-[(x-y)^2/t + y^2/s]/2}$, and

$$\frac{(x-y)^2}{t} + \frac{y^2}{s} = \frac{t+s}{st}\left(y - \frac{s}{t+s}x\right)^2 + \frac{1}{t+s}x^2 \ .$$

$$\int_{\mathbb{R}} \gamma_t(x-y)\gamma_s(y)\mathrm{d}y = \gamma_{s+t}(x)\sqrt{\frac{s+t}{2\pi st}} \int_{\mathbb{R}} e^{-(t+s)(y-sx/(t+s))^2/(2st)}\mathrm{d}y = \gamma_{s+t}(x) \ .$$

□

0.6 DEFINITION (Heat semigroup)**.** For each $t > 0$, define an operator P_t on $L^1 \cap L^\infty$ by

$$P_t f = \gamma_t * f \ . \tag{0.21}$$

For $t > 0$, define $P_0 = I$. Then the family of operators $\{P_t\}_{t \geq 0}$ constitute the *heat semigroup*. The name will be justified shortly.

0.7 THEOREM. *Let $\{P_t\}_{t\geq 0}$ be the heat semigroup on $L^1 \cap L^\infty$. Then for all $s,t \geq 0$,*

$$P_s P_t = P_{s+t} \ . \tag{0.22}$$

For all $p \in [1,\infty]$, P_t extends by continuity to an element of $\mathscr{B}(L^p)$, and, for all $f \in L^p$, $\|P_t f\|_p \leq \|f\|_p$. Moreover, for all $p \in [1,\infty)$ and all $f \in L^p$, $\lim_{t\to 0}\|P_t f - f\|_p = 0$. Finally, for all $r > p \in [1,\infty]$, and all $t > 0$, P_t extends by continuity to an element of $\mathscr{B}(L^p, L^r)$ with operator norm

$$\|P_t\|_{p\to r} \leq \left(\frac{1}{q}\right)^{\frac{r-p}{2qr}} (2\pi t)^{-\frac{r-p}{2qr}} \ . \tag{0.23}$$

Proof. By Lemma 0.4 and the definition of the heat semigroup, and then Lemma 0.5, for all $f \in L^1 \cap L^\infty$,

$$P_s P_t f = \gamma_s * (\gamma_t * f) = (\gamma_s * \gamma_t) * f = \gamma_{s+t} * f = P_{s+t} f \ .$$

This proves (0.22). The inequality $\|P_t f\|_p \leq \|f\|_p$ is a direct consequence of (0.5) for $f \in L^1 \cap L^\infty$, and then evidently P_t extends by continuity to an element of $\mathscr{B}(L^p)$. The statement concerning $\lim_{t\to 0}\|P_t f - f\|_p$ is a direct consequence of Theorem 0.3. finally, by (0.13),

$$\|P_t f\|_\infty \leq \|\gamma_t\|_{p/(p-1)}\|f\|_p = \left(\frac{1}{q}\right)^{1/2q} (2\pi t)^{-1/2q}\|f\|_p \ ,$$

which proves (0.23) for $r = \infty$. For $r \in (p,\infty)$ $|P_t f(x)|^r \leq |P_t f(x)|^p \|P_t f\|_\infty^{r-p}$, and hence

$$\|P_t f\|_r^r \leq \|P_t f\|_p^p \|P_t f\|_\infty^{r-p} \leq \|f\|_p^p \|P_t f\|_\infty^{r-p} \leq \left(\frac{1}{q}\right)^{\frac{r-p}{2q}} (2\pi t)^{\frac{r-p}{2q}} \|f\|_p^r \ ,$$

which proves (0.23) in general. □

The case $p = 2$ will be particularly important in what follows. Since for all $f, g \in L^2$, $t > 0$,

$$\langle f, P_t g\rangle_{L^2} = \int_{\mathbb{R}} \overline{f} P_t g(x)\mathrm{d}x = \int_{\mathbb{R}}\int_{\mathbb{R}} \overline{f(x)}\gamma_t(x-y)g(y)\mathrm{d}x\mathrm{d}y = \langle P_t f, g\rangle_{L^2} \ ,$$

it follows that P_t is self-adjoint on L^2.

Theorem 0.7 says that for each $p \in [1,\infty]$ and each $f \in L^p$, $P_t f$ converges to f in the L^p norm as $t \to 0$. as $t \to 0$. Later in this chapter we will have the means to easily generate examples showing that for all $t, \epsilon > 0$ and all $1 \leq p < \infty$, there exists $h \in L^p$, $\|h\|_p = 1$ and $\|P_t h - h\|_p > 1 - \epsilon$. This brings us to the notion of the *strong operator topology*

0.8 DEFINITION (The strong operator topology)**.** Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. For each $x \in X$, define the map ψ_x on $\mathscr{B}(X,Y)$ by $\psi_x(T) = \|Tx\|_Y$

The *strong operator topology* on $\mathscr{B}(X,Y)$ is the weakest topology on $\mathscr{B}(X,Y)$ making each of the map ψ_x continuous.

By what we have seen in our study of topological vector spaces, this topology makes $\mathscr{B}(X,Y)$ a topological vector space, an a neighborhood base at the origin is given by the sets

$$V_{x_1,\dots,x_n,\epsilon} = \bigcap_{j=1}^n \{T \in \mathscr{B}(X,Y) \ : \ \|Tx_j\| < \epsilon\}$$

for $\epsilon > 0$ and finite sets $\{x_1,\dots,x_n\} \subset X$.

0.9 DEFINITION (Strongly continuous semigroup)**.** Let $(X, \|\cdot\|)$ be a Banach space. A *strongly continuous semigroup* on X is a set $\{P_f \ : \ t \geq 0\} \subset \mathscr{B}(X)$ such that $P_0 = I$, $t \mapsto P_t$ is continuous with respect to the strong operator topology on $\mathscr{B}(X)$, and such that for all $s, t \geq 0$, $P_t P_s = P_{s+t}$.

0.10 EXAMPLE (The heat semigroup on L^p)**.** *For each $1 \leq p \leq \infty$, $t > 0$, let $P_t \in \mathscr{B}(L^p)$ be defined as in (0.23) and extended by continuity. Then by Theorem 0.7, $t \mapsto P_t$ is a strongly continuous semigroup on L^p. It is called the* heat semigroup on L^p.

The heat semigroup gets its name from the fact that, considered as a function of $t > 0$ and $x \in \mathbb{R}$, $\gamma_t(x)$ satisfies the *heat equation*:

$$\frac{\partial}{\partial t}\gamma_t(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\gamma_t(x) \ . \tag{0.24}$$

Indeed, a calculation gives $\frac{\partial}{\partial t}\gamma_t(x) = \frac{x^2}{2t^2}\gamma_t(x)$, which, as a function of x, belongs to L^q for all $1 \leq q \leq 1$, Moreover, for all $0 < s, t$, $s \neq t$,

$$\gamma_t(x) - \gamma_s(x) = \int_s^t \frac{x^2}{2r^2}\gamma_r(x)\mathrm{d}r = (t-s)\frac{x^2}{2t^2}\gamma_t(x) + \int_s^t \left[\frac{x^2}{2r^2}\gamma_r(x) - \frac{x^2}{2t^2}\gamma_t(x)\right]\mathrm{d}r$$

Simple estimates now yield, for all $1 \leq q \leq \infty$,

$$\left\|\gamma_t - \gamma_s - (t-s)\tfrac{\partial}{\partial t}\gamma_t\right\|_q = o(|t-s|) \ . \tag{0.25}$$

Using (0.25) for $q = 1$, we have that for all $f \in L^p$ if we define $f(t,x) = P_t f(x)$,

$$\lim_{s \to t}\frac{f(s,x) - f(t,x)}{s-t} = \frac{\partial}{\partial t}\gamma_t * f(x)$$

in the L^p norm, making use once more of (0.5). We may also use (0.25) with $q = p/(p-1)$ to prove point-wise convergence of this limit. This shows that $t \mapsto f(t,x)$ is differentiable for $t > 0$, and $\dfrac{\partial}{\partial t}f(t,x) = \dfrac{\partial}{\partial t}\gamma_t * f(x)$. The argument can be repeated to show that for all $m \in \mathbb{N}$,

$$\frac{\partial^m}{\partial t^m}f(t,x) = \frac{\partial^m}{\partial t^m}\gamma_t * f \ .$$

A similar argument shows that $x \mapsto f(t,x)$ is infinitely differentiable everywhere on $\mathbb{R}$, and in fact

$$\frac{\partial^m}{\partial x^m}f(t,x) = \frac{\partial^m}{\partial x^m}\gamma_t * f \ .$$

Together with (0.24) we have proved:

0.11 THEOREM. *Let $1 \leq p < \infty$, $f \in L^p$, and define $f(t,x) = P_t f(x)$. Then $f(\cdot,\cdot)$ is C^∞ on $(0,\infty) \times \mathbb{R}$, satisfies the heat equation*

$$\frac{\partial}{\partial t}f(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}f(t,x) \tag{0.26}$$

on $(0,\infty) \times \mathbb{R}$, and $\lim_{t \to 0}\|f(t,\cdot) - f\|_p = 0$.

We will come back to the question of uniqueness later in the chapter when we we have more tools to deal with this question.

0.12 LEMMA. *For all $t > 0$, and $1 \leq p \leq 2$, P_t is injective on L^p.*

Proof. Suppose that $f \in L^2$ and $P_t f = 0$. Then $P_{t/2} f \in L^2$, and by by Theorem 0.7, and the fact that each $P_{t/2}$ is self-adjoint,

$$0 = \langle f, P_t f \rangle = \langle f, P_{t/2} P_{t/2} f \rangle = \langle P_{t/2} f, P_{t/2} f \rangle = \|P_{t/2} f\|_2^2 \ .$$

Hence $P_{t/2} f = 0$ as well. A simple induction shows that $P_{t/2^n} f = 0$ for all n. Then by Theorem 0.7,

$$0 = \lim_{n\to\infty} \|P_{t/2^n} f - f\| = \|f\| \ ,$$

so that $f = 0$.

Now suppose that $f \in L^p$, $1 \leq p < 2$ and $P_t f = 0$. By the semigroup property, for all $0 < s < t$ $P_t f = P_{t-s}(P_s f)$, and by Then by Theorem 0.7, $P_s f \in L^2$. Therefore, by what was proved just above, $P_s f = 0$ for all $0 < s < t$, and so $f = \lim_{s\to 0} P_s f$. □

0.4 Hermite polynomials

Let ν be the normalized Gaussian probability measure that has the density $\gamma(x) := (2\pi)^{-1/2} e^{-|x|^2/2}$ with respect to Lebesgue measure $\mathrm{d}x$. Let $\mathcal{H} = L^2(\mathbb{R}, \mathcal{B}, \nu)$. In this section, $\|\cdot\|$ denotes the norm on $\mathcal{H}$, and $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathcal{H}$. Also, in this section, L^p denotes $L^p(\mathbb{R}, \mathcal{B}, \nu)$.

0.13 DEFINITION (Hermite polynomials)**.** For each $n \in \mathbb{N}$, let $V_n = \mathrm{span}(\{1, x, \dots, x^n\})$. Let P_n be the orthogonal projection in $\mathcal{H}$ onto V_n. For $n \in \mathbb{N}$, define the polynomial $p_n(x) = x^n$, and define $p_o(x) = 1$. Note that $\|H_n\| \neq 0$. For all $n \geq 0$, the nth *Hermite polynomial* H_n is given by

$$H_n := p_n - P_{n-1} p_n \tag{0.27}$$

and the nth *normalized Hermite polynomial* h_n is defined by

$$h_n := \|H_n\|^{-1} H_n \ . \tag{0.28}$$

By construction H_n is orthogonal to H_k for all $k \leq n$, and hence $\langle H_n, H_m \rangle = 0$ for all $n \neq m$. It follows that $\{h_n\}_{n\geq 0}$ is an orthonormal sequence in $\mathcal{H}$.

0.14 THEOREM. *For $1 \leq p < \infty$, the set $\mathscr{P}$ of polynomial functions is dense in $L^p(\mathbb{R}, \mathcal{B}, \nu)$.*

Proof. For $y \in \mathbb{R}$, define the function $g_y = e^{yx/2}$.

$$\int_{\mathbb{R}} |g_y|^p \gamma(x) \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{xyp/2 - x^2/2} = \frac{e^{y^2p^2/8}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-yp/2)/2} = e^{y^2p^2/8} \ .$$

Thus $g_y \in L^p(\mathbb{R}, \mathcal{B}, \nu)$. For $n \in \mathbb{N}$, let $p_n(x) = x^n$ and let $p_0 := 1$. Evidently, for all $n \geq 0$, $p_n \in L^p(\mathbb{R}, \mathcal{B}, \nu)$. We now claim that the point-wise series expansion $g_y(x) = \displaystyle\sum_{n=0}^{\infty} \frac{1}{n!} \frac{y^n}{2^n} p_n(x)$ converges in $L^p(\mathbb{R}, \mathcal{B}, \nu)$. To see this, we compute

$$\|p_n\|_p^p = \int_{\mathbb{R}} |x|^{np} \gamma(x) \mathrm{d}x = \int_0^\infty (x^2)^{(np-1)/2} \gamma(x) 2x \mathrm{d}x = \int_0^\infty y^{(np-1)/2} e^{-y} \mathrm{d}y = \Gamma((np+1)/2) \ .$$

Stirling's formula for the Gamma function says that

$$\Gamma((np+1)/2) \sim \sqrt{2\pi((np-1)/2)}\left(\frac{(np-1)/2}{e}\right)^{(np-1)/2} = \sqrt{\frac{2\pi}{e}}\left(\frac{np-1}{2e}\right)^{np/2} .$$

Therefore,

$$\|p_n\|_p \sim \left(\frac{2\pi}{e}\right)^{1/p}\left(\frac{np-1}{2e}\right)^{n/2} .$$

Since the power series $\displaystyle\sum_{n=0}^{\infty}\frac{n^{n/2}}{n!}z^n$ has an infinite radius of convergence, $\displaystyle\sum_{n=0}^{\infty}\frac{1}{n!}\frac{|y|^n}{2^n}\|p_n\|_p < \infty$. Therefore, by Minkowski's inequality, the series expansion $g_y = \displaystyle\sum_{n=0}^{\infty}\frac{1}{n!}\frac{y^n}{2^n}p_n$ converges in $L^p(\mathbb{R},\mathcal{B},\nu)$.

Suppose that $\mathscr{P}$ is not dense in $L^p(\mathbb{R},\mathcal{B},\nu)$. Then by the Hahn-Banach Theorem and the Riesz Representation Theorem, there exists a unit vector $h \in L^q$, $q = p/(p-1)$, such that $\int_{\mathbb{R}} hf\,\mathrm{d}\nu = 0$ for all $f \in \mathscr{P}$. Since g_y is the L^p norm limit of a sequence of polynomials, $\displaystyle\int_{\mathbb{R}} hg_y\,\mathrm{d}\nu = 0$ for all $y \in \mathbb{R}$. Define $\widetilde{h} := \gamma^{1/q}h$, and note that

$$\|\widetilde{h}\|_{L^q(\mathrm{d}x)} = \|h\|_{L^q(\nu)} \tag{0.29}$$

and also that

$$\begin{aligned}\int_{\mathbb{R}} h(x)g_y(x)\mathrm{d}\nu &= \int_{\mathbb{R}}\widetilde{h}(x)\gamma^{1/p}(x)g_y(x)\mathrm{d}x\\ &= (2\pi)^{-1/2p}\int_{\mathbb{R}}\widetilde{h}(x)e^{-x^2/2p+xy/2}\mathrm{d}x = (2\pi)^{-1/2p}e^{py^2/8}\int_{\mathbb{R}}\widetilde{h}(x)e^{-(x-py/2))^2/2p}\mathrm{d}x .\end{aligned}$$

Since $\displaystyle\frac{1}{4\sqrt{\pi p}}\int_{\mathbb{R}}\widetilde{h}(x)e^{-(x-py/2))^2/2p}\mathrm{d}x = P_{2p}\widetilde{h}(py)$ where $\{P_t\}_{t\geq 0}$ is the heat semigroup, $P_{2p}\widetilde{h} = 0$, when $q \in [1,2]$, Lemma 0.12 implies that $\widetilde{h} = 0$. By (0.29), $h = 0$, which is a contradiction. Therefore, $\mathscr{P}$ is dense in L^p for $p \in [2,\infty)$.

For $p \in [1,2)$, $f \in L^p$, and $\epsilon > 0$, pick $g \in L^2$ such that $\|g-f\|_p < \epsilon/2$. Then pick $h \in \mathscr{P}$ such that $\|g-h\|_2 < \epsilon/2$. Then since ν is a probability measure,

$$\|f-h\|_p \leq \|f-g\|_p + \|g-h\|_p \leq \|f-g\|_p + \|g-h\|_2 \leq \epsilon$$

This proves the density of $\mathscr{P}$ in L^p for $p \in [1,2)$. □

0.15 COROLLARY. *The normalized hermite polynomials $\{h_n\}_{n\geq 0}$ are an orthonormal basis for $L^2(\mathbb{R},\mathcal{B},\mathrm{d}\nu)$.*

0.5 The Mehler semigroup

Let ν be the normalized Gaussian probability measure that has the density $\gamma(x) := (2\pi)^{-1/2}e^{-|x|^2/2}$ with respect to Lebesgue measure $\mathrm{d}x$. Let $\mathcal{H} = L^2(\mathbb{R},\mathcal{B},\nu)$. A function f on $\mathbb{R}$ is *polynomially bounded* in case there exists a constant C and an $n \in \mathbb{N}$ such that $|f(x)| \leq C(1+|x|^n)$ for all x. Clearly any polynomial bounded function belongs to $L^p(\mathbb{R},\mathcal{B},\nu)$ for all $p \in [1,\infty)$.

For $\lambda \in (0,1)$, define a linear transformation S_λ on the continuous polynomially bounded functions on $\mathbb{R}$ by

$$S_\lambda f(x) = \int_{\mathbb{R}} f(\lambda x + \sqrt{1-\lambda^2}y)\gamma(y)\mathrm{d}y \ . \tag{0.30}$$

Define $z = \lambda x + \sqrt{1-\lambda^2}y$ so that $\begin{pmatrix} x \\ z \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda & \sqrt{1-\lambda^2} \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Computing the Jacobian of the transformation, we find $\mathrm{d}x\mathrm{d}y = (1-\lambda^2)^{-1/2}\mathrm{d}x\mathrm{d}z$. Also,

$$x^2 + y^2 = \frac{\lambda^2}{1-\lambda^2}\left(z^2 + x^2 - \frac{2}{\lambda}zx\right) + x^2 + z^2 \ ,$$

Hence

$$\gamma(x)\gamma(y) = \exp\left[-\frac{\lambda^2}{2(1-\lambda^2)}\left(z^2 + x^2 - \frac{2}{\lambda}zx\right)\right]\gamma(x)\gamma(z) \ .$$

and for $g \in \mathcal{H}$, $\int_{\mathbb{R}} g(x)S_\lambda(x)\gamma(x)\mathrm{d}x = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x)K_\lambda(x,z)f(z)\gamma(x)\gamma(z)\mathrm{d}z\mathrm{d}x$ where

$$K_\lambda(x,z) = \frac{1}{\sqrt{1-\lambda^2}}\exp\left[-\frac{\lambda^2}{2(1-\lambda^2)}\left(z^2 + x^2 - \frac{2}{\lambda}zx\right)\right] \ . \tag{0.31}$$

Therefore, we can write (0.30) in the alternate form

$$S_\lambda f(x) = \int_{\mathbb{R}} K_\lambda(x,z)f(z)\gamma(z)\mathrm{d}z \ . \tag{0.32}$$

Note that $K_\lambda(x,z)\gamma(z) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{1-\lambda^2}}\exp\left[-\frac{1}{2(1-\lambda^2)}|z-\lambda x|^2\right] = \gamma_{1-\lambda^2}(z-\lambda x)$. By (0.19), (**??**) and symmetry,

$$\int_{\mathbb{R}} K_\lambda(x,z)\gamma(z)\mathrm{d}z = \int_{\mathbb{R}} \gamma(x)K_\lambda(x,z)\mathrm{d}x = 1 \ . \tag{0.33}$$

This shows that $K_\lambda(x,y)$ is a symmetric (and therefore doubly stochastic) Markov kernel on $(\mathbb{R}, \mathcal{B}, \mathrm{d}\nu)$. Since S_λ is the operators associated to K_λ, Theorem 0.2 immediately yields:

0.16 LEMMA. *For all $p \in [1,\infty]$, all continuous polynomaily bounded functions f, and all $\lambda\ in(0,1)$,*

$$\|S_\lambda f\|_p \leq \|f\|_p \ . \tag{0.34}$$

Combining this with the density of $\mathscr{P}$ in L^p for $p \in [1,\infty)$, we have that for each $\lambda \in (0,1)$, S_λ extends by continuity to a bounded linear operator of norm one on L^p. We denote the extension also by S_λ.

0.17 THEOREM. *For each $\lambda \in (0,1)$, the Hermite basis $\{h_n\}_{n\geq 0}$ of $\mathcal{H}$ is a complete orthonormal set of eigenvectors for S_λ considered as an operator on $\mathcal{H}$: For each $\lambda \in (0,1)$ and each $n > 0$*

$$S_\lambda h_n = \lambda^n h_n \ . \tag{0.35}$$

Proof. Let $p_n(x) := x^n$. By the definition (0.30),

$$\begin{aligned} S_\lambda p_n(x) &= \sum_{m=0}^{n} x^{n-m}(1-\lambda^2)^{m/2}\lambda^{n-m}\binom{n}{m}\left(\int_{\mathbb{R}} y^m \gamma(y)\mathrm{d}y\right) \\ &= \sum_{m=0\ ,\ m \text{ even}}^{n} \left[(1-\lambda^2)^{m/2}\lambda^{n-m}\binom{n}{m}(m-1)!!\right] p_{n-m}(x)\ , \end{aligned} \tag{0.36}$$

where we have used the fact that $\int_{\mathbb{R}} y^m\gamma(y)\mathrm{d}y = (m-1)!! = (m-1)(m-3)\cdots 1$ for m even and is zero for m odd. The $m = 0$ term in the sum is $\lambda^n x^n$, and all other terms are multiples of lower powers of x. Therefore,

$$S_\lambda p_n(x) = \lambda^n x^n + \text{lower order}\ . \tag{0.37}$$

That is, for each $n \in \mathbb{N}$, $S_\lambda p_n$ is a polynomial of degree n at most. This shows that for each $k \in \mathbb{N}$, the subspace V_k of $\mathcal{H}$ spanned by polynomials of degree k or less is invariant under S_λ, and this is the crucial point upon which the proof rests.

Then since $H_n(x) = p_n(x) + \text{lower order}$, and since $S_\lambda V_{n-1} \subset V_{n-1}$,

$$S_\lambda H_n = S_\lambda p_n + \text{lower order} = \lambda^n p_n + \text{lower order} = \lambda^n H_n + \text{lower order}\ . \tag{0.38}$$

Since $\{h_0, \dots, h_n\}$ spans V_n, $S_\lambda H_n = \displaystyle\sum_{m=0}^{n}\langle h_m, S_\lambda H_n\rangle h_m = \sum_{m=0}^{n}\langle S_\lambda h_m, H_n\rangle$, using the fact that S_λ is self-adjoint. Then since $S_\lambda h_m \in V_m$, for $m < n$, $\langle S_\lambda h_m, H_n\rangle = 0$. Therefore, $S_\lambda H_n$ is a multiple of H_n, and the lower order terms on the right hand side of (0.38) must all be zero. □

0.18 LEMMA. *For all $\lambda, \mu \in (0,1)$, $p \in [1,\infty)$,*

$$S_\mu S_\lambda = S_{\lambda\mu}\ , \tag{0.39}$$

holds as an identity in $\mathscr{B}(L^p)$, and for all $f \in L^p$,

$$\lim_{\lambda\to 1}\|S_\lambda f - f\|_p = 0\ . \tag{0.40}$$

Proof. It is not hard to prove (0.39) directly from the definition, but now that we have an orthonormal basis for $\mathcal{H}$ consisting of eigenvectors, things are much easier. For every $f \in \mathcal{H}$, consider the expansion $f = \sum_{n=1}^{\infty}\langle h_n, f\rangle h_n$. Then

$$S_\lambda f = \sum_{n=0}^{\infty}\langle h_n, f\rangle S_\lambda h_n = \sum_{n=0}^{\infty}\langle h_n, f\rangle \lambda^n h_n\ .$$

Therefore $S_\mu S_\lambda f = \displaystyle\sum_{n=0}^{\infty}\langle h_n, f\rangle(\mu\lambda)^n h_n = S_{\mu\lambda} f$ Since $L^2 \cap L^p$ is dense in L^p for all $p \in [1,\infty)$, this proves (0.39).

To prove (0.40), we first consider the case in which f is a polynomial. Again let $p_n(x) := x^n$, and combine (0.36) with Minkowski's inequality to conclude

$$\|S_\lambda p_n - p_n\|_p \le (1-\lambda^n)\|p_n\| + \sum_{m=0, m \text{ even}}^{n} (1-\lambda^2)^{m/2}\binom{n}{m}(m-1)!!\|p_{n-m}\|_p$$

This shows that $\lim_{\lambda\to 1}\|S_\lambda p_n - p_n\|_p = 0$. It follows that for all $g \in \mathscr{P}$, $\lim_{\lambda\to 1}\|S_\lambda g - g\|_p = 1$.

Next fix $\epsilon > 0$, $p \in [1,\infty)$, and $f \in L^p$. Let $g \in \mathscr{P}$ be such that $\|g - f\|_p < \epsilon$. Then since $\|S_\lambda\| = 1$,

$$\|S_\lambda f - f\|_p \le \|S_\lambda(f-g)\|_p + \|S_\lambda g - g\|_p + \|t - g\| \le 2\epsilon + \|S_\lambda g - g\|_p \ .$$

For λ sufficiently close to 1, $\|S_\lambda g - g\| < \epsilon$, and then $\|S_\lambda f - f\| < 3\epsilon$. This proves (0.40). □

0.19 LEMMA. *For all $f \in \mathscr{P}$, define*

$$\mathscr{N} f(x) = \frac{\partial^2}{\partial x^2} f(x) - x\frac{\partial}{\partial x} f(x) \ . \tag{0.41}$$

Then for all $q \in [1,\infty)$, $\lambda \mapsto S_\lambda f$ is left differentiable in L^q at $\lambda = 1$, and

$$\lim_{\lambda\to 1}\frac{S_\lambda f - f}{\lambda - 1} = -\mathscr{N} f \ , \tag{0.42}$$

and the same formula is valid point-wise.

Proof. Consider (0.36), and explicitly evaluate the terms for $m = 0$ and $m = 2$ in the sum on the right to obtain

$$\begin{aligned} S_\lambda p_n(x) - p_n(x) &= (\lambda^n - 1)p_n(x) + (1-\lambda^2)\lambda^{n-2}\frac{n(n-1)}{2}p_{n-2}(x) \\ &+ \sum_{m=4\ ,\ m \text{ even}}^{n}\left[(1-\lambda^2)^{m/2}\lambda^{n-m}\binom{n}{m}(m-1)!!\right]p_{n-m}(x) \ . \end{aligned} \tag{0.43}$$

Now note that

$$(\lambda^n - 1)p_n(x) + (1-\lambda^2)\lambda^{n-2}\frac{n(n-1)}{2}p_{n-2}(x) = \left[np_n(x) - n(n-1)p_{n-2}(x)\right](\lambda - 1) + \mathcal{O}(\lambda-1)^2) \ .$$

Therefore, for all $q \in [1,\infty)$

$$\|S_\lambda p_n - p_n - [np_n - n(n-1)p_{n-2}](\lambda-1)\|_q = \mathcal{O}(\lambda-1)^2 \ .$$

This proves that $\lambda \mapsto S_\lambda p_n$ is left differentiable in L^q at $\lambda = 1$, and that

$$\lim_{\lambda\to 1}\frac{1}{\lambda - 1}(S_\lambda p_n - p_n) = np_n - n(n-1)p_{n-2} \ .$$

Since subsequences of L^q convergent sequences converge almost everywhere, and since the limiting function is continuous, the convergence also holds point-wise everywhere. Now note that

$$np_n(x) - n(n-1)p_{n-2}(x) = \frac{\partial^2}{\partial x^2}p_n(x) - x\frac{\partial}{\partial x}p_n(x) \ .$$

By linearity, these formulae and conclusions extend to general $f \in \mathscr{P}$. □

For what follows, it will be convenient to reparameterize the map $\lambda \mapsto S_\lambda$.

0.20 DEFINITION (Mehler semigroup)**.** For each $t > 0$, define the operators $M_t \in \mathscr{B}(L^p)$ by $M_t = S_{e^{-t}}$, where $S_\lambda \in \mathscr{B}(L^p)$ is the extension by continuity of the operator on $\mathscr{P}$ defined in (0.30). Define $M_0 = I$. Then for $t > 0$ and continuous polynomially bounded $f \in \mathcal{H}$,

$$P_t f(x) = \int_{\mathbb{R}} f(e^{-t}x + \sqrt{1-e^{-2t}}y)\gamma(y)\mathrm{d}y \ . \tag{0.44}$$

Since the Hermite basis $\{h_n\}_{n\geq 0}$ consists of eigenvectors of S_λ, with $S_\lambda h_n = \lambda^n h_n$, it follows that for all $t > 0$ and all $n \geq 0$,

$$M_t h_n = e^{-nt} h_n \ . \tag{0.45}$$

Moreover, (0.39) becomes

$$M_s M_t = M_{s+t} \tag{0.46}$$

so that $\{M_t\}_{t\geq 0}$ has the semigroup property. By (0.40), for all $p \in [1,\infty)$, and all $f \in L^p$, $\lim_{t\to 0} \|M_t f - f\|_p = 0$, and then by the semigroup property, for all $t > s$ $\|M_t f - M_s f\|_p = \|M_{t-s} M_s f = M_s f\|_p$, so that $t \mapsto M_t f$ is continuous into L^p. Thus, the Mehler semigroup is a strongly continuous semigroup on L^p for all $p \in [1,\infty)$.

In this notation, the differentiability formulae of Lemma 0.21 become

$$\lim_{h\downarrow 0} \frac{M_h f - f}{h} = -\mathscr{N} f \tag{0.47}$$

for all $f \in \mathscr{N}$, which is valid in L^q, $q \in [1,\infty)$, and point-wise. Since $\mathscr{P}$ is invariant under M, this and the semigroup property yield

$$\lim_{h\to 0} \frac{M_{t+h} f - M_t f}{h} = -\mathscr{N} f \tag{0.48}$$

for all $t > 0$, again in L^q for all $q \in [1,\infty)$ and point-wise. This proves:

0.21 LEMMA. *For all $f \in \mathscr{P}$, $t > 0$ and $x \in \mathbb{R}$, define $f(t,x) = M_t f(x)$. Then*

$$\frac{\partial}{\partial t} f(t,x) = \frac{\partial^2}{\partial x^2} f(t,x) - x\frac{\partial}{\partial x} f(t,x) \ . \tag{0.49}$$

The equation (0.21) is known as the *Fokker-Planck equation.* It can be written in terms of the operator $\mathscr{N}$ defined in (0.49) as $\dfrac{\partial}{\partial t} f(t,x) = -\mathscr{N} f(t,x)$. The operator $\mathscr{N}$ is known as the *number operator* for reasons that we now explain.

By (0.48), for all $f \in \mathscr{P}$, $\dfrac{\mathrm{d}}{\mathrm{d}t} M_t f\big|_{t=0} = -\mathscr{N} f$. Specializing to $f = h_n$ and using (0.45), we obtain

$$\mathscr{N} h_n = n h_n \ . \tag{0.50}$$

That is, the Hermite basis $\{h_n\}_{n\geq 0}$ is an orthonormal basis if $\mathcal{H}$ consisting of eigenfunctions of $\mathscr{N}$.

0.22 LEMMA. *For all $f, g \in \mathscr{P}$,*

$$\langle f, \mathscr{N} g\rangle_{\mathcal{H}} = \left\langle \tfrac{\mathrm{d}}{\mathrm{d}x} f, \tfrac{\mathrm{d}}{\mathrm{d}x} g\right\rangle_{\mathcal{H}} \tag{0.51}$$

and

$$\langle f, \tfrac{\mathrm{d}}{\mathrm{d}x} g\rangle_{\mathcal{H}} = \left\langle \left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) f, g\right\rangle_{\mathcal{H}} \tag{0.52}$$

Proof. This is simply an integration by parts calculation. □

0.23 THEOREM. *Define $h_{-1} = 0$. Then for all $n \geq 0$, and all $x \in \mathbb{R}$,*

$$\frac{\mathrm{d}}{\mathrm{d}x} h_n(x) = \sqrt{n} h_{n-1}(x) \qquad \text{and} \qquad \left(x - \frac{\mathrm{d}}{\mathrm{d}x}\right) h_n(x) = \sqrt{n+1} h_{n+1}(x) \ . \tag{0.53}$$

Moroever,

$$x h_n(x) = \sqrt{n} h_{n-1}(x) + \sqrt{n+1} h_{n+1}(x) \ . \tag{0.54}$$

Proof. Since $\frac{\mathrm{d}}{\mathrm{d}x} h_n$ is a polynomial of degree $n-1$, it may be expanded in the Hermite basis using only $\{h_0, \dots, h_{m-1}\}$:

$$\frac{\mathrm{d}}{\mathrm{d}x} h_n = \sum_{m=1}^{n-1} \langle h_m, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n \rangle_{\mathcal{H}} h_m \ . \tag{0.55}$$

By (0.52), for $m < n-1$,

$$\langle h_m, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n \rangle_{\mathcal{H}} = \left\langle \left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) h_m, h_n \right\rangle_{\mathcal{H}} = 0$$

since when $m < n-1$, $\left(x - \frac{\mathrm{d}}{\mathrm{d}x}\right) h_m$ is a polynomial of degree at most $n-1$, and is therefore orthogonal to h_n. Hence (0.56) simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}x} h_n = \langle h_{n-1}, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n \rangle_{\mathcal{H}} h_{n-1} \ . \tag{0.56}$$

Combining this with (0.51),

$$n = \langle h_n, \mathscr{N} h_n \rangle_{\mathcal{H}} = \left\langle \tfrac{\mathrm{d}}{\mathrm{d}x} h_n, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n \right\rangle_{\mathcal{H}} = \left| \langle h_{n-1}, \tfrac{\mathrm{d}}{\mathrm{d}x} h_n \rangle_{\mathcal{H}} \right|^2 \ . \tag{0.57}$$

Since the leading coefficient of h_n is positive for all n, combining (0.56) and (0.57) yields the first identity in (0.53).

It now follows from (0.52) that for all $m, n \geq 0$,

$$\sqrt{m} \delta_{n,m-1} = \langle h_n, \tfrac{\mathrm{d}}{\mathrm{d}x} h_m \rangle_{\mathcal{H}} = \left\langle \left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) h_n, h_m \right\rangle_{\mathcal{H}} \tag{0.58}$$

and then by the completeness of the Hermite polynomials, this means that for all $n \geq 0$,

$$\left(x - \tfrac{\mathrm{d}}{\mathrm{d}x}\right) h_n = \sqrt{n+1} h_{n+1} \ .$$

Combining the identities in (0.53), we obtain (0.54). □

0.6 The Fourier Transform

The map $U : f \mapsto \sqrt{\gamma} f$ is evidently a unitary transformation from $L^2(\mathbb{R}, \mathcal{B}, \nu)$ to $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. The image under U of the Hermite polynomial basis $\{h_n\}_{n \geq 0}$ for $L^2(\mathbb{R}, \mathcal{B}, \nu)$ is evidently an orthonormal basis for $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$

0.24 DEFINITION (Hermite functions)**.** For integers $n \geq 0$, define $g_n = \sqrt{\gamma} h_n$. Then g_n is the nth *Hermite function* and $\{g_n\}_{n \geq 0}$ is the *Hermite function basis* of $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$.

0.25 LEMMA. *For all $n \geq 0$,*

$$\frac{\mathrm{d}}{\mathrm{d}x} g_n(x) = \frac{1}{2}\left(\sqrt{n} g_{n-1} - \sqrt{n+1} g_{n+1}\right) \tag{0.59}$$

and

$$\frac{x}{2} g_n(x) = \frac{1}{2}\left(\sqrt{n} g_{n-1} + \sqrt{n+1} g_{n+1}\right) \tag{0.60}$$

Proof. We compute using (0.53) and (0.54):

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} g_n(x) &= \sqrt{\gamma(x)} \frac{\mathrm{d}}{\mathrm{d}x} h_n(x) + \left(\frac{\mathrm{d}}{\mathrm{d}x}\sqrt{\gamma(x)}\right) h_n(x) \\
&= \sqrt{\gamma(x)}\sqrt{n} h_{n-1}(x) - \frac{x}{2}\sqrt{\gamma(x)} h_n(x) \\
&= \sqrt{\gamma(x)}\sqrt{n} h_{n-1}(x) - \frac{1}{2}\sqrt{\gamma(x)}\left(\sqrt{n} h_{n-1}(x) + \sqrt{n+1} h_{n+1}(x)\right) \ ,
\end{aligned}$$

and this proves (0.59). Even more simply, (0.60) follows from (0.54). □

0.26 DEFINITION (Fourier Transform on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$)**.** Define a unitary transformation $\mathcal{F}$ on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ by and

$$\mathcal{F} g_n = (-i)^n g_n \tag{0.61}$$

for all $n \geq 0$. This is the *Fourier transform on* $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$

In what follows, we write $\dfrac{x}{2}$ to denote the operator of multiplication by $x/2$. Then by the definition of $\mathcal{F}$ and Lemma 0.25, for all $n \geq 0$

$$\frac{x}{2}\mathcal{F} g_n = (-i)^n \frac{1}{2}\left(\sqrt{n} g_{n-1} + \sqrt{n+1} g_{n+1}\right) = \frac{1}{2}\left(\sqrt{n}(-i)i\mathcal{F} g_{n-1} - \sqrt{n+1}(-i)i\mathcal{F} g_{n+1}\right) = -i\mathcal{F}\frac{\mathrm{d}}{\mathrm{d}x} g_n \ .$$

In other words,

$$\mathcal{F}^* \circ \frac{x}{2} \circ \mathcal{F} g_n = \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} g_n \qquad \text{and} \qquad \mathcal{F} \circ \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} \circ \mathcal{F}^* g_n = \frac{x}{2} g_n \ . \tag{0.62}$$

This identity is the primary source of the utility of the Fourier transform; it *diagonalizes differentiation* in the sense that it unitarily identifies differentiation with a multiplication operator. The formulae (0.62) would be more useful as a pair of identities between operators. To reformulate them as such, we have to take into account that neither multiplication by $x/2$ nor differentiation are defined on all of $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Therefore, we introduce a Hilbert space on which they are defined.

0.27 DEFINITION. The Hilbert space $\mathcal{H}_1$ consists of the functions $f \in Ł^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ having the Hermite function expansion

$$f = \sum_{n=0}^{\infty} \alpha_n g_n$$

such that $\sum_{n=0}^{\infty}(n+1)|\alpha_n|^2 < \infty$. For f, g in $\mathcal{H}_1$, the $\mathcal{H}_1$ inner product is defined by

$$\langle f, g\rangle_{\mathcal{H}_1} = \sum_{n=0}^{\infty}(n+1)\overline{\alpha_n}\beta_n \ ,$$

where $\alpha_n = \langle g_n, f\rangle_{L^2}$ and $\beta_n = \langle g_n, g\rangle_{L^2}$. We denote the norm on $\mathcal{H}_1$ by $\|\cdot\|_{\mathcal{H}_1}$. Observe that $\mathcal{H}_1$, considered as a subspace of L^2, is dense in L^2.

Consider any finite linear combination $f = \sum_{n=0}^{m} \alpha_n g_n$ for Hermite functions. Then

$$\frac{\mathrm{d}}{\mathrm{d}x} f = \frac{1}{2} \sum_{n=0}^{m} \alpha_n \left(\sqrt{n} g_{n-1} - \sqrt{n+1} g_{n+1} \right)$$

Now a simple computation shows that $\left\| \frac{\mathrm{d}}{\mathrm{d}x} f \right\|_{L^2} \leq \|f\|_{\mathcal{H}_1}$. Likewise, $\left\| \frac{x}{2} f \right\|_{L^2} \leq \|f\|_{\mathcal{H}_1}$. Therefore, both $\frac{\mathrm{d}}{\mathrm{d}x}$ and $\frac{x}{2}$ belong to $\mathscr{B}(\mathcal{H}_1, L^2)$. Since $\{(n+1)^{-1/2} g_n\}_{n \geq 0}$ is an orthonormal basis for $\mathcal{H}_1$, $\mathcal{F}$ is unitary on $\mathcal{H}_1$ as well as on L^2. Therefore we have:

0.28 THEOREM. *The operators* $\frac{\mathrm{d}}{\mathrm{d}x}$ *and* $\frac{x}{2}$ *in* $\mathscr{B}(\mathcal{H}_1, L^2)$ *are related through the Fourier transform* $\mathcal{F}$ *by*

$$\mathcal{F}^* \circ \frac{x}{2} \circ \mathcal{F} = \frac{1}{i} \frac{\mathrm{d}}{\mathrm{d}x} \qquad \text{and} \qquad \mathcal{F} \circ \frac{1}{i} \frac{\mathrm{d}}{\mathrm{d}x} \circ \mathcal{F}^* = \frac{x}{2} \ . \tag{0.63}$$

The Fourier transform was originally introduced in a rather different way. We now explain the connection with the original definition. As we have seen, for each $t > 0$, M_t is a compact operator on $L^2(\mathbb{R}, \mathcal{B}, \nu)$ with the eigenfunction expansion

$$M_t = \sum_{n=1}^{\infty} e^{-tn} |h_n\rangle\langle h_n| \ . \tag{0.64}$$

Let D be the closed unit disk in $\mathbb{C}$: $D := \{z \in \mathbb{C} \ : \ |z| \leq 1\}$. For $z \in D$ we define

$$S_z = \sum_{n=0}^{\infty} z^n |h_n\rangle\langle h_n| \ . \tag{0.65}$$

Note that with $z = e^{-t}$, (0.65) reduces to (0.64). For $z \in D^o$, the series in (0.65) converges in the operator norm, while for $z = e^{i\theta} \in \partial D$, the series converges in the strong operator topology and $S_{e^{i\theta}}$ is unitary.

This family of operators can be transplanted from $L^2(\mathbb{R}, \mathcal{B}, \nu)$ to $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ using the unitary transformation $U : f \mapsto \sqrt{\gamma} f$. For $z \in D$ we define $T_z = U M_z U^*$, which is the same as

$$T_z = \sum_{n=1}^{\infty} z^n |g_n\rangle\langle g_n| \ . \tag{0.66}$$

As before, the series in (0.66) converges in the operator norm, while for $z = e^{i\theta} \in \partial D$, the series converges in the strong operator topology and $T_{e^{i\theta}}$ is unitary. Now note that $\mathcal{F} = T_i$.

0.29 LEMMA. *The map* $z \mapsto T_z$ *is continuous from* D^o *into* $\mathscr{B}(L^2)$ *with the norm topology on the range, and it is continuous from* D *into* $\mathscr{B}(L^2)$ *with the strong operator topology on the range.*

Proof. Consider $z, w \in D^o$, Then for some $r \in (0,1)$, $|z|, |w| \leq r$. Since $\| |g_b\rangle\langle g_n| \| = 1$, Minkowski's inequality and the identity

$$z^n - w^n = (z - w) \sum_{m=0}^{n-1} z^{n-1-m} w^m \tag{0.67}$$

gives us

$$\|T_z - T_w\| \le \sum_{n=0}^{\infty} |z^n - w^n| \le |z-w| \left(\sum_{n=0}^{\infty} n r^n \right) = |z-w| \frac{r}{(1-r)^2} \ .$$

This proves the norm continuity into $\mathscr{B}(L^2)$.

Next, fix $f \in L^2$, and for $n \ge 0$, define $\alpha_n = \langle g_n, f \rangle_{L^2}$ so that $\|f\|_2^2 = \sum_{n=0}^{\infty} |\alpha_n|^2$. Fix $z \in D$, which we may as well assume to belong to ∂D. We must show that for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $w \in D$ with $|w - z| < \delta$, $\|T_w f - T_z f\|_2 < \epsilon$. Pick $\epsilon > 0$, and then N such that $\sum_{n=N+1}^{\infty} |\alpha_n|^2 < \epsilon/4$. Then, using (0.67) once more in the last line

$$\begin{aligned} \|T_w f - T_z f\|_2^2 &\le \sum_{n=1}^{N} |z^n - w^n||\alpha_n|^2 + \sum_{n=N+1}^{\infty} |z^n - w^n||\alpha_n|^2 \\ &\le \|f\|_2^2 \sum_{n=1}^{N} |z^n - w^n| + 2 \sum_{n=N+1}^{\infty} |\alpha_n|^2 \\ &\le |z-w|\|f\|_2^2 \sum_{n=1}^{N} (n-1) + \frac{\epsilon}{2} \ . \end{aligned}$$

Thus we may tale $\delta = \|f\|_2^2 N(N-1)\epsilon/4$.

□

For $z = \lambda \in (0,1)$, the operator S_z has the representation (0.31), in terms of the kernel $K_\lambda(x,y)$ defined there, and consequently, for any $f, g \in L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ and $\lambda \in (0,1)$,

$$\langle f, T_\lambda g \rangle_{L^2(\mathrm{d}x)} = \int_{\mathbb{R}} \int_{\mathbb{R}} \overline{f(x)} \sqrt{\gamma}(x) K_\lambda(x,y) \sqrt{\gamma}(y) g(y) \mathrm{d}x\mathrm{d}y = \int_{\mathbb{R}} \int_{\mathbb{R}} \overline{f(x)} M_\lambda(x,y) g(y) \mathrm{d}x\mathrm{d}y$$

where

$$M_\lambda(x,y) := \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\lambda^2}} \exp\left[-\frac{1+\lambda^2}{4(1-\lambda^2)} \left(y^2 + x^2\right) + \frac{\lambda}{1-\lambda^2} yx \right] \ . \tag{0.68}$$

The from (0.66) we obtain the identity

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\lambda^2}} \exp\left[-\frac{1+\lambda^2}{4(1-\lambda^2)} \left(y^2 + x^2\right) + \frac{\lambda}{1-\lambda^2} yx \right] = \sum_{n=0}^{\infty} \lambda^n g_n(x) g_n(y) \ . \tag{0.69}$$

Since $z \mapsto \sum_{n=0}^{\infty} z^n g_n(x) g_n(y)$ is evidently analytic from D^o into $L^2(\mathbb{R}^2, \mathcal{B}, \mathrm{d}x\mathrm{d}y)$, we may replace $\lambda \in (0,1)$ by $z \in D^o$ in (0.69) to define $M_z(x,y)$, and then with this definition, for all $f \in L^2$, $z \in D$,

$$T_z f(x) = \int_{\mathbb{R}} M_z(x,y) f(y) \mathrm{d}x\mathrm{d}y \ . \tag{0.70}$$

By the previous lemma, $T_{-i} f = \lim_{z \to -i} T_z f$ in the norm topology on L^2, Hence if we let $\{z_n\}_{n \in \mathbb{N}}$ be any sequence in D^o with $\lim_{n\to\infty} z_n = -i$, we have that

$$T_{-i} f(x) = \lim_{n \to \infty} \int_{\mathbb{R}} M_{z_n}(x,y) f(y) \mathrm{d}x\mathrm{d}y$$

Since $M_{-i}(x,y) = \dfrac{1}{2\sqrt{\pi}}e^{-ixy/2}$, if we formally take the limit under the integral sign, we obtain

$$\mathcal{F}f(x) = T_t f(x) = \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-ixy/2}f(y)\mathrm{d}y\ ,$$

If $f \in L^1 \cap L^2$, there is nothing formal: We may take the limit $n \to \infty$ two ways. Using the fact that $f \in L^2$, we have the L^2 norm convergence of $T_{z_n}f$ to $T_i f$, and hence we have convergence almost everywhere along a subsequence. Next, since $f \in L^1$, and $|M_z(x,y)|$ is uniformly bounded by a constant for z in a neighborhood of i, we may apply the Lebesgue Dominated Convergence Theorem to conclude the point-wise convergence

$$\lim_{z\to -i} T_z f(x) = \lim_{z\to -i}\int_{\mathbb{R}} M_{z_n}(x,y)f(y)\mathrm{d}x\mathrm{d}y = \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-ixy/2}f(y)\mathrm{d}y\ .$$

We have proved:

0.30 THEOREM. *For all $f \in L^1(\mathbb{R},\mathcal{B},\mathrm{d}x)\cap L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$, the Fourier transform $\mathcal{F}f$ of f is given by*

$$\mathcal{F}f(x) = \frac{1}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-ixy/2}f(y)\mathrm{d}y \tag{0.71}$$

There is a one-parameter family of unitary transformations on $L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$ by composing $\mathcal{F}$ with a unitary scale transformation: For λ^0, and $f \in L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$, define

$$\delta_\lambda f(x) = \sqrt{\lambda}f(\lambda x)\ . \tag{0.72}$$

Then δ_λ is unitary, and therefore so it $\delta_\lambda\mathcal{F}$, and we have

$$\delta_\lambda\mathcal{F}f(x) = \frac{\sqrt{\lambda}}{2\sqrt{\pi}}\int_{\mathbb{R}} e^{-i\lambda xy/2}f(y)\mathrm{d}y \tag{0.73}$$

The choices $\lambda = 2$ and $\lambda = 4\pi$ are traditional in various fields, and the resulting operators are also known as the Fourier transform. The latter has many advantages; For $f \in L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$, define $\widehat{f}$ by $\widehat{f} := \delta_{4\pi}\mathcal{F}f$. Then for $f \in L^1(\mathbb{R},\mathcal{B},\mathrm{d}x)\cap L^2(\mathbb{R},\mathcal{B},\mathrm{d}x)$

$$\widehat{f}(x) = \int_{\mathbb{R}} e^{-i2\pi xy}f(y)\mathrm{d}y\ . \tag{0.74}$$

0.31 Remark. We could have arrived directly at (0.74) had we used a different Gaussian density γ_t to define our reference measure ν, and thus the Hermite polynomials. The choice $t = 1$ for the density of ν makes the variance of ν equal to one, so that ν is what probabilists call the "normal distribution" on $\mathbb{R}$. This "normal" choice gives the "normal" formulas for the "normal" Hermite polynomials, though of course other Hermite polynomials can be defined as the orthonormal sequence in $L^2(\mathbb{R},\mathcal{B},\gamma_t(x)\mathrm{d}x)$ that one obtains by applying the Gram-Schmidt algorithm to the sequence of non-negative powers of x. Then one obtains identities analogous to those found in Theorem 0.23, which are the basis of the key properties of the Fourier transform given in Theorem 0.28. Had we use $\gamma_t e^{-x^2/2t}\mathrm{d}x$ for $t = \frac{1}{2\pi}$ as our reference measure; that is, $e^{-\pi|x^2}\mathrm{d}x$, we would have arrived directly at (0.74) as the definition of the Fourier transform.

Partly for the reasons explained above, it is useful to record the simple interaction of $\mathcal{F}$ with several other one parameter families of unitary operations on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$. Another is the unitary group of translation τ_y, $y \in \mathbb{R}$ where $\tau_y f(x) = f(x-y)$. Another is the group of *phase transformations* ϕ_y given

$$\phi_y f(x) = e^{-ixy} f(x) \ . \tag{0.75}$$

In the formulation of the following lemma, we will use the physics convention of writing k for the variable that is the argument of $\mathcal{F}f$. *In this context, k ranges over $\mathbb{R}$ and not $\mathbb{N}$ or $\mathbb{Z}$.*

0.32 LEMMA. *For $\lambda > 0$, let δ_λ be given by (0.72). Then*

$$\delta_\lambda \circ \mathcal{F} = \mathcal{F} \circ \delta_{\lambda^{-1}} \ . \tag{0.76}$$

Moreover for all $y \in \mathbb{R}$, let τ_y be the translation operator on $L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$, $\tau_y f(x) = f(x-y)$, and let ϕ_y be the phase transformation defined in (0.75). Then

$$\mathcal{F} \circ \tau_y = \phi_{y/2} \circ \mathcal{F} \ . \tag{0.77}$$

Proof. For $f \in L^1(\mathbb{R}, \mathcal{B}, \mathrm{d}x) \cap L^2(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$, making the change of variables $x = u/\lambda$,

$$\delta_\lambda \mathcal{F} f(x) = \frac{\sqrt{\lambda}}{2\sqrt{\pi}} \int_{\mathbb{R}} e^{-i\lambda kx/2} f(y)\mathrm{d}x = \delta_\lambda \mathcal{F} f(k) = \frac{\sqrt{\lambda}}{2\sqrt{\pi}} \int_{\mathbb{R}} e^{-i\lambda k(u/\lambda)/2} f(u/\lambda)\frac{1}{\lambda}\mathrm{d}u = \mathcal{F}\delta_{\lambda^{-1}} f(k) \ .$$

Likewise, making the change of variable $u = x - y$,

$$\mathcal{F}\tau_y f(k) = \frac{1}{2\sqrt{\pi}} \int_{\mathbb{R}} e^{-ikx/2} f(x-y)\mathrm{d}x = \frac{1}{2\sqrt{\pi}} \int_{\mathbb{R}} e^{-ik(u+y)/2} f(u)\mathrm{d}u = e^{-iky/2}\mathcal{F}f(k) = \phi_{y/2}\mathcal{F}f(k) \ .$$

□

Another way to rephrase (0.77) is

$$\mathcal{F} \circ \tau_y \circ F^* = \phi_{y/2} \ , \tag{0.78}$$

so that $\mathcal{F}$ "diagonalizes" translation, in the sense that it relates translation to a unitary "multiplication operator" (multiplication by $e^{-iky/2}$) as specified in (0.78).

0.33 THEOREM. *For any function $h \in L^1 \cap L^\infty$ define*

$$\widehat{f}(k) = \int_{\mathbb{R}} e^{-i2\pi kx} f(x)\mathrm{d}x \ . \tag{0.79}$$

Then for $f, g \in L^1 \cap L^\infty$,

$$\widehat{f * g}(k) = \widehat{f}(k)\widehat{g}(k) \ . \tag{0.80}$$

Proof. We compute

$$\begin{aligned}\widehat{f * g}(k) = \int_{\mathbb{R}} e^{-i2\pi kx} \left(\int_{\mathbb{R}} f(x-y)g(y)\mathrm{d}y \right) \mathrm{d}x \\ = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-i2\pi k(x-y)} f(x-y) e^{-i2\pi ky} g(y)\mathrm{d}x\mathrm{d}y = \widehat{f}(k)\widehat{g}(k) \ .\end{aligned}$$

□

Since $\gamma_t(x)e^{i2\pi kx} = \frac{1}{\sqrt{2\pi t}}\exp(-\frac{1}{2t}(x-2\pi ikt)^2)\exp(2\pi^2k^2t)$

$$\widehat{\gamma_t}(k) = \exp(2\pi^2k^2t) \ . \tag{0.81}$$

In particular, taking $t = (2\pi)^{-1}$ so that $\gamma_{1/2pi}(x) = e^{-\pi x^2}$,

$$\widehat{\gamma_{1/2\pi}}(k) = \gamma_{1/2\pi}(k) \ . \tag{0.82}$$

That is, $\gamma_{1/2\pi}$ is a fixed point of the map $f \mapsto \widehat{f}$, as we have seen earlier.

For $f \in L^1 \cap L^2$, recall that $P_tf(x) = \gamma_t * f * x)$, where $\{P_t\}_{t\geq 0}$ is the heat semigroup. Then by Theorem 0.33,

$$\widehat{P_tf}(k) = \widehat{\gamma_t}(k)\widehat{f}(k) = \exp(2\pi^2k^2t)\widehat{f}(k) \ .$$

Since $L^1 \cap L^2$ is dense in L^2, the formula $\widehat{P_tf}(k) = \exp(2\pi^2k^2t)\widehat{f}(k)$ extends by continuity to all $f \in L^2$. Since $f \mapsto \widehat{f}$ is unitary and P_t is contractive on L^2, $P_tf = 0$ implies that $\widehat{P_tf}(k) = 0$ for almost every k, and then the identity above shows that $\widehat{f}(k) = 0$ for almost every k. By the unitarity of $f \mapsto \widehat{f}$, this means that $f = 0$. This gives another proof of the infectivity of the heat semigroup operators on L^2, however, it relies on the unitarity of $f \mapsto \widehat{f}$, and in the approach to this taken her, we have relied on this infectivity to prove the unitarity of $f \mapsto \widehat{f}$, but there are other approaches in which the order can be inverted.

0.7 Higher dimensions

So far, we have discussed the Fourier transform and convolution for functions on $\mathbb{R}$. It is easy to extend our results to functions on $\mathbb{R}^n$. In this section, $L^p(\mathbb{R}^n)$ denotes $L^p(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x)$.

We begin with the Fourier transform and $n = 2$. For integers $n, m \geq 0$ define

$$g_m \otimes g_m(x_1, x_2) = g_m(x_1)g_n(x_2) \ . \tag{0.83}$$

It is then evident that $\{g_m \otimes g_m\}_{m,n\geq 0}$ is an orthonormal basis of $L^2(\mathbb{R}^2)$.

For a function $f \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ define

$$\widehat{f}(k) = \int_{\mathbb{R}^2} e^{-i2\pi k\cdot x}f(x)\mathrm{d}x \ , \tag{0.84}$$

where $k \cdot x$ denotes the standard inner product in $\mathbb{R}^2$, so that $k \cdot x = k_1x_1 + k_2x_2$ and

$$e^{-i2\pi k\cdot x} = e^{-i2\pi k_1x_1}e^{-i2\pi k_2x_2} \ .$$

It follows that

$$\begin{aligned} \widehat{g_m \otimes g_n}(k_1.k_2) &= \left(\int_{\mathbb{R}} e^{-i2\pi k_1x_1}g_m(x_1)\mathrm{d}x_1\right)\left(\int_{\mathbb{R}} e^{-i2\pi k_2x_2}g_n(x_2)\mathrm{d}x_2\right) \\ &= \widehat{g}_m(k_1)\widehat{g}_n(k_2) = (-i)^{m+n}g_m \otimes g_n(k_1.k_2) \ . \end{aligned}$$

Since $\{(-i)^{m+n}g_m \otimes g_m\}_{m,n\geq 0}$ is also an orthonormal basis of $L^2(\mathbb{R}^2)$, the map $f \mapsto \widehat{f}$ defined in (0.84) is unitary on $L^2(\mathbb{R}^n)$.

The same considerations extend in the obvious way to functions on $\mathbb{R}^n$ for arbitrary $n \in N$. For $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, we define

$$\widehat{f}(k) = \int_{\mathbb{R}^n} e^{-i2\pi k\cdot x} f(x)\mathrm{d}x \ , \tag{0.85}$$

and then $f \mapsto \widehat{f}$ extends by continuity to a unitary transformation on $L^2(\mathbb{R}^n)$. Even more simply,

$$|\widehat{f}(k)| \leq \int_{\mathbb{R}^n} |f(x)|\mathrm{d}x = \|f\|_1$$

so that $f \mapsto \widehat{f}$ extends by continuity to a contraction from $L^1(\mathbb{R}^n)$ to $L^\infty(\mathbb{R}^n)$. In summary, we have two operator norm bounds for the map $f \mapsto \widehat{f}$: It extends by continuity to an element of $\mathscr{B}(L^2(\mathbb{R}^n), L^2(\mathbb{R}^n))$ with operator norm 1, and it extends by continuity to an element of $\mathscr{B}(L^1(\mathbb{R}^n), L^\infty(\mathbb{R}^n))$ with operator norm 1

Likewise, if $f \in L^1(\mathbb{R}^n)$, $f \neq 0$, define $\rho(x) = \|f\|_1^{-1}|f(x)|$ so that ρ is a probability density on $\mathbb{R}^n$ and $K(x,y) := \rho(x-y)$ is a doubly stochastic Markov Kernel. It then follows directly from Theorem 0.2 that for all $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, that if we define

$$f * g(x) := \int_{\mathbb{R}^n} f(x-y)g(y)\mathrm{d}y = \|f\|_1 \int_{\mathbb{R}^n} K(x,y)g(y)\mathrm{d}y \ ,$$

then for all $p \in [1,\infty]$, then $\|f^* g\|_p \leq \|f\|_1\|g\|_p$. Even more simply if we assume that both f and g belong to $L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$, and therefore to $L^p(\mathbb{R}^n)$ for all $p \in [1,\infty]$, H "older's inequality gives us, for each x, and each p,

$$|f * g(x)| \leq \|f\|_{p/(p-1)}\|g\|_p \ .$$

We may rephrase the last two inequalities as a pair of operator bounds. For $g \in L^p(\mathbb{R}^n)$ define an operator T_g on $g \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ $T_g f = f * g$, Then we have $\|T_g f\|_p \leq \|g\|_p\|f\|_1$ and $\|T_g f\|_\infty \leq \|g\|_p\|f\|_{p/(p-1)}$. Therefore, T_g extends by continuity to an element of $\mathscr{B}(L^1(\mathbb{R}^n), L^p(\mathbb{R}^n))$ an its operator norm in this space is no more that $\|g\|_p$. Likewise, T_g extends by continuity to an element of $\mathscr{B}(L^{p/(p-1)}(\mathbb{R}^n), L^\infty(\mathbb{R}^n))$ an its operator norm in this space is no more that $\|g\|_p$.

Thus for both the Fourier transform and for convolution we have operator norm bounds in $\mathscr{B}(L^p(\mathbb{R}^n), L^q(\mathbb{R}^n))$ for two specific pair of values of p and q. In the next section we shall see how to "interpolate" between such pairs of inequalities producing a one parameter family of inequalities.

0.8 The Riesz-Thorin Interpolation Theorem

0.34 THEOREM. *Let $(\Omega, \mathcal{M}, \mu)$ be an arbitrary measure space, and for $p \in [1,\infty]$, let L^p denote $L^p(\Omega, \mathcal{M}, \mu)$. For $p_1, p_2 \in [1,\infty]$, The T be a linear operator defined on $L^{p_1} \cap L^{p_2}$. Suppose that for $q_1, q_2 \in [1,\infty]$ there exists finite constants C_1, C_2 such that for all $f \in L^{p_1} \cap L^{p_2}$,*

$$\|Tf\|_{q_1} \leq C_1\|f\|_{p_1} \quad \text{and} \quad \|Tf\|_{q_2} \leq C_1\|f\|_{p_2} \ . \tag{0.86}$$

Then for all $\lambda \in [0,1]$, T extends to a bounded linear transformation from $L^{p(\lambda)}$ to $L^{q(\lambda)}$ where

$$\frac{1}{p(\lambda)} = (1-\lambda)\frac{1}{p_1} + \lambda\frac{1}{p_2} \quad \text{and} \quad \frac{1}{q(\lambda)} = (1-\lambda)\frac{1}{q_1} + \lambda\frac{1}{q_2} \ , \tag{0.87}$$

and for all $f \in L^{p(\lambda)}$,

$$\|Tf\|_{q(\lambda)} \leq C_1^{1-\lambda}C_2^{\lambda}\|f\|_{p(\lambda)} \ . \tag{0.88}$$

This has a number of consequences; we give two of these in which the measure space is $(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x)$ before giving the proof.

0.35 THEOREM (Huassdoerff-Young-Titschmarsh)**.** *Define the Fourier transform $f \mapsto \widehat{f}$ on $L^1(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x) \cap L^2(\mathbb{R}^n, \mathcal{B}, \mathrm{d}x)$ by*

$$\widehat{f}(k) = \int_{\mathbb{R}^n} e^{-i2\pi k\cdot x} f(x)\mathrm{d}x \ . \tag{0.89}$$

Then for all $p \in [1, 2]$, $f \mapsto \widehat{f}$ extends to a contraction from $L^p(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ to $L^q(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ where $q = p/(p-1)$.

Proof. It is evident that $|\widehat{f}(k)| \leq \displaystyle\int_R |f(x)|\mathrm{d}x = \|f\|_1\mathrm{m}$ so that $\||\widehat{f}\|_\infty \leq \|f\|_1$. By the unitarity of the Fourier transform on L^2, $\|\widehat{f}\|_2 = \|f\|_2$. Therefore (0.86) is valid with $p_1 = 1$, $q_1 = \infty$, $C_1 = 1$, and $p_2 = q_2 = 1$, $C_2 = 1$.

For all $p \in (1, 2)$ write

$$\frac{1}{p} = \frac{1}{p(\lambda)} = (1-\lambda)1 + \lambda\frac{1}{2} = 1 + \frac{\lambda}{2} \quad \text{and hence} \quad \lambda = 2\left(1 - \frac{1}{p}\right) = \frac{2}{q} \ .$$

Then, with this value of λ, $1/q(\lambda) = (1-\lambda)(1/\infty) + \lambda(1/2) = 1/q$. Since $C_1 = C_2 = 1$, (**??**) becomes $\|\widehat{f}\|_q \leq \|f\|_p$. □

0.36 THEOREM (Young's Inequality for Convolution)**.** *For all $f, g \in L^1(\mathbb{R}, \mathcal{B}, \mathrm{d}x) \cap L^\infty(\mathbb{R}, \mathcal{B}, \mathrm{d}x)$ and all $r, s, t \in [1, \infty]$ such that*

$$\frac{1}{s} + \frac{1}{t} = 1 + \frac{1}{r} \ , \tag{0.90}$$

$$\|f * g\|_r \leq \|f\|_s \|g\|_t \ . \tag{0.91}$$

Proof. Fix $g \in L^1 \cap L^\infty$ and let T denote the operator on $L^1 \cap L^\infty$ defined by

$$Tf = g * f \ .$$

Then for $t \in (1, \infty)$, the inequality $\|g * f\|_t \leq \|f\|_1 \|g\|_t$ and be written as $\|Tf\|_t \leq \|g\|_p \|f\|_1$, while the inequality $\|g * f\|_\infty \leq \|f\|_{t/(t-1)} \|g\|_t$ can be written as $\|Tf\|_\infty \leq \|g\|_t \|f\|_{t/(t-1}$. Therefore (0.86) is valid with

$$p_1 = 1 \ , \ q_1 = t \ , \ C_1 = \|g\|_p \quad \text{and} \quad p_2 = t/(t-1) \ , \ q_2 = \infty \ , \ C_2 = \|g\|_t \ .$$

Note from (0.90) that $s \leq t/(t-1)$ with equality if and only if $r = \infty$. Therefore, $s \in [p_1, p_2]$. Define $\lambda \in [0, 1]$ so that

$$\frac{1}{s} = \frac{1}{p(\lambda)} = (1-\lambda) + \lambda\frac{t-1}{t} \quad \text{and hence} \quad 1 - \lambda = \frac{1}{t}\left(\frac{1}{s} + \frac{1}{t} - 1\right) \ .$$

Then for this λ,

$$\frac{1}{q(\lambda)} = (1-\lambda)\frac{1}{t} + \lambda\frac{1}{\infty} = \frac{1}{s} + \frac{1}{t} - 1 = \frac{1}{r} \ .$$

It follows from the Riesz-Thorin Theorem that T extends to map from L^s to L^r with norm $\|g\|_t$, and this yields (0.91). □

We now turn to the basic lemma upon which the Riesz-Thorin Theorem rests. For $x \in \mathbb{C}^n$, $p \in [1, \infty]$, let $\|x\|_p$ denote the ℓ^p norm on $\mathbb{C}^n$. Consider the bilinear form $A(x, y)$ on C^n given by

$$A(x, y) = \sum_{j,k=1}^{n} x_j a_{j,k} y_k \ .$$

For $1 \leq p, q \leq 1$, define

$$C(1/p, 1/q) = \sup \{|A(x, y)| \ : \ \|x\|_p = \|y\|_q = 1\} \ .$$

It will be convenient to write $\alpha = \dfrac{1}{p}$ and $\beta = \dfrac{1}{q}$. Notice then that $0 \leq \alpha, \beta \leq 1$.

0.37 LEMMA. *[Riesz–Thorin Interpolation Lemma] The function* $(\alpha, \beta) \mapsto \ln(C(\alpha, \beta))$ *is convex on* $[0, 1] \times [0, 1]$.

0.38 Remark. Marcel Riesz proved this in 1927 for $\alpha + \beta \geq 1$; i.e., in the upper right half of the unit square. His method was a direct assault on the maximization problem using Lagrange multipliers. This approach worked only in the restricted domain $\alpha + \beta \geq 1$, and he even conjectured that this wasn't die to his method, but that the result wasn't true in general except in this case. About a decade later, his student Thorin proved the full result by a completely different method. It is the lower–left half of the square that is of most interest to us.

Proof. Fix any two pairs of numbers (α_1, β_1) and (α_2, β_2) in $[0, 1] \times [0, 1]$. Let $p_1 = 1/\alpha_1$, $q_1 = 1/\beta_1$, $p_2 = 1/\alpha_2$ and $q_2 = 1/\beta_2$ with the obvious meaning if any of the denominators vanish. Next, any x with $\|x\|_{p_1} = 1$ can be written as

$$(x_1, x_2, \ldots, x_n) = (e^{i\phi_1} b_1^{\alpha_1}, e^{i\phi_2} b_2^{\alpha_1}, \ldots, e^{i\phi_1} b_n^{\alpha_1}) \ ,$$

where each $b_j \geq 0$ and $\sum_{j=1}^{n} b_j = 1$. Let $\mathcal{P}_n$ denote the set of all n–tuples $(d_1, d_2, \ldots, d_n)$ of non–negative numbers such that $\sum_{j=1}^{n} d_j = 1$, so that we may express the condition on the finite sequence $\{b_j\}$ as $\{b_j\} \in \mathcal{P}_n$. Similarly y with $\|y\|_{q_1} = 1$ can be written as

$$(y_1, y_2, \ldots, y_n) = (e^{i\psi_1} c_1^{\beta_1}, e^{i\psi_2} c_2^{\beta_1}, \ldots, e^{i\psi_1} c_n^{\beta_1}) \ ,$$

where $\{c_j\} \in \mathcal{P}_n$.

Then we can rewrite the definition of $C(\alpha_1, \beta_1)$ as

$$C(\alpha_1, \beta_1) = \sup \left\{ \sum_{j,k=1}^{n} a_{j,k} e^{i(\psi_k - \phi_j)} b_j^{\alpha_1} c_k^{\beta_1} \ : \ \{b_j\}, \{c_k\} \in \mathcal{P}_n, \{\phi_j\}, \{\psi_k\} \right\} \ .$$

It will be convenient to take the supremum in two parts, as follows:

$$C(\alpha_1, \beta_1) = \sup \{C(\alpha_1, \beta_1, \{\phi_j\}, \{\psi_k\}) \ : \ \{\phi_j\}, \{\psi_k\}\}$$

where

$$C(\alpha_1, \beta_1, \{\phi_j\}, \{\psi_k\}) = \sup \left\{ \sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{\alpha_1} c_k^{\beta_1} \ | \ \{b_j\}, \{c_k\} \in \mathcal{P}_n \right\} \quad \text{and} \quad \tilde{a}_{j,k} = a_{j,k} e^{i(\psi_k - \phi_j)} \ .$$

We can do the same for (α_2) and (β_2), and all we are changing is the exponents in the sum $\sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{\alpha} c_k^{\beta}$. In particular, we optimize over the *same* sets of sequences $\{b_j\}$, $\{c_k\}$, $\{\phi_j\}$ and $\{\psi_k\}$ in each case. Thorin's idea is now to interpolate between (α_1, β_1) and (α_2, β_2) with a *complex variable* z as follows: For all z belonging to the infinite strip given by

$$0 \leq \mathcal{R}(z) \leq 1$$

define

$$f(z) = \frac{\sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{(1-z)\alpha_1 + z\alpha_2} c_k^{(1-z)\beta_1 + z\beta_2}}{C(\alpha_1, \beta_1)^{1-z} C(\alpha_2, \beta_2)^z} \ .$$

Now for any positive number a, $a^z = e^{(\ln a)z}$ is an analytic function on the entire complex plane, and has no zeros. For this reason, $f(z)$ is an analytic function everywhere on its domain $0 \leq \mathcal{R}(z) \leq 1$. Also, on the left boundary of this domain; i.e., where $\mathcal{R}(z) = 0$,

$$|f(z)| = \frac{\left| \sum_{j,k=1}^{n} a_{j,k} e^{i(\tilde{\psi}_k - \tilde{\phi}_j)} b_j^{\alpha_1} c_k^{\beta_1} \right|}{C(\alpha_1, \beta_1)}$$

where the $\tilde{\phi}_j$ and $\tilde{\psi}_k$ include additional phases involving the imaginary part of z. By definition, the numerator is no more than $C(\alpha_1, \beta_1)$, and hence we arrive at the conclusion that $|f(z)| \leq 1$ for all z with $\mathcal{R}(z) = 0$. That is,

$$\sup_{y \in R} |f(iy)| \leq 1 \ .$$

In the exact same way, we see that $|f(z)| \leq 1$ for all z with $\mathcal{R}(z) = 1$. That is,

$$\sup_{y \in R} |f(1 + iy)| \leq 1 \ .$$

Now we are in a position to apply the *maximum modulus principle* of complex analysis to conclude that $|f(z)|$ is largest on the boundary of the strip $0 \leq \mathcal{R}(z) \leq 1$, and hence that $|f(z)| \leq 1$ for all such z.

We will come back to the maximum modulus principle argument shortly to make the proof self contained, but let us suppose for the moment that it has been established that $|f(z)| \leq 1$ whenever $0 \leq \mathcal{R}(z) \leq 1$. We are now done with the complex interpolation; take $z = \lambda$ with $0 < \lambda < 1$. Then $|f(\lambda)| \leq 1$ implies that

$$\left| \sum_{j,k=1}^{n} \tilde{a}_{j,k} b_j^{(1-\lambda)\alpha_1 + \lambda\alpha_2} c_k^{(1-\lambda)\beta_1 + \lambda\beta_2} \right| \leq C(\alpha_1, \beta_1)^{1-\lambda} C(\alpha_2, \beta_2)^{\lambda} \ .$$

Since this is true for each $\{b_j\}, \{c_k\} \in \mathcal{P}_n$, and each $\{\phi_j\}$ and $\{\psi_k\}$, we can take the supremum to arrive at

$$C((1-\lambda)\alpha_1 + \lambda\alpha_2, (1-t)\beta_1 + t\beta_2) \leq C(\alpha_1, \beta_1)^{1-\lambda} C(\alpha_2, \beta_2)^{\lambda}$$

and taking the natural logarithm, we obtain the stated convexity result.

We now return to the part of the argument involving the maximum modulus principle. First modify $f(z)$ by multiplying by $e^{z^2/n - 1/n}$. For $z = x + iy$ with $0 \leq x \leq 1$, the real part of z^2 is

$x^2 - y^2 \leq 1$, and so for all z with $0 \leq \mathcal{R}(z) \leq 1$, $|e^{z^2/n-1/n}| \leq 1$. In fact, the real part of z^2 becomes strongly negative for y large, and so $g(z) = f(z)e^{z^2/n-1/n}$ vanishes as z tends to infinity in $0 \leq \mathcal{R}(z) \leq 1$, and it is also clearly analytic there. In particular, it is continuous, and so there is a point z_0 in this domain so that

$$|g(z_0)| \geq |g(z)| \tag{0.92}$$

for all z with $0 \leq \mathcal{R}(z) \leq 1$. We claim that a point z_0 satisfying (0.92) lies on the boundary of the strip. To see this, consider *any* point z_0 satisfying 0.92). If it is not already on the boundary, appeal to the Cauchy integral formula

$$g(z_0) = \frac{1}{2\pi i}\int_0^{2\pi} g(z_0 + re^{i\phi})\mathrm{d}\phi$$

where r is the distance from z_0 to the boundary; i.e., $r = \min\{\mathcal{R}(z), 1 - \mathcal{R}(z)\}$. Then

$$|g(z_0)| \leq \frac{1}{2\pi}\int_0^{2\pi} |g(z_0 + re^{i\phi})|\mathrm{d}\phi \leq \frac{1}{2\pi}\int_0^{2\pi} |g(z_0)|\mathrm{d}\phi = |g(z_0)|$$

where in the first inequality we simply to absolute values, and in the second one we used (0.92) with $z = z_0 + re^{i\phi}$. The conclusion is that $|g(z_0 + re^{i\phi})| = |g(z_0)|$ for almost every ϕ, and then by continuity, for all ϕ. For either $\phi = 0$ or $\phi = \pi$, $z_0 + re^{i\phi}$ lies on the boundary, and the claim is proved. Hence $|g(z)|$ is maximized on the boundary, and so

$$|g(z)| = |f(z)e^{z^2/n-1/n}| \leq 1$$

for all $0 \leq \mathcal{R}(z) \leq 1$. Taking the limit in which n tends to infinity, we obtain the desired result. □

Proof of Theorem 0.34. Given any t with $0 < \lambda < 1$, and any simple function f, we we first observe that $f \in L^{p_1} \cap L^{p_2}$, and hence Tf is defined. By the variational characterization of the L^p norms,

$$\|T_f\|_{q(t)} = \sup\left\{\left|\int_{R^d} gTf\mathrm{d}\mu\right| \ : \ \|g\|_{q(t)'} = 1\right\}$$

where $1/q(t)' + 1/q(t) = 1$.

By the density of simple functions, for any $\epsilon > 0$, there is a simple function g with $\|g\|_{q(t)'} = 1$ so that

$$\|T_f\|_{q(t)} \leq \left|\int_{R^d} gTf\mathrm{d}\mu\right| + \epsilon\ .$$

Now by further partitioning the subsets on which the simple functions f and g take on their various values, we may write them in the form

$$f(x) = \sum_{j=1}^n \frac{f_j}{\mu(A_j)^{1/p(t)}} 1_{A_j}(x) \qquad \text{and} \qquad g(x) = \sum_{k=1}^n \frac{g_k}{\mu(A_k)^{1/q'(t)}} 1_{A_k}(x)$$

with the same n and the same family of *disjoint* measurable sets A_j. Notice that with the way the coefficients are defined,

$$\|f\|_{p(t)} = \left(\sum_{j=1}^n |f_j|^{p(t)}\right)^{1/p(t)} \qquad \text{and} \qquad \|g\|_{q'(t)} = \left(\sum_{k=1}^n |g_k|^{q'(t)}\right)^{1/q'(t)}$$

Then $\int_{R^d} gTf\,\mathrm{d}\mu = \sum_{j,k=1}^{n} a_{j,k} f_j g_k$ where $a_{j,k} = \int_{R^d} 1_{A_k} T 1_{A_j}\,\mathrm{d}\mu$. This reduces the Corollary to the Reisz–Thorin Lemma and, since $\epsilon > 0$ is arbitrary, the theorem is proved. $\square$