

Bakry - Emery Criterion

Reference: ① Ramon van Handel, Probability in High Dimension, Lecture notes, Princeton University, 2016 (RVH book)

② Solution sets to problems in Prob. in High Dim, Available online at: tylertkf.github.io/prob-reading/

③ Sinho Chewi, Log-concave sampling, Online notes, Nov. 2024

(I closely follow Chaps. 2, 3 in ①, and solutions in ②)

The goal in this note is to study the convergence of the Langevin Dynamic (Markov Process, with $\mu = \pi^*$)

$$dx_t = \nabla \ln \pi^*(x_t) dt + \sqrt{2} dB_t, \quad t \geq 0, \quad x_t \in \mathbb{R}^d$$

when $\pi^*(x)$ is strongly log-concave, i.e.,
if $\pi^*(x) = \frac{e^{-V(x)}}{Z}$, then $V(x)$ is ρ -strongly convex.

for some $\rho > 0$.

In this case, we will develop the theory to show that the Poincaré Inequality holds with $C = \frac{1}{\rho}$,

i.e.,
$$\text{Var}_\mu(f) \leq \frac{1}{\rho} \mathcal{E}(f, f)$$

Since this ρ does not depend on d (ambient dimension), we have from Theorem 1 in previous notes that convergence occurs in $\chi^2(\mu)$ exponentially fast, with no dimension dependence.

In the next note, we will strengthen this result through the Log-Sobolev Inequality to show a similar property for the KL Divergence.

The key constructions are the **carré du champ** operator and the iterated carré du champ operator, which together allow an algebraic way to verify if the Poincaré inequality holds.

("square of the field")

Carré du Champ Operator: As before $\{P_t\}$ semigroup and (d, μ) are the associated generator and invariant measure. Then, $f, g \in \mathcal{F}$,

$$\Gamma(f, g) = \frac{1}{2} (d(fg) - f d(g) - g d(f)).$$

↓
Carré du champ operator.

Note: If D is the derivative operator, then chain rule is $D(fg) = D(f)g + f D(g)$. For the generator, an analogous statement does not hold, i.e., $d(fg) \neq d(f)g + f d(g)$. The carré du champ is the "quadratic correction" to the above linear approximation.

For the OU process, $\mu \sim N(0, I)$, and

$$\Gamma(f, f) = \|\nabla f\|_2^2 = \sum_{i=1}^d \left(\frac{\partial f(x)}{\partial x_i} \right)^2, \quad x \in \mathbb{R}^d$$

In fact for any Langevin Dynamic, $\Gamma(f, f) = \|\nabla f\|_2^2$

(see Thm 2 in this note for the proof)

Lemma 1 (Properties of Carre du Champ) (Problem 2.7 in RvH)

① $\mathcal{E}(f, f) = \int \Gamma(f, f) d\mu$, and further, if the M.P. is reversible, $\mathcal{E}(f, g) = \int \Gamma(f, g) d\mu$.

Pf: $\mathcal{E}(f, f) = -\langle f, df \rangle_\mu = -\mu(f df)$

$$= \frac{1}{2} \mu(df^2) - \mu(f df)$$

$$\mu(df^2) = 0 \quad \because \mu(df) = 0 \quad \forall f \in \mathcal{F}$$

$$= \mu\left(\frac{1}{2} df^2 - f df\right)$$

$$= \mu(\Gamma(f, f)) \stackrel{\text{def}}{=} \int \Gamma(f, f) d\mu.$$

$$(\alpha, \mu) \text{ reversible} \Rightarrow \langle df, g \rangle_\mu = \langle f, dg \rangle_\mu$$
$$\mu(df g) = \mu(f dg)$$

$$\therefore \int \Gamma(f, g) d\mu = \frac{1}{2} \left(\underbrace{\mu(df g)}_{=0} - \mu(f dg) - \mu(df g) \right)$$

$$= -\mu(f dg) = \mathcal{E}(f, g) \quad \square$$

⑥ $\Gamma(f, f) \geq 0 \quad \forall f \in \mathcal{F}$

We use the following fact: (i) ϕ convex, then $\int \phi(f) \geq \phi'(f) \int f$.
 (Here: $\phi: \mathbb{R} \rightarrow \mathbb{R}$; $\phi(f)(x) = \phi(f(x))$)

(ii) ϕ convex, then $P_t \phi(f) \geq \phi(P_t f)$.
 (see problem 2.6, RvH.)

$$\int f^2 \stackrel{(i)}{\geq} 2f \int f \Rightarrow \frac{1}{2} \int f^2 - f \int f \triangleq \Gamma(f, f) \geq 0$$

⑦ Cauchy-Schwartz : $\Gamma(f, g)^2 \leq \Gamma(f, f) \Gamma(g, g)$

$\Gamma(f+tg, f+tg) \geq 0$ for any $f, g \in \mathcal{F}, t \in \mathbb{R}$.

$$\therefore \Gamma(f+tg, f+tg) = \frac{1}{2} \int (f^2 + t^2 g^2 + 2tfg) - (f+tg) \int (f+tg).$$

Combine terms as above colored dots:

$$\Gamma(f+tg, f+tg) = \Gamma(f, f) + 2\Gamma(f, g)t + \Gamma(g, g)t^2$$

$$\geq 0$$

$\therefore$ For the quadratic eqn (in t), the discriminant is ≤ 0 , i.e.,

$$(2\Gamma(f,g))^2 - 4\Gamma(f,f)\Gamma(g,g) \leq 0$$

□

Recall the Poincaré inequality: $\underbrace{\text{Var}_\mu(f)}_{\substack{= \mu(f^2) - \mu(f)^2 \\ = P_\infty(f^2) - P_\infty(f)^2}} \leq c \underbrace{\mathcal{E}(f,f)}_{\substack{= \int \Gamma(f,f) d\mu \\ = P_\infty \Gamma(f,f)}}$

A non-asymptotic, local Poincaré inequality we will next derive is:

$$P_t(f^2) - (P_t f)^2 \leq c(t) P_t \Gamma(f,f)$$

for a suitably defined $c(t)$.

Lemma 2 (Problem 2.12 in RvH): Setting as before.

$$(a) \quad P_t(f^2) - (P_t f)^2 = 2 \int_0^t P_{t-s} \Gamma(P_s f, P_s f) ds$$

Pf: $\frac{d}{ds} P_{t-s}((P_s f)^2) = -\alpha P_{t-s}(P_s f)^2$

$$\frac{d}{dt} P_t f = \alpha P_t f$$

$$+ P_{t-s}(2(P_s f)(\alpha P_s f))$$

(α, P_t commute)

$$= -P_{t-s}(\alpha(P_s f)^2 - 2(P_s f)(\alpha P_s f))$$

$$= -2 P_{t-s}(\Gamma(P_s f, P_s f))$$

$\therefore$ From fundamental Thm of Calculus

$$\text{LHS} = \int_0^t \frac{d}{ds} P_{t-s}(P_s f)^2 ds = (P_t f)^2 - P_t(f^2)$$

// (integrate both sides from 0 to t)

$$\text{RHS} = -2 \int_0^t P_{t-s}(\Gamma(P_s f, P_s f)) ds.$$

⑥ Suppose $\Gamma(P_s f, P_s f) \leq \alpha(s) P_s \Gamma(f, f)$

Then,

$$P_t(f^2) - (P_t f)^2 \leq c(t) P_t \Gamma(f, f)$$

where $c(t) = \int_0^t 2\alpha(s) ds.$

Pf: From ②, we have:

$$P_t(f^2) - (P_t f)^2 = 2 \int_0^t P_{t-s} \Gamma(P_s f, P_s f) ds$$

Substituting:

$$\Gamma(P_s f, P_s f) \leq \alpha(s) P_s \Gamma(f, f)$$

We have: $LHS \leq 2 \int_0^t \alpha(s) P_{t-s} P_s \Gamma(f, f) ds.$

$$= 2 \int_0^t \alpha(s) \underbrace{P_t \Gamma(f, f)}_{\text{does not depend on } s} ds$$

$$= 2 P_t \Gamma(f, f) \int_0^t \alpha(s) ds. \quad \square$$

With the above lemmas, we now prove the main theorem.

Theorem 2 (Thm 2.35 in RvH) The following are equivalent:

(i) $c \Gamma_2(f, f) \geq \Gamma(f, f) \quad \forall f \in \mathcal{F}$

(Bakry Emery Criterion)

where $\Gamma_2(f, f) = \frac{1}{2} \left(d \Gamma(f, g) - \Gamma(f, dg) - \Gamma(df, g) \right)$

is the iterated carré du champ operator.

(ii) $\Gamma(P_t f, P_t f) \leq e^{-2t/c} P_t \Gamma(f, f), \quad \forall f \in \mathcal{F}, t \geq 0$
(Local Ergodicity)

(iii) $P_t(f^2) - (P_t f)^2 \leq c(1 - e^{-2t/c}) P_t \Gamma(f, f)$

(Local Poincaré Inequality)

Remark: Suppose we can verify that (i) holds, i.e., $c \Gamma_2(f, f) \geq \Gamma(f, f)$, which is an algebraic

property. This implies (iii) above, i.e., the Local Poincaré Inequality. By letting $t \rightarrow \infty$, we get the Poincaré inequality, which from Theorem 1 implies exponential convergence in time. $\hookrightarrow$ (Note 5.)

Thus, the Bakry-Emery criterion provides a sufficient condition for deriving convergence rates for a M.P.

Pf: (ii) $\Rightarrow$ (iii) (Local ergodicity $\Rightarrow$ Local Poincaré)

From (ii) $\Gamma(P_t f, P_t f) \leq e^{-2t/c} P_t \Gamma(f, f)$.

$\therefore$ From Lemma 2 (b), we have

$$P_t(f^2) - (P_t f)^2 \leq \left(\int_0^t 2e^{-2s/c} ds \right) P_t \Gamma(f, f)$$

$$= c(1 - e^{-2t/c}) P_t \Gamma(f, f)$$

Suppose $\Gamma(P_s f, P_s f) \leq \alpha(s) P_s \Gamma(f, f)$
 Then,
 $P_t(f^2) - (P_t f)^2 \leq c(t) P_t \Gamma(f, f)$
 where $c(t) = \int_0^t 2\alpha(s) ds$.
 Lemma 2, (b)

□

(i) $\Rightarrow$ (ii)

(Bakry-Erery crit. $\Rightarrow$ Local ergodicity)

$$\Gamma(f, g) = \frac{1}{2} (\Delta(fg) - f\Delta(g) - g\Delta(f)).$$

$$\frac{d}{ds} P_{t-s} \Gamma(P_s f, P_s f) = \frac{d}{ds} P_{t-s} \left(\frac{1}{2} \Delta(P_s f)^2 - P_s f \Delta(P_s f) \right)$$

$$= -\Delta P_{t-s} \left(\frac{1}{2} \Delta(P_s f)^2 - P_s f \Delta(P_s f) \right)$$

$$\frac{d}{ds} P_s(f) = \Delta P_s(f)$$

$$+ P_{t-s} \left(\frac{1}{2} \Delta \left(2 P_s f \frac{d}{ds} (P_s f) \right) - \frac{d}{ds} P_s f \Delta(P_s f) - P_s f \frac{d}{ds} (\Delta(P_s f)) \right)$$

$$= P_{t-s} \left(-\Delta \Gamma(P_s f, P_s f) + \Delta(P_s f \Delta(P_s f)) - (\Delta(P_s f))^2 - (P_s f \Delta^2(P_s f)) \right)$$

$$= P_{t-s} \left(-\Delta \Gamma(P_s f, P_s f) + 2 \Gamma(P_s f, \Delta(P_s f)) \right)$$

$$= -2 P_{t-s} \Gamma_2(P_s f, P_s f)$$

$$\Gamma_2(f, f) = \frac{1}{2} (\Delta \Gamma(f, g) - \Gamma(f, \Delta g) - \Gamma(\Delta f, g))$$

$\rightarrow$ ~~XX~~

Here, we assume that (i) (Bakry-Erery criterion):

$c \Gamma_2(f, f) \geq \Gamma(f, f)$ holds. Thus, we have

$$\frac{d}{ds} P_{t-s} \Gamma(P_s f, P_s f) \leq -\frac{2}{c} P_{t-s} \Gamma(P_s f, P_s f).$$

Also, recall that $\Gamma(f, f) \geq 0 \forall f \in \mathcal{F}$. Therefore, we have:

$$\frac{\frac{d}{ds} P_{t-s} \Gamma(P_s f, P_s f)}{P_{t-s} \Gamma(P_s f, P_s f)} \leq -\frac{2}{c}$$

i.e.,

$$\frac{d}{ds} \ln P_{t-s} \Gamma(P_s f, P_s f) \leq -\frac{2}{c}$$

$$\therefore \int_0^t \frac{d}{ds} \ln P_{t-s} \Gamma(P_s f, P_s f) ds \leq \int_0^t \left(-\frac{2}{c}\right) ds$$

(when $s=0$, $P_0 = \mathbb{I}$)

$$\text{i.e., } \ln P_0 \Gamma(P_t f, P_t f) - \ln P_t \Gamma(P_0 f, P_0 f) \leq -2t/c.$$

$$\text{i.e., } \Gamma(P_t f, P_t f) \leq e^{-2t/c} \Gamma(f, f). \quad \square$$

$$(iii) \Rightarrow (i)$$

(Local Poincaré $\Rightarrow$ Bakry Emery crit.)

Recall Lemma 2 (a):

$$(a) \quad P_t(f^2) - (P_t f)^2 = 2 \int_0^t P_{t-s} \Gamma(P_s f, P_s f) ds$$

From (iii), we have:

$$(ii) \quad P_t(f^2) - (P_t f)^2 \leq c(1 - e^{-2t/c}) P_t \Gamma(f, f)$$

$$\begin{aligned} 0 &\geq P_t(f^2) - (P_t f)^2 - c(1 - e^{-2t/c}) P_t \Gamma(f, f) \\ &= 2 \int_0^t P_{t-s} \Gamma(P_s f, P_s f) ds - c(1 - e^{-2t/c}) P_t \Gamma(f, f) \end{aligned}$$

Now, $h(s) = h(0) + \int_0^s \left(\frac{d}{du} h(u) \right) du$ (fundamental thm of calculus)

$$\begin{aligned} &= 2 \int_{s=0}^t \left(P_t \Gamma(f, f) + \int_{u=0}^s \frac{d}{du} P_{t-u} \Gamma(P_u f, P_u f) du \right) ds \\ &\quad - c(1 - e^{-2t/c}) P_t \Gamma(f, f) \end{aligned}$$

From $(**) (two \text{ pages back})$, we have:

$$\frac{d}{ds} P_{t-s} \Gamma(P_s f, P_s f) = -2 P_{t-s} \Gamma_2(P_s f, P_s f)$$

$$\begin{aligned}
& \downarrow = 2t P_t \Gamma(f, f) - 4 \int_{s=0}^t \int_{u=0}^s P_{t-u} \Gamma_2(P_u f, P_u f) du ds \\
& \quad - c(1 - e^{-2t/c}) P_t \Gamma(f, f) \\
& = 2t P_t \Gamma(f, f) - 4 \int_{s=0}^t \int_{u=0}^s \left(P_t \Gamma_2(f, f) + \int_{r=0}^u \frac{d}{dr} P_{t-r} \Gamma_2(P_r f, P_r f) dr \right) du ds \\
& \quad - c(1 - e^{-2t/c}) P_t \Gamma(f, f)
\end{aligned}$$

$$\begin{aligned}
& = 2t P_t \Gamma(f, f) - 4 P_t \Gamma_2(f, f) \frac{t^2}{2} - 2t P_t \Gamma(f, f) \\
& \quad + c\left(\frac{1}{2}\right) \left(\frac{2t}{c}\right)^2 P_t \Gamma(f, f) + o(t^2)
\end{aligned}$$

$\vdots$

$$= \frac{2}{c} \Gamma(f, f) t^2 - 2 \Gamma_2(f, f) t^2 + o(t^2).$$

$$\text{i.e., } \frac{2}{c} \Gamma(f, f) t^2 - 2 \Gamma_2(f, f) t^2 + o(t^2) \leq 0$$

$\hookrightarrow \textcircled{x^*}$

Above true for any $t \geq 0$. Divide by t^2 and taking limit as $t \rightarrow 0$, we have the Bakry-Emery criterion:

$$c \Gamma_2(f, f) \geq \Gamma(f, f).$$

$\square$

Theorem 3 (Problem 2.13 (d) in RvH book). Let $\mu(x) \propto e^{-W(x)}$, $x \in \mathbb{R}^d$ be a density with $W(x)$ being ρ -strongly convex, i.e.,

$$\sum_{i,j} v_i v_j \frac{\partial^2 W(x)}{\partial x_i \partial x_j} \geq \rho \sum_{i=1}^d v_i^2 \quad \forall v \in \mathbb{R}^d$$

(Note: μ is then a strongly log-concave dist.)

Consider the Langevin Dynamics sampler associated with μ :

$$dX_t = -\nabla W(X_t)dt + \sqrt{2} dB_t, \quad X_t \in \mathbb{R}^d$$

Recall that for this process,

$$\mathcal{L}(f) = -\sum_{i=1}^d \frac{\partial W(x)}{\partial x_i} \cdot \frac{\partial f(x)}{\partial x_i} + \sum_{i=1}^d \frac{\partial^2 f(x)}{\partial x_i^2}$$

$$\mu(x) = \frac{e^{-W(x)}}{Z} \quad \xrightarrow{\text{normalizing constant}} \quad x \in \mathbb{R}^d$$

Then,

$$\text{Var}_\mu(f) \leq \frac{1}{\rho} \int \|\nabla f\|^2 d\mu$$

Proof: We need some preliminary calculations: $\Gamma(f, f)$, $d\Gamma(f, f)$, $d(f df)$, $\Gamma(f, df)$, $\Gamma_2(f, f)$.

$$2\Gamma(f, f) = d(f^2) - 2f df$$

$$= - \sum_i \left(\cancel{\frac{\partial W}{\partial x_i}} \cancel{2f} \cancel{\frac{\partial f}{\partial x_i}} \right) + 2 \sum_i \left(\cancel{f} \cancel{\frac{\partial^2 f}{\partial x_i^2}} + \left(\frac{\partial f}{\partial x_i} \right)^2 \right)$$

$$- 2f \left(\sum_i \cancel{\frac{\partial W}{\partial x_i}} \cdot \cancel{\frac{\partial f}{\partial x_i}} + \sum_{i=1}^d \cancel{\frac{\partial^2 f}{\partial x_i^2}} \right)$$

$$= 2 \|\nabla f\|^2 \Rightarrow \boxed{\Gamma(f, f) = \|\nabla f\|^2}$$

$$d\Gamma(f, f) = d\|\nabla f\|^2$$

$$= - \sum_i \frac{\partial W}{\partial x_i} \frac{\partial}{\partial x_i} \left(\sum_i \left(\frac{\partial f}{\partial x_i} \right)^2 \right) + \sum_i \frac{\partial^2}{\partial x_i^2} \left(\sum_i \left(\frac{\partial f}{\partial x_i} \right)^2 \right)$$

$$= - \sum_{i,j} \frac{\partial W}{\partial x_i} \frac{\partial f}{\partial x_j} \frac{\partial^2 f}{\partial x_i \partial x_j} + 2 \sum_{i,j} \left(\left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 + \frac{\partial f}{\partial x_j} \frac{\partial^3 f}{\partial^2 x_i \partial x_j} \right)$$

$$\Delta(f df) = \sum_i \frac{\partial w}{\partial x_i} \frac{\partial}{\partial x_i} (f df) + \sum_i \frac{\partial^2}{\partial x_i^2} (f df)$$

$$= -\Delta f \sum_i \frac{\partial w}{\partial x_i} \frac{\partial f}{\partial x_i} - f \sum_i \frac{\partial w}{\partial x_i} \frac{\partial^2 f}{\partial x_i^2}$$

$$+ \sum_i \left(f \frac{\partial^2 df}{\partial x_i^2} + 2 \frac{\partial f}{\partial x_i} \frac{\partial df}{\partial x_i} + df \frac{\partial^2 f}{\partial x_i^2} \right)$$

$$\therefore \Delta(f df) = (df)^2 + f \Delta^2 f + 2 \sum_i \frac{\partial f}{\partial x_i} \frac{\partial df}{\partial x_i} = 2 \langle \nabla f, \nabla df \rangle$$

$$\Gamma(f, df) = \frac{1}{2} \left(\Delta(f df) - (df)^2 - f \Delta^2 f \right)$$

$$= \frac{1}{2} \left(\cancel{(df)^2} + \cancel{f \Delta^2 f} + 2 \langle \nabla f, \nabla df \rangle - \cancel{(df)^2} - \cancel{f \Delta^2 f} \right)$$

$$= \langle \nabla f, \nabla df \rangle$$

$$= - \sum_{i,j} \left(\frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \frac{\partial^2 w}{\partial x_i \partial x_j} \right) - \sum_{i,j} \left(\frac{\partial w}{\partial x_i} \frac{\partial f}{\partial x_i} \frac{\partial^2 f}{\partial x_j \partial x_j} \right)$$

$$+ \sum_{i,j} \left(\frac{\partial f}{\partial x_i} \frac{\partial^3 f}{\partial^2 x_i \partial x_j} \right)$$

$$2 \Gamma_2(f, f) = 2 \Gamma(f, f) - 2 \Gamma(f, \Delta f)$$

$$= -2 \sum_i \frac{\partial W}{\partial x_i} \frac{\partial f}{\partial x_j} \frac{\partial^2 f}{\partial x_i \partial x_j} + 2 \sum_{i,j} \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2 + \frac{\partial f}{\partial x_j} \frac{\partial^3 f}{\partial^2 x_i \partial x_j}$$

$$+ 2 \sum_{i,j} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \frac{\partial^2 W}{\partial x_i \partial x_j} + 2 \sum_{i,j} \frac{\partial W}{\partial x_i} \frac{\partial f}{\partial x_j} \frac{\partial^2 f}{\partial x_i \partial x_j}$$

$$- 2 \sum_{i,j} \frac{\partial f}{\partial x_j} \frac{\partial^3 f}{\partial^2 x_i \partial x_j}$$

Simplify:

$$2 \Gamma_2(f, f) = 2 \sum_{i,j} \left(\underbrace{\left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)^2}_{\geq 0} + \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \frac{\partial^2 W}{\partial x_i \partial x_j} \right)$$

$$\geq 2 \langle \nabla f, \Delta W \nabla f \rangle$$

ρ -strong convexity
of W

$$\geq 2 \rho \|\nabla f\|^2$$

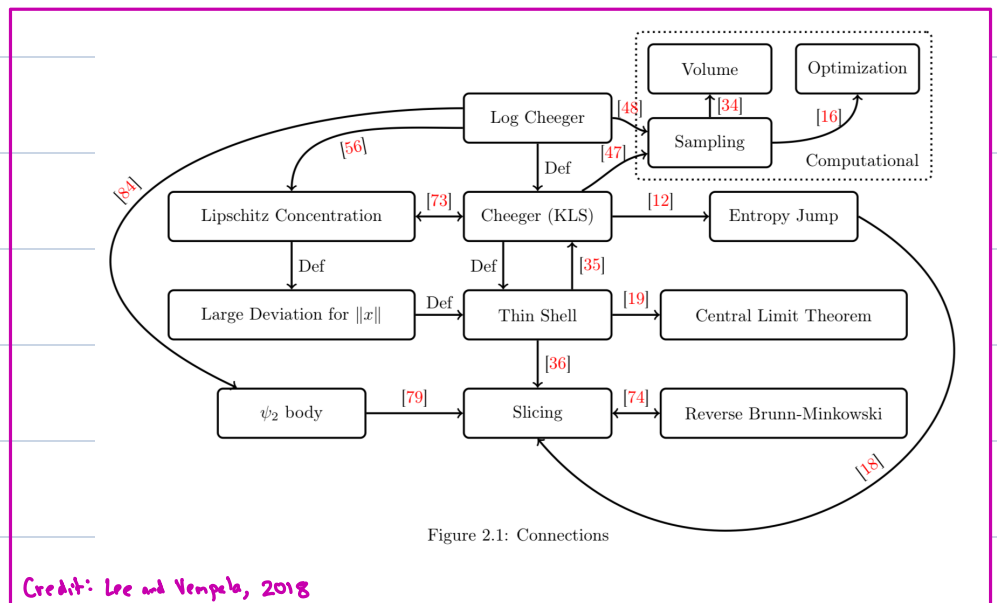


Corollary: The Langevin dynamics converges exponentially fast in time, with no dependence on ambient dimension d , for ρ -strongly concave dist.

Remark: The Kannan-Lovasz-Simonovits (KLS) conjecture states that even if $\rho=0$, (i.e., μ is log-concave), $\exists c>0$, s.t. the Poincaré inequality holds with parameter c . This conjecture, if true, would have many implications in probability theory CS; see:

Y.-T. Lee, S. Vempala, "The KLS conjecture",
arXiv 1807.03465, July 2018.

The following image is taken from the tutorial above:



Addendum: Details on Lemma 2.

Lemma 2, First Step

$$\frac{d}{ds} P_{t-s}((P_s f)^2)$$

$$= \lim_{h \rightarrow 0} \frac{P_{t-s-h}((P_{s+h} f)^2) - P_{t-s}((P_s f)^2)}{h}$$

$$= \lim_{h \rightarrow 0} \frac{P_{t-s-h}((P_{s+h} f)^2) - P_{t-s}((P_{s+h} f)^2) + P_{t-s}((P_{s+h} f)^2) - P_{t-s}((P_s f)^2)}{h}$$

$$= -d P_{t-s}((P_s f)^2) + \lim_{h \rightarrow 0} P_{t-s} \left(\frac{(P_{s+h} f)^2 - (P_s f)^2}{h} \right)$$

$$= -d P_{t-s}((P_s f)^2) + P_{t-s}(2(P_s f) d P_s f).$$