CHAPTER 13

Entropy and the Logarithmic Sobolev Inequality (LSI)

13.1. Basic Properties of the Entropy

In this section, we prove a few well-known inequalities for the entropy that will be used in next section for the logarithmic Sobolev inequality (LSI). In this section we work on an arbitrary probability space \((\Omega, \mathcal{B})\) with an appropriate \(\sigma\)-algebra. We will use the probabilistic and the analysis notation and concepts in parallel. In particular, we interchangeably use the concept of random variables and measurable functions on \(\Omega\). Similarly, both the expectation \(E^\nu[f]\) and the integral notation \(\int_{\Omega} f \, d\mu\) will be used. We will drop the \(\Omega\) in the notation.

**Definition 13.1.** For any two probability measures \(\nu, \mu\) on the same probability space, we define the \textit{relative entropy} of \(\nu\) w.r.t. \(\mu\) by

\[
S(\nu|\mu) := \int \log \frac{d\nu}{d\mu} \, d\mu = \int \log \frac{d\nu}{d\mu} \, d\mu
\]

if \(\nu\) is absolutely continuous with respect to \(\mu\), and we set \(S(\nu|\mu) := \infty\) otherwise.

Applying the definition (13.1) to \(\nu = f\mu\) for any probability density \(f\) with \(\int f \, d\mu = 1\), we may also write

\[
S_\mu(f) := S(f\mu|\mu) = \int f(\log f) \, d\mu.
\]

In most cases, the reference measure \(\mu\) will be canonical (e.g., the natural equilibrium measure), so often we will drop the subscript \(\mu\) if there is no confusion. We may then call \(S_\mu(f) = S(f)\) the \textit{entropy} of \(f\). Using the convexity of the function \(x \mapsto x \log x\) on \(\mathbb{R}_+\), a simple Jensen’s inequality,

\[
0 = \left( \int f \, d\mu \right) \log \left( \int f \, d\mu \right) \leq \int f \log f \, d\mu,
\]

shows that the relative entropy is always nonnegative, \(S(\nu|\mu) \geq 0\).

**Proposition 13.2 (Entropy inequality or Gibbs inequality).** Let \(\mu, \nu\) be two probability measures on a common probability space and let \(X\) be a random variable. Then, for any positive number \(\alpha > 0\), the following inequality holds as long as the right-hand side is finite:

\[
E^\nu[X] \leq \alpha^{-1} S(\nu|\mu) + \alpha^{-1} \log E^\mu e^{\alpha X}.
\]

(13.2)
Proof. First note that we only have to prove the case \( \alpha = 1 \) since we can redefine \( \alpha X \rightarrow X \). From the concavity of the logarithm and Jensen’s inequality, we have

\[
\int X \, d\nu - S(\nu|\mu) = \int \log \left( e^{X} \frac{d\mu}{d\nu} \right) d\nu \leq \log \mathbb{E}^\nu e^X,
\]

and this proves (13.2).

Although (13.2) is just an inequality, there is a variational characterization of the relative entropy behind it. Namely,

\[
S(\nu|\mu) = \sup_X \left[ \mathbb{E}^\nu [X] - \log \mathbb{E}^\mu e^X \right],
\]

where the supremum is over all bounded random variables. As we will not need this relation in this book, we leave it to the interested reader to prove it.

As a corollary of (13.2), we mention that for any set \( A \) we have the bound

\[
\mathbb{P}^\nu(A) \leq \frac{\log 2 + S(\nu|\mu)}{\log \frac{1}{\mathbb{P}^\mu(A)}}.
\]

The proof is left as an exercise. This inequality will be used in the following context. Suppose that the relative entropy of \( \nu \) with respect to \( \mu \) is finite. Then in order to show that a set has a small probability w.r.t. \( \nu \), we need to verify that this set is exponentially small w.r.t. \( \mu \). In this sense, entropy provides only a relatively weak link between two measures. However, it is still stronger than the total variational norm, which we will show now.

For any two probability measures \( f\mu \) and \( \mu \), the \( L^p \)-distance between \( f\mu \) and \( \mu \) is defined by

\[
\left[ \int |f - 1|^p \, d\mu \right]^{1/p}.
\]

When \( p = 1 \), it is called the total variational norm between \( f\mu \) and \( \mu \). Entropy is a weaker measure of distance between two probability measures than the \( L^p \)-distance for any \( p > 1 \) but stronger than the total variational norm. The former statement can be expressed, for example, by the elementary inequality

\[
\int f \log f \, d\mu \leq 2 \left[ \int |f - 1|^p \, d\mu \right]^{1/p} + \frac{2}{p-1} \int |f - 1|^p \, d\mu, \quad p > 1,
\]

whose proof is left as an exercise. (There is a more natural way to compare \( L^p \)-distance and the relative entropy in terms of the \( L \log L \) Orlicz norm, but we will not use this norm in this book.) The latter statement will be made precise in the following proposition. Furthermore, we also remark the following easy relation among \( L^p \)-norms and the entropy:

\[
\frac{d}{dp} \bigg|_{p=1} \left[ \int f^p \, d\mu \right]^{1/p} = \int f \log f \, d\mu
\]

holds for any probability density \( f \) w.r.t. \( \mu \).
Proposition 13.3 (Entropy and total variation norm, Pinsker inequality).
Suppose that \( \int f \, d\mu = 1 \) and \( f \geq 0 \). Then, we have
\[
2 \left[ \int |f - 1| \, d\mu \right]^2 \leq \int f \log f \, d\mu.
\]

Proof. By the variational principle, we first rewrite
\[
\int |f - 1| \, d\mu = \sup_{|g| \leq 1} \int fg \, d\mu - \int g \, d\mu.
\]
For any such function \( g \), we have by the entropy inequality (13.2) that for any \( t > 0 \)
\[
\int fg \, d\mu - \int g \, d\mu \leq t^{-1} \log \int e^{tg} \, d\mu - \int g \, d\mu + t^{-1} \int f \log f \, d\mu.
\]
We now define the function
\[
h(t) := \log \int e^{tg} \, d\mu
\]
for any \( t \geq 0 \). A simple calculation shows that the second derivative of \( h \) is given by
\[
h''(t) = \langle g; g \rangle_{\omega_t}, \quad \omega_t := \frac{e^{tg} \mu}{\int e^{tg} \, d\mu},
\]
where \( \omega_t \) is a probability measure, and
\[
\langle f; g \rangle_{\omega} := \int fg \, d\omega - \left( \int f \, d\omega \right) \left( \int g \, d\omega \right)
\]
denotes the covariance. Recall that the covariance is positive definite; i.e., it satisfies the usual Schwarz inequality,
\[
|\langle f; g \rangle_{\omega}|^2 \leq \langle f; f \rangle_{\omega} \langle g; g \rangle_{\omega}.
\]
Since \( |g| \leq 1 \), we have \( 0 \leq \langle g; g \rangle_{\omega_t} \leq 1 \), so \( h''(t) \leq 1 \). Thus by Taylor’s theorem, \( h(t) \leq h(0) + th'(0) + t^2/2 \), i.e.,
\[
t^{-1} \log \int e^{tg} \, d\mu \leq \int g \, d\mu + \frac{t}{2}.
\]
Together with (13.10), we have
\[
\int fg \, d\mu - \int g \, d\mu \leq \frac{t}{2} + t^{-1} \int f \log f \, d\mu \leq \sqrt{\frac{1}{2} \int f \log f \, d\mu},
\]
where we optimized \( t \) in the last step. Since this bound holds for any \( g \) with \( |g| \leq 1 \), using (13.9) we have proved (13.8). \( \square \)
13.2. Entropy on Product Spaces and Conditioning

Now we consider a product measure; i.e., we assume that the probability space has a product structure, $\Omega = \Omega_1 \times \Omega_2$ and $\mu = \mu_1 \otimes \mu_2$, $\nu = \nu_1 \otimes \nu_2$ where $\mu_j$ and $\nu_j$ are probability measures on $\Omega_j$, $j = 1, 2$. For simplicity, we denote the elements of $\Omega$ by $(x, y)$ with $x \in \Omega_1$, $y \in \Omega_2$. Writing $\nu_j = f_j \mu_j$, clearly we have

$$S(\nu_1 \otimes \nu_2 | \mu_1 \otimes \mu_2) = S_{\mu_1 \otimes \mu_2}(f_1 \otimes f_2)$$

\[= \int_{\Omega_1 \times \Omega_2} f_1(x)f_2(y)[\log f_1(x) + \log f_2(y)]\mu_1(dx)\mu_2(dy) \]

\[= S_{\mu_1}(f_1) + S_{\mu_2}(f_2) = S(\nu_1 | \mu_1) + S(\nu_2 | \mu_1); \]

i.e., the entropy is additive for product measures.

This property of the entropy makes it an especially suitable tool to measure closeness in very high-dimensional analysis. For example, if the probability space is a large product space $\Omega = \Omega_1 \otimes \Omega_1^N$ equipped with two product measures $\mu = \mu_1 \otimes \mu_1^N$ and $\nu = f \otimes \mu_1^N$, then their relative entropy

$$S_{\mu}(f \otimes \mu_1^N) = NS_{\mu_1}(f)$$

grows only linearly in $N$, the number of degrees of freedom. It is easy to check that any $L^P$-distance (13.5) for $p > 1$ grows exponentially in $N$, which usually renders it useless. Thus the entropy is still stronger than the $L^1$-norm, but its growth in $N$ is much more manageable. An additional advantage is that the entropy is often easier to compute than any other $L^P$-norm (with the exception of $L^2$).

Next we discuss how the entropy decomposes under conditioning. For simplicity, we assume the product structure, $\Omega = \Omega_1 \times \Omega_2$ as before. Let $\omega$ be a probability measure on the product space $\Omega$. For any integrable function (random variable) $u(x, y)$ on $\Omega$ we denote its conditional expectation by

$$\hat{u} = E_{\omega}[u | F_1]$$

where $F_1$ is the $\sigma$-algebra of $\Omega_1$ canonically lifted to $\Omega$. Note that $\hat{u} = \hat{u}(x)$ depends only on the $x$-variable.

The conditional expectation $\hat{u}$ may also be characterized by the following relation: it is the unique measurable function on $\Omega_1$ such that for any bounded, measurable (w.r.t. $F_1$) function $O(x)$ the following identity holds:

$$E_{\omega}[\hat{u}O] = E_{\omega}[uO].$$

Written out in coordinates, it means that

$$\int_{\Omega} \hat{u}(x)xO(x)\omega(dx \ dy) = \int_{\Omega} u(x, y)xO(x)\omega(dx \ dy).$$

In particular, if $\omega(dx \ dy) = \omega(x, y)dx \ dy$ is absolutely continuous with respect to some reference product measure $dx \ dy$ on $\Omega$, then we have, somewhat formally,

$$\hat{u}(x) = \frac{\int_{\Omega_2} u(x, y)x\omega(x, y)dy}{\int_{\Omega_2} \omega(x, y)dy}. $$
The notation $dx\,dy$ for the reference measure already indicates that in our applications $\Omega_j$ will be euclidean spaces $\mathbb{R}^{n_j}$ of some dimension $n_j$, and the reference measure will be the Lebesgue measure.

We remark that the concept of conditioning can be defined in full generality with respect to any sub-$\sigma$-algebra; the product structure of $\Omega$ is not essential. However, we will not need the general definition in this book and the above definition is conceptually simpler.

The conditional expectation gives rise to a trivial martingale decomposition:

$$u = \hat{u} + (u - \hat{u}),$$

where $\hat{u}$ is $\mathcal{F}_1$-measurable, while $u - \hat{u}$ has zero expectation on any $\mathcal{F}_1$-measurable set. Subtracting the expectation $\bar{u} := \mathbb{E}\omega u = \mathbb{E}\omega \hat{u}$ and squaring this formula, we have the martingale decomposition of the variance of $u$:

$$\text{Var} \omega(u) := \mathbb{E}\omega(u - \bar{u})^2 = \mathbb{E}\omega(u - \hat{u})^2 + \mathbb{E}\omega(\hat{u} - \bar{u})^2$$

$$= \mathbb{E}\omega \text{Var}(u(x, \cdot)) + \text{Var} \omega(\hat{u})$$

where we defined the conditional variance

$$\text{Var}(u(x, \cdot)) := \mathbb{E}[u - \hat{u}]^2|\mathcal{F}_1] = \mathbb{E}[u^2|\mathcal{F}_1] - [\hat{u}]^2.$$

The identity (13.17) is a triviality, but its interpretation is important. It means that the variance is additive w.r.t. the martingale decomposition (13.16). The first term $\mathbb{E}\omega \text{Var}(u(x, \cdot))$ is the expectation of the variance w.r.t. $y$ conditioned on $x$; the second term $\text{Var}(\hat{u} - \bar{u})^2$ is the variance of the marginal w.r.t. $x$. In other words, we can compute the variance one by one.

The martingale decomposition has an analogue for the entropy. For simplicity, we assume that $\omega$ has a density $\omega(x, y)$ w.r.t. a reference measure $dx\,dy$. Denote by $\tilde{\omega}$ the marginal $\tilde{\omega}$ probability density on $\Omega_1$,

$$\tilde{\omega}(x) = \int_{\Omega_2} \omega(x, y) \, dy,$$

and by $\hat{f}$ the marginal density of $f\omega$ w.r.t. $\tilde{\omega}$, i.e.,

$$\hat{f}(x) = \frac{\int_{\Omega_2} f(x, y) \omega(x, y) \, dy}{\tilde{\omega}(x)}.$$

In particular, for any test function $O(x)$ we have

$$\iint_{\Omega} O(x)f(x, y)\omega(x, y)\,dx\,dy = \int_{\Omega_1} O(x)\hat{f}(x)\tilde{\omega}(x)\,dx.$$

Let

$$\omega_x(y) := \frac{\omega(x, y)}{\tilde{\omega}(x)}$$

be the probability density on $\Omega_2$ conditioned on a fixed $x \in \Omega_1$. Define

$$f_x(y) := \frac{f(x, y)}{\hat{f}(x)}.$$
to be the corresponding density of the measure $f \omega$ conditioned on a fixed $x \in \Omega_1$. Note that with these definitions, we have

$$(f \omega)_x(y) = \frac{f(x, y)\omega(x, y)}{\int_{\Omega_2} f(x, y)\omega(x, y)dy} = f_x(y)\omega_x(y).$$

Now we are ready to state the following proposition on the additivity of entropy w.r.t. martingale decomposition:

**Proposition 13.4.** Using the notation above, we have

$$(13.20) S_\omega(f) = \mathbb{E}\hat{f}\tilde{\omega}S_{\omega_x}(f_x) + S_{\hat{\omega}}(\hat{f}).$$

In particular, the marginal entropy is bounded by the total entropy, i.e.,

$$(13.21) S_{\hat{\omega}}(\hat{f}) \leq S_\omega(f).$$

In other words, the entropy is additive w.r.t. the martingale decomposition; i.e., the entropy is the sum of the expectation of the entropy in $y$ conditioned on $x$ and the entropy of the marginal in $x$. The additivity of entropy in this sense is an important tool in the application of the LSI, as we will demonstrate shortly. It also indicates that entropy is an extensive quantity; i.e., the entropy of two probabilities on a product space $\Omega^N$ is often of order $N$, generalizing the formulas (13.11)-(13.12) to nonproduct measures.

**Proof.** We can decompose the entropy as follows:

$$(13.22) S_\omega(f) = \int_{\Omega} f \log f \, d\omega = \int_{\Omega} f \log(f/\hat{f})d\omega + \int \hat{f} \log \hat{f} \, d\omega$$

where in the last step we used (13.18). The first term we can rewrite as

$$(13.23) \int_{\Omega} f \log(f/\hat{f})\omega(x, y)dx \, dy$$

$$= \int dx \omega(x)\hat{f}(x)\left[ \int f_x(y) \log f_x(y)\omega_x(y)dy \right]$$

$$= \mathbb{E}\hat{f}\tilde{\omega}S_{\omega_x}(f_x).$$

We have thus proved (13.20). □

### 13.3. Logarithmic Sobolev Inequality

The logarithmic Sobolev inequality requires that the underlying probability space admit a concept of differentiation. While the theory can be developed for more general spaces, we restrict our attention to the probability space $\Omega = \mathbb{R}^N$ in this subsection. We will work with probability measures $\mu$ on $\mathbb{R}^N$ that are defined by a Hamiltonian $\mathcal{H}$:

$$(13.24) d\mu(x) = \frac{e^{-\mathcal{H}(x)}}{Z} \, dx,$$
where \( Z \) is a normalization factor. In Section 13.7 we will comment on how to extend the results of this section to certain subsets of \( \mathbb{R}^N \), most importantly, to the simplex \( \Sigma_N = \{ x \in \mathbb{R}^N : x_1 < \cdots < x_N \} \).

Note that for simplicity in this presentation we neglect the \( \beta N \) prefactor compared with (12.13). Let \( \mathcal{L} \) be the generator of the dynamics associated with the Dirichlet form

\[
D(\mathcal{L}) = D\mu(f) = \int f(\mathcal{L}f) d\mu
\]

(13.25)

\[
:= \int |\nabla f|^2 d\mu = \sum_{j=1}^N \int (\partial_j f)^2 d\mu, \quad \partial_j = \partial_{x_j}.
\]

Formally, we have \( \mathcal{L} = \Delta - (\nabla \mathcal{H}) \cdot \nabla \). The operator \( \mathcal{L} \) is symmetric with respect to the measure \( d\mu \), i.e.,

\[
\int f(\mathcal{L}g) d\mu = \int (\mathcal{L}f)g d\mu = -\int \nabla f \cdot \nabla g d\mu.
\]

(13.26)

We remark that in many books on probability, e.g., [43], the Dirichlet form (13.25) is defined with a factor \( \frac{1}{2} \), but this convention is not compatible with the \( \frac{1}{\beta N} \) prefactor in (12.14). The lack of this \( \frac{1}{2} \) factor in (13.25) causes slight deviations from their customary form in the following theorems.

**Definition 13.5.** The probability measure \( \mu \) on \( \mathbb{R}^N \) satisfies the logarithmic Sobolev inequality if there exists a constant \( \gamma \) such that

\[
S(f) = \int f \log f \, d\mu \leq \gamma \int |\nabla \sqrt{f}|^2 \, d\mu = \gamma D(\sqrt{f})
\]

(13.27)

holds for any smooth density function \( f \geq 0 \) with \( \int f \, d\mu = 1 \). The smallest such \( \gamma \) is called the logarithmic Sobolev inequality constant of the measure \( \mu \).

A simple density argument shows that (13.27) extends from smooth functions to any nonnegative function \( f \in C^\infty_0(\mathbb{R}^N) \), the space of smooth functions with compact support. In fact, it is easy to extend it to all \( \sqrt{f} \in H^1(\mu) \). Since we will not use the space \( H^1(\mu) \) in this book, we will just use \( f \in C^\infty_0(\mathbb{R}^N) \) in the LSI.

**Theorem 13.6 (Bakry-Émery [13]).** Consider a probability measure \( \mu \) on \( \mathbb{R}^N \) of the form (13.24), i.e., \( \mu = e^{-\mathcal{H}}/Z \). Suppose that a convexity bound holds for the Hamiltonian; i.e., with some positive constant \( K \) we have

\[
\nabla^2 \mathcal{H}(x) \geq K
\]

(13.28)

for any \( x \) (in the sense of quadratic forms). Then, the logarithmic Sobolev inequality (13.27) holds with an LSI constant \( \gamma \leq 2/K \), i.e.,

\[
S(f) \leq \frac{2}{K} D(\sqrt{f}) \quad \text{for any density } f \text{ with } \int f \, d\mu = 1.
\]

(13.29)
Furthermore, the dynamics

\begin{equation}
\partial_t f_t = \mathcal{L} f_t, \quad t > 0,
\end{equation}

relaxes to equilibrium on the time scale \( t \approx 1/K \), both in the sense of entropy and Dirichlet form:

\begin{equation}
S(f_t) \leq e^{-2tK}S(f_0), \quad D(\sqrt{f_t}) \leq \frac{2}{t} e^{-tK}S(f_0).
\end{equation}

**Proof.** Let \( f_t \) be the solution to the evolution equation (13.30) with a given smooth initial condition \( f_0 \). Simple calculation shows that the derivative of the entropy \( S(f_t) \) is given by

\begin{equation}
\partial_t S(f_t) = \int (\mathcal{L} f_t) \log f_t \, d\mu + \int f_t \frac{\mathcal{L} f_t}{f_t} \, d\mu
\end{equation}

\begin{equation}
= -\int \frac{(\nabla f_t)^2}{f_t} \, d\mu = -4D(\sqrt{f_t}),
\end{equation}

where we used that \( \int \mathcal{L} f_t \, d\mu = 0 \) by (13.26). Similarly, we can compute the evolution of the Dirichlet form. Let \( h_t := \sqrt{f_t} \) for simplicity; then

\begin{equation}
\partial_t h_t = \frac{1}{2h_t} \partial_i h_t^2 = \frac{1}{2h_t} \mathcal{L} h_t^2 = \mathcal{L} h_t + \frac{1}{h_t} (\nabla h_t)^2.
\end{equation}

We compute (dropping the \( t \) subscript for brevity)

\begin{equation}
\partial_t D(\sqrt{f}) = \partial_t \int (\nabla h)^2 \, d\mu
\end{equation}

\begin{equation}
= 2 \int \nabla h \cdot \nabla \partial_t h \, d\mu
\end{equation}

\begin{equation}
= 2 \int (\nabla h) \cdot (\nabla \mathcal{L} h) \, d\mu + 2 \int (\nabla h) \cdot \nabla (\frac{(\nabla h)^2}{h}) \, d\mu
\end{equation}

\begin{equation}
= 2 \int (\nabla h) \cdot [\nabla, \mathcal{L}] h \, d\mu + 2 \int (\nabla h) \cdot \mathcal{L}(\nabla h) \, d\mu
\end{equation}

\begin{equation}
+ 2 \int \sum_{ij} \partial_i h \left[ 2(\partial_j h)(\partial_j h) - \frac{(\partial_j h)^2}{h^2} \right] \, d\mu
\end{equation}

\begin{equation}
= -2 \int (\nabla h) \cdot (\nabla^2 \mathcal{L}) \nabla h \, d\mu - 2 \int \sum_{ij} (\partial_i \partial_j h)^2 \, d\mu
\end{equation}

\begin{equation}
+ 2 \int \sum_{ij} \left[ 2(\partial_j h)(\partial_j h) - \frac{(\partial_j h)^2}{h^2} \right] \, d\mu
\end{equation}

\begin{equation}
= -2 \int (\nabla h) \cdot (\nabla^2 \mathcal{L}) \nabla h \, d\mu
\end{equation}

\begin{equation}
- 2 \int \sum_{ij} \left( \partial_i \partial_j h - \frac{(\partial_i h)(\partial_j h)}{h} \right)^2 \, d\mu
\end{equation}
where we used the commutator 
\[ [\nabla, \mathcal{L}] = -(\nabla^2 \mathcal{H})\nabla. \]

Therefore, under the convexity condition (13.28), we have

(13.34) 
\[ \partial_t D(\sqrt{f_t}) \leq -2KD(\sqrt{f_t}). \]

Integrating (13.34), we have

(13.35) 
\[ D(\sqrt{f_t}) \leq e^{-2tK}D(\sqrt{f_0}). \]

This proves that the equilibrium is achieved at \( t = \infty \) with \( f_\infty = 1 \), and both the entropy and the Dirichlet form are zero. Integrating (13.32) from \( t = 0 \) to \( t = \infty \), and using the monotonicity of \( D(\sqrt{f_t}) \) from (13.34), we obtain

(13.36) 
\[ -S(f_0) = -4 \int_0^\infty D(\sqrt{f_t})dt \geq -4D(\sqrt{f_0}) \int_0^\infty e^{-2tK} dt = -\frac{2}{K}D(\sqrt{f_0}). \]

This proves the LSI (13.29) for any smooth function \( f = f_0 \). By a standard density argument, it can be extended to any function \( f \) with finite Dirichlet form, \( D(\sqrt{f}) < \infty \). For functions with unbounded Dirichlet form we interpret the LSI (13.29) as a tautology. In particular, (13.29) holds for \( f_t \) as well, \( S(f_t) \leq (2/K)D(\sqrt{f_t}) \). Inserting this back into (13.32), we have

\[ \partial_t S(f_t) \leq -2KS(f_t). \]

Integrating this inequality from time 0, we obtain the exponential relaxation of the entropy on time scale \( t \simeq 1/K \)

(13.37) 
\[ S(f_t) \leq e^{-2tK}S(f_0). \]

Finally, we can integrate (13.32) from time \( t/2 \) to \( t \) to get

\[ S(f_t) - S(f_{t/2}) = -4 \int_{t/2}^t D(\sqrt{f_r})dr. \]

Using the positivity of the entropy \( S(f_t) \geq 0 \) on the left side and the monotonicity of the Dirichlet form (from (13.34)) on the right side, we get

(13.38) 
\[ D(\sqrt{f_t}) \leq \frac{2}{t} S(f_{t/2}); \]

thus, using (13.37), we obtain \textit{exponential relaxation of the Dirichlet form on time scale} \( t \simeq 1/K \),

\[ D(\sqrt{f_t}) \leq \frac{2}{t} e^{-tK}S(f_0). \]
Standard example. Let \( \mu \) be the centered Gaussian measure with variance \( a^2 \) on \( \mathbb{R}^n \), i.e.,
\[
d\mu(x) = \frac{1}{(2\pi a^2)^{n/2}} e^{-x^2/2a^2} \, dx.
\]

Written in the form \( e^{-\mathcal{H}/Z} \) with the quadratic Hamiltonian \( \mathcal{H}(x) = x^2/2a^2 \), we see that (13.28) holds with \( K = a^{-2} \). Then, (13.29) with the replacement \( f \to f^2 \) yields that
\[
\int f^2 \log f^2 \, d\mu \leq 2a^2 \int (\nabla f)^2 \, d\mu
\]
for any normalized \( f \) with \( \int f^2 \, d\mu = 1 \). The constant of this inequality is optimal.

Alternative formulation of LSI. We remark that there is another formulation of the LSI (see [100]). To make this connection, let
\[
g(x) := f(x) \cdot \frac{1}{(2\pi a^2)^{n/4}} e^{-x^2/4a^2},
\]
and notice that \( \int g^2(x) \, dx = \int f^2 \, d\mu \) where \( dx \) is the Lebesgue measure and \( d\mu \) is of the form (13.39). Rewriting (13.40) to an inequality for \( g \) and with the replacement \( 2\pi a^2 \to a^2 \), we get the following version of the LSI:
\[
\int_{\mathbb{R}^n} g^2 \log(g^2/\|g\|^2) \, dx + n[1 + \log a] \int_{\mathbb{R}^n} g^2 \, dx \leq (a^2/\pi) \int_{\mathbb{R}^n} |\nabla g|^2 \, dx,
\]
which holds for any \( a > 0 \) and any function \( g \), where \( \|g\| = (\int g^2 \, dx)^{1/2} \) is the \( L^2 \)-norm with respect to the Lebesgue measure.

**Proposition 13.7 (LSI implies spectral gap).** Let \( \mu \) satisfy the LSI (13.27) with an LSI constant \( \gamma \). Then, for any \( v \in L^2(\mu) \) with \( \int v \, d\mu = 0 \), we have
\[
\int v^2 \, d\mu \leq \frac{\gamma}{2} \int |\nabla v|^2 \, d\mu = \frac{\gamma}{2} D(v);
\]
i.e., \( \mu \) has a spectral gap of size at least \( \gamma/2 \).

**Proof.** By definition of the LSI constant, we have
\[
\int u \log u \, d\mu \leq \gamma D(\sqrt{u})
\]
for any \( u \) with \( \int u \, d\mu = 1 \). For any bounded, smooth function \( v \) with \( \int v \, d\mu = 0 \), define \( u = 1 + \varepsilon v \). Then, we have
\[
\varepsilon^{-2} \int (1 + \varepsilon v) \log(1 + \varepsilon v) \, d\mu \leq \frac{\gamma}{4} \int \frac{|\nabla v|^2}{1 + \varepsilon v} \, d\mu.
\]
Taking the limit \( \varepsilon \to 0 \), we get that the right-hand side converges to \( \frac{1}{2} \int v^2 \, d\mu \) by dominated convergence. This proves the proposition. \( \square \)
PROPOSITION 13.8 (Concentration inequality (Herbst bound)). Suppose that the measure $\mu$ satisfies the LSI with a constant $\gamma$. Let $F$ be a function with $\mathbb{E}^{\mu} F = 0$. Then, we have

\begin{equation}
\mathbb{E}^{\mu} e^F \leq \exp \left( \frac{\gamma}{4} \| \nabla F \|_\infty^2 \right)
\end{equation}

where

\[ \| \nabla F \|_\infty := \sup_x \sqrt{\sum_i | \partial_i F(x) |^2}. \]

In particular, we have

\begin{equation}
\mathbb{P}^{\mu}(|F| \geq \alpha) \leq \exp \left( - \frac{\alpha^2}{\gamma \| \nabla F \|_\infty^2} \right)
\end{equation}

for any $\alpha > 0$.

Notice that we get an exponential tail estimate from the LSI. If we only have the spectral gap estimate, (13.43), we can only bound the variance of $F$ that yields a quadratic tail estimate

\[ \mathbb{P}^{\mu}(|F| \geq \alpha) \leq \frac{\gamma \| \nabla F \|_\infty^2}{2\alpha^2}. \]

In our typical applications, we have $\gamma \| \nabla F \|_\infty^2 \asymp N^{-1}$. We often need to control the concentration of many ($\asymp N^C$) different functions $F$ in parallel. Thus the simple union bound is applicable with the LSI but not with the spectral gap estimate.

PROOF. Denote by

\[ u = u(t) := \frac{\exp(e^t F)}{\mathbb{E}^{\mu} \exp(e^t F)}. \]

By differentiation and the LSI, we have

\begin{equation}
\frac{d}{dt} [e^{-t} \log \mathbb{E}^{\mu} \exp(e^t F)] = e^{-t} \mathbb{E}^{\mu} u \log u \leq e^{-t} \gamma \mathbb{E}^{\mu} |\nabla \sqrt{u}|^2.
\end{equation}

Clearly,

\begin{equation}
\mathbb{E}^{\mu} |\nabla \sqrt{u}|^2 \leq \frac{e^{2t}}{4} \| \nabla F \|_\infty^2.
\end{equation}

Integrating from any $t < 0$ to 0 yields that

\begin{equation}
[\log \mathbb{E}^{\mu} \exp(F)] - [e^{-t} \log \mathbb{E}^{\mu} \exp(e^t F)] = \frac{\gamma}{4} \| \nabla F \|_\infty^2 \int_t^0 ds \, e^s.
\end{equation}

From the condition $\mathbb{E}^{\mu} F = 0$, we have

\[ \lim_{t \to -\infty} [e^{-t} \log \mathbb{E}^{\mu} \exp(e^t F)] = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \log \mathbb{E}^{\mu} e^{\epsilon F} \]

\[ = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \log \mathbb{E}^{\mu} [1 + \epsilon F + O(\epsilon^2 e^{\epsilon F})] = 0 \]
by dominated convergence. This proves the inequality (13.45). The concentration bound (13.46) follows from an exponential Markov inequality:

\[ \mathbb{P}_\mu(F \geq \alpha) \leq \mathbb{E}_\mu e^{(F-\alpha)} \leq \exp \left( -\alpha t + \frac{t^2}{4} \| \nabla F \|_\infty^2 \right), \]

where we chose the optimal \( t = \frac{2\alpha}{\| \nabla F \|_\infty^2} \) in the last step. Changing \( F \to -F \), we obtain the opposite bound and thus we have proved (13.46). □

**Proposition 13.9 (Stability of the LSI constant).** Consider two measures \( \mu, \nu \) on \( \mathbb{R}^N \) that are related by \( \nu = g\mu \) with some bounded function \( g \). Let \( \gamma_\mu \) denote the LSI constant for \( \mu \). Then

\[ \gamma_\nu \leq \| g \|_\infty \| g^{-1} \|_\infty \gamma_\mu. \]

**Proof.** Take an arbitrary function \( f \geq 0 \) with \( \int f \, d\nu = 1 \). Let \( \alpha := \int f \, d\mu \leq \| g^{-1} \|_\infty \) and by definition of the entropy we have

\[ S_\mu(f/\alpha) = \int (f/\alpha) \log(f/\alpha) \, d\mu. \]

The following inequality for any two nonnegative numbers \( a, b \) can be checked by elementary calculus:

\[ a \log a - b \log b - (1 + \log b)(a - b) \geq 0. \]

Hence,

\[ \int [f \log f - \alpha \log \alpha - (1 + \log \alpha)(f - \alpha)] \, d\nu \leq \| g \|_\infty \int [f \log f - \alpha \log \alpha - (1 + \log \alpha)(f - \alpha)] \, d\mu \]

\[ = \| g \|_\infty \alpha S_\mu(f/\alpha). \]

The left-hand side of the last inequality is equal to

\[ S_\nu(f) - [\log \alpha - (\alpha - 1)] \geq S_\nu(f), \]

where we have used the concavity of log. This proves that

\[ S_\nu(f) \leq \| g \|_\infty \alpha S_\mu(f/\alpha). \]

Suppose that the LSI holds for \( \mu \) with constant \( \gamma_\mu \). Then, we have

\[ S_\nu(f) \leq \| g \|_\infty \gamma_\mu \int (\nabla \sqrt{f/\alpha})^2 \, d\mu = \| g \|_\infty \gamma_\mu \int (\nabla \sqrt{f})^2 g^{-1} \, d\nu \]

\[ \leq \| g \|_\infty \| g^{-1} \|_\infty \gamma_\mu \int (\nabla \sqrt{f})^2 \, d\nu, \]

and this proves the lemma. □

**Proposition 13.10 (Tensorial property of the LSI).** Consider two probability measures \( \mu, \nu \) on \( \mathbb{R}^n, \mathbb{R}^m \), respectively. Suppose that the LSI holds for them with LSI constants \( \gamma_\mu \) and \( \gamma_\nu \), respectively. Let \( \omega = \mu \otimes \nu \) be the product measure on \( \mathbb{R}^{m+n} \). Then, the LSI holds for \( \omega \) with LSI constant \( \gamma_\omega \leq \max \{ \gamma_\mu, \gamma_\nu \}. \)
Proof. We will use the notation introduced in Section 13.2 with \( \Omega_1 = \mathbb{R}^n \), \( \Omega_2 = \mathbb{R}^m \). Recall the additivity of entropy (13.20) w.r.t. martingale decomposition. In the current situation, \( \omega = \mu \otimes \nu \) and thus \( \hat{\omega} = \mu \) and \( \omega_x = \nu \) for any \( x \). Furthermore,

\[
\hat{f}(x) = \hat{\omega}(x)^{-1} \int f(x, y)\omega(x, y)dy = \int f(x, y)\nu(y)dy.
\]

By the additivity of entropy and the LSI w.r.t. \( \mu \) and \( \nu \), we have

\[
S_\omega(f) = \mathbb{E}\hat{f}\mu S_\nu(f_x) + S_\mu(\hat{f}) \leq \gamma_\nu \int \hat{f}(x)\mu(x)dx \int \frac{|\nabla_y \sqrt{f(x, y)}|^2}{\hat{f}(x)}\nu(y)dy
\]

\[
+ \gamma_\mu \int \left| \nabla_x \sqrt{\int f(x, y)\nu(y)dy} \right| \mu(x)dx.
\]

The integral in the first term on the right-hand side is equal to

\[
\iint |\nabla_y \sqrt{f(x, y)}|^2 \mu(x)\nu(y)dx\,dy.
\]

The integral of the second term on the right-hand side is bounded by

\[
\int \frac{\left| \nabla_x f(x, y)\nu(y)dy \right|^2}{4f(x, y)\nu(y)dy} \mu(x)dx \leq \iint |\nabla_x \sqrt{f(x, y)}|^2 \mu(x)\nu(y)dx\,dy,
\]

where we have written \( \nabla_x f = 2(\nabla_x \sqrt{f})\sqrt{f} \) and used the Schwarz inequality. Summarizing, we have proved that

\[
S_\omega(f) \leq \gamma_\nu \int |\nabla_y \sqrt{f(x, y)}|^2 \omega(x, y)dx\,dy + \gamma_\mu \int |\nabla_x \sqrt{f(x, y)}|^2 \omega(x, y)dx\,dy,
\]

and this proves the proposition. \( \square \)

The following lemma is a useful tool to control the entropy flow w.r.t. non-equilibrium measure.

**Lemma 13.11** ([140]). Suppose we have evolution equation \( \partial_t f_t = \mathcal{L} f_t \) with \( \mathcal{L} \) defined via the Dirichlet form \( \langle f, (-\mathcal{L})f \rangle_\mu = \sum_j f(\partial_j f)^2 d\mu \). Then, for any time-dependent probability density \( \psi_t \) w.r.t. \( \mu \), we have the entropy flow identity

\[
\partial_t S_\mu(f_t|\psi_t) = -4 \sum_j \int (\partial_j \sqrt{g_t})^2 \psi_t d\mu + \int g_t(\mathcal{L} - \partial_t)\psi_t d\mu
\]

where \( g_t := f_t/\psi_t \) and

\[
S_\mu(f_t|\psi_t) := \int f_t \log \frac{f_t}{\psi_t} d\mu = S(f_t\mu|\psi_t\mu)
\]

is the relative entropy of \( f_t\mu \) w.r.t. \( \psi_t\mu \).
PROOF. A simple computation then yields that
\[
\frac{d}{dt} S_\mu(f_t | \psi_t) = \int (\mathcal{L} f_t)(\log g_t) d\mu + \int f_t \frac{\mathcal{L} f_t}{f_t} d\mu - \int (\partial_t \psi_t) g_t \, d\mu
\]
\[
= \int (\mathcal{L} f_t)(\log g_t) d\mu - \int (\partial_t \psi_t) g_t \, d\mu
\]
\[
= \int (g_t \psi_t) \mathcal{L} (\log g_t) d\mu - \int g_t \partial_t \psi_t \, d\mu
\]
\[
= \int \psi_t \left[ g_t \mathcal{L} (\log g_t) - g_t \frac{\mathcal{L} g_t}{g_t} \right] d\mu + \int g_t (\mathcal{L} - \partial_t) \psi_t \, d\mu.
\]
By definition of $\mathcal{L}$, we have
\[
(13.52) \quad \mathcal{L} (\log g) - \frac{\mathcal{L} g}{g} = - \sum \left( \frac{\partial_j g}{g} \right)^2 = -4 \sum \left( \frac{\partial_j \sqrt{g}}{g} \right)^2,
\]
and we have proved Lemma 13.11.

We remark that stochastic processes and their generators can be defined in more general setups, e.g., the underlying probability spaces may be different from $\mathbb{R}^N$ and the generators may involve discrete jumps. In this case, the stochastic generator $\mathcal{L}$ is defined directly, without an a priori Dirichlet form. It can, however, be proved that $-g[\mathcal{L} (\log g) - (\mathcal{L} g)/g] \geq 0$, which can then be viewed as a generalization of the “Dirichlet form” associated with a generator.

### 13.4. Hypercontractivity

We now present an interesting connection between the LSI of a probability measure $\mu$ and the hypercontractivity properties of the semigroup generated by $\mathcal{L} = \mathcal{L}_\mu$. Since this result will not be used later in this book, this section can be skipped.

To state the result, we define the semigroup $\{P_t\}_{t \geq 0}$ by $P_t f := f_t$, where $f_t$ solves the equation $\partial_t f_t = \mathcal{L} f_t$ with initial condition $f_0 = f$.

**Theorem 13.12 (L. Gross [79]).** For a measure $\mu$ on $\mathbb{R}^n$ and for any fixed constants $\beta \geq 0$ and $\gamma > 0$ the following two statements are equivalent:

(i) **The generalized LSI**

\[
(13.53) \quad \int f \log f \, d\mu \leq \gamma \int |\nabla \sqrt{f}|^2 \, d\mu + \beta \quad \text{for any } f \geq 0 \text{ with } \int f \, d\mu = 1
\]
holds.

(ii) **The hypercontractivity estimate**

\[
(13.54) \quad \|P_t f\|_{L^q(\mu)} \leq \exp \left\{ \beta \left[ \frac{1}{p} - \frac{1}{q} \right] \right\} \|f\|_{L^p(\mu)}
\]
holds for all exponents satisfying
\[
\frac{q-1}{p-1} \leq e^{4t/\gamma}, \quad 1 < p \leq q < \infty.
\]
13.4. HYPERCONTRACTIVITY

**Proof.** We will only prove (i) \(\Rightarrow\) (ii), i.e., that the generalized LSI implies the decay estimate, the proof of the converse statement can be found in [43]. First we assume that \(f \geq 0\), hence \(f_t \geq 0\). Direct differentiation yields the identity

\[
\frac{d}{dt} \log \|f_t\|_{p(t)} = \frac{\dot{p}(t)}{p(t)^2} \left[ -\frac{4(p(t) - 1)}{\dot{p}(t)} D(u(t)) + \int u(t)^2 \log(u(t)^2) d\mu \right]
\]

with \(\dot{p}(t) = \frac{d}{dt} p(t)\) and where we defined

\[
u(t) := f_t^{p(t)/2} \|f_t\|_{p(t)}^{-p(t)/2}, \quad D(u) = \int (\nabla u)^2 d\mu.
\]

Now choose \(p(t)\) to solve the differential equation

\[
\gamma = \frac{4(p(t) - 1)}{\dot{p}(t)} \text{ with } p(0) = p, \text{ i.e., } p(t) = 1 + (p - 1)e^{4t/\gamma},
\]

where \(\gamma\) is the constant given in the theorem. Using (13.53) for the \(L^1(\mu)\)-normalized function \(u(t)^2\), we have

\[
\frac{d}{dt} \log \|f_t\|_{p(t)} \leq \beta \frac{\dot{p}(t)}{p(t)^2}.
\]

Integrating both sides, from \(t = 0\) to some \(T\) we have

\[
\log \|f_T\|_{p(T)} - \log \|f\|_p \leq \beta \left[ \frac{1}{p(T)} - \frac{1}{p} \right].
\]

Choosing \(T\) such that \(p(T) = q\), we have proved (13.54) for \(f \geq 0\). The general case follows from separating the positive and negative parts of \(f\). \(\square\)

**Exercise.** In this exercise, we show that the idea of the LSI can be useful even if the invariant measure is not a probability measure. The sketch below follows the paper by Carlen-Loss [33], and it works for any parabolic equation of the type

\[
\partial_t f_t = [\nabla \cdot (D(x,t)\nabla) + b(x,t) \cdot \nabla] f_t
\]

for any divergence free \(b\) and \(D(x,t) \geq c > 0\). For simplicity of notation, we consider only the heat equation on \(\mathbb{R}^n\)

\[
(13.56) \quad \partial_t f_t = \Delta f_t.
\]
The invariant measure of this flow is the standard Lebesgue measure on $\mathbb{R}^n$.

(i) Check the formula (13.55) in this setup, i.e., show that the following identity holds:

$$
\frac{d}{dt} \log \|f_t\|_p(t) = \frac{1}{p(t)} \|f_t\|_p^{-p(t)} \frac{d}{dt} \int f_t^p(t) \, dx \cdot \frac{\dot{p}(t)}{p(t)^2} \log(\|f_t\|_p^p) \\
= \frac{\dot{p}(t)}{p(t)^2} \left[ -\frac{4(p(t) - 1)}{\dot{p}(t)} \int (\nabla u(t))^2 \, dx + \int u(t)^2 \log(u(t)^2) \, dx \right],
$$

where all norms are w.r.t. the Lebesgue measure and

$$
u(t) = f_t^{p/2} \|f_t\|_p^{-p/2}, \quad p = p(t).$$

(ii) For any $t$ fixed, choose the constant in (13.42) by setting

$$a^2 = \frac{4(p(t) - 1)}{\dot{p}(t)}.$$

Thus we have

$$\frac{d}{dt} \log \|f_t\|_p(t) \leq -\frac{n \dot{p}(t)}{p(t)^2} \left[ 1 + \frac{1}{2} \log \left( \frac{4\pi(p(t) - 1)}{\dot{p}(t)} \right) \right].$$

For a suitable choice of the function $p(t)$ with $p(0) = 1$, $p(T) = \infty$, show that

$$\|f_T\|_\infty \leq (4\pi T)^{-n/2} \|f_0\|_1.$$

We remark that in the case of $D \equiv 1$ and $b \equiv 0$, this heat kernel estimate is also a trivial consequence of the explicit formula for the heat kernel $e^{t \Delta}(x, y)$. The point is, as remarked earlier, that this proof works for a general class of second-order parabolic equations.

### 13.5. Brascamp-Lieb Inequality

The following inequality will not be needed for the main result of this book, so the reader may skip it. Nevertheless, we included it here since it is an important tool in the analysis of probability measures in very high dimensions and was used in universality proofs for the invariant ensembles [25]. We will formulate it on $\mathbb{R}^N$, and in Section 13.7 we comment on how to extend it to certain probability measures on the simplex $\Sigma_N$ defined in (12.3).

**Theorem 13.13 (Brascamp-Lieb inequality [29]).** Consider a probability measure $\mu$ on $\mathbb{R}^N$ of the form (13.24), i.e., $\mu = e^{-\mathcal{H}}/Z$. Suppose that the Hamiltonian is strictly convex, i.e.,

$$\nabla^2 \mathcal{H}(x) \geq K > 0$$

(13.59)
as a matrix inequality for some positive constant $K$. Then, for any smooth function $f \in L^2(\mu)$, we have

\[(13.60) \quad \langle f; f \rangle_\mu \leq \langle \nabla f, [\nabla^2 \mathcal{H}]^{-1} \nabla f \rangle_\mu.\]

Recall that $\langle f, g \rangle_\mu = \int f g \, d\mu$ denotes the scalar product and $\langle f; g \rangle_\mu = \langle f, g \rangle_\mu - \langle 1, f \rangle_\mu \langle 1, g \rangle_\mu$ is the covariance. With a slight abuse of notation, we also use the notation $\langle F, G \rangle_\mu = \sum_{i=1}^N \langle F_i, G_i \rangle_\mu$ for the scalar product of any two vector-valued functions $F, G : \mathbb{R}^N \to \mathbb{R}^N$; this extended scalar product is used in the right-hand side of (13.60).

**Remark.** Notice that the Brascamp-Lieb inequality is optimal in the sense that, if $\mu$ is a Gaussian measure, then we can find $f$ so that the inequality becomes equality. Moreover, if we use the matrix inequality $\nabla^2 \mathcal{H} \geq K$, then (13.60) is reduced to the spectral gap inequality (13.43), but of course (13.60) is stronger. The following proof is due to Helffer-Sjöstrand [81] and Naddaf-Spencer [109].

**Proof of Theorem 13.13.** Recall that $\mathcal{L}$ is the generator for a reversible dynamics with reversible measure $\mu$ defined in (13.26). We have for the dynamics (13.30)

\[
\frac{d}{dt} \langle f, e^{t\mathcal{L}} g \rangle = \langle f, \mathcal{L} e^{t\mathcal{L}} g \rangle = -\sum_j \langle \partial_{x_j} f, \partial_{x_j} e^{t\mathcal{L}} g \rangle.
\]

Define

\[
G(t,x) := (G_1(t,x), \ldots, G_N(t,x)), \quad G_j(t,x) := \partial_{x_j} [e^{t\mathcal{L}} g(x)].
\]

Recall the explicit form of $\mathcal{L}$:

\[
\mathcal{L} = \Delta - (\nabla \mathcal{H}) \cdot \nabla = \Delta + \sum_k b_k \partial_{x_k}, \quad b_k := -\partial_{x_k} \mathcal{H}.
\]

Define a new generator $\mathcal{L}$ on any vector-valued function $G : \mathbb{R}^N \to \mathbb{R}^N$ by

\[
(\mathcal{L}G)_j(x) := \mathcal{L}G_j(x) + \sum_k (\partial_{x_j} b_k) G_k(x).
\]

Clearly, by explicit computation, we have

\[
\partial_t G(t,x) = \mathcal{L}G(t,x).
\]

From the decay estimate (13.31), it follows that the dynamics are mixing, i.e., $\lim_{t \to \infty} \langle f, e^{t\mathcal{L}} g \rangle = 0$ for any smooth function $f$ with $\langle f, f \rangle d\mu = 0$. Thus for
any such function $f$ we have
\[
\langle f, g \rangle_\mu = -\int_0^\infty \frac{d}{dt} \langle f, e^{tL} g \rangle_\mu \, dt = \sum_j \int_0^\infty \langle \partial_x^j f, G_j(t, x) \rangle_\mu \, dt = \int_0^\infty \langle \nabla f, e^{tL} \nabla g \rangle_\mu \, dt = \langle \nabla f, (-L)^{-1} \nabla g \rangle_\mu.
\]

Taking $g = f$ and from the definition of $L$, we have
\[-(L)_{i,j} = -\mathcal{L} \delta_{i,j} - (\partial_x^i b_j) \geq (\partial_x^i \partial_x^j \mathcal{J})\]
as an operator inequality, and we have proved that
\[
\langle f, f \rangle_\mu = \langle \nabla f, (-L)^{-1} \nabla f \rangle_\mu \leq \langle \nabla f, [\nabla^2 \mathcal{J}]^{-1} \nabla f \rangle_\mu.
\]
\[\square\]

13.6. Remarks on the Applications of the LSI to Random Matrices

The Herbst bound provides strong concentration results for probability measures that satisfy the LSI. In this section we exploit this connection for random matrices. For definiteness, we will consider the Gaussian orthogonal ensemble (GOE), although the arguments below can be extended to more general Wigner as well as invariant ensembles. There are two ways to view Gaussian matrix ensembles in the context of the LSI: we can either work on the probability space of the matrix $H$ with Gaussian matrix elements, or we can work directly on the space of the eigenvalues $\lambda$ by using the explicit formula (4.12) for invariant ensembles. Both measures satisfy the LSI (in Section 13.7 we will comment on how to generalize the LSI from $\mathbb{R}^N$ to the simplex $\Sigma_N$, (12.3), a version of the LSI that we will actually use). We will explore both possibilities, and we start with the matrix point of view.

13.6.1. LSI from the Wigner matrix point of view. Let $\mu = \mu_G$ be the standard Gaussian measure $\mu \sim \exp (-N \text{Tr} H^2)$ on symmetric matrices. Notice that for this measure the family \{\(x_{ij} = N^{1/2} h_{ij}, i \leq j\)\} is a collection of independent standard Gaussian random variables (up to a factor 2). Hence the LSI holds for every matrix element and by its tensorial property, i.e., Proposition 13.10, the LSI holds for any function of the full matrix considered as a function of this collection \{\(x_{ij} = N^{1/2} h_{ij}, i \leq j\)\}. In particular, using the spectral gap estimate (13.43) for the Gaussian variables $x_{ij}$, we have for any function $F = F(H)$

\[
\langle F(H); F(H) \rangle_\mu \leq C \sum_{i \leq j} \int |\partial_x^i F(H)|^2 \, d\mu = \frac{C}{N} \sum_{i \leq j} \int |\partial_h^i F(H)|^2 \, d\mu,
\]

where the additional $N^{-1}$ factor comes from the scaling $h_{ij} = N^{-1/2} x_{ij}$. 

Suppose that \( F(H) = R(\lambda(H)) \) is a real function of the eigenvalues \( \lambda = (\lambda_1, \ldots, \lambda_N) \); then by the chain rule we have

\[
\frac{C}{N} \sum_{i \leq j} \int |\partial_{h_{ij}} R(\lambda(H))|^2 \, d\mu = \frac{C}{N} \sum_{i \leq j} \left| \sum_{\alpha} \partial_{\lambda_{\alpha}} R(\lambda) \frac{\partial \lambda_{\alpha}}{\partial h_{ij}} \right|^2 \, d\mu.
\]

Expanding the square and using the perturbation formula (12.9) with real eigenvectors, we can compute

\[
\frac{1}{N} \sum_{i \leq j} \left| \sum_{\alpha} \partial_{\lambda_{\alpha}} R(\lambda) \frac{\partial \lambda_{\alpha}}{\partial h_{ij}} \right|^2 \, d\mu \\
\leq \frac{1}{N} \sum_{i \leq j} \frac{2}{2 - \delta_{ij}} \int \left| \sum_{\alpha} \partial_{\lambda_{\alpha}} R(\lambda) \frac{\partial \lambda_{\alpha}}{\partial h_{ij}} \right|^2 \, d\mu = \\
\frac{1}{N} \sum_{i \leq j} \frac{2}{2 - \delta_{ij}} \sum_{\alpha, \beta} \int \partial_{\lambda_{\alpha}} R(\lambda) \partial_{\lambda_{\beta}} R(\lambda) \frac{\partial \lambda_{\alpha}}{\partial h_{ij}} \frac{\partial \lambda_{\beta}}{\partial h_{ij}} \, d\mu = \\
\frac{2}{N} \sum_{i \leq j} \int \partial_{\lambda_{\alpha}} R(\lambda) \partial_{\lambda_{\beta}} R(\lambda) u_{\alpha}(i) u_{\alpha}(j) u_{\beta}(i) u_{\beta}(j) \, d\mu = \\
\frac{2}{N} \sum_{i, j} \sum_{\alpha, \beta} \int \partial_{\lambda_{\alpha}} R(\lambda) \partial_{\lambda_{\beta}} R(\lambda) u_{\alpha}(i) u_{\alpha}(j) \frac{\partial \lambda_{\alpha}}{\partial h_{ij}} \frac{\partial \lambda_{\beta}}{\partial h_{ij}} \, d\mu = \\
\frac{2}{N} \sum_{\alpha} \int |\partial_{\lambda_{\alpha}} R(\lambda)|^2 \, d\mu,
\]

where we have used the orthogonality property and the normalization convention of the eigenvectors in the last step.

We remark that the annoying factor \([2 - \delta_{ij}]\) can be avoided if we first consider \( \lambda_{\alpha} \) as a function of all \( \{x_{ij} : 1 \leq i, j \leq N\} \) as independent variables. Then the perturbation formula (12.9) becomes

\[
\frac{\partial \lambda_{\alpha}}{\partial h_{ij}} \bigg|_{h_{ij} = h_{ji}} = u_{\alpha}(i) u_{\alpha}(j),
\]

i.e., the derivative evaluated on the submanifold of Hermitian matrices. In this way, we can keep the summation in (13.61) unrestricted, and up to a constant factor, we will get the same final result as in (13.62).

In summary, we proved

\[
\langle F; F \rangle_{\mu} = \langle R(\lambda(H)); R(\lambda(H)) \rangle_{\mu} \leq \frac{C}{N} \sum_{\alpha} \int |\partial_{\lambda_{\alpha}} R(\lambda)|^2 \, d\mu.
\]

Notice that this argument holds for any generalized Wigner matrix as long as a spectral gap estimate (13.43) holds for the distribution of every rescaled matrix element \( N^{1/2}h_{ij} \). Furthermore, it can be generalized for Wigner matrices with Bernoulli random variables for which there is a spectral gap (and LSI) in discrete form. Similar remarks apply to the following LSI estimates.
To see how good this estimate is, consider a local linear statistics of eigenvalues, i.e., the local average of $K$ consecutive eigenvalues with label near $M$, by setting

$$R(\lambda) := \frac{1}{K} \sum_{\alpha} A \left( \frac{\alpha - M}{K} \right) \lambda_{\alpha},$$

where $A$ is a smooth function of compact support and $1 \leq M, K \leq N$. Computing the right-hand side of (13.64) and combining it with (13.61), we get

$$\langle F; F \rangle_\mu \leq \frac{C}{NK^2} \sum_{\alpha} A^2 \left( \frac{\alpha - M}{K} \right) \leq \frac{C_A}{NK}$$

where the last constant $C_A$ depends on the function $A$. For $K = 1$, this inequality estimates the square of the fluctuation of a single eigenvalue (choosing $A$ appropriately). The bound (13.65) is off by a factor $N$ since the true fluctuation is of order almost $1/N$ by rigidity, Theorem 11.5, at least in the bulk, i.e., if $\delta N \leq M \leq (1 - \delta)N$. On the other hand, for $K = N$ the bound (13.65) is much more precise; it shows that the variance of a macroscopic average of the eigenvalues is at most of order $N^{-2}$. This is the correct order of magnitude; in fact, it is known that $\sum_{\alpha} A \left( \frac{\alpha}{N} \right) \lambda_{\alpha}$ converges to a Gaussian random variable (see, e.g., [103, 118] and references therein). Hence, the spectral gap argument yields the optimal (up to a constant) result for any macroscopic average of eigenvalues.

Another common quantity of interest is the Stieltjes transform of the empirical eigenvalue distribution, i.e.,

$$F(H) = G(\lambda(H)) \quad \text{with} \quad G(\lambda) = m_N(z) = \frac{1}{N} \sum_{\alpha} \frac{1}{\lambda_{\alpha} - z}, \quad z = E + i\eta.$$

In this case, using (13.64), we can estimate

$$\langle m_N(z); m_N(z) \rangle_\mu \leq \frac{C}{N^2} \sum_{\alpha} \int |\partial_{\lambda_{\alpha}} G(\lambda)|^2 d\mu$$

$$= \frac{C}{N^3} \sum_{\alpha} \int \frac{1}{|\lambda_{\alpha} - z|^4} d\mu \leq \frac{C}{N^2 \eta^4}$$

where we used $|\lambda_{\alpha} - z| \geq \text{Im } z = \eta$.

This shows that the variance of $m_N(z)$ vanishes if $\eta \gg N^{-1/2}$. The last estimate can be improved if we know that the density of the empirical measure is bounded. Roughly speaking, if we know that in the summation over $\alpha$, the number of eigenvalues in an interval of size $\eta$ near $E$ is at most of order $N \eta$; then the last estimate can be improved to

$$\frac{1}{N^3} \sum_{\alpha} \int \frac{1}{|\lambda_{\alpha} - z|^4} d\mu \leq \frac{1}{N^3} \sum_{|\lambda_{\alpha} - E| \leq \eta} \int \frac{1}{|\lambda_{\alpha} - z|^4} d\mu$$

$$\leq \frac{C}{N^3 (N \eta)} \frac{1}{\eta^4} = \frac{C}{N^2 \eta^3}.$$
This shows that the variance of $m_N(z)$ vanishes if $\eta \gg N^{-2/3}$. Since vanishing fluctuations can be used to estimate the density, this argument can actually be made rigorous by a bootstrapping argument [61]. Notice that the scale $\eta \gg N^{-2/3}$ is still far from the resolution demonstrated in the local semicircle law, Theorem 6.7.

Whenever the LSI is available, the variance bounds can be easily lifted to a concentration estimate with exponential tail. We now demonstrate this mechanism for the fluctuation of a single eigenvalue. In other words, we will apply (13.46) with $F(H) = \lambda_\alpha(H) - \mathbb{E}\lambda_\alpha(H)$ for a fixed $\alpha$. From (12.9), we have

$$\sum_{i \leq j} |\nabla x_{ij} F|^2 \leq \frac{2}{N};$$

thus (13.46) implies

$$\mathbb{P}^{\mu}(|\lambda_\alpha - \mathbb{E}\lambda_\alpha| \geq t) \leq \exp(-ct^2)$$

for any $\alpha$ fixed. This inequality has two deficiencies compared with the rigidity bounds in Theorem 11.5. First, the control of $|\lambda_\alpha - \mathbb{E}\lambda_\alpha|$ is only up to $N^{-1/2}$, which is still far from the optimal $N^{-1}$ (in the bulk). Second, it does not give information on the difference between the expectation $\mathbb{E}\lambda_\alpha$ and the corresponding classical location $\gamma_\alpha$ defined as the $\alpha$th $N$-quantile of the limiting density (see (11.31)).

### 13.6.2. LSI from the invariant ensemble point of view.

Now we pass to the second point of view, where the basic measure $\mu_G$ is the invariant ensemble on the eigenvalues. One might hope that the situation can be improved since we look directly at the Gaussian eigenvalue ensemble defined in (12.13) with the Gaussian choice for $V(\lambda) = \frac{1}{2}\lambda^2$. Notice that the role of $\mathcal{H}$ in Theorem 13.6 will be played by $N\mathcal{H}_N$ defined in (12.13) (with $\beta = 1$). The Hessian of $\mathcal{H}_N$ is given by (all inner products and norms in the following equation are w.r.t. the standard inner product in $\mathbb{R}^N$)

$$\langle \nabla^2 \mathcal{H}_N(x)v, v \rangle \geq \frac{2}{N} \|v\|^2 + \frac{1}{4N} \sum_{i<j} \frac{(x_i - x_j)^2}{(x_i - x_j)^2} \geq \frac{1}{2} \|v\|^2, \quad v \in \mathbb{R}^N;$$

thus the convexity bound (13.28) holds with a constant $K = N/2$ for $N\mathcal{H}_N$. Hence, the spectral gap from (13.43) implies that for any function $R(\lambda)$ we have

$$\langle R(\lambda); R(\lambda) \rangle_{\mu_G} \leq \frac{C}{N} \sum_\alpha \int |\delta_{\lambda_\alpha} R(\lambda)|^2 d\mu_G.$$

Notice that this bound is in the same form as in (13.64). A similar statement holds for the LSI; i.e., we have

$$\int R \log R \ d\mu_G \leq \frac{C}{N} \sum_\alpha \int |\delta_{\lambda_\alpha} \sqrt{R(\lambda)}|^2 d\mu_G.$$
We can apply the concentration inequality (13.46) to the function \( R(\lambda) = \lambda^\alpha \) to get (13.69). Once again, we get exactly the same result as considered earlier, so there is no advantage of using the explicit invariant measure \( \mu_G \).

We have seen that in both the matrix and the invariant ensemble setup, even for Gaussian ensembles, concentration estimates based on spectral gap or LSI do not yield optimal results for the eigenvalues on short scales. In the matrix setup, the proof of the optimal rigidity bounds for Wigner matrices, Theorem 11.5, uses a completely different approach independent of the LSI. In the invariant ensemble setup, while the convexity bound (13.70) cannot be improved for a general vector \( v \), it becomes much stronger if we consider it only for vectors \( v \) satisfying \( \sum_j v_j = 0 \) due to the structure of the Hessian (13.70). In the next section, we will explore this idea to improve estimates on certain functions of the eigenvalues.

### 13.7. Extensions to the Simplex; Regularization of the DBM

In the above application to either spectral gap or LSI, the measure \( \mu_G \) was restricted to the subset \( \Sigma = \Sigma_N = \{ x \in \mathbb{R}^N : x_1 < \cdots < x_N \} \). It is absolutely continuous with respect to the Lebesgue measure on \( \Sigma_N \), but if we write it in the form (13.24), then the Hamiltonian \( \mathcal{H} \) has to be infinite outside of \( \Sigma_N \); in particular, it is not differentiable, a property that is implicitly assumed in Section 13.3. This issue is closely related to the boundary terms in the integration by parts, especially in (13.33), that arise if one formally extends the proofs to \( \Sigma \). Therefore, the results of Theorem 13.6 do not apply directly. The proper procedure involves an approximation and extension of the measure from \( \Sigma_N \) to \( \mathbb{R}^N \) by using the results of Section 13.3 for the regularized measure on \( \mathbb{R}^N \) and then removing the regularization. Whether the regularization can be removed for any \( \beta > 0 \) or only for \( \beta \geq 1 \) depends on the statement, as we explain now.

For simplicity, we work in the setup of the Gaussian \( \beta \)-ensemble; i.e., we set

\[
\mathcal{H}(x) = N \sum_{i=1}^{N} \frac{1}{4} x_i^2 - \frac{1}{N} \sum_{i<j} \log(x_j - x_i)
\]

and define \( \mu_\beta(x) = Z_\mu^{-1} e^{-N\beta V(x)} \) on the simplex \( \Sigma_N \), exactly as in (12.13) with the Gaussian potential \( V(x) = \frac{1}{2} x^2 \) and with any parameter \( \beta > 0 \). As usual, \( Z_\mu \) is a normalization constant.

Recall from (12.4) that the DBM on the simplex \( \Sigma_N \) is defined via the stochastic differential equation

\[
(13.73) \quad dx_i = \frac{\sqrt{2}}{\sqrt{\beta N}} dB_i + \left( -\frac{1}{2} x_i + \frac{1}{N} \sum_{j \neq i} \frac{1}{x_i - x_j} \right) dt \quad \text{for } i = 1, \ldots, N,
\]

where \( B_1, \ldots, B_N \) is a family of independent standard Brownian motions. The unique strong solution exists only if \( \beta \geq 1 \) (Theorem 12.2) with equilibrium measure \( \mu_\beta \). Since DBM does not exist on \( \Sigma_N \) unless \( \beta \geq 1 \), we cannot use DBM either in the SDE form (12.4) or in the PDE form (12.17) when \( \beta < 1 \) (see a
remark below (12.17)). On the other hand, certain results may be extended to any \( \beta > 0 \) if their final formulations do not involve DBM.

In this section, we present a regularization procedure to show that substantial parts of the main results of Theorem 13.6, i.e., the LSI for any \( \beta > 0 \) and exponential relaxation decay of the entropy for \( \beta \geq 1 \), remain valid on the simplex \( \Sigma_N \). A similar generalization holds for the Brascamp-Lieb inequality. In the next section, the same regularization will be used to show that the key Dirichlet form inequality (Theorem 14.3) also holds for \( \beta > 0 \).

For later applications, we work with a slightly bigger class of measures than just \( \mu_\beta \). We consider measures on \( \Sigma = \Sigma_N \) of the form

\[
\omega = Z_\omega^{-1} e^{-\beta N \hat{H}_\omega} \mu_\beta,
\]

where \( \hat{H}(x) = \sum_j U_j(x_j) \) for some convex real valued functions \( U_j \) on \( \mathbb{R} \). The total Hamiltonian of \( \omega \) is \( H_\omega := H + \hat{H} \). Note that \( U_j \) are defined and convex on the entire \( \mathbb{R}^N \). The entropy and the Dirichlet form are defined as before:

\[
S_\omega(f) = \int_{\Sigma} f \log f \, d\omega, \quad D_\omega(f) = \frac{1}{\beta N} \int_{\Sigma} |\nabla f|^2 \, d\omega.
\]

The corresponding DBM is given by

\[
(dx_i = \frac{\sqrt{2}}{\sqrt{\beta N}} dB_i + \left(-\frac{1}{2} x_i - U'_i(x_i) + \frac{1}{N} \sum_{j \neq i} \frac{1}{x_i - x_j}\right) dt).
\]

We assume a lower bound on the Hessian of the total Hamiltonian,

\[
H''_\omega = H'' + \hat{H}'' \geq \frac{1}{2} + \min \min_{x \in \mathbb{R}} U''_j(x) \geq K
\]

for some positive constant \( K \) on the entire \( \mathbb{R}^N \). This bound (13.75) plays the role of (13.28). Let \( D_\omega, S_\omega, \) and \( L_\omega \) denote the Dirichlet form, entropy, and generator corresponding to the measure \( \omega \). Now we claim that Theorem 13.6 holds for the measure \( \omega \) on \( \Sigma_N \) in the following form:

**Theorem 13.14.** Assume (13.75). Then for \( \beta > 0 \), the LSI holds, i.e.,

\[
S_\omega(f) \leq 2 \frac{D_\omega(\sqrt{f})}{K}
\]

for any nonnegative normalized density \( f \) on \( \Sigma_N \), \( \int f \, d\omega = 1 \), that satisfies \( f \in L^\infty \) and \( \nabla \sqrt{f} \in L^\infty \). For \( \beta \geq 1 \) the requirement that \( f \) and \( \nabla \sqrt{f} \) are bounded can be removed.

Moreover, the Brascamp-Lieb inequality also holds for any \( \beta > 0 \); i.e., for any bounded function \( f \in L^2(\Sigma_N, d\omega) \) we have

\[
\langle f; f \rangle_\omega \leq \langle \nabla f, [H''_\omega]^{-1} \nabla f \rangle_\omega.
\]

For \( \beta \geq 1 \) the requirement that \( f \) be bounded can be removed.
Furthermore, for any $\beta \geq 1$ the dynamics $\delta_{t} f_t = \mathcal{L}_\omega f_t$ with initial condition $f_0 = f$ is well-defined on $\Sigma_N$, and it approaches to equilibrium in the sense that

$$S_\omega(f_t) \leq e^{-2tK}S_\omega(f).$$

(13.78)

The inequalities above are understood in the usual sense that they are relevant only when the right-hand side is finite. Moreover, by a standard density argument, they extend to the closure of the corresponding spaces. For example, (13.76) holds for any $f$ that can be approximated by a sequence of bounded normalized densities $f_n \in L^\infty$ with $\nabla \sqrt{f_n} \in L^\infty$ such that $D_\omega(\sqrt{f_n} - \sqrt{f}) \to 0$.

Before starting the formal proof, we explain the key idea. For the proofs of (13.76) and (13.77), we extend the measure $\omega$ from $\Sigma$ to the entire $\mathbb{R}^N$ in a continuous way by relaxing the strict ordering $x_i < x_{i+1}$ imposed by $\Sigma$ but heavily penalizing the opposite order. In this way we can use Theorem 13.6 for the regularized measure and avoid the problematic boundary terms in the integration by parts in (13.33). At the end we remove the regularization using the additional boundedness assumptions on $f$ and $\nabla \sqrt{f}$. For $\beta \geq 1$ these additional assumptions are not necessary using that $C^\infty_0(\Sigma)$ functions are dense in $H^1(\Sigma, d\omega)$. The entropy decay (13.78) will follow from the LSI and the time-integral version of the entropy dissipation (13.32), which can be proven directly on $\Sigma$ if $\beta \geq 1$.

We remark that there is an alternative regularization method that mimics the proof of Theorem 13.6 directly on $\Sigma$ for $\beta \geq 1$. This is based on introducing carefully selected cutoff functions in order to approximate $f_t$ by a function compactly supported on $\Sigma$. The compact support renders the boundary terms in the integration by parts zero, but the cutoff does not commute with the dynamics; the error has to be tracked carefully. The advantage of this alternative method is that it also gives the exponential decay of the Dirichlet form, and it is also the closest in spirit to the strategy of the proof of Theorem 13.6 on $\mathbb{R}^N$. The disadvantage is that it works only for $\beta \geq 1$; in particular, it does not yield the LSI for $\beta \in (0,1)$. We will not discuss this approach in this book; the interested reader may find details in appendix B of [64].

**Proof of Theorem 13.14.** Define the extension $\mu_\beta^\delta = Z_{\mu_\beta}^{-1}\delta e^{-\beta N^2 K}$ of the measure $\mu_\beta$ from $\Sigma_N$ to $\mathbb{R}^N$ for any $\delta > 0$ by replacing the singular logarithm with a $C^2$-function. To that end, we introduce the approximation parameter $\delta > 0$ and define for $x \in \mathbb{R}^N$

$$J_\delta(x) := \sum_i \frac{1}{4} x_i^2 - \frac{1}{N} \sum_{i<j} \log_\delta(x_j - x_i)$$

(13.79)

where we set

$$\log_\delta(x) := 1(x \geq \delta) \log x + 1(x < \delta) \left( \log \delta + \frac{x - \delta}{\delta} - \frac{1}{2\delta^2} (x - \delta)^2 \right), \quad x \in \mathbb{R}.$$
It is easy to check that $\log_\delta \in C^2(\mathbb{R})$ is concave and satisfies

$$\lim_{\delta \to 0} \log_\delta(x) = \begin{cases} \log x & \text{if } x > 0, \\ -\infty & \text{if } x \leq 0. \end{cases}$$

The convergence is monotone decreasing, i.e.,

$$\log_\delta(x) \geq \log_{\delta'}(x), \quad x \in \mathbb{R}, \quad \text{for any } \delta \geq \delta'.$$

Furthermore, we have

$$\partial_x \log_\delta(x) = \begin{cases} \frac{1}{x} & \text{if } x > \delta, \\ \frac{2\delta - x}{\delta^2} & \text{if } x \leq \delta, \end{cases}$$

and the lower bound

$$\partial_x^2 \log_\delta(x) \geq \begin{cases} -\frac{1}{x^2} & \text{if } x > \delta, \\ -\frac{1}{\delta^2} & \text{if } x \leq \delta; \end{cases}$$

in particular, $\mathcal{H}_\delta$ is convex with $\mathcal{H}'_\delta \geq \frac{1}{2}$. Similarly, we can extend the measure $\omega$ to $\mathbb{R}^N$ by setting

$$\omega_\delta := Z_{\omega, \delta} e^{-\beta NH_\delta} \mu_\delta^\delta,$$

and its Hamiltonian still satisfies (13.75). By the monotonicity (13.80), we know that $e^{-\beta NH_\delta(x)}$ is pointwise monotonically decreasing in $\delta$ and clearly $Z_{\mu, \delta}$ and $Z_{\omega, \delta}$ converge to 1 as $\delta \to 0$. Clearly, $\omega_\delta \to \omega$ as $\delta \to 0$ in the sense that for any set $A \subset \mathbb{R}^N$ we have $\omega_\delta(A) \to \omega(A \cap \Sigma_N)$. We remark that the regularized DBM corresponding to the measure $\omega_\delta$ is given by

$$dx_i = \sqrt{\frac{2}{\beta N}} dB_i + \left(-\frac{1}{2} x_i - U_i'(x_i) + \frac{1}{N} \sum_{j \neq i} \log_\delta(x_i - x_j) - \frac{1}{N} \sum_{j > i} \log_\delta(x_j - x_i) \right) dt.$$

Given a function $f$ on $\Sigma_N$ such that $\int_{\Sigma_N} f \, d\omega = 1$ and $D_\omega(\sqrt{f}) = \int_{\Sigma_N} |\nabla \sqrt{f}|^2 \, d\omega < \infty$, we extend it by symmetry to $\mathbb{R}^N$. To do so, we first define the symmetrized version of $\Sigma_N$, i.e.,

$$\Sigma_N' := \mathbb{R}^N \setminus \{x : \exists i \neq j, x_i = x_j\},$$

which has the disjoint union structure

$$\Sigma_N = \bigcup_{\pi \in S_N} \pi(\Sigma_N)$$

where $\pi$ runs through all $N$-element permutations and acts by permuting the coordinates of any point $x \in \mathbb{R}^N$. For any $x \in \Sigma_N'$ there is a unique $\pi$ so that $x \in \pi(\Sigma_N)$, and we then define the extension $\tilde{f}$ by $\tilde{f}(x) := f(\pi^{-1}(x))$. Clearly,
\( \pi^{-1}(x) \) is simply the coordinates of \( x = (x_1, \ldots, x_N) \) permuted in increasing order. Thus \( f \) is defined on \( \mathbb{R}^N \) apart from a zero measure set and is bounded since \( f \in L^\infty \). Furthermore, \( \nabla[(\tilde{f})^{1/2}] \) is also bounded since \( \nabla \sqrt{f} \in L^\infty \).

Since \( \tilde{f} \) is bounded, we have \( f_{\delta} \) is a probability density, i.e., \( \int f_{\delta} \, d\omega = 1 \), so there is a constant \( C_{\delta} \) such that \( f_{\delta} := C_{\delta} \tilde{f} \) is a probability density, i.e., \( \int f_{\delta} \, d\omega = 1 \), and clearly \( C_{\delta} \to 1 \) as \( \delta \to 0 \). Applying Theorem 13.6 to the measure \( \omega_{\delta} \) and the function \( f_{\delta} \) on \( \mathbb{R}^N \), we see that the LSI holds for \( \omega_{\delta} \), i.e.,

\[
S_{\omega_{\delta}}(f_{\delta}) \leq \frac{2}{K} D_{\omega_{\delta}}(\sqrt{f_{\delta}}),
\]

or, equivalently,

\[
C_{\delta} \int_{\mathbb{R}^N} \tilde{f} \log \tilde{f} \, d\omega_{\delta} + \log C_{\delta} \leq \frac{2C_{\delta}}{K} D_{\omega_{\delta}}(\sqrt{f_{\delta}}).
\]

Now we let \( \delta \to 0 \). Using the boundedness of \( \nabla[(\tilde{f})^{1/2}] \), the weak convergence of \( \omega_{\delta} \) to \( \omega \), and that \( Z_{\omega,\delta} \) and \( C_{\delta} \) converge to 1, the Dirichlet form on the right side of the last inequality converges to \( D_\omega(\sqrt{f}) \). The first term on the left side of the inequality converges to \( \int f \log f \, d\omega \) by dominated convergence, and the second term converges to 0. Thus we arrive at (13.76).

In the above argument, the boundedness of \( f \) and \( \nabla \sqrt{f} \) were only used to ensure that \( f \) or rather its extension \( \tilde{f} \) has finite integral, and the Dirichlet form w.r.t. the regularized measure \( \omega_{\delta} \), \( D_{\omega_{\delta}}(\sqrt{f_{\delta}}) \), converges to \( D_\omega(\sqrt{f}) \). For \( \beta \geq 1 \) we can remove these conditions by using a different extension of \( f \) to \( \mathbb{R}^N \) if \( f \in H^1(\Sigma, d\omega) \). We may assume that \( D_\omega(\sqrt{f}) < \infty \); otherwise (13.76) is a tautology. We first still assume that \( f \in L^\infty(\Sigma) \). We smoothly cut off \( f \) to be 0 at the boundary of \( \Sigma \); i.e., we find a nonnegative sequence \( f_\varepsilon \in C_0^\infty(\Sigma) \) such that \( \sqrt{f_\varepsilon} \to \sqrt{f} \) in \( H^1(\Sigma, d\omega) \) and \( \int f_\varepsilon \, d\omega = 1 \). For \( \beta \geq 1 \) the existence of a similar sequence but \( f_\varepsilon \) in \( H^1(\Sigma, d\omega) \) was shown in Section 12.4. In fact, the same construction shows that we can also guarantee \( \sqrt{f_\varepsilon} \to \sqrt{f} \) in \( H^1(\Sigma, d\omega) \). Now we use the LSI for the smooth functions \( f_\varepsilon \), i.e.,

\[
S_{\omega}(f_\varepsilon) \leq \frac{2}{K} D_\omega(\sqrt{f_\varepsilon}),
\]

and we let \( \varepsilon \to 0 \). The right-hand side converges to \( D_\omega(\sqrt{f}) \) by the above choice of \( f_\varepsilon \). For the left-hand side, recall that apart from a smoothing that can be dealt with via standard approximation arguments, the cutoff function was constructed in the form \( f_\varepsilon(x) = C_\varepsilon \phi_\varepsilon(x) f(x) \), where \( \phi_\varepsilon(x) \in (0, 1) \) with \( \phi_\varepsilon \not\to 1 \) monotonically pointwise and \( C_\varepsilon \) is a normalization such that \( C_\varepsilon \to 1 \) as \( \varepsilon \to 0 \). Clearly,

\[
S_{\omega}(f_\varepsilon) = C_\varepsilon \log C_\varepsilon \int_\Sigma \phi_\varepsilon f \, d\omega + C_\varepsilon \int_\Sigma f \phi_\varepsilon \log \phi_\varepsilon \, d\omega + C_\varepsilon \int_\Sigma \phi_\varepsilon f \log f \, d\omega.
\]
The first term converges to 0 since \( \int \phi_\varepsilon f \, d\omega \leq 1 \) and \( C_\varepsilon \log C_\varepsilon \to 0 \). The second term also converges to 0 by dominated convergence since \( |\phi_\varepsilon \log \phi_\varepsilon| \) is bounded by 1 and goes to 0 pointwise. Finally, the last term converges to \( S_\omega(f) \) by monotone convergence. This proves (13.76) for \( \beta \geq 1 \) for any bounded \( f \). Finally, we remove this last condition. Given any \( f \) with \( D_\omega(\sqrt{f}) < \infty \), we define \( \bar{f}_M := \min\{f, M\} \) and \( \tilde{f}_M := C_M \bar{f}_M \) where \( C_M \) is the normalization. Clearly, \( C_M \to 1 \) and \( \bar{f}_M \to f \) pointwise. Since (13.76) holds for \( \tilde{f}_M \), we have

\[
C_M \log C_M \int f_M \, d\omega + C_M \int f_M \log f_M \, d\omega \leq \frac{2}{K} C_M \int |\nabla \sqrt{f_M}|^2 \, d\omega.
\]

Now we let \( M \to \infty \). The first term on the left is just \( \log C_M \to 0 \). The second term on the left converges to \( S_\omega(f) \) by monotone convergence and \( C_M \to 1 \) and similarly the right-hand side converges to \( (2/K)D_\omega(\sqrt{f}) \) by monotone convergence. This proves (13.76) for \( \beta \geq 1 \) without any additional condition on \( f \).

The Brascamp-Lieb inequality, (13.77), is proved similarly, starting from its regularized version on \( \mathbb{R}^N \),

\[
\langle f; f \rangle_{\omega_\delta} \leq \langle \nabla f, [\mathcal{H}\delta'' + \tilde{\mathcal{H}}'']^{-1}\nabla f \rangle_{\omega_\delta}
\]

that follows directly from Theorem 13.13. We can then take the limit \( \delta \to 0 \) using monotone convergence on the left and the dominated convergence on the right, using that \( \mathcal{H}\delta'' \to \mathcal{H} \) and the inverse \( [\mathcal{H}\delta'' + \tilde{\mathcal{H}}'']^{-1} \) is uniformly bounded.

For the third part of the theorem, for the proof of (13.78), we first note that the remark after (12.17) applies to the generator \( \mathcal{L}_\omega \) as well; i.e., \( \beta \geq 1 \) is necessary for the well-posedness of the equation \( \partial_t f_t = \mathcal{L}_\omega f_t \) on \( \Sigma_N \) with initial condition \( f_0 \) supported on \( \Sigma_N \). The construction of the dynamics in Section 12.4 also implies that \( f_t \in H^1(\omega) \) for any \( t > 0 \) if \( f \in L^2(\omega) \).

We now mimic the proof of the entropy dissipation (13.32) in our setup. Since we do not know that \( D_\omega(\sqrt{f}) < \infty \), we have to introduce a regularization \( c > 0 \) to keep \( f_t \) away from 0. We compute

\[
\frac{d}{dt} \int f_t \log(f_t + c) \, d\omega = \int (\mathcal{L} f_t) \log(f_t + c) \, d\omega + \int f_t \frac{\mathcal{L} f_t}{f_t + c} \, d\omega
\]

\[
= -\int \frac{[\nabla f_t]^2}{f_t + c} \, d\omega - \int c \mathcal{L} f_t \frac{f_t}{f_t + c} \, d\omega
\]

\[
= -\int \frac{[\nabla f_t]^2}{f_t + c} \, d\omega - \int c \frac{[\nabla f_t]^2}{(f_t + c)^2} \, d\omega.
\]

Owing to the regularization \( c > 0 \), we used integration by parts for \( H^1(\omega) \) functions only. Dropping the last term that is negative and integrating from 0 to \( t \), we obtain

\[
\int f_t \log(f_t + c) \, d\omega + \int_0^t \int [\nabla f_s]^2 \frac{1}{f_s + c} \, d\omega \leq \int f_0 \log(f_0 + c) \, d\omega.
\]
Since $f_t \log \frac{f_t + c}{f_t} \geq 0$, we have

$$\int f_t \log f_t \, d\omega + \int_0^t \int |\nabla f_s|^2 \frac{f_s}{f_s + c} \, ds \leq \int f_0 \log (f_0 + c) \, d\omega. \quad (13.87)$$

Note that both terms on the left-hand side are nonnegative. Now we let $c \to 0$. By monotone convergence and $S_\omega(f_0) < \infty$, we get

$$\int f_t \log f_t \, d\mu + \int_0^t ds \int |\nabla f_s|^2 \frac{f_s}{f_s} \, d\omega \leq \int f_0 \log f_0 \, d\omega \quad (13.88)$$
or

$$S_\omega(f_t) + 4 \int_0^t D_\omega(\sqrt{f_s}) \, ds \leq S_\omega(f_0). \quad (13.89)$$

This is the entropy dissipation inequality in a time integral form. Notice that neither equality nor the differential version as in (13.32) is claimed. Note that by integrating (13.85) between $t$ and $\tau$, a similar argument yields

$$S_\omega(f_t) + 4 \int_\tau^t D_\omega(\sqrt{f_s}) \, ds \leq S_\omega(f_\tau), \quad t \geq \tau \geq 0. \quad (13.90)$$

In particular, the entropy decays:

$$0 \leq S_\omega(f_t) \leq S_\omega(f_\tau), \quad t \geq \tau \geq 0. \quad (13.91)$$

Now we use the LSI (13.76) to estimate $D_\omega(\sqrt{f_s})$ in (13.90) and recall that the LSI holds for any $f_s$ since $\beta \geq 1$. We get

$$S_\omega(f_t) + 2K \int_\tau^t S_\omega(f_s) \, ds \leq S_\omega(f_\tau), \quad t \geq \tau \geq 0. \quad (13.92)$$

A standard calculus exercise shows that $S_\omega(f_t) \leq e^{-2Kt} S_\omega(f_0)$ for all $t \geq 0$. One possible argument is to fix any $\delta > 0$ and choose $\tau = (n - 1)\delta$, $t = n\delta$ with $n = 1, 2, \ldots$ in (13.92). By monotonicity of the entropy, we have

$$(1 + 2K\delta)S_\omega(f_{n\delta}) \leq S_\omega(f_{(n-1)\delta})$$

for any $n$, and by iteration we obtain

$$S_\omega(f_{n\delta}) \leq (1 + 2K\delta)^{-n} S_\omega(f_0).$$

Setting $\delta = t/n$ and letting $n \to \infty$ we get (13.78). This completes the proof of Theorem 13.14. \qed