THermodynamic Formalism and Uncertainty Quantification

Luc Rey-Bellet
University of Massachusetts Amherst

Quantissima III, Venice, August 2019

Work supported by NSF and AFOSR
Collaborators on related projects

- **Paul Dupuis** (Brown University),
- **Markos Katsoulakis** (UMass Amherst)
- **Sung-Ha Hwang** (KAIST)
- **Peter Plechac** (U. of Delaware)
- **Yannis Pantazis** (FORTH Crete)
- **Jeremiah Birrell** (UMass Amherst)
- **Panagiota Birmpa** (UMass Amherst)
- **Konstantinos Gourgoulias** (UMass Amherst)
- **Jinchao Feng** (UMass Amherst)
- **Jie Wang** (UMass Amherst)
- **Sosung Baek** (KAIST)
Some references:


• and several more to come.
UQ framework: Baseline model

Baseline model $P \ (\equiv \text{probability measure on } \mathcal{X}).$

Think of it as a (tractable) model you use to compute or do analysis.

Maybe obtained after inference and/or model reduction, and so on....

Mots interesting you should think of $P$ is high-dimensional, e.g,

\[ P_\nu \text{ is the distribution of a process } \{X_t\}_{0 \leq t \leq \infty} \text{ with } X_0 \sim \nu. \]

$P$ is a Gibbs measure on $\Omega^{\mathbb{Z}^d}$

In any case, we think there are possibly lots of and large uncertainties in the model (model-form uncertainties)

$P$ IS NOT TO BE TRUSTED!!
Specific observables/statistics/quantities of interest = QoI

- $E_P[f]$ (Expectation)
- $\text{Var}_P(f)$ (Variance) or $\frac{\text{Cov}_P(f,g)}{\sqrt{\text{Var}_P(f)\text{Var}_P(g)}}$ (correlation), or
- $\Lambda_{P,f}(c) = \log E_P[e^{cf}]$ (risk sensitive functional)
- $\log P(A) \sim \log e^{-I(A)/\epsilon}$ (probability of some rare event)

or maybe path-space QoI

- $E_{P_{\nu}} \left[ \int_0^T f(x_t) \, dt \right]$ where $\tau$ is a stopping time.
- $E_{P_{\nu}} \left[ \frac{1}{T} \int_0^T f(x_s) \, dt \right]$ that is ergodic averages.
- $E_{P_{\nu}} \left[ \int_0^\infty e^{-\lambda s} f(x_s) \, dt \right]$ that is discounted observables.
- and so on....
UQ framework: Non Parametric Stress tests

→ Family of **alternative models** $Q$. Think of it as describing the true but "unknowable" or partially known models. Set

$$Q_\eta = \{Q \text{ is } \eta "\text{close}" \text{ to } P\}$$

Given a QoI $f$ can one find **uncertainty bounds** or performance guarantees

$$\inf_{Q \in Q_\eta} E_Q[f] \leq E_P[f] \leq \sup_{Q \in Q_\eta} E_Q[f]?$$

and similarly for other quantities. The bounds should be **tight** and **computable** (numerically or analytically).

→ **Robustness**, cf book by Hansen (Nobel 2011) and Sargent (Nobel 2013)
→ **Stress tests** in Operation research, Finance, etc....
UQ framework: distances and divergences

Which measure of distance or pseudo-distance divergence should one use?

→ Use Information Theory concepts to measure information loss between $Q$ and $P$.

- Relative entropy (a.k.a Kullback-Leibler divergence)

\[
R(Q||P) = E_Q \left[ \log \frac{dQ}{dP} \right]
\]

- Relative Renyi entropy (a.k.a Renyi divergence): For $\alpha \neq 0, 1$

\[
R_\alpha(Q||P) = \frac{1}{\alpha(\alpha - 1)} \log E_P \left[ \frac{dQ^\alpha}{dP} \right] = \frac{1}{\alpha(\alpha - 1)} \log E_P \left[ e^{\alpha \log \frac{dQ}{dP}} \right]
\]

Note that

\[
R_\alpha(Q||P) \rightarrow \begin{cases} 
R(Q||P) & \text{as } \alpha \rightarrow 1 \\
R(P||Q) & \text{as } \alpha \rightarrow 0
\end{cases}
\]
UQ framework: distances and divergences

• **Scalability:** If $Q^{0:T}$ and $P^{0:T}$ are the distribution of the process restricted to the time window 0 to $T$ then, typically,

$$R_\alpha(Q^{0:T} \| P^{0:T}). = O(T) \text{ as } T \to \infty$$

i.e. Information is additive. For the relative entropy we have the chain rule for relative entropy which is even better (not asymptotic in $T$).

• **Information processing inequality:** If $\mathcal{F}$ is a sub $\sigma$-algebra then

$$R_\alpha(Q|\mathcal{F} \| P|\mathcal{F}) \leq R_\alpha(Q \| P)$$

• **What is the right divergence for the QoI?**

• **Not the whole story:**

→ Heavy tailed observable may require other entropies (f-divergences)
→ Wasserstein type distances— needed if $Q \ll P$...
What is wrong with CKP? Scalability

Cziszar-Kullback-Pinsker

\[ |E_Q[f] - E_P[f]| \leq \sqrt{2R(Q||P)} \|f - E_P[f]\|_\infty \]

Take e.g. Markov measures \( P = P^{0:T} \) and \( Q = Q^{0:T} \) and

\[ F_T = \frac{1}{T} \int_0^T f(X_s) \, ds. \]

Then \( \|F_T\|_\infty = \|f\|_\infty = O(1) \) and \( R(Q^{0:T}||P^{0:T}) = O(T) \) and so

\[ |E_{Q^{0:T}}[F_T] - E_{P^{0:T}}[F_T]| \leq \sqrt{2R(Q^{0:T}||P^{0:T})} \|F_T - E_P[F_T]\|_\infty = O(1) \]

CKP does not scale correctly!

Note though that

\[ \text{Var}_{P^{0:T}}[F_T] = O\left(\frac{1}{T}\right) \]

so one would need the variance instead of the sup norm.
**Gibbs Variational principle a.k.a.** \( F = U - TS \)

- **Relative entropy** (a.k.a Kullback-Leibler divergence).

\[
R(Q \ll P) = \begin{cases} 
E_Q \left[ \log \frac{dQ}{dP} \right] & \text{if } Q \ll P \\
+\infty & \text{otherwise}
\end{cases}
\]

\( R(Q \ll P) \) is a divergence, that is \( R(Q \ll P) \geq 0 \) and \( R(Q \ll P) = 0 \) if and only if \( Q = P \).

- **Gibbs variational principle** for the relative entropy: (convex duality).

\[
\log E_P \left[ e^f \right] = \sup_Q \left\{ E_Q[f] - R(Q||P) \right\}
\]

with the supremum attained if and only if

\[
dQ = dQ^f = \frac{e^f dP}{E_P[e^f]}
\]

Play a **central role** in statistical mechanics, in large deviation theory and in dynamical systems.
Gibbs information inequality

From the Gibbs variational principle, for any $Q$ and $c \geq 0$
\[
E_Q[\pm cf] \leq \log E_P[e^{\pm cf}] + R(Q||P).
\]

**Theorem** (Gibbs Information inequality)

\[
- \inf_{c>0} \left\{ \frac{\Lambda(-c) + R(Q||P)}{c} \right\} \leq E_Q[f] - E_P[f] \leq \inf_{c>0} \left\{ \frac{\Lambda(c) + R(Q||P)}{c} \right\}
\]

\[
= \Xi_{P,f}(R(Q||P))
\]

\[
\Xi_{P,f}(\eta) \equiv \inf_{c>0} \left\{ \frac{\Lambda(c) + \eta}{c} \right\}
\]

\[
\Lambda(c) = \log E_P[e^{c(f-E_P[f])}] = \log E_P[e^{cf}] - E_P[f]
\]

How good is it? (Long history... Dupuis; Bobkov; Boucheron, Lugosi. Massart; Breuer, Cziszhar, etc...)
Properties of the Gibbs information inequality

\[ \Xi_{P,f}(R(Q\|P)) \] is a divergence, i.e.

\[ \Xi_{P,f}(\eta) \geq 0 \text{ and } \Xi_{P,f}(\eta) = 0 \iff \begin{cases} \eta = 0 \text{ i.e. } Q = P \\ \text{or } f = \text{const} \end{cases} \]

Moreover the Gibbs information inequality is tight: Given the family of alternative models \( Q_\eta = \{ Q; R(Q\|P) \leq \eta \} \) we have

\[ \Xi_{P,f}(\eta) = \max_{Q \in Q_\eta} \{ E_Q[f] - E_P[f] \} \]

and the maximum is attained at \( Q_\eta \in Q_\eta \) with

\[ \frac{dQ_\eta}{dP} = \frac{e^{c(\eta)f}}{E_P[e^{c(\eta)f}]} \text{ with } c \text{ such that } R(Q_\eta\|Q) = \eta \]

and of course similarly for min.
Concentration / UQ duality

Recall: If \(X_1, X_2, \cdots\) are IID copies with (centered) MGF \(\Lambda(c)\) for \(f(X)\) then by Chernov bound

\[
P\left(\frac{1}{N} \sum_{k=1}^{N} f(X_i) - E_P[f] > x\right) \leq e^{-N\Lambda^*(x)}
\]

Concentration

and by Cramer and Sanov Theorem and the contraction principle

\[
\Lambda^*(x) = \sup_c \{xc - \Lambda(c)\} \quad \text{(Legendre transform)}
\]

\[
= \inf_Q \{R(Q||P) ; E_Q[f] - E_P[f] = x\} \quad "\text{(Entropy maximization)}"
\]


versus (duality of optimization problems)

\[
(\Lambda^*)^{-1}(\eta) = \inf_{c \geq 0} \left\{ \frac{\Lambda(\pm c) + \eta}{c} \right\} \quad \text{(Fenchel-Young)}
\]

\[
= \sup_Q \{\pm(E_Q[f] - E_P[f]) ; R(Q||P) = \eta\} \quad \text{(UQ bounds)}
\]
**Linearization/Variance**

**Linearization:** For small $\eta = R(Q||P)$ one has the asymptotic expansion

$$\Xi_{P,f}(\eta) = \sqrt{2\text{Var}_P[f]}\eta + \frac{1}{3}\sqrt{\text{Var}_P[f]}\gamma_P(f)\eta + O(\eta^{3/2})$$

where $\gamma_P(f) = \frac{E[(f-E_P[f])^3]}{\text{Var}_P[f]^{3/2}}$ is the skewness.

→ For small perturbation of $P$ UQ is driven by CLT fluctuations, in the linear regime.

→ For large perturbations of $P$ UQ is driven by rare events or rather concentration of measure.
Markov process: choosing the right path space entropy

Baselines: Markov process $X_t$ with path-space measure $P^{0:T}$

Alternative: Stochastic process $Y_t$ with path-space measure $Q^{0:T}$ (not necessarily Markovian!) and

$$Q^{0:T} \ll P^{0:T}$$

Idea is to restrict the relative entropy to a sub $\sigma$-algebra tailored to the observables at hand

- Ergodic averages. Apply the inequality to $F_T = \int_0^T f(X_t) \, dt$

$$E_Q \left( \frac{F_T}{T} \right) - E_P \left( \frac{F_T}{T} \right) \leq \inf_{c > 0} \left\{ \frac{1}{T} \log E_P \left[ e^{c(F_T - E_P[F_T])} \right] + \frac{1}{T} R(Q^{0:T}_{\nu_0} || P^{0:T}_{\mu_0}) \right\}$$

Under suitable ergodicity assumptions for $X_t$ the bounds scale as $T \to \infty$. The important quantity is the relative entropy rate (it scales nicely with $T$ as we shall see later)...
• Ergodic averages: statistical mechanics.

\( P = \) Gibbs measure on \( \Omega^{\mathbb{Z}^d} \) (\( \Omega \) finite set) with potential \( \Phi \).

\( Q = \) any translation invariant measure on \( \Omega^{\mathbb{Z}^d} \).

\[
\frac{1}{|V|} \lim_{V \nearrow \mathbb{Z}^d} R(Q\|P) = \lim_{V \nearrow \mathbb{Z}^d} \frac{1}{|V|} R(Q\|P) \] always exist and is finite

Theorem: For (quasilocal) observable \( f \)

\[
\inf_{c > 0} \left\{ \frac{\lambda(-c) + r(Q\|P)}{c} \right\} \leq E_Q[f] - E_P[f] \leq \inf_{c > 0} \left\{ \frac{\lambda(c) + r(Q\|P)}{c} \right\}
\]

\( \lambda(c) = P(\Phi + c\Psi_f) - P(\phi) \) translated pressure

(that is local Hamiltonian \( H_V + c \sum_{x \in V} \tau_x(f) \))
• Stopping time $\tau$ and QoI $F_\tau = \int_0^\tau f(X_t)\,dt$. It is natural to restrict the relative entropy to the $\sigma$-algebra $\mathcal{F}_\tau$.

\[
E_Q[F_\tau] - E_P[F_\tau] \leq \inf_{c>0} \left\{ \frac{\log E_P[e^{cF_\tau} - E_P[F_\tau]] + R(Q^{0:\tau}||P^{0:\tau})}{c} \right\}
\]

Just stop the process....

• Discounted observable QoI $G_\lambda(f) = \int_0^\infty f(X_t)\lambda e^{-\lambda t}\,dt$.

Define a new measure $P_\lambda$: $X_t$ runs up to a random time $T$ with exponential distribution with mean $1/\lambda$. Then

\[
R(Q_\lambda||P_\lambda) = \int_0^\infty R(Q^{0:t}||P^{0:t})\lambda e^{-\lambda t}\,dt \quad \text{discounted entropy}
\]

\[
E_Q[G_\lambda(f)] - E_P[G_\lambda(f)] \leq \inf_{c>0} \left\{ \frac{G_\lambda(e^{cf}) + R_\lambda(Q||P)}{c} \right\}
\]
UQ for statistical estimators/ mean field formalism

How do we get UQ bounds for non-linear functionals of $P$, for example variance or skewness

$$\text{Var}_P[f(X)] \quad \text{or} \quad \gamma_P[f] = \frac{E_P[(f - E_P[f(X)])^3]}{\text{Var}_P[f(X)]^{3/2}}$$

or more general statistical estimators?

A fundamental result in large deviations

**Laplace principle:** (Varadhan, Bryc, Dupuis-Ellis)

The sequence $S_N$ taking value in $Y$ satisfy a LDP with rate function $I(Y)$ if and only if for all $\Phi : Y \rightarrow \mathbb{R}$ bounded and continuous

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log E_P[e^{N\Phi(S_N)}] = \sup_y \{\Phi(y) - I(y)\}$$
Example: UQ for the variance

Build a statistical estimator for the variance

\[ \frac{1}{N} \sum_{i=1}^{N} f(X_i)^2 - \left( \frac{1}{N} \sum_{i=1}^{N} f(X_i) \right)^2 \rightarrow \text{Var}_P[f] \]

where \( X_i \) are IID copies of \( X \).

Apply the Gibbs information inequality to statistical estimator, to find

**Theorem** Gibbs UQ Bounds for the variance

\[ - \inf_{c > 0} \left\{ \frac{H(-c) + R(Q||P)}{c} \right\} \leq \text{Var}_Q[f] \leq \inf_{c > 0} \left\{ \frac{H(c) + R(Q||P)}{c} \right\} \]

where

\[ H(c) = \lim_{N \to \infty} \frac{1}{N} \log E_{P_0^N} \left[ e^{\sum_{i=1}^{N} f(X_i)^2 - \frac{1}{N} \left( \sum_{i=1}^{N} f(X_i) \right)^2} \right] \]
Using the Laplace principle for the joint \((f(X), f^2(X))\) one finds the convex function

\[ H(c) = \sup_{(u,v) \in \mathbb{R}^2} \left\{ c(v - u^2) - I(u, v) \right\} \]

where

\[ \Lambda(\alpha, \beta) = \log E_P \left[ e^{\alpha f(X) + \beta f^2(X)} \right] \] (cumulant generating function)

\[ I(u, v) = \sup_{\alpha, \beta} \{ \alpha u + \beta v - \Lambda(\alpha, \beta) \} \] (rate function in Cramer’s Theorem)

The inequality is tight with optimizer

\[ dQ^{\alpha, \beta} = \frac{e^{\alpha f + \beta f^2}}{E_P \left[ e^{\alpha f + \beta f^2} \right]} dP \]

for suitable \(\alpha\) and \(\beta\) such that

\[ R(Q^{\alpha, \beta} \| P) = \eta \]

This generalizes to general statistical estimators.
Rare events and risk sensitive functionals

UQ for rare events:

\[ P(A) \sim e^{-I(A)/\epsilon} \text{ (rare event probability)} \]

We really want to control \( I(A) = -\epsilon \log P(A) \).

More generally we consider risk sensitive functionals

\[ \log E_P[e^{c f}] \text{ if } c \text{ large (free energy)} \]

Relative Renyi entropy (a.k.a Renyi divergence): For \( \alpha \neq 0, 1 \)

\[
R\alpha(Q||P) = \frac{1}{\alpha(\alpha - 1)} \log E_P \left[ \frac{dQ^\alpha}{dP} \right] = \frac{1}{\alpha(\alpha - 1)} \log E_P \left[ e^{\alpha \log \frac{dQ}{dP}} \right]
\]
Variational principle for the Relative Reny entropy:
(Dupuis et al.)

Extension of the Gibbs Variational Principle proved by Atar, Chowdhary, and Dupuis.

Relative Renyi entropy (a.k.a Renyi divergence): For $\alpha \neq 0, 1$

$$R_\alpha(Q \| P) = \frac{1}{\alpha(\alpha - 1)} \log E_P \left[ \frac{dQ^\alpha}{dP} \right] = \frac{1}{\alpha(\alpha - 1)} \log E_P \left[ e^{\alpha \log \frac{dQ}{dP}} \right]$$

Renyi Variational Principle proved by Atar, Chowdhary, and Dupuis.

$$\frac{1}{\beta} \log E_Q \left[ e^{\beta g} \right] = \inf_Q \left\{ \frac{1}{\gamma} \log E_P \left[ e^{\gamma g} \right] + \frac{1}{\gamma - \beta} R_{\gamma - \beta} (Q \| P) \right\} \quad \gamma > \beta$$

$$\frac{1}{\beta} \log E_Q \left[ e^{\beta g} \right] = \sup_Q \left\{ \frac{1}{\gamma} \log E_P \left[ e^{\gamma g} \right] - \frac{1}{\beta - \gamma} R_{\beta - \gamma} (Q \| P) \right\} \quad \gamma < \beta$$
UQ bounds for risk sensitive functionals
\[
\sup_{\beta < \gamma} \left\{ \frac{1}{\beta} \log \mathbb{E}_P[e^{\beta g}] + \frac{1}{\beta - \gamma} \mathcal{R}_{\gamma-\beta}(Q \| P) \right\} \leq \frac{1}{\gamma} \log \mathbb{E}_Q[e^{\gamma g}]
\]
\[
\frac{1}{\gamma} \log \mathbb{E}_Q[e^{\gamma g}] \leq \inf_{\beta > \gamma} \left\{ \frac{1}{\beta} \log \mathbb{E}_P[e^{\beta g}] + \frac{1}{\gamma - \beta} \mathcal{R}_{\gamma-\beta}(Q \| P) \right\}
\]

You can prove similar tightness properties as well.

To treat rare events you take \( g = -M 1_{A^c} \) and take \( M \to \infty \) and relabeling the indices

UQ bounds for rare events
\[
- \inf_{\alpha > 0} \left\{ \frac{\log \mathbb{E}_P \left[ e^{-\alpha \log \frac{dQ}{dP}} \right] - \log P(A)}{\alpha} \right\} \leq \log Q(A) - \log P(A)
\]

\[
\log Q(A) - \log P(A) \leq \inf_{\alpha > 1} \left\{ \frac{\log \mathbb{E}_P \left[ e^{\alpha \log \frac{dQ}{dP}} \right] - \log P(A)}{\alpha} \right\}
\]

Similar optimization problem as before.
Making it computable with concentration inequalities

Some examples: (Much more in Gourgoulias, Katsoulakis, R.-B., Wang).

• If $a \leq f \leq b$ we have Hoeffding’s inequality

$$\Lambda(c) \leq \frac{c^2(b-a)^2}{8} \leq \frac{c^2\|f - \mathbf{E}_P[f]\|_\infty}{2}$$

and then

$$\Xi_{P,f}(\eta) \leq \sqrt{2\eta\|f - \mathbf{E}_P[f]\|_\infty} \quad \text{(Cziszar-Kullback Pinsker)}.$$  

• If $f$ is bounded and $\text{Var}_P[f] = \sigma^2$ then we have Bernstein inequality

$$\Lambda(c) \leq \frac{c^2\sigma^2}{2(1 - c\|f - \mathbf{E}_P[f]\|_\infty)}$$

and then

$$\Xi_{P,f}(\eta) \leq \sqrt{2\text{Var}_P[f]\eta + \|f - \mathbf{E}_P[f]\|_\infty\eta}$$

This beats Pinsker if $\eta$ is not too big (especially if $\sigma^2$ is small) and captures the exact small $\eta$ asymptotics.

• Many more: Sharper inequalities for bounded $f$ and other for Poissonian, Gaussian, exponential tails....

25
Steady state UQ bounds for ergodic Markov processes

Consider ergodic averages $\frac{1}{T} \int_0^T f(X_s) \, ds$ then using the Gibbs UQ bound one obtains the steady state bias bound

\[ \xi_{P,f}(r(Q\|P)) \leq \lim_{T \to \infty} \frac{1}{T} \int_0^T f(Y_s) \, ds - E_{\mu}[f] \leq \xi_{P,f}(r(Q\|P)) \]

where

\[ \xi_{P,f}(\eta) = \inf_{c > 0} \left\{ \frac{\lambda(c) + \eta}{c} \right\} \]

\[ \lambda(c) = \lim_{T \to \infty} \frac{1}{T} \log EP_{\mu^n} \left[ e^{c \int_0^T (f(X_s) - E_{\mu}[f]) \, ds} \right] \] (CGF)

\[ \eta = \lim_{T \to \infty} \frac{1}{T} R(Q^{0:T} \| P^{0:T}) \] (relative entropy rate)
Coercive Dynamics

Langevin equation:

\[ dX = -\nabla V + J\nabla V + \sqrt{2}dW_t \]

for any any antisymmetric \( J \) has invariant measure \( d\mu = Z^{-1}e^{-V}dx \)
and we have

\[
\mathcal{L} = \underbrace{\Delta - \nabla V \nabla}_\text{symmetric} + \underbrace{J\nabla V \nabla}_\text{antisymmetric}
\]

• **Main idea (from Liming Wu):** Bound the Feynmann-Kac semi group

\[
e^T(\mathcal{L} + V)h(x) = E_{P_{\delta_x}^{0:T}} \left[ e^{\int_0^T V(X_s)ds} h(X_t) \right]
\]

using **Lumer-Philips Theorem**

\[
\frac{1}{T} \log \|e^T(\mathcal{L}+V)\|_{L^2(\mu)} \leq \sup \left\{ \langle g, \mathcal{L}g \rangle_{L^2(\mu)} + \int V|g|^2d\mu, \|g\|^2 = 1 \right\}.
\]

See the works on concentration inequalities by **Lezeaud, Wu, Catiaux, Guillin** and collaborators on which we rely here.
Poincaré inequalities and bounded $f$

**Theorem:** If we have a Poincaré inequality (spectral gap)

$$\text{Var}_\mu[f] \leq -\alpha \langle f, \mathcal{L}f \rangle_{L^2(\mu)}, \quad f \in D(\mathcal{L})$$

then for bounded $f$ and general $\mathcal{L}$

$$\lambda(c) \leq \frac{c^2 \alpha \text{Var}_\mu[f]}{1 - \alpha c \|f - E_\mu[f]\|_\infty} \quad \text{Bernstein type bound}$$

$$\xi_{P,f}(\eta) \leq 2\sqrt{\alpha \text{Var}_\mu[f]}\eta + \alpha \|f - E_\mu[f]\|_\infty \eta$$

**Theorem:** For symmetric $\mathcal{L}$ we we have the sharper bound

$$\lambda(c) \leq \frac{c^2 \sigma^2(f)}{2(1 - \alpha c \|\tilde{f}\|_\infty)} \quad \text{Bernstein type bound}$$

$$\xi_{P,f}(\eta) \leq \sqrt{2\sigma^2(f)}\eta + \alpha \|f - E_\mu[f]\|_\infty \eta$$

*(sharp for small $\eta$).*
Log-Sobolev inequalities and unbounded $f$

Assume a stronger Log-Sobolev inequality

$$E_\mu[f^2 \log(f^2)] - E_\mu[f^2] \log E_\mu[f^2] \leq -\beta \langle f, \mathcal{L} f \rangle \quad f \in D(\mathcal{L})$$

Then using the Gibbs variational principle get the bound

$$\xi_{P,f}(\eta) = \inf_{c > 0} \left\{ \log E_\mu \left[ e^{c(f - E_\mu[f])} \right] + \frac{\beta \eta}{c} \right\}$$

The only trace of the dynamics is left in the constant $\beta$.
The tail behavior of $f$ in the stationary distribution determines the UQ. Use another concentration inequality

If $V(x) \sim |x|^\beta$ (usual bounds on $\nabla V$ and $\Delta V$...)
- Poincaré for $\beta > 1$
- Log Sobolev for $\beta > 2$ so UQ bounds for $V(X)$ itself.
For $1 < b \leq 2$ we can use $F$- Sobolev inequalities to consider unbounded $f$.  

29
Hypocoercive samplers

Goal: To sample from $\nu(dq) \propto e^{-\beta V(q)} dq$ extending the phase space and sample from the measure

$$\mu(dp, dq) = \nu(dq) \pi(dp) \propto e^{-\beta (V(q) + p^2/2m)} dp dq$$

You can use other distribution of $p$ too.

Why?: Add extra dimensions to escape your bad karma.... Make the dynamics irreversible to get faster (maybe).

• Ex1: Langevin equation

$$dq_t = \frac{p_t}{m} dt, \quad dp_t = \left( -\nabla V(q_t) - \gamma \frac{p_t}{m} \right) dt + \sqrt{\frac{2 \gamma}{\beta}} dW_t$$

\[
\mathcal{L} = \left( \frac{p^T}{m} \right) \nabla q - \nabla V^T \nabla p + \frac{1}{\beta} \left( \Delta p - \gamma \left( \frac{p}{M} \right)^T \nabla p \right)
\]

\[T = -T^* \quad S = S^*\]
Ex2: Randomized Hamiltonian Monte-Carlo.

The particle follows Hamiltonian equation of motions

\[ dq_t = \frac{p_t}{m} dt, \quad dp_t = -\nabla V(q_t) \]

without noise or dissipation for a random amount of time at which we resample the momentum according to the stationary measure.

With the projection \( \Pi f = \int f(p, q) d\pi(p) \) the generator is

\[
\mathcal{L} = \left( \frac{p^T}{m} \right) \nabla_q - \nabla V^T \nabla_p + \lambda (\Pi - I) \]

\( T = -T^* \)  \( S = S^* \)

\( 31 \)
• **EX 3: Bouncy particle sampler.**

The particle follow **straight lines** for a random time. At updating time one either **resample the momentum** according to the stationary measure or the particle "**bounces**", i.e., it undergoes a Newtonian elastic collision on the hyperplane tangential to the gradient of the energy and the momentum is updated according to the rule

\[
    r(q)p = p - \frac{p^T \nabla V(q)}{\| \nabla V \|^2} \nabla V \quad Rf(p, q) = f(q, r(q)p)
\]

\[
    \mathcal{L} = \left( \frac{p}{m} \right)^T \nabla q + \left[ \left( \frac{p}{m} \right)^T \nabla V(q) \right]^+ (R - I) + \lambda (\Pi - I)
\]

- free motion
- bouncing
- noise

• Zig-zag sampler..... etc...
• Temperature accelerated molecular dynamics
• Ask **Gabriel Stoltz.**
Hypocoercivity

Dolbeaut-Mouhot-Schmeiser
Andrieu-Durmus-Nüsken-Roussel
Rouset-Stoltz-Trstanova, Olla, ...
after many other works (Villani, Hereau-Nier, Hairer-Eckmann).

Idea: The dynamics is not coercive (no Poincaré inequality in \( L^2(\mu) \) for \( \mathcal{L} \)), but there exists a scalar product equivalent to \( L^2(\mu) \) where a Poincaré inequality holds.

\[
\langle f, g \rangle_\epsilon = \langle f, g \rangle + \epsilon \langle f, (B + B^*)g \rangle.
\]

\[
B = (1 + (T\Pi)^*(T\Pi))^{-1}(-T\Pi)^*
\]
and \( T \) is the antisymmetric part of the generator

Modified Poincaré inequality:

\[
\langle -\mathcal{L}g, g \rangle_\epsilon \geq \Lambda(\epsilon) \text{Var}_\mu(f)
\]

and \( \Lambda(\epsilon) \) is explicitly expressed in terms of the Poincaré constant for \( \nu(dq) \) the spectral gap of the noise operator and the potential \( V \).
Performance guarantees for hypocoercive samplers

New results (Jeremiah Birell and L. R.-B.)

Theorem (Bernstein type inequalities for hypocoercive samplers)

For bounded \( f \) we have

\[
P_{\mu_0} \left( \left\| \frac{1}{T} \int_0^T f(X_t) dt - \int f d\mu \right\| \geq r \right) \\
\leq a(\epsilon) \left\| \frac{d\mu_0}{d\mu} \right\|_{L^2(\mu)} \exp \left( -T \frac{b(\epsilon)\Lambda(\epsilon)r^2}{4\text{Var}_\mu[f] + 2c(\epsilon)\|f - E_\mu[f]\|_r} \right)
\]

where \( a(\epsilon), b(\epsilon), c(\epsilon) \) only depends on \( \epsilon \).

\rightarrow \text{Explicit non asymptotic confidence intervals for } \int f d\mu, \text{ i.e.}

\rightarrow \text{UQ bounds for alternative processes}

\[
\xi_{P,f}(\eta) \leq \sqrt{2d(\epsilon)\Lambda(\epsilon)\text{Var}_\mu[f]\eta + e(\epsilon)\Lambda(\epsilon)\|f - E_\mu[f]\|_\infty\eta}
\]