HYPERCONTRACTIVITY, THE ORNSTEIN-UHLENBECK PROCESS, AND GAUSSIAN CONCENTRATION INEQUALITIES

ADAM QUINN JAFFE

MATH 258: Harmonic Analysis

1. Introduction

The notion of hypercontractivity first arose in quantum field theory [3] and underwent much development in functional analysis and mathematical physics throughout the subsequent decade [4, 7]. While the concept had existed in some form or another for some time before this, E. Nelson is credited with casting this collection of ideas into the same setting [6]. It was later discovered that this concept is closely related to certain concentration of measure phenomena, and this has sparked a great interest in this concept from probabilists; the monograph [2] contains a comprehensive account of these developments.

In this note we explore the basic principles of hypercontractivity and its relationship with concentration of measure, focusing on the setting of Gaussian measures where explicit computations can be made. Our exposition mainly follows the accounts in [1] and [2, Chapter 5 and Appendix B], but we have filled in much of the technical detail. We also provide citations to other sources where helpful.

2. Hypercontractivity

In this section we introduce the basic notions that will be used in this note. We mainly follow the abstract setting of [7, Section 2A].

Let $(S, \mathcal{S})$ denote a measurable space, and write $\mathcal{M}(S)$ be the vector space of all measurable functions from Ω to $\mathbb{R}$. For a measure μ on $(S, \mathcal{S})$, write $L^p(\mu)$ for the subspace of $\mathcal{M}(S)$ consisting of all functions with finite p-th moment, and write $||\cdot||_{L^p(\mu)}$ for the usual norm that makes this into a Banach space. Throughout this note, we will always assume that $(S, \mathcal{S}, \mu)$ is a probability space.

Write $C(\mathbb{R}^n)$ for the space of continuous functions from $\mathbb{R}^n$ to $\mathbb{R}$. We include subscripts c, b, and 0, to denote the subspaces of such functions which are compactly-supported, bounded, and vanishing at infinity, respectively. We include the superscript k to denote the subspaces of such functions which are k-times continuously differentiable. We of course combine these notations, so that, for example, $C_c^2(\mathbb{R}^n)$ denotes the space of compactly-supported, twice continuously differerentiable functions. Often these spaces are endowed with a supremum norm (or a combination of supremum norms on some of its derivatives), but we will not do so in this note unless explicitly stated.

By a *semigroup* $\{P_t\}_{t\geq 0}$ on a Banach space $(\mathcal{B}, \|\cdot\|)$ we mean a family $\{P_t\}_{t\geq 0}$ of (possibly unbounded) linear maps satisying the composition rule $P_tP_s = P_{t+s}$

for all $s, t \geq 0$. A semigroup is called *strongly continuous* if for all $x \in \mathcal{B}$, we have $\|P_t x - x\| \to 0$ as $t \to 0$.

Assumptions. Suppose that $(S, \mathcal{S}, \mu)$ is a probability space and that $\{P_t\}_{t\geq 0}$ is a family of linear maps on $L^\infty(\mu)$ which extends to a family of linear maps on $L^p(\mu)$ for all $p \in [1, \infty]$ such that the following properties are satisfied:

(A1) For each $p \in [1, \infty]$, the family $\{P_t\}_{t\geq 0}$ is a semigroup on $(L^p(\mu), \|\cdot\|_{L^p(\mu)})$.
(A2) For each $p \in [1, \infty]$ and $t \geq 0$, the map $P_t : L^p(\mu) \to L^p(\mu)$ is self-adjoint.
(A3) For each $t \geq 0$, the map $P_t : L^2(\mu) \to L^2(\mu)$ is unitary.

When we say $\{P_t\}_{t\geq 0}$ *is a semigroup on* $(S, \mathcal{S}, \mu)$, it will be understood that it satisfies the assumptions above.

We will later specialize to the case of the transition semigroup of the Ornstein-Uhlenbeck process along with the standard Gaussian measure on finite-dimensional Euclidean space, and we will see that this pair satisfies the assumptions above. However, for now, we choose to deal with the abstract case.

Definition 2.1. A semigroup $\{P_t\}_{t\geq 0}$ on $(S, \mathcal{S}, \mu)$ is called *hypercontractive* if the following conditions hold:

(1) We have $\|P_t f\|_{L^1(\mu)} \leq \|f\|_{L^1(\mu)}$ for all $t \geq 0$ and $f \in L^1(\mu)$, and
(2) There exists some $T \geq 0$ and some $C > 0$ such that we have $\|P_T f\|_{L^4(\mu)} \leq C\|f\|_{L^2(\mu)}$ for all $f \in L^2(\mu)$.

This notion is rather strong: the hypothesis (1) is called "contractivity" for obvious reasons, and the hypothesis (2) guarantees that one can bootstrap a priori bounds on $L^2(\mu)$-norms (quantitatively!) to further bounds on $L^4(\mu)$ norms, hence the name "hypercontractivity". (Nelson himself suggests [6] that a better term for (2) would actually be "hyperboundedness" since the constant in need not be equal to 1.)

It turns out that hypercontractivity is equivalent to another set of conditions, which, a priori, appears to be much stronger.

Lemma 2.2. A semigroup $\{P_t\}_{t\geq 0}$ on $(S, \mathcal{S}, \mu)$ is hypercontractive if and only if the following conditions hold:

(1') For any $p \in [1, \infty]$, we have $\|P_t f\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)}$ for all $t \geq 0$ and $f \in L^p(\mu)$, and
(2') For any $p, q \in (1, \infty)$, there is some $T \geq 0$ and some $C > 0$ such that $\|P_t f\|_{L^q(\mu)} \leq C\|f\|_{L^p(\mu)}$ for all $t > T$ and $f \in L^p(\mu)$.

Proof. The "if" direction is immediate, so we only need to show the "only if" direction, that is, that $\{P_t\}_{t\geq 0}$ being hypercontractive implies (1') and (2'). For (1'), note that for any $t \geq 0$, and any $f \in L^\infty(\mu)$, the duality between the L^1 and L^∞ norms and assumption (A2) together imply:

$$\begin{aligned}
\|P_t f\|_{L^\infty(\mu)} &= \sup\left\{\int_S |P_t f g| d\mu : g \in L^1(\mu), \|g\|_{L^1(\mu)} \le 1\right\} \\
&= \sup\left\{\int_S P_t f g \, d\mu : g \in L^1(\mu), \|g\|_{L^1(\mu)} \le 1\right\} \\
&= \sup\left\{\int_S f P_t g \, d\mu : g \in L^1(\mu), \|g\|_{L^1(\mu)} \le 1\right\} \\
&= \sup\left\{\int_S |f P_t g| d\mu : g \in L^1(\mu), \|g\|_{L^1(\mu)} \le 1\right\} \\
&\le \sup\left\{\|f\|_{L^\infty(\mu)} \|P_t g\|_{L^1(\mu)} : g \in L^1(\mu), \|g\|_{L^1(\mu)} \le 1\right\} \\
&= \|f\|_{L^\infty(\mu)} \sup\left\{\|P_t g\|_{L^1(\mu)} : g \in L^1(\mu), \|g\|_{L^1(\mu)} \le 1\right\} \\
&\le \|f\|_{L^\infty(\mu)} \sup\left\{\|g\|_{L^1(\mu)} : g \in L^1(\mu), \|g\|_{L^1(\mu)} \le 1\right\} \\
&\le \|f\|_{L^\infty(\mu)}.
\end{aligned}$$

This shows that $P_t : L^\infty(\mu) \to L^\infty(\mu)$ is contractive. By the Riesz-Thorin theorem, this implies that $P_t : L^p(\mu) \to L^p(\mu)$ is contractive for all $p \in [1, \infty]$, as claimed.

For (2'), note by assumption that P_T is bounded from $L^2(\mu) \to L^4(\mu)$, and that it is bounded from $L^\infty(\mu)$ to $L^\infty(\mu)$ by (1'). Applying the Riesz-Thorin theorem gives that P_T is bounded from $L^m(\mu) \to L^{2m}(\mu)$ for all integers $m \ge 2$. Composing these maps shows that, for each $n \in \mathbb{N}$, the map $P_T : L^2(\mu) \to L^{2^n}(\mu)$ is bounded. Now let $p, q \in (1, \infty)$ be arbitrary, assume without loss of generality that $p < q$. Since $2^n/(2^n - 1) \downarrow 1$ and $2^n \to \infty$ as $n \to \infty$, we can choose some $n \in \mathbb{N}$ large enough such that we have

$$p_0 = \frac{2^n}{2^n - 1} \le p < q \le 2^n = q_0$$

Now observe that P_T is bounded from $L^2(\mu)$ to $L^{q_0}(\mu)$, hence $P_T^* = P_T$ is bounded from $L^{p_0}(\mu)$ to $L^2(\mu)$. Since P_T is unitary, this implies that for all $f \in L^p(\mu)$ we have

$$\begin{aligned}
\|P_T f\|_{L^q(\mu)} \le \|P_T f\|_{L^{q_0}(\mu)} &\lesssim \|f\|_{L^2(\mu)} \\
&= \|P_T f\|_{L^2(\mu)} \lesssim \|f\|_{L^{p_0}(\mu)} \le \|f\|_{L^p(\mu)}.
\end{aligned}$$

This shows that there exists some $C > 0$ such that $\|P_T f\|_{L^q(\mu)} \le C\|f\|_{L^p(\mu)}$ holds for all$f \in L^p(\mu)$. Then for any $t > T$ the semigroup property and (1') combine to give

$$\begin{aligned}
\|P_t f\|_{L^q(\mu)} &= \|P_{T+(t-T)} f\|_{L^q(\mu)} \\
&= \|P_{t-T} P_T f\|_{L^q(\mu)} = \|P_T f\|_{L^q(\mu)} \le C\|f\|_{L^p(\mu)}
\end{aligned}$$

as claimed. □

3. Background on Stochastic Processes

In this section we will develop some abstract theory about stochastic processes which is also of independent interest. We mainly follow [1, Chapter 1].

Suppose that S is a locally compact topological space with a countable basis and that $\mathcal{S}$ is its Borel σ-algebra, and let $\{X_t\}_{t\geq 0}$ be a right-continuous Markov process in S. That is, let Ω denote the space of all right-continuous paths from $[0,\infty)$ to S, and let $\mathcal{F}$ denote the Borel σ-algebra generated by Skorokhod topology on Ω. For each $x \in S$, the law of $\{X_t\}_{t\geq 0}$ conditioned on $X_0 = x$ determines a probability measure $\mathbb{P}_x$ on Ω, and the corresponding expectation operator is denoted $\mathbb{E}_x$.

Also suppose that μ is a Borel probability measure on S which is stationary for $\{X_t\}_{t\geq 0}$. For any $f \in L^\infty(\mu)$ and $t \geq 0$, note that we have

$$\int_S \mathbb{E}_x[|f(X_t)|]d\mu(x) = \int_S |f(x)|d\mu(x) \leq \|f\|_{L^\infty(\mu)} < \infty,$$

and hence that $\mathbb{E}_x[|f(X_t)|]$ is finite for μ-almost every $x \in S$. This proves that $\mathbb{E}_x[f(X_t)]$ is well-defined for μ-almost every $x \in S$; we define the *transition semigroup* $\{P_t\}_{t\geq 0}$ of $\{X_t\}_{t\geq 0}$ via $P_t f(x) = \mathbb{E}_x[f(X_t)]$, and it follows that $P_t : L^\infty(\mu) \to L^\infty(\mu)$ is a well-defined linear map.

Lemma 3.1. If μ is a reversible probability measure for $\{X_t\}_{t\geq 0}$, then the collection $\{P_t\}_{t\geq 0}$ satisfies the semigroup assumptions of Section 2.

Proof. For (A1), let $p \in [1,\infty]$ and $t \geq 0$ be arbitrary. Then take any $f \in L^\infty(S)$ and recall that μ being reversible for $\{X_t\}_{t\geq 0}$ implies that it is also stationary for $\{X_t\}_{t\geq 0}$. Then Jensen's inequality gives

$$\begin{aligned}
\|P_t f\|_{L^p(\mu)}^p &= \int_S |\mathbb{E}_x[f(X_t)]|^p d\mu(x) \\
&\leq \int_S \mathbb{E}_x\left[|f(X_t)|^p\right] d\mu(x) \\
&= \int_S |f(x)|^p d\mu(x) = \|f\|_{L^p(\mu)}^p,
\end{aligned}$$

which shows that $P_t : L^\infty(\mu) \to L^\infty(\mu)$ is bounded linear, with respect to the $L^p(\mu)$ norm. To extend this, note that $L^\infty(\mu)$ is dense in $L^p(\mu)$, so for $f \in L^p(\mu)$ we can get $\{f_n\}_{n\in\mathbb{N}}$ in $L^\infty(\mu)$ with $f_n \to f$ in $L^p(\mu)$. Then by the above and Fatou's lemma we have

$$\begin{aligned}
\|P_t f\|_{L^p(\mu)} &\leq \liminf_{n\to\infty} \|P_t f_n\|_{L^p(\mu)} \\
&\leq \liminf_{n\to\infty} \|f_n\|_{L^p(\mu)} = \|f\|_{L^p(\mu)},
\end{aligned}$$

which shows that P_t extends uniquely to a linear contraction $P_t : L^p(\mu) \to L^p(\mu)$. To see that the composition rule is also satisfied, note that for any $s,t \geq 0$ and $f \in L^\infty(\mu)$ we have, by the Markov property of $\{X_t\}_{t\geq 0}$:

$$P_t P_s f(x) = \mathbb{E}_x[\mathbb{E}_{X_s}[f(X_t)]] = \mathbb{E}_x[f(X_{t+s})] = P_{t+s} f(x).$$

This proves that we have $P_t P_s = P_{t+s}$ as operators from $L^\infty(\mu)$ to itself, and by density, also as operators from $L^p(\mu)$ to itself. Thus, $\{P_t\}_{t\geq 0}$ is indeed a semigroup on $L^p(\mu)$.

For (A2), let $f, g \in L^\infty(\mu)$ be arbitrary, and note that, since μ is reversible for $\{X_t\}_{t\geq 0}$, we have

$$\begin{aligned}
\int_S P_t f(x) g(x) d\mu(x) &= \int_S \mathbb{E}_x[f(X_t)g(X_0)] d\mu(x) \\
&= \int_S \mathbb{E}_x[f(X_0)g(X_t)] d\mu(x) \\
&= \int_S f(x) P_t g(x) d\mu(x).
\end{aligned}$$

Now let $p, q \in [1, \infty]$ be conjugate exponents, and take any $f \in L^p(\mu)$ and $g \in L^q(\mu)$. Get sequences $\{f_n\}_{n\in\mathbb{N}}$ in $L^\infty(\mu)$ with $f_n \to f$ in $L^p(\mu)$ and $\{g_n\}_{n\in\mathbb{N}}$ in $L^\infty(\mu)$ with $g_n \to g$ in $L^q(\mu)$. Then by Hölder's inequality, the contraction property of P_t, and the fact that self-adjointness already holds for elements of $L^\infty(\mu)$, we have

$$\begin{aligned}
&\left| \int_S P_t f(x) g(x) d\mu(x) - \int_S f(x) P_t g(x) d\mu(x) \right| \\
&\leq \int_S |P_t(f - f_n)(x) g(x)| d\mu(x) + \int_S |P_t f_n(x)(g - g_n)(x)| d\mu(x) \\
&\qquad + \int_S |f_n(x) P_t(g - g_n)(x)| d\mu(x) + \int_S |(f - f_n)(x) P_t g(x)| d\mu(x) \\
&\leq \|f - f_n\|_{L^p(\mu)} \|g\|_{L^q(\mu)} + \|f_n\|_{L^p(\mu)} \|g - g_n\|_{L^q(\mu)} \\
&\qquad + \|f_n\|_{L^p(\mu)} \|g - g_n\|_{L^q(\mu)} + \|f - f_n\|_{L^p(\mu)} \|g_n\|_{L^q(\mu)}.
\end{aligned}$$

Of course, the right side goes to zero as $n \to \infty$, so self-adjointness holds.

For (A3), for any $x \in S$, write $\tilde{\mathbb{P}}_x = \mathbb{P}_x \otimes \mathbb{P}_x$ for the product measure on Ω^2, and write $\tilde{\mathbb{E}}_x$ for its expectation operator. In other words, $\tilde{\mathbb{P}}_x$ is the joint law of two independent copies of the same process $\{X_t\}_{t\geq 0}$ and $\{Y\}_{t\geq 0}$, which are coupled to have the same starting point. Then for $f \in L^\infty(\mu)$ we have

$$\begin{aligned}
\int_S (P_t f(x))^2 d\mu(x) &= \int_S \mathbb{E}_x[f(X_t)] \mathbb{E}_x[f(Y_t)] d\mu(x) \\
&= \int_S \tilde{\mathbb{E}}_x[f(X_t) f(Y_t)] d\mu(x) \\
&= \int_S (f(x))^2 d\mu(x).
\end{aligned}$$

To extend this, we use a similar density argument as before. For $f \in L^2(\mu)$, get a sequence $\{f_n\}_{n\in\mathbb{N}}$ in $L^\infty(\mu)$ with $f_n \to f$ in $L^2(\mu)$. Then by Cauchy-Schwarz, the contraction property of P_t, and the fact the unitarity already holds for elements of $L^\infty(\mu)$, we have

$$\begin{aligned}
&\left|\int_S (P_t f(x))^2 d\mu(x) - \int_S (f(x))^2 d\mu(x)\right| \\
&\leq \int_S |P_t f(x) P_t(f-f_n)(x)| d\mu(x) + \int_S |P_t(f-f_n)(x) P_t f_n(x)| d\mu(x) \\
&\quad + \int_S |(f-f_n)(x) f_n(x)| d\mu(x) + \int_S |f(x)(f-f_n)(x)| d\mu(x) \\
&\leq 2\|f\|_{L^2(\mu)} \|f-f_n\|_{L^2(\mu)} + 2\|f_n\|_{L^2(\mu)} \|f-f_n\|_{L^2(\mu)}.
\end{aligned}$$

Since the right side goes to zero as $n \to \infty$, it follows that $P_t : L^2(\mu) \to L^2(\mu)$ is unitary, and the result is proved. □

Now return to the general setting that μ is stationary for $\{X_t\}_{t\geq 0}$. It is not hard to show that the semigroup $\{P_t\}_{t\geq 0}$ is strongly continuous on $L^2(\mu)$, and we can take this further by defining the set

$$D(\mathcal{L}) = \left\{ f \in L^2(\mu) : \lim_{t\to 0} \frac{1}{t}(P_t f - f) \text{ exists in } L^2(\mu) \right\},$$

and the operator $\mathcal{L} : D(\mathcal{L}) \to L^2(\mu)$ to be the value of the limit; we say that $\mathcal{L}$ is the $L^2(\mu)$-*generator* of $\{X_t\}_{t\geq 0}$ and that $D(\mathcal{L})$ is its *domain*. (This setting is slightly different from the more familiar setting of "Feller processes" in which $\{P_t\}_{t\geq 0}$ is assumed to be strongly continuous on a suitable space of continuous functions endowed with the supremum norm.) In this case we have that, for each $f \in D(\mathcal{L})$, the map $P_\cdot f : [0,\infty) \to L^2(\mu)$ is strongly differentiable, with $\partial_t P_t f = \mathcal{L} P_t f = P_t \mathcal{L} f$. These identities are referred to as the *Kolmogorov equations*.

The ideas above also give rise to an important bilinear form. For $f, g \in D(\mathcal{L})$, we define the value

$$\mathcal{E}(f,g) = -\int_{\mathbb{R}^n} f(x) \mathcal{L} g(x) d\mu(x).$$

We refer to $\mathcal{E}$ as the *Dirichlet form* of $\{X_t\}_{t\geq 0}$.

4. The Ornstein-Uhlenbeck Process

We will derive the Gaussian inequalities of the next section by studying a concrete stochastic process of interest. We rely somewhat on a background in stochastic calculus, as outlined in [5].

Fix $n \in \mathbb{N}$, and let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space satisfying Doob's "usual conditions" [5, Chapter 1, Definition 4.13] on which a standard n-dimensional Brownian motion $\{B_t\}_{t\geq 0}$ is defined. On this space, we let the standard *Ornstein-Uhlenbeck (OU) process* $\{X_t\}_{t\geq 0}$ be defined via

$$X_t = X_0 e^{-t} + \sqrt{2} e^{-t} \int_0^t e^s dB_s. \tag{1}$$

As before, write $\mathbb{P}_x$ for the probability measure $\mathbb{P}$ conditioned on the (possibly singular) event $X_0 = x$. We can also compute

$$\begin{aligned}
dX_t &= -X_0 e^{-t} dt - \sqrt{2} e^{-t} \int_0^t e^s dB_s + \sqrt{2} dB_t \\
&= -X_t dt + \sqrt{2} dB_t,
\end{aligned}$$

so it follows by strong uniqueness [5, Chapter IX, Section 1] that the law $\mathbb{P}_x$ is also characterized as the law of solution to the SDE $dX_t = -X_t dt + \sqrt{2} dB_t$ with initial condition $X_0 = x$.

Write $\mathbb{E}_x$ and Var_x for the expectation and variance with respect to the measure $\mathbb{P}_x$. From (1) we can see that the mean of the OU process is $\mathbb{E}_x[X_t] = xe^{-t}$, and that its variance, by the Itô isometry, is

$$\begin{aligned}\mathrm{Var}_x(X_t) &= \mathbb{E}\left[\left(\sqrt{2}e^{-t}\int_0^t e^s dB_s\right)^2\right] \\ &= 2e^{-2t}\mathbb{E}\left[\left(\int_0^t e^s dB_s\right)^2\right] \\ &= 2e^{-2t}\int_0^t e^{2s} ds = 1 - e^{-2t}.\end{aligned}$$

This implies that, for any $f \in L^\infty(\gamma^n)$, we have

$$\mathbb{E}_x[f(X_t)] = \int_{\mathbb{R}^n} f\left(e^{-t}x + \sqrt{1 - e^{-2t}}z\right) d\gamma^n(z). \tag{2}$$

where γ^n denotes the standard n-dimensional Gaussian measure on $\mathbb{R}^n$. In particular, this implies that γ^n is a reversible measure for the OU semigroup $\{P_t\}_{t \geq 0}$. So, by Lemma 3.1, the OU semigroup satisfies the needed assumptions. The formula (2) also shows that for $f \in C_b^1(\mathbb{R}^n)$ we can differentiate under the inegral to get $\nabla P_t f = e^{-t} P_t \nabla f$, called the *commutation identity*.

We also note that much can be said about the $L^2(\mu)$-generator of the OU process [1, Section 2.7]. Its domain $D(\mathcal{L})$ consists exactly of functions $f \in C^1(\mathbb{R}^d)$ such that the distributional Laplacian Δf is a function $g : \mathbb{R}^n \to \mathbb{R}$ which satisfies $g(x) - x \cdot \nabla f(x) \in L^2(\gamma^n)$; in particular, if $f \in C^2(\mathbb{R}^n)$ is such that $|x|\Delta f(x)$ vanishes at infinity, and $a \in \mathbb{R}$ is arbitrary, then we have $f(x) + a \in D(\mathcal{L})$. On this domain the generator is given exactly by

$$\mathcal{L}f(x) = \Delta f(x) - x \cdot \nabla f(x).$$

A simple application of Gaussian integration by parts shows that for $f, g \in D(\mathcal{L})$, the Dirichlet form can be equivalently written as

$$\mathcal{E}(f, g) = \int_{\mathbb{R}^n} \nabla f(x) \cdot \nabla g(x) d\gamma^n(x)$$

In particular, this shows that $\mathcal{E}$ is symmetric on $D(\mathcal{L})$. (The collection $D(\mathcal{E})$ of all functions $f \in L^2(\gamma^n)$ such that $-\frac{1}{t}\int_{\mathbb{R}^n}(P_t f(x) - f(x))f(x)d\gamma^n$ converges as $t \to 0$ is called the *domain of* $\mathcal{E}$, and it is strictly larger than $D(\mathcal{L})$.)

5. Gaussian Concentration Inequalities

In this last part, we use the previous results about the OU semigroup to prove some concentration results for the Gaussian measure, eventually leading to the proof of the hypercontractivity of the OU semigroup. For convenience, we set up a bit of notation. For $f \in \mathcal{M}(\mathbb{R}^n)$, write $\mathbb{E}[f] = \int_{\mathbb{R}^n} f d\gamma^n$ and $\mathrm{Var}(f) = \int_{\mathbb{R}^n} f^2 d\gamma^n - (\int_{\mathbb{R}^n} f d\gamma^n)^2$ for the expectation and variance on this space, whenever they are well-defined. Also define

$$\operatorname{Ent}(f) = \int_{\mathbb{R}^n} f \log f \, d\gamma^n - \left(\int_{\mathbb{R}^n} f d\gamma^n \right) \log \left(\int_{\mathbb{R}^n} f d\gamma^n \right)$$

called the *entropy* of f. Note by Jensen's inequality that we have $\operatorname{Ent}(f) \geq 0$, and that $\operatorname{Ent}(f) < \infty$ iff $\int_{\mathbb{R}^n} f \log f \, d\gamma^n < \infty$. (The space of measurable functions with finite entropy with respect to the Gaussian measure forms a Banach space under a suitable norm. This is an example of a so-called "Orlicz space" but we do not develop this idea further in the present work.) For concreteness, observe that, if $f \in L^p(\gamma^n)$ for some $p > 1$, then $\int_{\mathbb{R}^n} f \log f \, d\gamma^n < \infty$,

Lemma 5.1 (logarithmic Sobolev inequality)**.** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is absolutely continuous, with $\operatorname{Var}(f) < \infty$. Then we have

$$\operatorname{Ent}\left(f^2\right) \leq 2 \int_{\mathbb{R}^n} \|\nabla f(x)\|^2 d\gamma^n(x)$$

Proof. First consider the case that $f \in C_c^\infty(\mathbb{R}^n)$. Then define the function $v(x) = (f(x))^2$, and, for $\varepsilon > 0$, the function $v^\varepsilon(x) = v + \varepsilon$. It is clear that v and v^ε are in $D(\mathcal{L})$, and that v^ε is uniformly bounded away from both zero and infinity. Now for $t \geq 0$, define $v_t = P_t v$ and $v_t^\varepsilon = P_t v^\varepsilon = P_t(v + \varepsilon) = v_t + \varepsilon$. By using the boundedness of v and the mean value theorem to apply dominated convergence, it is easy to differentiate under the inegral and deduce that $v_t \in C^\infty(\mathbb{R}^n)$, and also that $|x|\Delta v_t(x)$ vanishes at infinity. Of course, this implies $v_t^\varepsilon \in D(\mathcal{L})$. Likewise, we have $\log(v_t^\varepsilon) \in D(\mathcal{L})$. In particular, the map $t \mapsto v_t^\varepsilon \log(v_t^\varepsilon)$ is strongly differentiable in $L^2(\gamma^n)$, hence, by passing to a subnet if necessary, we can assume the convergence occurs Lebesgue almost everywhere. This implies that the derivative $\partial_t(v_t^\varepsilon \log(v_t^\varepsilon))$ exists (as an equivalence class of functions up to Lebesgue almost everywhere agreement).

Also note that, for any $x \in \mathbb{R}^n$, we have $v_t^\varepsilon(x) \to \int_{\mathbb{R}^n} v^\varepsilon(x) d\gamma^n(x)$ as $t \to \infty$. So by the fundamental theorem of calculus, and then by using the dominated convergence theorem to differentiate under the integral, we have

$$\begin{aligned}
&\int_{\mathbb{R}^n} v^\varepsilon \log(v^\varepsilon) d\gamma^n - \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n \right) \log \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n \right) \\
&= \int_{\mathbb{R}^n} v_0^\varepsilon \log(v_0^\varepsilon) d\gamma^n - \int_{\mathbb{R}^n} v_\infty^\varepsilon \log(v_\infty^\varepsilon) d\gamma^n \\
&= - \int_0^\infty \partial_t \int_{\mathbb{R}^n} v_t^\varepsilon \log(v_t^\varepsilon) d\gamma^n dt \\
&= - \int_0^\infty \int_{\mathbb{R}^n} \partial_t \left(v_t^\varepsilon \log(v_t^\varepsilon) \right) d\gamma^n dt \\
&= - \int_0^\infty \int_{\mathbb{R}^n} \partial_t v_t \left(1 + \log(v_t^\varepsilon) \right) d\gamma^n dt \\
&= \int_0^\infty \mathcal{E}(\partial_t v_t^\varepsilon, 1 + \log(v_t^\varepsilon)) dt
\end{aligned}$$

Now recall that by the Kolmogorov equation we have $\partial_t v_t^\varepsilon = \partial_t P_t v^\varepsilon = \mathcal{L} P_t v^\varepsilon = \mathcal{L} v_t^\varepsilon$, and by the alternative formula for the Dirichlet form, we have

$$\begin{aligned}
\mathcal{E}(\partial_t v_t^\varepsilon, 1+\log(v_t^\varepsilon)) &= \mathcal{E}(\mathcal{L} v_t^\varepsilon, 1+\log(v_t^\varepsilon)) \\
&= \int_{\mathbb{R}^n} \nabla v_t^\varepsilon \cdot \nabla(1+\log(v_t^\varepsilon)) d\gamma^n \\
&= \int_{\mathbb{R}^n} \frac{\|\nabla v_t^\varepsilon\|^2}{v_t^\varepsilon} d\gamma^n
\end{aligned}$$

By the commutation identity $\nabla v_t^\varepsilon(x) = e^{-t} P_t \nabla v^\varepsilon(x)$ and Cauchy-Schwarz, we bound this value as follows:

$$\begin{aligned}
\frac{\|\nabla v_t^\varepsilon(x)\|^2}{v_t^\varepsilon(x)} &= \frac{e^{-2t}}{v_t^\varepsilon(x)} \|P_t \nabla v^\varepsilon(x)\|^2 \\
&= \frac{e^{-2t}}{v_t^\varepsilon(x)} \|\mathbb{E}_x[\nabla v^\varepsilon(X_t)]\|^2 \\
&= \frac{e^{-2t}}{v_t^\varepsilon(x)} \left\| \mathbb{E}_x \left[\sqrt{v^\varepsilon(X_t)} \frac{\nabla v^\varepsilon(X_t)}{\sqrt{v^\varepsilon(X_t)}} \right] \right\|^2 \\
&\le \frac{e^{-2t}}{v_t^\varepsilon(x)} \mathbb{E}_x[v^\varepsilon(X_t)] \mathbb{E}_x \left[\frac{\|\nabla v^\varepsilon(X_t)\|^2}{v^\varepsilon(X_t)} \right] \\
&= e^{-2t} P_t \left(\frac{\|v^\varepsilon\|^2}{v^\varepsilon} \right)(x).
\end{aligned}$$

Combining these last results and using the fact that γ^n is stationary for $\{X_t\}_{t\ge 0}$, we have shown

$$\begin{aligned}
&\int_{\mathbb{R}^n} v^\varepsilon \log(v^\varepsilon) d\gamma^n - \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n\right) \log\left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n\right) \\
&\qquad \le \int_0^\infty e^{-2t} \int_{\mathbb{R}^n} P_t \left(\frac{\|v^\varepsilon\|^2}{v^\varepsilon} \right) d\gamma^n dt \\
&\qquad = \int_0^\infty e^{-2t} \int_{\mathbb{R}^n} \frac{\|v^\varepsilon\|^2}{v^\varepsilon} d\gamma^n dt \\
&\qquad = \frac{1}{2} \int_{\mathbb{R}^n} \frac{\|v^\varepsilon\|^2}{v^\varepsilon} d\gamma^n.
\end{aligned}$$

Recalling that $v^\varepsilon(x) = (f(x))^2 + \varepsilon$, we have $\nabla v^\varepsilon(x) = 2f(x)\nabla f(x)$, hence

$$\begin{aligned}
&\int_{\mathbb{R}^n} v^\varepsilon \log(v^\varepsilon) d\gamma^n - \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n\right) \log\left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n\right) \\
&\qquad \le 2 \int_{\mathbb{R}^n} \frac{f^2}{f^2+\varepsilon} \|\nabla f\|^2 d\gamma^n.
\end{aligned}$$

Now consider taking $\varepsilon \to 0$. First notice that we have $f^2/(f^2+\varepsilon) \uparrow 1$ for all $x \in \mathbb{R}^n$, so monotone convergence applies to the right side above. Then observe that $\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n = \|f\|_{L^2(\gamma^n)}^2 + \varepsilon$, so $\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n \downarrow \|f\|_{L^2(\gamma^n)}^2$. Finally, the map $a \mapsto a^2 \log(a^2)$ is uniformly bounded below, so Fatou's lemma applies to the first term on the left side above. Combining this all yields

$$\begin{aligned}
&\int_{\mathbb{R}^n} f^2 \log(f^2) d\gamma^n - \|f\|^2_{L^2(\gamma^n)} \log \|f\|^2_{L^2(\gamma^n)} \\
&\quad \leq \liminf_{\varepsilon\to 0} \int_{\mathbb{R}^n} v^\varepsilon \log(v^\varepsilon) d\gamma^n - \lim_{\varepsilon \to 0} \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n \right) \log \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n \right) \\
&\quad = \liminf_{\varepsilon\to 0} \left(\int_{\mathbb{R}^n} v^\varepsilon \log(v^\varepsilon) d\gamma^n - \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n \right) \log \left(\int_{\mathbb{R}^n} v^\varepsilon d\gamma^n \right) \right) \\
&\quad \leq \liminf_{\varepsilon\to 0} 2 \int_{\mathbb{R}^n} \frac{f^2}{f^2+\varepsilon} \|\nabla f\|^2 d\gamma^n \\
&\quad = 2 \int_{\mathbb{R}^n} \|\nabla f\|^2 d\gamma^n.
\end{aligned}$$

Finally, we use a standard density argument to extend this to the full result. Let $f : \mathbb{R}^n \to \mathbb{R}$ be absolutely continuous so that, in particular, the gradient ∇f is well-defined for Lebesgue almost every $x \in \mathbb{R}^n$. If $\int_{\mathbb{R}^n} \|\nabla f\|^2 d\gamma^n = \infty$, then there is nothing to prove, so we can assume $\int_{\mathbb{R}^n} \|\nabla f\|^2 d\gamma^n < \infty$. Combining this with our assumption $\mathrm{Var}(f) < \infty$ implies that f lies in the Sobolev space $H^1(\gamma^n)$. Since $C_c^\infty(\mathbb{R}^n)$ is dense in $H^1(\gamma^n)$, we can get a sequence $\{f_k\}_{k\in\mathbb{N}}$ in $C_c^\infty(\mathbb{R}^n)$ with $f_k \to f$ in $H^1(\gamma^n)$. This implies $f_k \to f$ in $L^2(\gamma^n)$, so, by passing to a subsequence if necessary, we can also assume $f_k \to f$ holds Lebesgue almost everywhere. Now use Fatou's lemma and the fact that we have already shown the inequality for $\{f_k\}_{k\in\mathbb{N}}$ to get:

$$\begin{aligned}
\mathrm{Ent}(f) &= \int_{\mathbb{R}^n} f^2 \log(f^2) d\gamma^n - \left(\int_{\mathbb{R}^n} f^2 d\gamma^n \right) \log \left(\int_{\mathbb{R}^n} f^2 d\gamma^n \right) \\
&\leq \liminf_{k\to\infty} \int_{\mathbb{R}^n} f_k^2 \log(f_k^2) d\gamma^n - \lim_{k\to\infty} \left(\int_{\mathbb{R}^n} f_k^2 d\gamma^n \right) \log \left(\int_{\mathbb{R}^n} f_k^2 d\gamma^n \right) \\
&= \liminf_{k\to\infty} \left(\int_{\mathbb{R}^n} f_k^2 \log(f_k^2) d\gamma^n - \left(\int_{\mathbb{R}^n} f_k^2 d\gamma^n \right) \log \left(\int_{\mathbb{R}^n} f_k^2 d\gamma^n \right) \right) \\
&= \liminf_{k\to\infty} \mathrm{Ent}(f_k) \\
&\leq \liminf_{k\to\infty} 2 \int_{\mathbb{R}^n} \|\nabla f_k\|^2 d\gamma^n \\
&= 2 \int_{\mathbb{R}^n} \|\nabla f\|^2 d\gamma^n.
\end{aligned}$$

This finishes the proof.

□

Theorem 5.2. The Ornstein-Uhlenbeck semigroup is hypercontractive.

Proof. That (1') is satisfied is implicit. We claim that a strong version of (2') holds, that for any $1 < p < q < \infty$ and $t > T = \frac{1}{2} \log \frac{q-1}{p-1}$ we have $\|P_t f\|_{L^q(\mu)} \leq \|f\|_{L^p(\mu)}$ for all $f \in L^p(\mu)$. (So, in this case, "hypercontractive" is actually a quite appropriate term!) To do this, define the monotonic function $q(t) = 1 + (p-1)e^{2t}$ and observe that we have $q(0) = p$ and $q(T) = q$. Let $f \in C_c^2(\mathbb{R}^n)$ be arbitrary and assume that f is non-negative. Then set $f_t = P_t f$ as well as $r(t) = \int_{\mathbb{R}^n} (f_t(x))^{q(t)} d\gamma^n(x)$.

Next we claim that we can differentiate r by differentiating under the integral. By dominated convergence and the mean value theorem, it suffices to show that, for each $s > 0$, the derivative $\partial_t \left((f_t(x))^{q(t)}\right)$ is bounded above, uniformly for all $(t, x) \in [0, s] \times \mathbb{R}^n$. To do this, note that, for any $x \in \mathbb{R}^n$, we can compute:

$$\frac{d}{dt}\left((f_t(x))^{q(t)}\right) = (f_t(x))^{q(t)} q'(t) \log(f_t(x)) + (f_t(x))^{q(t)-1} \partial_t f_t q(t).$$

Observe that we have $\sup_{x\in\mathbb{R}^n} |f_t(x)| \leq \sup_{x\in\mathbb{R}^n} |f(x)|$ by Jensen's inequality. Then since $f \in C_c^2(\mathbb{R}^n)$ is clearly bounded above uniformly in $x \in \mathbb{R}^n$, this implies that $f_t(x)$ is bounded above uniformly in $(t, x) \in [0, \infty) \times \mathbb{R}^n$. Moreover, q is bounded above uniformly for $t \in [0, s]$ by inspection. Therefore, it suffices to show that $\partial_t f_t(x)$ has the necessary bound. Recall that the heat equation says $\partial_t f_t(x) = P_t \mathcal{L} f(x)$, so we have $\sup_{x\in\mathbb{R}^n} |\frac{d}{dt} f_t(x)| \leq \sup_{x\in\mathbb{R}^n} |\Delta f(x) - x \cdot \nabla f(x)|$ again by Jensen. But $f \in C_c^2(\mathbb{R}^n)$ implies that $\sup_{x\in\mathbb{R}^n} |\Delta f(x) - x \cdot \nabla f(x)|$ is finite, as needed.

Now we differentiate under the integral and apply the Kolmogorov equation to compute the derivative of r to be:

$$\begin{aligned}
r'(t) &= \int_{\mathbb{R}^n} \left(f_t^{q(t)} q'(t) \log(f_t) + f_t^{q(t)-1} \partial_t f_t q(t)\right) d\gamma^n \\
&= q'(t) \int_{\mathbb{R}^n} f_t^{q(t)} \log(f_t) d\gamma^n + q(t) \int_{\mathbb{R}^n} f_t^{q(t)-1} \mathcal{L} f_t d\gamma^n \\
&= \frac{q'(t)}{q(t)} \int_{\mathbb{R}^n} f_t^{q(t)} \log\left(f_t^{q(t)}\right) d\gamma^n - q(t)(q(t)-1) \int_{\mathbb{R}^n} f_t^{q(t)-2} \|\nabla f_t\|^2 d\gamma^n
\end{aligned}$$

Now consider the value $\log \|f_t\|_{L^{q(t)}(\gamma^n)} = \frac{1}{q(t)} \log(r(t))$. Differentiating this, and using the observation that $q'(t) = 2(q(t) - 1)$, yields:

$$\begin{aligned}
&\frac{d}{dt} \log \|f_t\|_{L^{q(t)}(\gamma^n)} \\
&= \frac{r'(t)}{q(t)r(t)} - \frac{q'(t) \log r(t)}{(q(t))^2} \\
&= \frac{q'(t)}{(q(t))^2 r(t)} \int_{\mathbb{R}^n} f_t^{q(t)} \log\left(f_t^{q(t)}\right) d\gamma^n - \frac{q(t)-1}{r(t)} \int_{\mathbb{R}^n} f_t^{q(t)-2} \|\nabla f_t\|^2 d\gamma^n - \frac{q'(t) \log r(t)}{(q(t))^2} \\
&= \frac{q'(t)}{(q(t))^2 r(t)} \int_{\mathbb{R}^n} f_t^{q(t)} \log\left(\frac{f_t^{q(t)}}{r(t)}\right) d\gamma^n - \frac{q(t)-1}{r(t)} \int_{\mathbb{R}^n} f_t^{q(t)-2} \|\nabla f_t\|^2 d\gamma^n \\
&= \frac{q'(t)}{(q(t))^2 r(t)} \int_{\mathbb{R}^n} f_t^{q(t)} \log\left(\frac{f_t^{q(t)}}{r(t)}\right) d\gamma^n - \frac{q'(t)}{2r(t)} \int_{\mathbb{R}^n} f_t^{q(t)-2} \|\nabla f_t\|^2 d\gamma^n \\
&= \frac{q'(t)}{(q(t))^2 r(t)} \left(\int_{\mathbb{R}^n} f_t^{q(t)} \log\left(\frac{f_t^{q(t)}}{r(t)}\right) d\gamma^n - \frac{(q(t))^2}{2} \int_{\mathbb{R}^n} f_t^{q(t)-2} \|\nabla f_t\|^2 d\gamma^n\right).
\end{aligned}$$

Finally, note that the logarithmic Sobolev inequality applied to the function $f_t^{q(t)/2}$ shows that the term in parenthesis is non-positive. Therefore, we have shown that $\log \|f_t\|_{L^{q(t)}(\gamma^n)}$ is non-increasing as a function of t, whence the result.

If $f \in C_c^2(\mathbb{R}^n)$ but not necessarily non-negative, then we use Jensen to get

$$\|P_t f\|_{L^q(\gamma^n)} \leq \|P_t |f|\|_{L^q(\gamma^n)} \leq \||f|\|_{L^p(\gamma^n)} = \|f\|_{L^p(\gamma^n)},$$

so the result holds.

Now let $f \in L^p(\gamma^n)$ be arbitrary, and use the density of $C_c^2(\mathbb{R}^n)$ in $L^p(\gamma^n)$ to get $\{f_n\}_{n\in\mathbb{N}}$ in $C_c^2(\mathbb{R}^n)$ such that $f_n \to f$ holds in $L^p(\gamma^n)$. Then the result for $C_c^2(\mathbb{R}^n)$ combined with Fatou's lemma gives:

$$\begin{aligned} \|P_t f\|_{L^q(\gamma^n)} &\leq \liminf_{n\to\infty} \|P_t f_n\|_{L^q(\gamma^n)} \\ &\leq \liminf_{n\to\infty} \|f_n\|_{L^p(\gamma^n)} = \|f\|_{L^p(\gamma^n)}, \end{aligned}$$

as claimed. □

As the proof above shows, the main property of the OU semigroup that was needed to show hypercontractivity was the logarithmic Sobolev inequality. More generally, it is know that any semigroup satisfying a logarithmic Sobolev inequality must be hypercontractive. A remarkable discovery of L. Gross is that the converse is also true: any hypercontractive semigroup satisfies a certain generalized form of the logarithmic Sobolev inequality. See [1, Theorem 5.2.3 of Chapter 5.2.2] for a precise description of this equivalence.

We also note that the logarithmic Sobolev inequality implies the well-known Gaussian Poincaré inequality [1, Proposition 5.1.3 of Chapter 5.1.2]. (Of course, there also exist elementary proofs that do not rely on the analysis of semigroups we have undertaken here.) In general, a logarithmic Sobolev inequality is a much stronger statement than a Poincaré inequality, as can be intuited from the extra logarithmic factor present on the left side.

We finally note the remarkable fact that the constant in the logarithmic Sobolev inequality (and, consequently, in the Gaussian Poincaré inequality) is independent of the dimension n. This is key to understanding concentration properties of infinite-dimensional Gaussian measures, and it has important consequences in diverse fields as mathematical physics, theoretical computer science, and high-dimensional statistics.

References

[1] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[2] Sourav Chatterjee. *Superconcentration and related topics*. Springer Monographs in Mathematics. Springer, Cham, 2014.

[3] Edward Nelson. A quartic interaction in two dimensions. In *Mathematical Theory of Elementary Particles (Proc. Conf., Dedham, Mass., 1965)*, pages 69–73. M.I.T. Press, Cambridge, Mass., 1966.

[4] Edward Nelson. The free Markoff field. *J. Functional Analysis*, 12:211–227, 1973.

[5] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2004.

[6] Barry Simon. Ed Nelson's work in quantum theory. In *Diffusion, quantum theory, and radically elementary mathematics*, volume 47 of *Math. Notes*, pages 75–93. Princeton Univ. Press, Princeton, NJ, 2006.

[7] Barry Simon and Raphael Høegh-Krohn. Hypercontractive semigroups and two dimensional self-coupled Bose fields. *J. Functional Analysis*, 9:121–180, 1972.