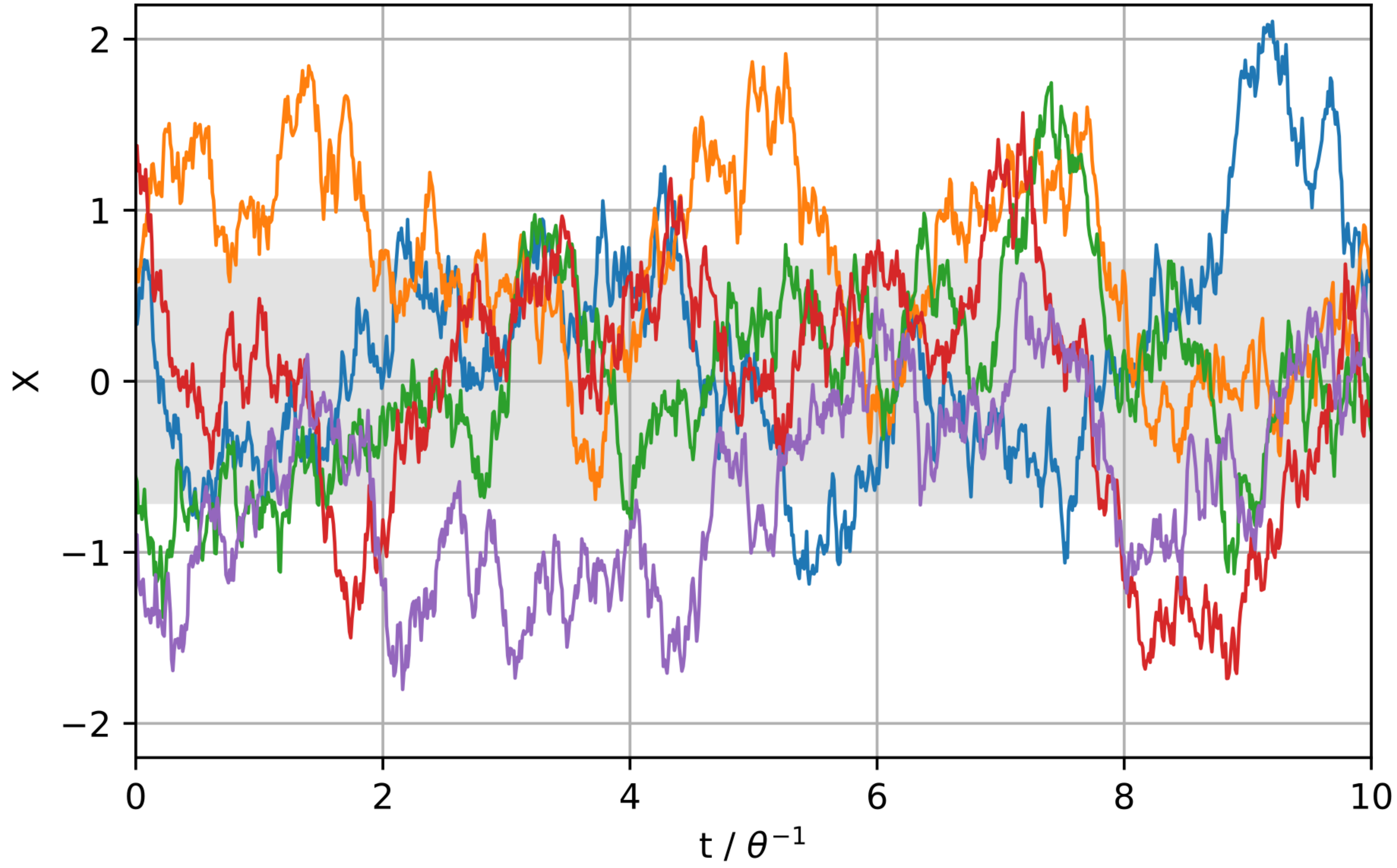
Mini-Course:
Geometric Functional Inequalities for Markov Chains

Matrix Institute, “BLoVHDPPM” Program
24 June 2025

Sam Power, University of Bristol

A Prompt for the Audience



"how do *you* like to study the long-time behaviour of a Markov process"

A Personal Perspective

- (some) principles for analysing the long-time behaviour of Markov processes
  - ‘<u>probabilistic</u>’: { pathwise, couplings, regenerations, local, … }
  - ‘<u>functional-analytic</u>’: { semigroup, divergences, entropies, global, … }
- i have a lot of appreciation for the former approach, when it applies well
- some specific applications led me to become interested in the latter approach
- i hope to share the reasons for my enthusiasm

Some Motivating Questions

- what is a ‘functional inequality’, and what are the ‘standard’ functional inequalities?
  - when ‘should’ they hold, and what are their consequences?
- to which settings is the functional-analytic approach well-suited?
  - highlight that it can be *easy to use* and *robust to (certain) details*
- how do probabilistic and functional-analytic techniques interface?
  - explain how to obtain functional-analytic consequences from probabilistic insights

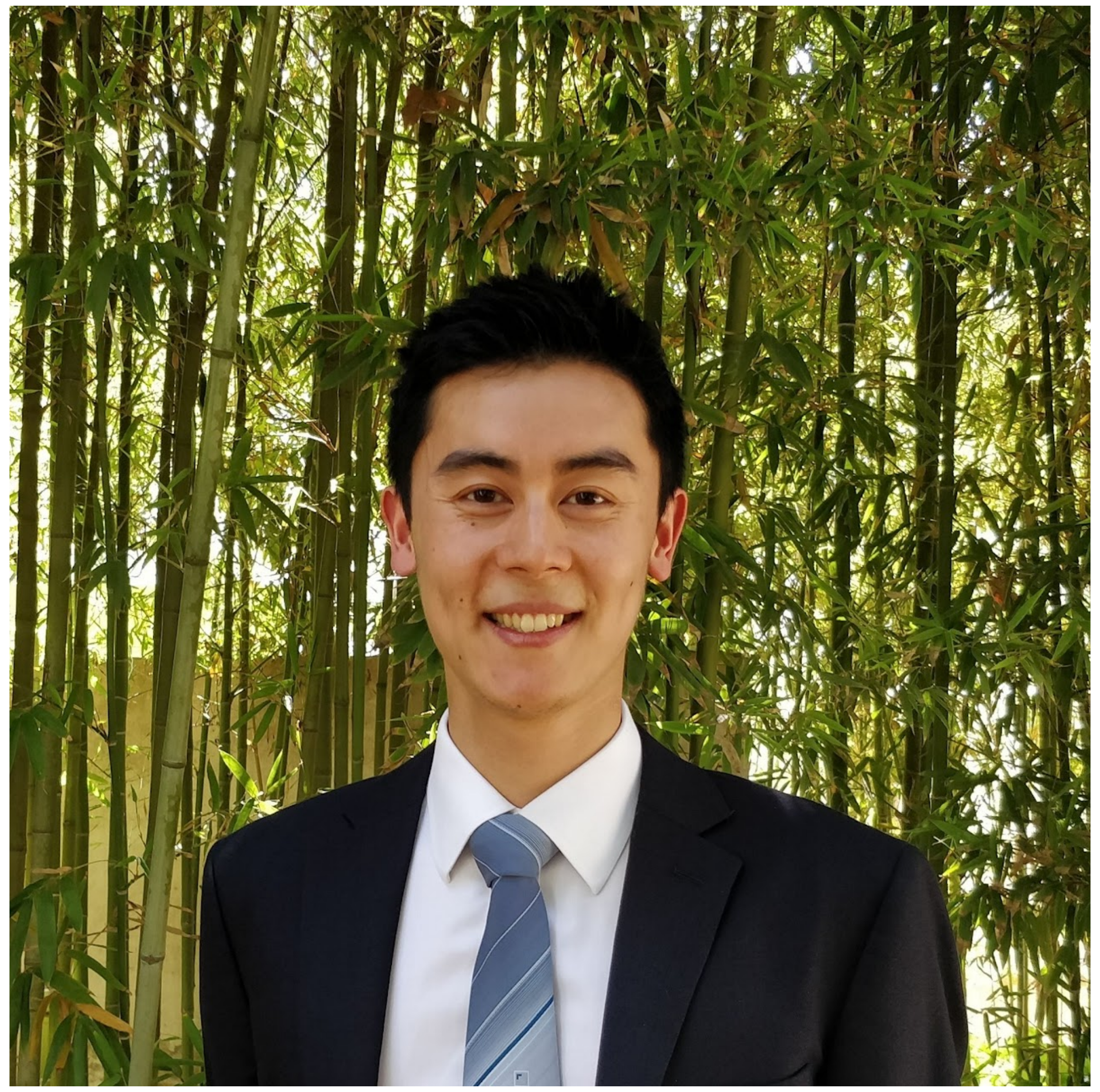
Coarse-Grained Plan

- Prelude: Why Take a Functional Approach?
- Part 1: The Story for Reversible Diffusion Processes // OLD (π)
- Part 2: The Story for Reversible Discrete-Time Markov Chains

Some Compromises

- **time-limited**: broad strokes, intuition, generous with references
- **Markov-centric**: limited discussion of geometry, concentration, etc.
- **reversibility-centric**: story is cleaner, though more is possible
- $\mathbf{L^2}$**-centric**: related to the above; will touch upon entropies somewhat
- **MCMC-centric**: my examples will inevitably reflect my own interests

this journey would not have started without:
Christophe Andrieu, Anthony Lee (Bristol), Andi Wang (Warwick)



$\boxed{\text{feel free to stop me at any point}}$

Before I Say Anything Interesting:

- P is a π-invariant Markov semigroup, often reversible and positive
- we initialise at $X_0 \sim \mu_0$, and then propagate to μ_t
- we are looking to quantify how $\mu_t \to \pi$ as $t \to \infty$

Prelude: Why Take a Functional Approach?

or, ‘How On Earth Shall I Quantify Convergence?’

(Some) Notions of Long-Time Convergence

- $\mathsf{TV}\left(p, q\right) = \frac{1}{2} \cdot \int \left| p - q \right| \left(\mathrm{d}x\right)$
- $\mathsf{TV}_{\varphi}\left(p, q\right) = \int \varphi\left(x\right) \cdot \left| p - q \right| \left(\mathrm{d}x\right)$
- $\mathscr{T}_1\left(p, q\right) = \inf \left\{ \mathbf{E}_{\gamma}\left[\mathsf{d}\left(X, Y\right)\right] : \gamma \in \mathsf{Couplings}\left(p, q\right) \right\}$
- ‘information-theoretic convergence’

Information-Theoretic Divergences

- for me: ‘Density Ratio Divergences’; also ‘f-divergence’, ‘Φ-entropy’, …
- let Φ be convex, non-negative, $\Phi(1) = 0$, and define

$$\mathsf{D}_{\Phi}\left(p, q\right) = \int q\left(\mathrm{d}x\right) \cdot \Phi\left(\frac{\mathrm{d}p}{\mathrm{d}q}(x)\right)$$

- examples: { TV, KL, Chi-Squared / Pearson, Hellinger, … }
- many well-understood inter-relations; TV convergence usually follows

Information-Theoretic Contraction

- if all goes well, then we should hope to write that

$$\mathsf{D}_{\Phi}\left(\mu_t, \pi\right) \leq \exp\left(-c \cdot t\right) \cdot \mathsf{D}_{\Phi}\left(\mu_0, \pi\right)$$

  - not very a useful assumption as written; need something ‘checkable’

- one idea: ‘reverse-Grönwall’

Information-Theoretic Contraction

- one idea: ‘reverse-Grönwall’
  - reverse-Grönwall: differentiate at $t = 0$, and interpret this

$$\mathtt{I}_{\Phi}\left(\mu_0, \pi\right) := \left[-\partial_t \mathsf{D}_{\Phi}\left(\mu_t, \pi\right)\right]_{t=0}$$

$$\mathtt{I}_{\Phi}\left(\mu_0, \pi\right) \geq c \cdot \mathsf{D}_{\Phi}\left(\mu_0, \pi\right)$$

  - a priori: a bit optimistic to start with the desired conclusion?
    - it will turn out here to be a very fruitful strategy
    - let us first study it in a more concrete setting

Part 1: Reversible Diffusions

or, ‘The Tale of the Overdamped Langevin Diffusion’

The Overdamped Langevin Diffusion

- let π be a probability density on $\mathbf{R}^d$
- then, $\mathsf{OLD}(\pi)$ is the Ito diffusion

$$\mathrm{d}X_t = \nabla \log \pi \left(X_t \right) \, \mathrm{d}t + \sqrt{2} \, \mathrm{d}W_t$$

- define the semigroup $\left(P_t \right)_{t \geq 0}$ by $P_t f(x) = \mathbf{E}^x \left[f \left(X_t \right) \right]$
    - very interesting, not particularly tractable

From Semigroup to Generator

- define the infinitesimal generator $\mathscr{L}$ by

$$\mathscr{L}f = \lim_{t \to 0^+} \frac{P_t f - f}{t}, \qquad \rightsquigarrow P_t = \exp\left(t \cdot \mathscr{L}\right)$$

- equally: $\partial_t P_t = \mathscr{L} P_t = P_t \mathscr{L}$

- concretely, this writes as

$$\mathscr{L}f(x) = \langle \nabla \log \pi(x), \nabla f(x) \rangle + \Delta f(x)$$

The Dirichlet Energy Form

- another useful object is the bilinear form

$$
\begin{aligned}
\mathscr{E}\left(f, g\right) &= \mathbf{E}\left[f \cdot \left(-\mathscr{L} g\right)\right] \\
&= \mathbf{E}\left[\langle \nabla f, \nabla g \rangle\right]
\end{aligned}
$$

(integration by parts)

- provides useful language for describing dissipation of various functionals
- particularly pertinent for <u>reversible</u> processes

Functional Forms

- we want to make statements of the form $\mathsf{D}_\Phi\left(\mu_t, \pi\right) \lesssim_t \mathsf{D}_\Phi\left(\mu_0, \pi\right)$
- by reversibility, can show that $\mathrm{d}\mu_t/\mathrm{d}\pi = P_t\left(\mathrm{d}\mu_0/\mathrm{d}\pi\right)$
- so, defining $\mathsf{Ent}_\Phi(F) = \mathbf{E}\left[\Phi \circ F\right] - \Phi \circ \mathbf{E}\left[F\right]$, it holds that

$$\mathsf{D}_\Phi\left(\mu_t, \pi\right) = \mathsf{Ent}_\Phi\left(P_t \rho_0\right), \qquad \rho_0 = \mathrm{d}\mu_0/\mathrm{d}\pi$$

- hence, sufficient to show that $\mathsf{Ent}_\Phi\left(P_t F\right) \lesssim_t \mathsf{Ent}_\Phi(F)$ for suitable F

Let Us Compute

- compute explicitly that

$$
\begin{aligned}
-\partial_t \mathrm{Ent}_\Phi\left(P_t F\right) &= -\,\partial_t \mathbf{E}\left[\Phi\left(P_t F\right)\right] \\
&= \mathbf{E}\left[\Phi'\left(P_t F\right)\cdot\left(-\mathscr{L} P_t F\right)\right] \\
&= \mathscr{E}\left(\Phi'\left(P_t F\right), P_t F\right) \\
&= \mathbf{E}\left[\Phi''\left(P_t F\right)\cdot\left|\,\nabla\left(P_t F\right)\right|^2\right]
\end{aligned}
$$

- this is a sort of 'Φ-information' functional

A Generic Energy-Entropy Inequality

- say that π satisfies a Φ-Sobolev inequality with constant c if
  - for all suitable F, there holds the estimate

$$\mathbf{E}\left[\Phi''(F) \cdot \left| \nabla F \right|^2\right] \geq c \cdot \mathsf{Ent}_\Phi(F)$$

- if this holds, then apply Grönwall, and we are done
- my convention: big constant $\sim$ good mixing (not universal)

For more, see …

- Chafaï, "Entropies, Convexity, and Functional Inequalities"
- studies general convex Φ, for which many conclusions persist
- bonus results appear particularly for Φ which are "Bregman-convex" (my term)
  - includes $\Phi(r) \sim \left\{ r^p : 1^+ < p \leq 2 \right\}$

J. Math. Kyoto Univ. (JMKYAZ)
44-2 (2004), 325–363

Entropies, convexity, and functional inequalities

On Φ-entropies and Φ-Sobolev inequalities

By

Djalil Chafaï

Abstract

Our aim is to provide a short and self contained synthesis which generalise and unify various related and unrelated works involving what we call Φ-Sobolev functional inequalities. Such inequalities related to Φ-entropies can be seen in particular as an inclusive interpolation between Poincaré and Gross logarithmic Sobolev inequalities. In addition to the known material, extensions are provided and improvements are given for some aspects. Stability by tensor products, convolution, and bounded perturbations are addressed. We show that under simple convexity assumptions on Φ, such inequalities hold in a lot of situations, including hyper-contractive diffusions, uniformly strictly log-concave measures, Wiener measure (paths space of Brownian Motion on Riemannian Manifolds) and generic Poisson space (includes paths space of some pure jumps Lévy processes and related infinitely divisible laws). Proofs are simple and relies essentially on convexity. We end up by a short parallel inspired by the analogy with Boltzmann-Shannon entropy appearing in Kinetic Gases and Information Theories.

Two Famous Characters

- $\Phi(r) = (r-1)^2$
  - the 'Poincaré inequality' (PI)

$$\gamma \cdot \mathrm{var}_\pi(F) \leq \mathbf{E}\left[\left|\nabla F\right|^2\right]$$

- $\Phi(r) = r \cdot \log r - r + 1$
  - the 'Logarithmic Sobolev Inequality' (LSI)

$$2 \cdot \lambda \cdot \mathsf{Ent}(F) \leq \mathbf{E}\left[F^{-1} \cdot \left|\nabla F\right|^2\right]$$

$$\rightsquigarrow \frac{\lambda}{2} \cdot \mathsf{Ent}\left(F^2\right) \leq \mathbf{E}\left[\left|\nabla F\right|^2\right] \text{ (2-hom)}$$

Poincaré and Log-Sobolev

- LSI implies PI with same effective rate
  - rough idea: $r \cdot \log r - r + 1 \asymp (r-1)^2/2$ as $r \to 1$
  - equally, look at $\mathsf{Ent}\,(1 + \varepsilon \cdot F)$ as $\varepsilon \to 0^+$
- LSI is strictly stronger (examples to come)
- inequalities weaker than PI map onto slower-than-exponential ergodicity
- inequalities stronger than LSI cannot tensorise; “too good to be true”

Illustrative Examples (1)

- let π be ρ-strongly log-concave, with $\rho > 0$, i.e.

$$\nabla^2 \left(-\log \pi\right)(x) \succeq \rho \cdot \mathbf{I}_d$$

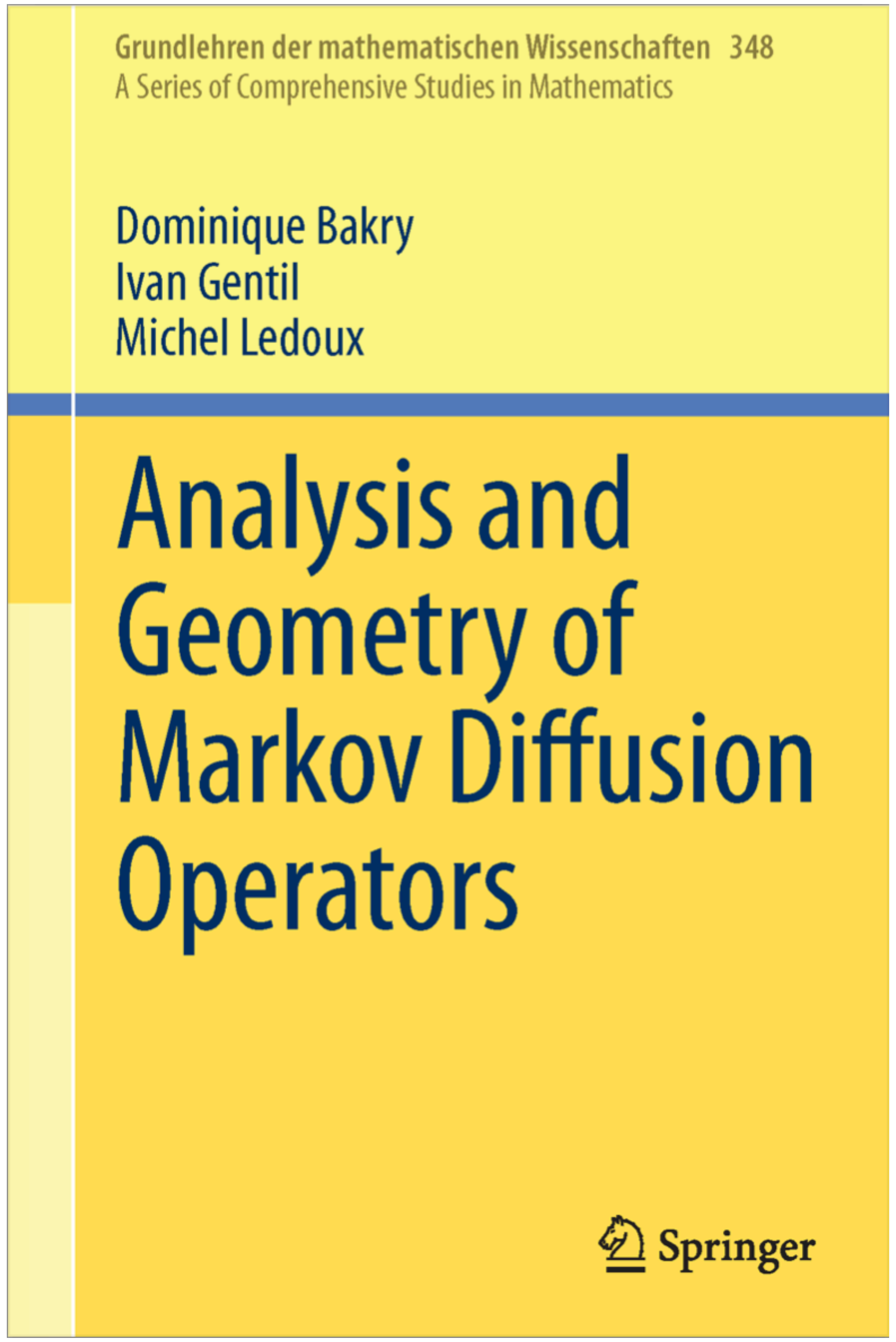
- then, LSI (hence PI) holds with constant ρ, i.e.

$$\frac{\rho}{2} \cdot \text{Ent}\left(F^2\right) \leq \mathbf{E}\left[\left|\nabla F\right|^2\right]$$

  - many ways to establish this, will comment shortly
  - the ‘Gamma Calculus’ of Bakry-Émery yields even more

For the curious, see …

- Bakry, Gentil, Ledoux - “Analysis and Geometry of Markov Diffusion Operators”
- pursues ‘all possible’ consequences of this assumption
  - can be somewhat brittle



Illustrative Examples (2)

- let π be the $\mathrm{Exp}(1)$ distribution on $\mathbf{R}^+$, i.e. $\pi(x) = \exp(-x) \cdot \mathbf{1}_{x>0}$
- then, PI holds with constant $1/4$, i.e. $\mathrm{var}_\pi(F) \leq 4 \cdot \mathbf{E}\left[\left|\nabla F\right|^2\right]$
  - can study eigenproblem more-or-less directly; try $F(x) = \exp(\lambda \cdot x)$
  - LSI **cannot** hold (will comment on why shortly)

Transfer Principles

- how can we prove that such inequalities hold for more interesting π?
- develop an ‘algebra’ (very loosely) for functional inequalities
  - { tensorisation, transport, convolution, bounded change of measure, mixtures }
- i will state (and perhaps prove) for the PI, but they will also hold for LSI
  - and often also general Φ-Sobolev inequalities

Tensorisation

- suppose that for $i \in [N]$, μ_i satisfies a PI with constant $\gamma_i > 0$
- upon defining $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_N$,
  - it holds that μ satisfies a PI with constant $\gamma \geq \min_{i \in \mathcal{I}} \gamma_i > 0$
- "you are only as bad as your worst one-dimensional component"
- enables application to high-dimensional problems

Probability in High Dimension

- many good references on this topic
- useful for concentration of measure, but also far beyond
- van Handel does a particularly good job of emphasising the significance of obtaining ‘dimension-free’ bounds

Ramon van Handel

Probability in High Dimension

APC 550 Lecture Notes
Princeton University

Transport

- suppose that μ satisfies a PI with constant $\gamma > 0$
- let $T : \mathbf{R}^d \to \mathbf{R}^{d'}$ be L-Lipschitz-continuous, and set $\pi = T_\# \mu$
- then, with $g = f \circ T$, see that

$$\mathrm{var}_\mu\left(g\right) = \mathrm{var}_\pi\left(f\right), \qquad \mathbf{E}_\mu\left[\left|\nabla g\right|^2\right] \leq L^2 \cdot \mathbf{E}_\pi\left[\left|\nabla f\right|^2\right]$$

- corollary: π satisfies a PI with constant $\gamma \cdot L^{-2} > 0$

Constructive Lipschitz Transport

- very active and exciting area
- *related* to 'optimal transport'; distinct goals
- recent developments use 'heat flow'
    - pioneered by Kim-Milman
    - greatly advanced by Mikulincer-Shenfeld
    - surprising (?) links to diffusion models

Math. Ann.
DOI 10.1007/s00208-011-0749-x

Mathematische Annalen

A generalization of Caffarelli's contraction theorem via (reverse) heat flow

Young-Heon Kim · Emanuel Milman

Received: 1 April 2010 / Revised: 14 April 2011

Abstract A theorem of L. Caffarelli implies the existence of a map, pushing forward a source Gaussian measure to a target measure which is more log-concave than the source one, which contracts Euclidean distance (in fact, Caffarelli showed that the optimal-transport Brenier map T_{opt} is a contraction in this case). We generalize this result to more general source and target measures, using a condition on the third derivative of the potential, by providing two different proofs. The first uses a map T, whose inverse is constructed as a flow along an advection field associated to an appropriate heat-diffusion process. The contraction property is then reduced to showing that log-concavity is preserved along the corresponding diffusion semi-group, by using a maximum principle for parabolic PDE. In particular, Caffarelli's original result immediately follows by using the Ornstein–Uhlenbeck process and the Prékopa–Leindler Theorem. The second uses the map T_{opt} by generalizing Caffarelli's argument, employing in addition further results of Caffarelli. As applications, we obtain new correlation and isoperimetric inequalities.

Micro-Application: Convolution

- let $X_i \sim \mu_i$ be independent, each μ_i satisfies a PI with constant $\gamma_i > 0$
- let $Y = X_1 + X_2 + \cdots \cdot X_N$
- then the law of Y satisfies a P
- proof sketch: tensorisation, followed by transport / projection

More on Tensorisation

- actually, to get 'correct' constants for convolution, this is sub-optimal
- the natural fix: 'tensorisation of entropies'
  - for 'Bregman-convex' Φ, it holds that

$$\mathrm{Ent}^{\Phi}_{\mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n}(F) \leq \mathbf{E}\left[\sum_{1 \leq i \leq n} \mathrm{Ent}^{\Phi}_{\mu_i}(F)\right]$$

  - apply Φ-Sobolev 'on the inside', and then collapse

Bounded Change-of-Measure

- suppose that μ satisfies a PI with constant $\gamma > 0$
- suppose that $\pi \equiv \mu$ strongly, i.e. $0 < \sup \mathrm{d}\pi/\mathrm{d}\mu, \sup \mathrm{d}\mu/\mathrm{d}\pi < \infty$
  - then π satisfies a PI with constant $\gamma \cdot \kappa^{-1}$, where

$$\kappa := \left(\sup \mathrm{d}\mu/\mathrm{d}\pi\right) \cdot \left(\sup \mathrm{d}\pi/\mathrm{d}\mu\right) \geq 1$$

- proof sketch:

$$\mathrm{var}_\pi\left(f\right) \leq \left(\sup \mathrm{d}\pi/\mathrm{d}\mu\right) \cdot \mathrm{var}_\mu\left(f\right)$$

$$\mathbf{E}_\mu\left[\left|\nabla f\right|^2\right] \leq \left(\sup \mathrm{d}\mu/\mathrm{d}\pi\right) \cdot \mathbf{E}_\pi\left[\left|\nabla f\right|^2\right]$$

Mixtures

- a bit more particular:
  - let (P_x) all satisfy PI (γ),
  - suppose that $\sup_{x,x'} \chi^2(P_x, P_{x'}) \le \bar{\chi}$
    - (actual assumptions weaker)
- then the mixture $\pi = \mu P$ also satisfies a PI, with constant depending only on $(\gamma, \bar{\chi})$
- same holds for LSI

Dimension-free log-Sobolev inequalities for mixture distributions

Hong-Bin Chen [a], Sinho Chewi [b], Jonathan Niles-Weed [a,*]

[a] *Courant Institute of Mathematical Sciences, New York University, New York, NY, USA*
[b] *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA*

A R T I C L E   I N F O

Article history:
Received 23 February 2021
Accepted 7 September 2021
Available online 14 September 2021
Communicated by L. Gross

Keywords:
Log-Sobolev inequality
Dimension-free
Mixture distribution

A B S T R A C T

We prove that if $(P_x)_{x\in\mathscr{X}}$ is a family of probability measures which satisfy the log-Sobolev inequality and whose pairwise chi-squared divergences are uniformly bounded, and μ is any mixing distribution on $\mathscr{X}$, then the mixture $\int P_x \, \mathrm{d}\mu(x)$ satisfies a log-Sobolev inequality. In various settings of interest, the resulting log-Sobolev constant is dimension-free. In particular, our result implies a conjecture of Zimmermann and Bardet et al. that Gaussian convolutions of measures with bounded support enjoy dimension-free log-Sobolev inequalities.

Other Routes to { PI, LSI, … }

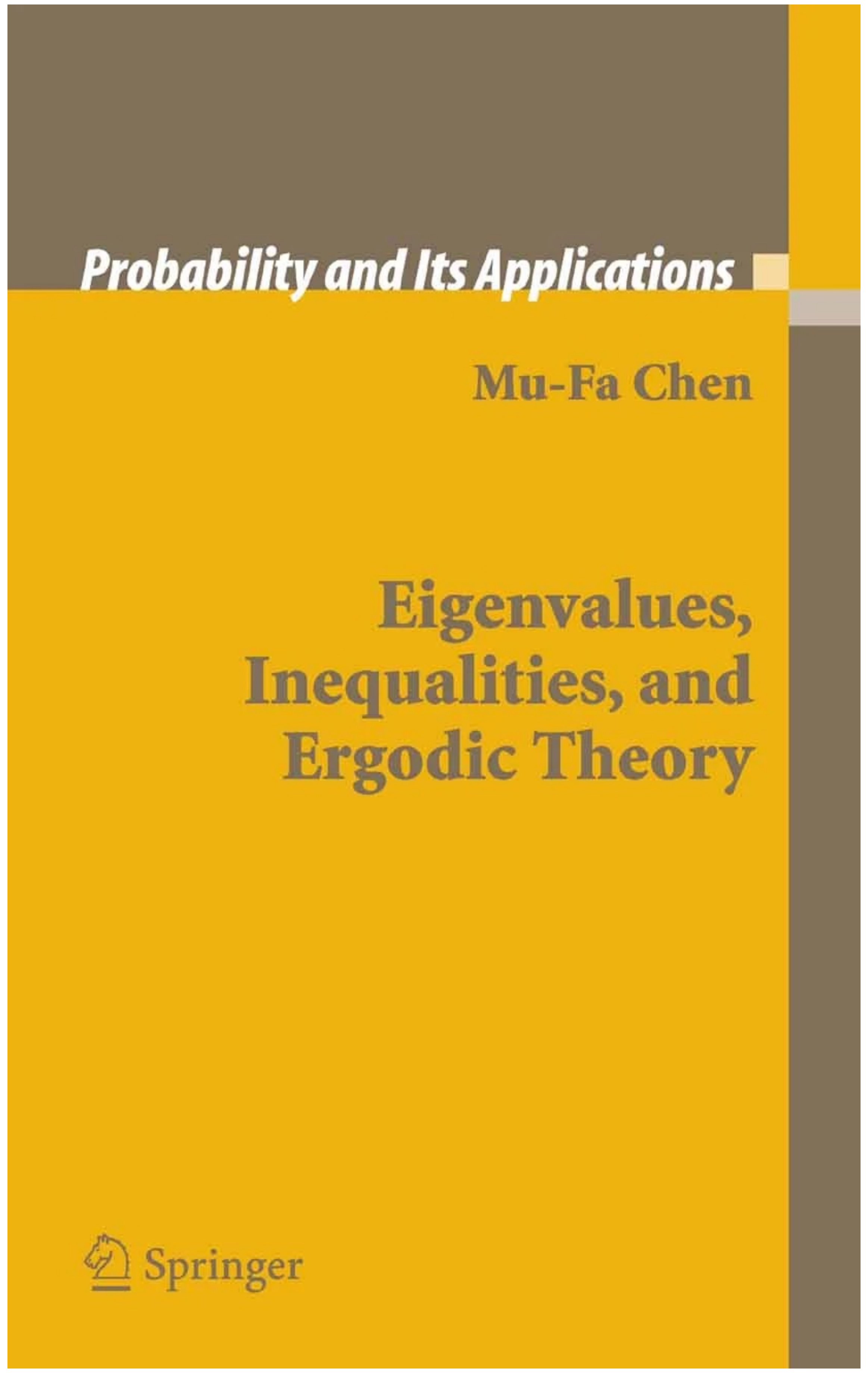
A Taster Menu

{ PI, LSI, …} from Coupling

- contraction in $\mathcal{T}_1$ yields PI with same constant
  - if dual is *non-expansive* in $\mathcal{T}_\infty$, then LSI follows with same constant
    - actually, one can even accommodate a little bit of expansion
- can hence solve strongly log-concave setting probabilistically
  - and by being creative with the metric, can solve yet more cases
    - read { Chen-Wang, Eberle, Monmarché, … }

Transport and Functional Inequalities

- that contraction implies PI is known for some years
  - (e.g. to MF Chen, FY Wang, LM Wu, etc.)
- extension to LSI was conjectured, and recently resolved by Caputo-Salez-Münch
  - an additional ingredient is genuinely needed; counter-example of Münch
- other impacts of contractivity assumption are numerous and remarkable; well-documented by Ollivier, Djellout-Guillin-Wu



A ‘Robust’ Curvature Approach

- suppose that we have a non-uniform curvature $\rho(x) > 0$, i.e. for all x,

$$\nabla^2 \left(-\log \pi\right)(x) \succeq \rho(x) \cdot \mathbf{I}_d$$

- Veysseire shows that π enjoys a spectral gap of $\gamma \geq \mathbf{E}\left[\rho(X)^{-1}\right]^{-1}$
  - it is compelling to seek $\gamma \geq \mathbf{E}\left[\rho(X)\right]$; unfortunately, this *does not hold*
- Cattiaux-Guillin-Fathi pursue same setting, obtain new (less clean) bounds

{ PI, LSI, … } from Drift (1)

- well-known that drift and minorisation yields exponential ergodicity
  - that is, assuming
    - $\mathscr{L}V \leq -\gamma \cdot V + K$,
    - minorisation on $C_\iota := \left\{x : V(x) \leq \iota \cdot K/\gamma\right\}$ for some $\iota > 1$
- also well-known that such strategies can be quantitatively poor
  - one hint: lack of tensorisation property
  - the culprit is not drift, rather *minorisation*

{ PI, LSI, … } from Drift (2)

- circa 2008, a group of French authors took a new perspective:
  - keep the drift condition, but *replace the minorisation condition*
  - instead, require that the process satisfies a 'restricted PI' on C_ι, i.e.

$$\mathrm{var}_{\pi_{C_\iota}}(f) \leq \kappa_\iota \cdot \pi\left(\left| \nabla f \right|^2 \right)$$

- this strategy can yield reasonable (though not excellent) estimates
- <u>Taghvaei-Mehta</u> recently brought this story full circle

Lyapunov Functions and Functional Inequalities

- Bakry, Barthe, Cattiaux, Guillin - "A Simple Proof of the Poincaré Inequality for a Large Class of Probability Measures, Including the Log-Concave Case"

- actually, the strategy applies more generally, and allows to prove inequalities which are { stronger, weaker } than the basic Poincaré Inequality under { stronger, weaker } conditions; see also book chapter

ELECTRONIC COMMUNICATIONS in PROBABILITY

A SIMPLE PROOF OF THE POINCARÉ INEQUALITY FOR A LARGE CLASS OF PROBABILITY MEASURES INCLUDING THE LOG-CONCAVE CASE

DOMINIQUE BAKRY
Institut de Mathématiques de Toulouse. CNRS UMR 5219. Université Paul Sabatier, Laboratoire de Statistique et Probabilités, 118 route de Narbonne, F-31062 Toulouse cedex 09
email: `bakry@math.univ-toulouse.fr`

FRANCK BARTHE
Institut de Mathématiques de Toulouse. CNRS UMR 5219. Université Paul Sabatier, Laboratoire de Statistique et Probabilités, 118 route de Narbonne, F-31062 Toulouse cedex 09
email: `barthe@math.univ-toulouse.fr`

PATRICK CATTIAUX
Institut de Mathématiques de Toulouse. CNRS UMR 5219. Université Paul Sabatier, Laboratoire de Statistique et Probabilités, 118 route de Narbonne, F-31062 Toulouse cedex 09
email: `cattiaux@math.univ-toulouse.f`

ARNAUD GUILLIN
Ecole Centrale Marseille et LATP, Université de Provence, Technopole Château-Gombert, 39 rue F. Joliot Curie, F-13453 Marseille Cedex 13
email: `guillin@cmi.univ-mrs.fr`

Submitted November 28, 2007, *accepted in final form* January 24, 2008

AMS 2000 Subject classification: Lyapunov functions, Poincaré inequality, log-concave measure
Keywords: 26D10, 47D07, 60G10, 60J60

Some Miscellanea

- Roberts-Tweedie: "Geometric L2 and L1 convergence are equivalent for reversible Markov chains"
- "localisation" methods (e.g. Kannan-Lovasz-Simonovits)
- symmetry arguments
    - reflection symmetry by Barthe-Klartag
    - spherical symmetry by Bobkov, Huet, Cattiaux-Guillin-Wu

From Checking to Using

Implications of PI (1)

- suppose that $\gamma \cdot \mathrm{var}_\pi [F] \leq \mathbf{E}\left[\left|\nabla F\right|^2\right]$; write $\mathsf{C} = \mathrm{cov}_\pi\left[\mathrm{Id}\right]$
- taking $F(x) = v^\top x$, see that $\gamma \cdot \mathsf{C} \preceq \mathbf{I}_d$, and so $\gamma \leq \|\mathsf{C}\|_{\mathrm{op}}^{-1}$
- KLS Conjecture: under log-concavity, get $\gamma \geq c_{\mathrm{KLS}} \cdot \|\mathsf{C}\|_{\mathrm{op}}^{-1}$
    - interpretation: only impediment to mixing is ‘long’ directions
    - proven up to $\sqrt{\log d}$ by Klartag, see also Y. Chen, Klartag-Lehec

Implications of PI (2)

- let G be 1-Lipschitz-continuous, and take $F = \exp(\theta \cdot G)$
  - elementary inductive argument shows that

$$\mathbf{E}\left[\exp\left(\theta \cdot \left(G - \mathbf{E}\left[G\right]\right)\right)\right] \leq \left(1 - \frac{\theta^2}{4 \cdot \gamma}\right)^{-2}$$

  - consequence: π has lighter-than-exponential tails
    - exponential example: can't expect better in general

Implications of LSI (1)

- suppose that $\mathbf{E}\left[F^{-1} \cdot \left| \nabla F \right|^2\right] \geq 2 \cdot \lambda \cdot \mathrm{Ent}\,(F)$
- taking $F(x) = \exp\left(v^\top x\right)$, see that $\lambda \cdot \log \mathbf{E}\left[\exp\left(v^\top x\right)\right] \leq \dfrac{|v|^2}{2}$
- <u>Bizeul KLS-Type Conjecture</u>: under log-concavity,
  - if $\log \mathbf{E}\left[\exp\left(v^\top x\right)\right] \leq \dfrac{v^\top \mathrm{C} v}{2}$, then $\lambda \geq c_{\mathrm{Biz}} \cdot \|\mathrm{C}\|_{\mathrm{op}}^{-1}$

Implications of LSI (2)

- let G be 1-Lipschitz-continuous, and take $F = \exp(\theta \cdot G)$
    - similar elementary ('Herbst') argument shows that

$$\log\left(\mathbf{E}\left[\exp\left(\theta \cdot \left(G - \mathbf{E}\left[G\right]\right)\right)\right]\right) \leq \frac{\theta^2}{2 \cdot \lambda}$$

- consequence: π has lighter-than-Gaussian tails
    - Gaussian example: can't expect better in general

Implications of LSI (3)

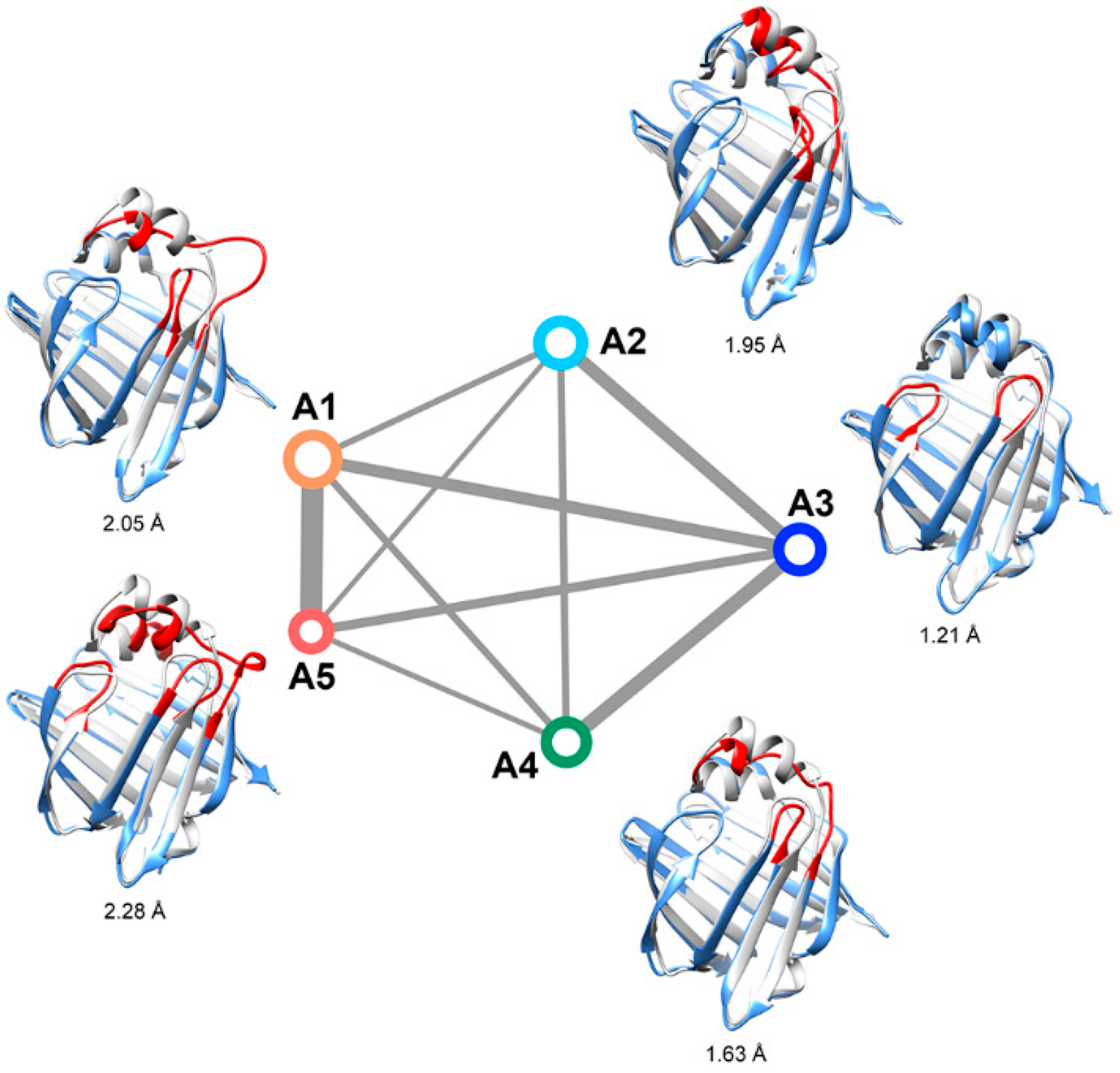
- the basic inequality shows us that we contract from one entropy to itself
- in fact, under LSI, we can contract into *better* norms:
    - let $p > 1, t > 0, p(t) = 1 + \left(p - 1\right) \cdot \exp(2 \cdot \lambda \cdot t)$. then

$$\left|P_t f\right|_{p(t)} \leq \left|f\right|_p$$

    - this is Nelson's "hypercontractivity"
    - actually, this also implies back the LSI

Some Related Inequalities of Interest

- { Cheeger-type / “L^1 Poincaré” } Inequalities
- { Weak, Super } Poincaré Inequalities
- { Weak, Super } Log-Sobolev Inequalities
- Weighted { Poincaré, Log-Sobolev } Inequalities
- Polyak-Łojasiewicz Inequalities in Optimisation
- A Glimpse Beyond “Energy-Entropy” Inequalities



Rapid-Fire

some directions, some references

Cheeger-Type Inequalities

- Poincaré asks that if $\pi(F) = 0$, it holds

$$\mathbf{E}\left[|F|^2\right] \lesssim \mathbf{E}\left[|\nabla F|^2\right]$$

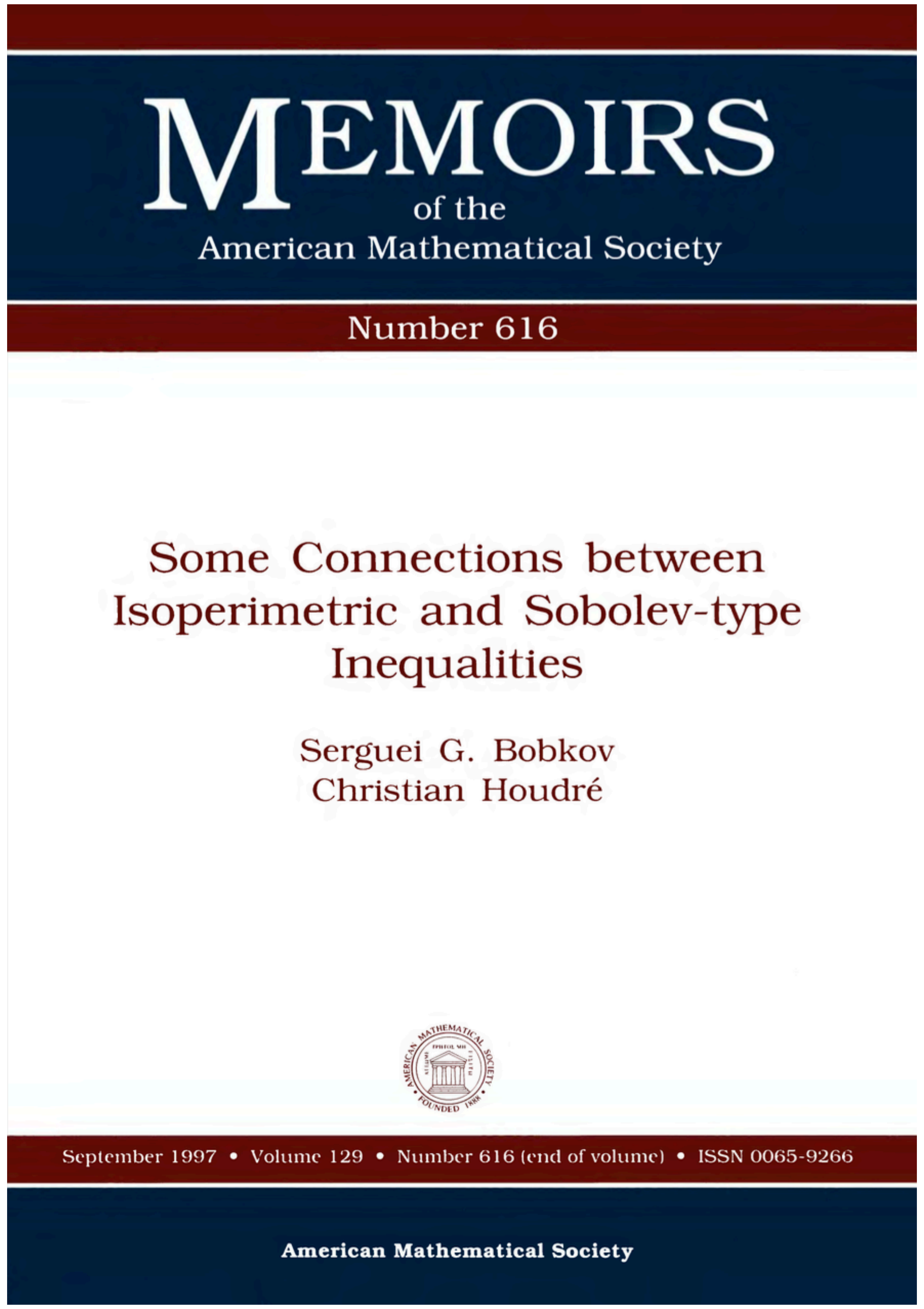
- Cheeger asks that if $\operatorname{med}_\pi(F) = 0$, it holds

$$\mathbf{E}\left[|F|\right] \lesssim \mathbf{E}\left[|\nabla F|\right]$$

- Weak Cheeger refines this to

$$\mathbf{E}\left[|F|\right] \lesssim \mathbf{E}\left[|\nabla F|\right] + \operatorname{Osc}(F)$$

  - equivalent to isoperimetric profile



Refined Poincaré Inequalities

- Poincaré asks that if $\pi(F) = 0$, it holds

$$\mathbf{E}\left[|F|^2\right] \lesssim \mathbf{E}\left[|\nabla F|^2\right]$$

- Weak Poincaré refines this to

$$\mathbf{E}\left[|F|^2\right] \lesssim \mathbf{E}\left[|\nabla F|^2\right] + \mathrm{Osc}(F)^2$$

- Super Poincaré refines this to

$$\mathbf{E}\left[|F|^2\right] \lesssim \mathbf{E}\left[|\nabla F|^2\right] + |F|_1^2$$

Journal of Functional Analysis **185**, 564–603 (2001)
doi:10.1006/jfan.2001.3776, available online at http://www.idealibrary.com on IDEAL®

Weak Poincaré Inequalities and L^2-Convergence Rates of Markov Semigroups[1]

Michael Röckner

Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany

and

Feng-Yu Wang

Department of Mathematics, Beijing Normal University, Beijing 100975, People's Republic of China

Communicated by L. Gross

Received September 12, 2000; revised March 2, 2001; accepted March 11, 2001

In order to describe L^2-convergence rates slower than exponential, the weak Poincaré inequality is introduced. It is shown that the convergence rate of a Markov semigroup and the corresponding weak Poincaré inequality can be determined by each other. Conditions for the weak Poincaré inequality to hold are presented, which are easy to check and which hold in many applications. The weak Poincaré inequality is also studied by using isoperimetric inequalities for diffusion and jump processes. Some typical examples are given to illustrate the general results. In particular, our results are applied to the stochastic quantization of field theory in finite volume. Moreover, a sharp criterion of weak Poincaré inequalities is presented for Poisson measures on configuration spaces.

Key Words: weak Poincaré inequality; isoperimetric inequality; L^2-convergence rate; Markov semigroup.

Refined Entropic Inequalities

- Log-Sobolev asks that

$$\mathrm{Ent}\left(F^2\right) \lesssim \mathbf{E}\left[\left|\nabla F\right|^2\right]$$

- Weak Log-Sobolev refines this to

$$\mathrm{Ent}\left(F^2\right) \lesssim \mathbf{E}\left[\left|\nabla F\right|^2\right] + \mathrm{Osc}\left(F\right)^2$$

- Super Log-Sobolev refines this to

$$\mathrm{Ent}\left(F^2\right) \lesssim \mathbf{E}\left[\left|\nabla F\right|^2\right] + \mathbf{E}\left[\left|F\right|^2\right]$$

Probab. Theory Relat. Fields (2007) 139:563–603
DOI 10.1007/s00440-007-0054-5

Weak logarithmic Sobolev inequalities and entropic convergence

P. Cattiaux · I. Gentil · A. Guillin

Received: 17 November 2005 / Revised: 4 December 2006 / Published online: 24 March 2007

Abstract In this paper we introduce and study a weakened form of logarithmic Sobolev inequalities in connection with various others functional inequalities (weak Poincaré inequalities, general Beckner inequalities, etc.). We also discuss the quantitative behaviour of relative entropy along a symmetric diffusion semi-group. In particular, we exhibit an example where Poincaré inequality can not be used for deriving entropic convergence whence weak logarithmic Sobolev inequality ensures the result.

Keywords Logarithmic Sobolev inequalities · Concentration inequalities · Entropy

Mathematics Subject Classification (2000) 26D10 · 60E15

Weighted Inequalities

- Poincaré asks that if $\pi(F) = 0$, it holds

$$\mathbf{E}\left[|F|^2 \right] \lesssim \mathbf{E}\left[|\nabla F|^2 \right]$$

- Weighted Poincaré instead asks for

$$\mathbf{E}\left[|F|^2 \right] \lesssim \mathbf{E}\left[\omega \cdot |\nabla F|^2 \right]$$

- ω relates to Langevin w/ multiplicative noise
    - can imagine 'preconditioned' Poincaré
- can seek the same for { LSI, etc. }

The Annals of Probability
2009, Vol. 37, No. 2, 403–427
DOI: 10.1214/08-AOP407

WEIGHTED POINCARÉ-TYPE INEQUALITIES FOR CAUCHY AND OTHER CONVEX MEASURES[1]

BY SERGEY G. BOBKOV AND MICHEL LEDOUX

University of Minnesota and Université Paul-Sabatier

Brascamp–Lieb-type, weighted Poincaré-type and related analytic inequalities are studied for multidimensional Cauchy distributions and more general κ-concave probability measures (in the hierarchy of convex measures). In analogy with the limiting (infinite-dimensional log-concave) Gaussian model, the weighted inequalities fully describe the measure concentration and large deviation properties of this family of measures. Cheeger-type isoperimetric inequalities are investigated similarly, giving rise to a common weight in the class of concave probability measures under consideration.

Functional Inequalities in Optimisation

- our inequalities look like

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{F}\left(x\left(t\right)\right) \gtrsim \mathscr{F}\left(x\left(t\right)\right) - \mathscr{F}_{\star}$$

- for deterministic gradient flows, this reads as

$$\left|\nabla F\right|^{2} \gtrsim F - F_{\star}$$

- 'Polyak-Łojasiewicz Inequality'
- can hold for non-convex F
  - e.g. nonlinear least-squares, deep learning

Linear Convergence of Gradient and Proximal-Gradient Methods Under the Polyak-Łojasiewicz Condition

Hamed Karimi, Julie Nutini, and Mark Schmidt(✉)

Department of Computer Science, University of British Columbia,
Vancouver, BC, Canada
hamedkarim@gmail.com, {jnutini,schmidtm}@cs.ubc.ca

Abstract. In 1963, Polyak proposed a simple condition that is sufficient to show a global linear convergence rate for gradient descent. This condition is a special case of the Łojasiewicz inequality proposed in the same year, and it does not require strong convexity (or even convexity). In this work, we show that this much-older Polyak-Łojasiewicz (PL) inequality is actually weaker than the main conditions that have been explored to show linear convergence rates without strong convexity over the last 25 years. We also use the PL inequality to give new analyses of coordinate descent and stochastic gradient for many non-strongly-convex (and some non-convex) functions. We further propose a generalization that applies to proximal-gradient methods for non-smooth optimization, leading to simple proofs of linear convergence for support vector machines and L1-regularized least squares without additional assumptions.

Keywords: Gradient descent · Coordinate descent · Stochastic gradient · Variance-reduction · Boosting · Support vector machines · L1-regularization

Beyond “Energy-Entropy” Inequalities

- i focused on “energy-entropy” inequalities
- there is an extended hierarchy of related inequalities, each of great independent interest in concentration and geometry
- in general, the implications ‘go downwards’
- under a curvature lower bound, these implications can also be reversed to ‘go upwards’

Russian Math. Surveys **67**:5 785–890

Uspekhi Mat. Nauk **67**:5 3–110
DOI 10.1070/RM2012v067n05ABEH004808

Dedicated to the centenary of the birth of Leonid Vital’evich Kantorovich

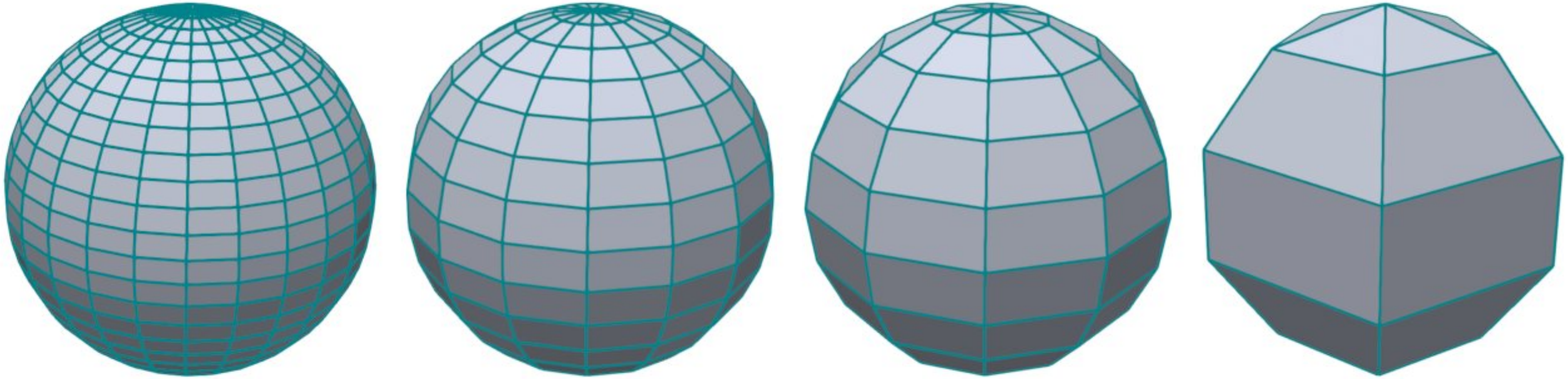
The Monge–Kantorovich problem: achievements, connections, and perspectives

V. I. Bogachev and A. V. Kolesnikov

Abstract. This article gives a survey of recent research related to the Monge–Kantorovich problem. Principle results are presented on the existence of solutions and their properties both in the Monge optimal transportation problem and the Kantorovich optimal plan problem, along with results on the connections between both problems and the cases when they are equivalent. Diverse applications of these problems in non-linear analysis, probability theory, and differential geometry are discussed.

Bibliography: 196 titles.

Keywords: Monge problem, Kantorovich problem, optimal transportation, transport inequality, Kantorovich–Rubinshtein metric.

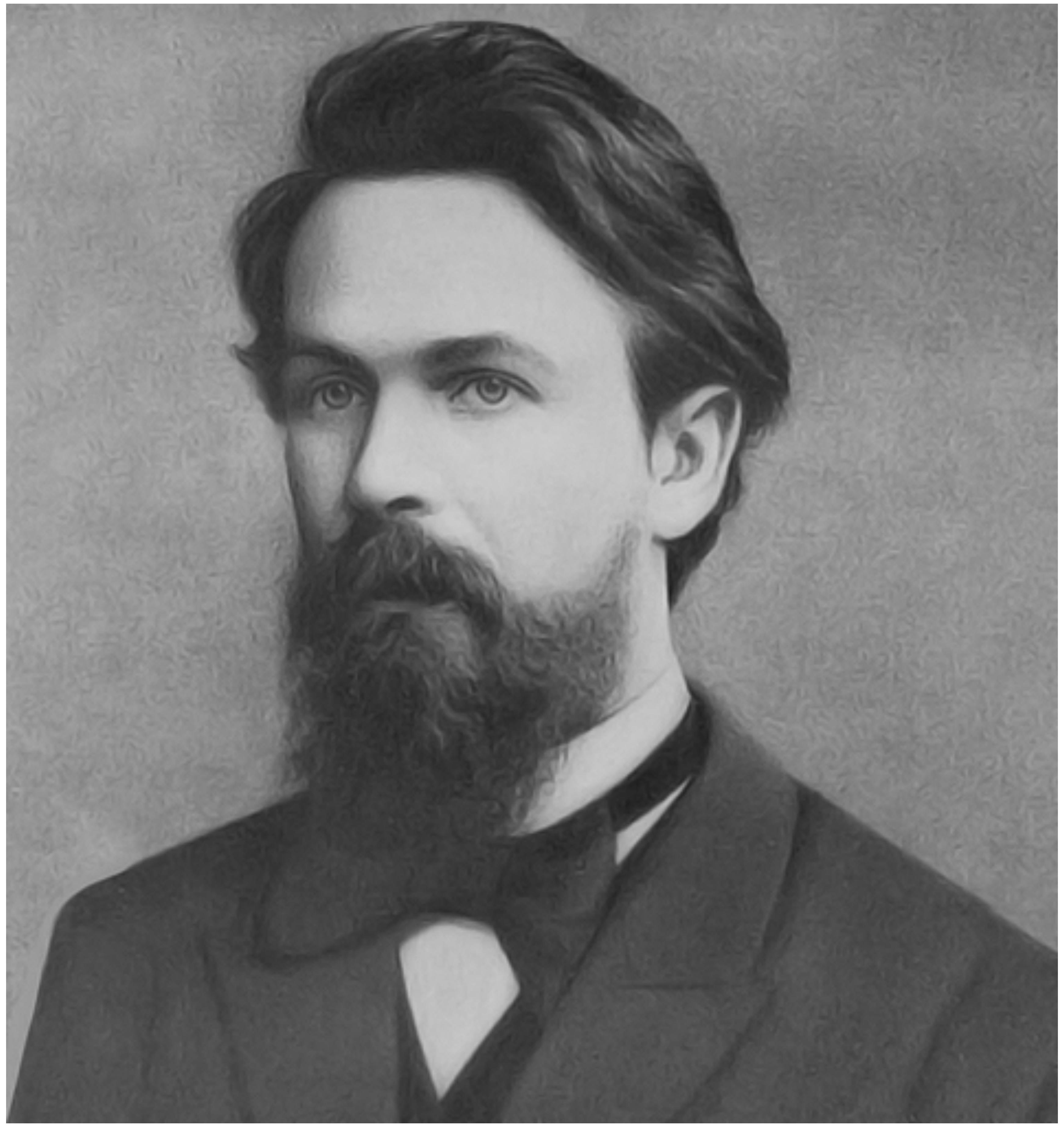


A Moment to Breathe

Shall We Continue?

<u>more content</u>

<u>wrap up</u>



Part 2: Discrete-Time Markov Chains

or, ‘Functional Analysis of Practical MCMC’

Setting the Scene

- some of the Langevin picture is quite particular
    - … but luckily, much of what we have discussed is quite general!
- let us work with a positive, π-reversible Markov kernel P
- let us also focus on the L^2 or ‘Poincaré’ picture

"… somehow, Grönwall returned"

Contraction in $\mathrm{L}^2(\pi)$

- as before, we might aim to show that whenever $\pi\left(f\right) = 0$,

$$\|Pf\|_2 \leq \left(1 - \gamma\right) \cdot \|f\|_2$$

- squaring and expanding, this is tantamount to the claim that

$$\mathscr{E}_{P^\star P}\left(f, f\right) \gtrsim \mathrm{var}_\pi\left(f\right)$$

for some 'interesting' object $\mathscr{E}_{P^\star P}$

Dirichlet Energy Forms in Discrete Time

- in our general setting, define

$$\begin{aligned} \mathscr{E}_T(f, g) &= \left\langle f, \left(\mathrm{Id} - T\right) g \right\rangle \\ &= \frac{1}{2} \int \pi\left(\mathrm{d}x\right) \cdot T\left(x, \mathrm{d}y\right) \cdot \Delta_{f,g}\left(x, y\right) \end{aligned}$$

where $\quad \Delta_{f,g}\left(x, y\right) = \left(f(x) - f\left(y\right)\right) \cdot \left(g(x) - g\left(y\right)\right)$

- for reversible, positive chains, $\mathscr{E}_P$ and $\mathscr{E}_{P \star P}$ are directly comparable

- $\mathscr{E}_P$ will be easier for us to work with

Connecting to the Diffusion Picture

- let $t > 0$, let P_t be the transition kernel for OLD (π)
- as $t \to 0^+$, one checks that $t^{-1} \cdot \mathscr{E}_{P_t}\left(f, g\right) \to \mathbf{E}\left[\langle \nabla f, \nabla g \rangle\right]$
  - 'proof': $\mathrm{Id} - P_t \approx - t \cdot \mathscr{L}$
- so, we have 'genuinely' generalised the diffusion picture
- our 'Poincaré inequality' is now

$$\mathscr{E}_P\left(f,f\right) \geq \gamma \cdot \mathrm{var}_\pi\left(f\right)$$

Back to Square One?

- instructive to revisit earlier principles, and assess what remains
    - tensorisation, bounded change of measure, mixtures
    - metric contractivity, ‘typical’ curvature, Lyapunov arguments
- when do we lose?
    - usually: when the *chain rule* was important

An Example: What *is* the LSI?

- we might reasonably propose either of

$$\frac{\lambda}{2} \cdot \text{Ent}\left(F^2\right) \leq \mathscr{E}(F, F)$$

$$2 \cdot \lambda \cdot \text{Ent}(F) \leq \mathscr{E}\left(F, \log F\right)$$

- in general, the former is *strictly stronger*; see e.g. Bobkov-Tetali
- Salez-Tikhomirov-Youssef clarify when the latter can ~imply the former

Application to MCMC: Case Studies (1)

- some warm-ups in the propose-accept-reject paradigm
  - changing the acceptance probability from Metropolis to Barker (to …)
  - bounded change-of-measure for target
  - bounded change-of-measure for proposal kernel

Application to MCMC: Case Studies (2)

- exact-approximate Metropolis-Hastings methods
  - safely approximating the accept-reject decision
- exact-approximate Gibbs sampling
  - safely approximating the conditional simulation step

Set-Up: Metropolis-Hastings kernels

- protocol: at x,
  - propose move to $y \sim Q\left(x, \mathrm{d}y\right)$
  - compute $r\left(x, y\right) = \pi\left(y\right) \cdot Q\left(y, x\right) / \pi\left(x\right) \cdot Q\left(x, y\right)$
  - with probability $\alpha\left(x, y\right) = \min\left\{1, r\left(x, y\right)\right\}$, move to y
    - otherwise, stay at x

The Dirichlet Energy Form

- in this paradigm, the Dirichlet form writes as

$$\mathscr{E}_P\left(f, g\right) = \frac{1}{2} \int \pi\left(\mathrm{d}x\right) \cdot Q\left(x, \mathrm{d}y\right) \cdot \alpha\left(x, y\right) \cdot \Delta_{f,g}\left(x, y\right)$$

- note that rejections are ‘ignored’, rather than ‘separated’

Some Warm-Up Applications

- these are *a little bit* interesting, but mostly illustrate the ‘tricks of the trade’
- qualitatively, they resemble the earlier “bounded change-of-measure”
- the emphasis is that this trick can now be applied in a few different places

Warm-Up: Acceptance Functions (1)

- fix π, Q, but consider accepting moves with either of

$$\alpha^{\mathrm{MH}} = \min\{1, r\} \qquad \alpha^{\mathrm{B}} = \frac{r}{1+r}.$$

- both generate π-reversible chains, so fit for purpose
    - more generally, if $\beta : \mathbf{R}^{+} \to [0,1]$ satisfies $\beta(r) = r \cdot \beta\left(r^{-1}\right)$, then we call it a 'balancing function', and can use $\alpha^{\beta} = \beta(r)$
- how shall we choose between the two options?

Warm-Up: Acceptance Functions (2)

- some painless verification:

$$\frac{1}{2} \cdot \min\{1, r\} \leq \frac{r}{1+r} \leq \min\{1, r\}$$

- consequence:

$$\frac{1}{2} \cdot \mathscr{E}_P^{\mathrm{MH}}\left(f, f\right) \leq \mathscr{E}_P^{\mathrm{B}}\left(f, f\right) \leq \mathscr{E}_P^{\mathrm{MH}}\left(f, f\right)$$

- conclusion: other things being equal, no worse off by using MH
    - … but if you do use the Barker acceptance, you won't lose too much

Warm-Up: Change of Proposal (1)

- fix π, and consider choosing between different proposal kernels Q_1, Q_2
  - suppose that they are each symmetric, so that $\alpha\left(x, y\right) = \min\left\{1, \pi\left(y\right)/\pi\left(x\right)\right\}$
- suppose that uniformly in x, y, we can bound

$$0 < \delta^- \leq \frac{Q_2\left(x, y\right)}{Q_1\left(x, y\right)} \leq \delta^+ < \infty$$

  - LHS: tails of Q_2 are at least as heavy as those of Q_1 (similar on RHS)

Warm-Up: Change of Proposal (2)

- suppose that uniformly in x, y, we can bound

$$\delta^{-} \leq \frac{Q_2\left(x, y\right)}{Q_1\left(x, y\right)} \leq \delta^{+}$$

- consequence:

$$\delta^{-} \cdot \mathscr{E}_{P_1}\left(f, f\right) \leq \mathscr{E}_{P_2}\left(f, f\right) \leq \delta^{+} \cdot \mathscr{E}_{P_1}\left(f, f\right)$$

- one interpretation: slightly heavier-tailed proposals can't hurt too much

Warm-Up: Change of Target (1)

- fix some symmetric Q, and consider changing target from π to π'
  - as earlier, assume that $0 < \sup \mathrm{d}\pi/\mathrm{d}\pi', \sup \mathrm{d}\pi'/\mathrm{d}\pi < \infty$
- want to compare

$$\mathscr{E}_P\left(f,f\right) = \frac{1}{2}\int \pi\left(\mathrm{d}x\right)\cdot Q\left(x,\mathrm{d}y\right)\cdot\alpha\left(x,y\right)\cdot\Delta_{f,f}\left(x,y\right)$$

$$\mathscr{E}_{P'}\left(f,f\right) = \frac{1}{2}\int \pi'\left(\mathrm{d}x\right)\cdot Q\left(x,\mathrm{d}y\right)\cdot\alpha'\left(x,y\right)\cdot\Delta_{f,f}\left(x,y\right)$$

Warm-Up: Change of Target (2)

- first, study acceptance probabilities

$$\begin{aligned}\alpha'\left(x,y\right) &= \min\left\{1, \pi'\left(y\right)/\pi'(x)\right\}\\ &\geq \min\left\{1, \kappa^{-1}\cdot\pi\left(y\right)/\pi(x)\right\}\\ &\geq \kappa^{-1}\cdot\min\left\{1, \pi\left(y\right)/\pi(x)\right\}\\ &= \kappa^{-1}\cdot\alpha\left(x,y\right)\end{aligned}$$

where $\kappa = \left(\sup \mathrm{d}\pi/\mathrm{d}\pi'\right)\left(\sup \mathrm{d}\pi'/\mathrm{d}\pi\right)$

Warm-Up: Change of Target (3)

- follows that

$$\mathscr{E}_P\left(f,f\right) = \frac{1}{2}\int \pi\left(\mathrm{d}x\right)\cdot Q\left(x,\mathrm{d}y\right)\cdot\alpha\left(x,y\right)\cdot\Delta_{f,f}\left(x,y\right)$$

$$\leq \left(\sup \mathrm{d}\pi/\mathrm{d}\pi'\right)\cdot\kappa\cdot\frac{1}{2}\int \pi'\left(\mathrm{d}x\right)\cdot Q\left(x,\mathrm{d}y\right)\cdot\alpha'\left(x,y\right)\cdot\Delta_{f,f}\left(x,y\right)$$

$$= \left(\sup \mathrm{d}\pi/\mathrm{d}\pi'\right)\cdot\kappa\cdot\mathscr{E}_{P'}\left(f,f\right)$$

- old argument still works to show that $\mathrm{var}_{\pi'}\left(f\right) \leq \left(\sup \mathrm{d}\pi/\mathrm{d}\pi'\right)\cdot\mathrm{var}_{\pi}\left(f\right)$

Warm-Up: Change of Target (4)

- so, assuming that for all suitable f,

$$\gamma \cdot \mathrm{var}_{\pi}\left(f\right) \leq \mathscr{E}_P\left(f, f\right),$$

we can argue that

$$\begin{aligned}
\mathrm{var}_{\pi'}\left(f\right) &\leq \left(\sup \mathrm{d}\pi / \mathrm{d}\pi'\right) \cdot \mathrm{var}_{\pi}\left(f\right) \\
&\leq \left(\sup \mathrm{d}\pi / \mathrm{d}\pi'\right) \cdot \gamma^{-1} \cdot \mathscr{E}_P\left(f, f\right) \\
&\leq \left(\sup \mathrm{d}\pi / \mathrm{d}\pi'\right) \cdot \gamma^{-1} \cdot \left(\sup \mathrm{d}\pi' / \mathrm{d}\pi\right) \cdot \kappa \cdot \mathscr{E}_{P'}\left(f, f\right) \\
&= \gamma^{-1} \cdot \kappa^2 \cdot \mathscr{E}_{P'}\left(f, f\right)
\end{aligned}$$

i.e. we preserve the spectral gap, up to a factor of κ^2

Improved Argument (1)

- studying target and acceptance separately is actually lossy
- useful to jointly study

$$J\left(\mathrm{d}\left(x,y\right)\right):=\pi\left(\mathrm{d}x\right)\cdot Q\left(x,\mathrm{d}y\right)\cdot\alpha\left(x,y\right)$$
$$=\min\left\{\pi\left(\mathrm{d}x\right)\cdot Q\left(x,\mathrm{d}y\right),\pi\left(\mathrm{d}y\right)\cdot Q\left(y,\mathrm{d}x\right)\right\}$$

  - changing π to π' incurs a factor of $\left(\sup \mathrm{d}\pi/\mathrm{d}\pi'\right)$
- sharpens to $\mathscr{E}_P\left(f,f\right)\leq\left(\sup \mathrm{d}\pi/\mathrm{d}\pi'\right)\cdot\mathscr{E}_{P'}\left(f,f\right)$
  - sharpens to $\mathrm{var}_{\pi'}\left(f\right)\leq\gamma^{-1}\cdot\kappa\cdot\mathscr{E}_{P'}\left(f,f\right)$

“Exact-Approximate” Methods

- the upcoming applications are instances of “exact-approximate” Monte Carlo methods
- these arise in problems where some ‘standard’ method is not possible to implement directly, due to some computational intractability
- the general idea is to replace the intractable part with a suitable Monte Carlo approximation, defining a new ‘approximate’ process
    - … which nevertheless *has the right limiting behaviour*
- these can be of substantial interest for various statistical models

Application: Exact-Approximate MH (1)

- in some settings, we cannot evaluate $\pi(x)$, even up to a constant,
  - … but we can estimate it unbiasedly
- formalisation:
  - draw $W \sim \omega_x$ s.t. $\mathbf{E}^x[W] = 1$, and observe $\widehat{\pi(x; w)} = \pi(x) \cdot w$
- using this estimator in an MH accept-reject decision *remains valid*

'the pseudo-marginal approach'

Set-Up: PMMH kernels

- protocol: at (x, w),
  - propose move to $\left(y, u\right) \sim Q\left(x, \mathrm{d}y\right) \cdot \omega_y\left(\mathrm{d}u\right)$
  - compute $r\left((x, w), \left(y, u\right)\right) = \widehat{\pi\left(y; u\right)} \cdot Q\left(y, x\right) / \widehat{\pi\left(x; w\right)} \cdot Q\left(x, y\right)$
  - with probability $\alpha\left((x, w), \left(y, u\right)\right) = \min\left\{1, r\right\}$, move to $\left(y, u\right)$
    - otherwise, stay at (x, w)

Application: Exact-Approximate MH (2)

- for PMMH, the Dirichlet Energy Form reads as

$$\mathscr{E}_{\tilde{P}}\left(f,g\right)=\frac{1}{2}\int\pi\left(\mathrm{d}x\right)\cdot\omega_{x}\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_{y}\left(\mathrm{d}u\right)\cdot\alpha\left(x,w,y,u\right)\cdot\Delta_{f,g}$$

- the ‘ideal’ algorithm would instead have Dirichlet Energy Form

$$\mathscr{E}_{P}\left(f,g\right)=\frac{1}{2}\int\pi\left(\mathrm{d}x\right)\cdot\omega_{x}\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_{y}\left(\mathrm{d}u\right)\cdot u\cdot\alpha\left(x,y\right)\cdot\Delta_{f,g}$$

Negative Result

- suppose that f depends only on x, not on w. then,

$$
\begin{aligned}
\mathscr{E}_{\tilde{P}}\left(f,f\right) &= \frac{1}{2}\int\pi\left(\mathrm{d}x\right)\cdot\omega_{x}\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_{y}\left(\mathrm{d}u\right)\cdot\alpha\left(x,w,y,u\right)\cdot\Delta_{f,f}\\
&= \frac{1}{2}\int\pi\left(\mathrm{d}x\right)\cdot\omega_{x}\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_{y}\left(\mathrm{d}u\right)\cdot\min\left\{1,r\left(x,y\right)\cdot\frac{u}{w}\right\}\cdot\Delta_{f,f}\\
&\overset{\mathrm{Jen}}{\leq} \frac{1}{2}\int\pi\left(\mathrm{d}x\right)\cdot\omega_{x}\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\min\left\{1,r\left(x,y\right)\cdot\frac{1}{w}\right\}\cdot\Delta_{f,f}\\
&= \frac{1}{2}\int\pi\left(\mathrm{d}x\right)\cdot\omega_{x}\left(\mathrm{d}w\right)\cdot Q\left(x,\mathrm{d}y\right)\cdot\min\left\{w,r\left(x,y\right)\right\}\cdot\Delta_{f,f}\\
&\overset{\mathrm{Jen}}{\leq} \frac{1}{2}\int\pi\left(\mathrm{d}x\right)\cdot Q\left(x,\mathrm{d}y\right)\cdot\min\left\{1,r\left(x,y\right)\right\}\cdot\Delta_{f,f}\\
&= \mathscr{E}_{P}\left(f,f\right)
\end{aligned}
$$

Positive Result

- suppose that $w \leq \bar{w}$ almost surely. then,

$$\begin{aligned}
\mathscr{E}_{\tilde{P}}\left(f,f\right) &= \frac{1}{2}\int \pi\left(\mathrm{d}x\right)\cdot\omega_x\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_y\left(\mathrm{d}u\right)\cdot\alpha\left(x,w,y,u\right)\cdot\Delta_{f,f} \\
&= \frac{1}{2}\int \pi\left(\mathrm{d}x\right)\cdot\omega_x\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_y\left(\mathrm{d}u\right)\cdot\min\left\{1,r\left(x,y\right)\cdot\frac{u}{w}\right\}\cdot\Delta_{f,f} \\
&= \frac{1}{2}\int \pi\left(\mathrm{d}x\right)\cdot\omega_x\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_y\left(\mathrm{d}u\right)\cdot u\cdot\min\left\{\frac{1}{u},r\left(x,y\right)\cdot\frac{1}{w}\right\}\cdot\Delta_{f,f} \\
&\geq \bar{w}^{-1}\cdot\left(\frac{1}{2}\int \pi\left(\mathrm{d}x\right)\cdot\omega_x\left(\mathrm{d}w\right)\cdot w\cdot Q\left(x,\mathrm{d}y\right)\cdot\omega_y\left(\mathrm{d}u\right)\cdot u\cdot\min\left\{1,r\left(x,y\right)\right\}\cdot\Delta_{f,f}\right) \\
&= \bar{w}^{-1}\cdot\mathscr{E}_P\left(f,f\right)
\end{aligned}$$

Comments

- negative result doesn't easily hold for all functions, but is still useful
- positive result is illustrative of how to handle 'strong' comparisons
  - start by understanding 'uniformly-bounded perturbation' setting
- for more advanced versions, need to use 'Metropolis lemma'

$$\min\{1, a \cdot b\} \geq \min\{1, a\} \cdot \min\{1, b\}$$

  - often well-suited to MCMC applications
- in each case, important to identify good target of comparison

Application: Exact-Approximate Gibbs (1)

- exact-approximate Gibbs sampling
  - safely approximating the conditional simulation step
  - think of Metropolis-within-Gibbs, Hybrid Slice Sampling, etc.

Set-Up: (EA-)RSGS kernels

- protocol: at x,
  - sample $i \sim \text{Categorical}\left(\underline{\lambda}\right)$
  - sample $x_i' \sim \pi_i\left(\mathrm{d}x_i' \mid x_{-i}\right)$; overwrite x_i
- exact-approximate setting:
  - for each $\left(i, x_{-i}\right)$, kernel $K_i\left(\,\cdot\,,\,\cdot\,\cdot \mid x_{-i}\right)$ is { reversible, positive } for $\pi_i\left(\,\cdot \mid x_{-i}\right)$
  - replace "sample $x_i' \sim \pi_i\left(\mathrm{d}x_i' \mid x_{-i}\right)$" with "sample $x_i' \sim K_i\left(x_i, \mathrm{d}x_i' \mid x_{-i}\right)$"

Application: Exact-Approximate Gibbs (2)

- for RSGS, the Dirichlet Energy Form reads as

$$\mathscr{E}_P\left(f,g\right) = \sum_{i \in \mathscr{I}} \lambda_i \cdot \int \pi_{-i}\left(\mathrm{d}x_{-i}\right) \cdot \mathrm{cov}_{\pi_i\left(\,\cdot\,\mid x_{-i}\right)}\left(f,g\right)$$

- for EA-RSGS, it instead reads as

$$\mathscr{E}_{\tilde{P}}\left(f,g\right) = \sum_{i \in \mathscr{I}} \lambda_i \cdot \int \pi_{-i}\left(\mathrm{d}x_{-i}\right) \cdot \mathscr{E}_{K_i\left(\,\cdot\,\mid x_{-i}\right)}\left(f,g\right)$$

Application: Exact-Approximate Gibbs (3)

- for 'easy' comparison, assume that
    - for each $\left(i, x_{-i}\right)$, $K_i\left(\cdot \mid x_{-i}\right)$ has a spectral gap $\geq \gamma$
- then, write

$$\begin{aligned}
\mathscr{E}_{\tilde{P}}\left(f, f\right) &= \sum_{i \in \mathscr{I}} \lambda_i \cdot \int \pi_{-i}\left(\mathrm{d}x_{-i}\right) \cdot \mathscr{E}_{K_i\left(\cdot \mid x_{-i}\right)}\left(f, f\right) \\
&\geq \sum_{i \in \mathscr{I}} \lambda_i \cdot \int \pi_{-i}\left(\mathrm{d}x_{-i}\right) \cdot \gamma \cdot \mathrm{var}_{\pi_i\left(\cdot \mid x_{-i}\right)}\left(f\right) \\
&= \gamma \cdot \mathscr{E}_P\left(f, f\right)
\end{aligned}$$

Comments

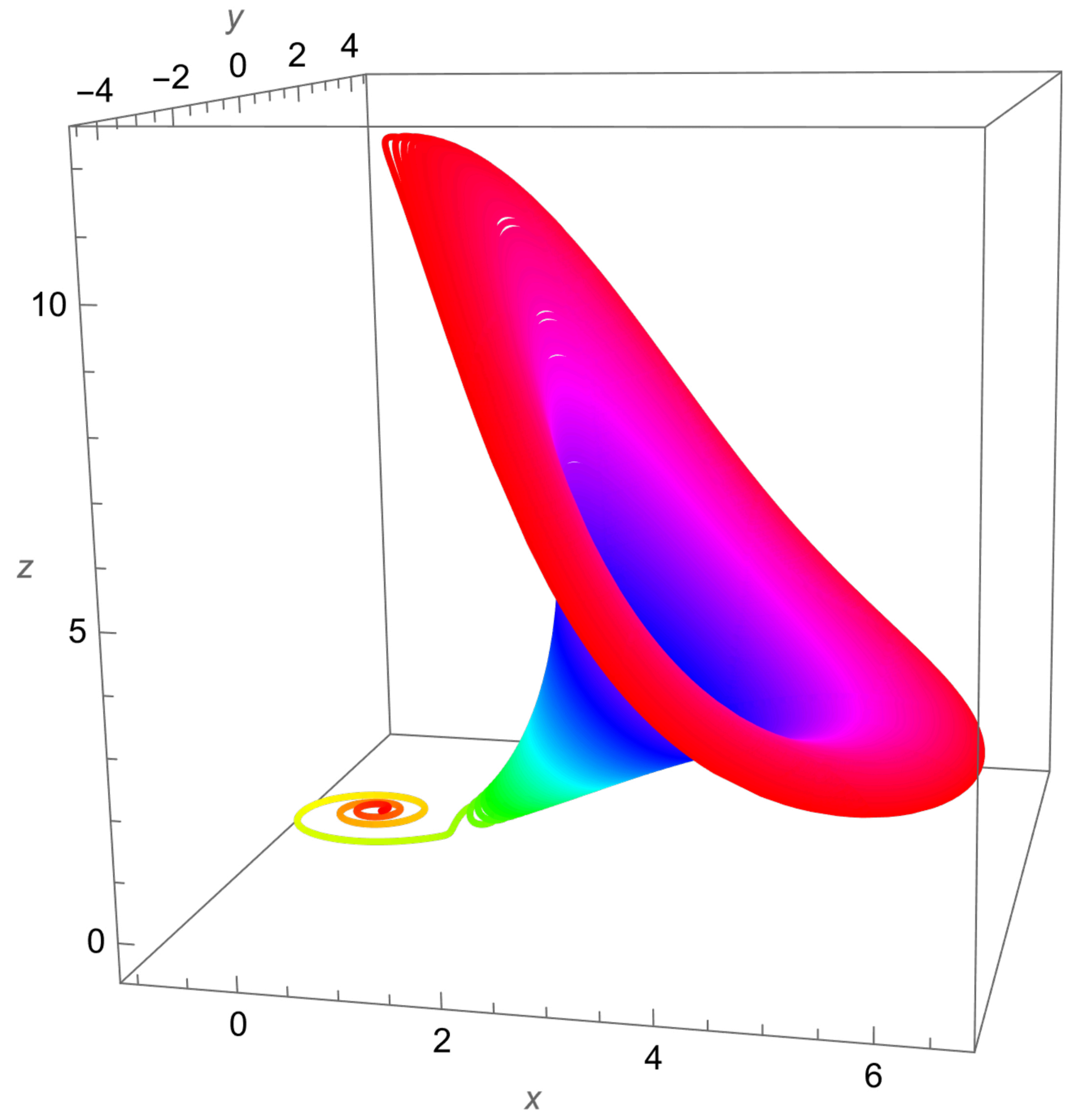
- can get negative result fairly easily through same arguments
  - (if exact conditional is available, then you should probably use it)
- immediately opens up numerous applications
  - Hit-and-Run, Slice Sampling, Particle Gibbs, …
- contrast { small-block, large-block } Gibbs

Taking Stock

- when is the functional approach applicable and easy-to-use?
  - ‘exact approximation’: still π-invariant, reversible, but replacing

    “*do [XXX] exactly*” $\leftarrow$ “*do [the Monte Carlo approximation to XXX]*”

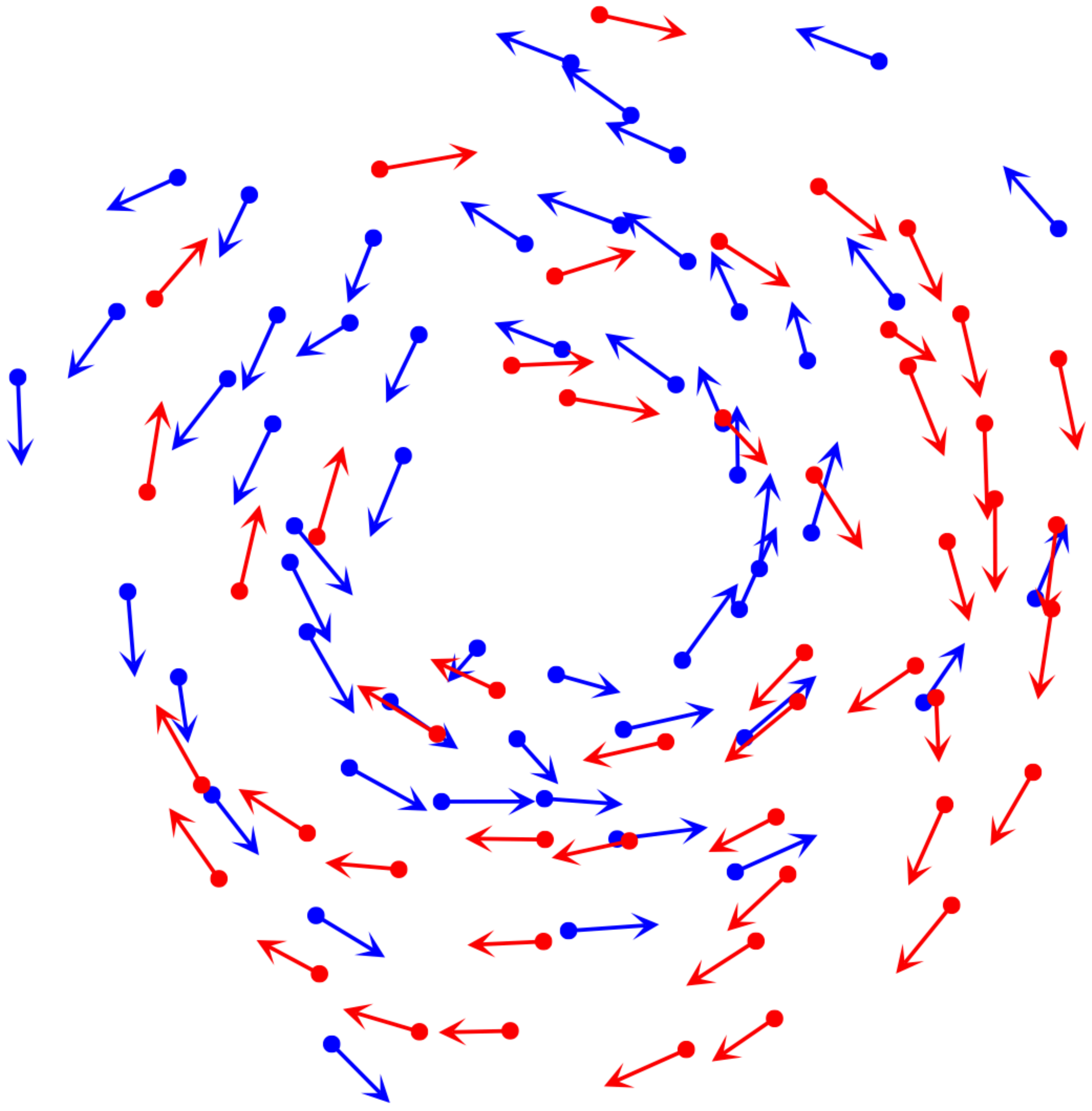
    - difference between ‘exact’ and ‘approximate’ is a probability density
- note that the fine details of the approximation rarely matter
  - instead: { uniform bounds, moment bounds, etc. } tell quite a clear story



Taking Stock Briefly

Revisiting our Goals

- ‘functional inequalities’ describe the convergence behaviour of Markov processes ‘at large scales’
- such tools are appealing in high dimension, when uniform curvature is not easily available
- probabilistic techniques are often a good route to a first functional inequality for ‘nice’ problems



Some Good and Some Bad

- a functional inequality is a sensible, robust way to say that

  'this { measure / process } is nice'

- the 'global' nature of functional inequalities is instructive

- functional inequalities compose very nicely

  - particularly useful for analysis of 'meta-algorithms'

- proving a functional inequality from scratch can be some work

- working with processes which are { inhomogeneous, non-reversible processes, of unknown invariant measure, … } may not be obvious

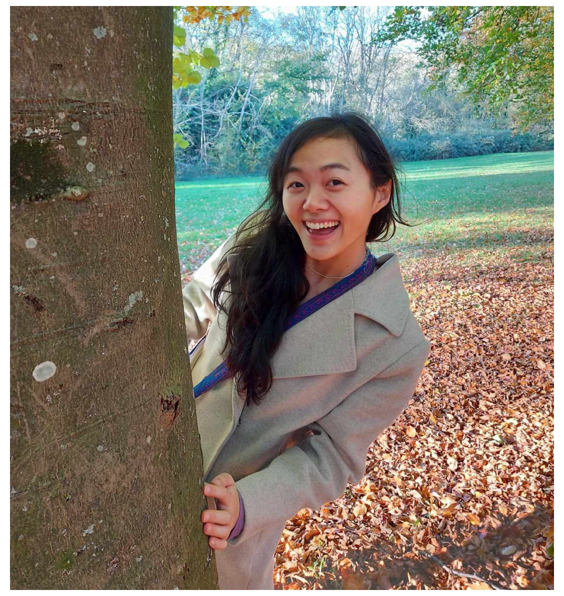
- building intuition for functional results can require some guidance

- 'conservation of (mathematical) difficulty' holds at some level

Some Recap and Outlook

- the toolbox of Markov processes and their long-time behaviour is vast
  - many nice frameworks exist, and many are somewhat compatible
- the functional-analytic framework seems rather accommodating
  - … and much is known about it already
- when studying your next Markov process, concrete or abstract,
  - … perhaps the functional-analytic perspective can offer you something!

Fin

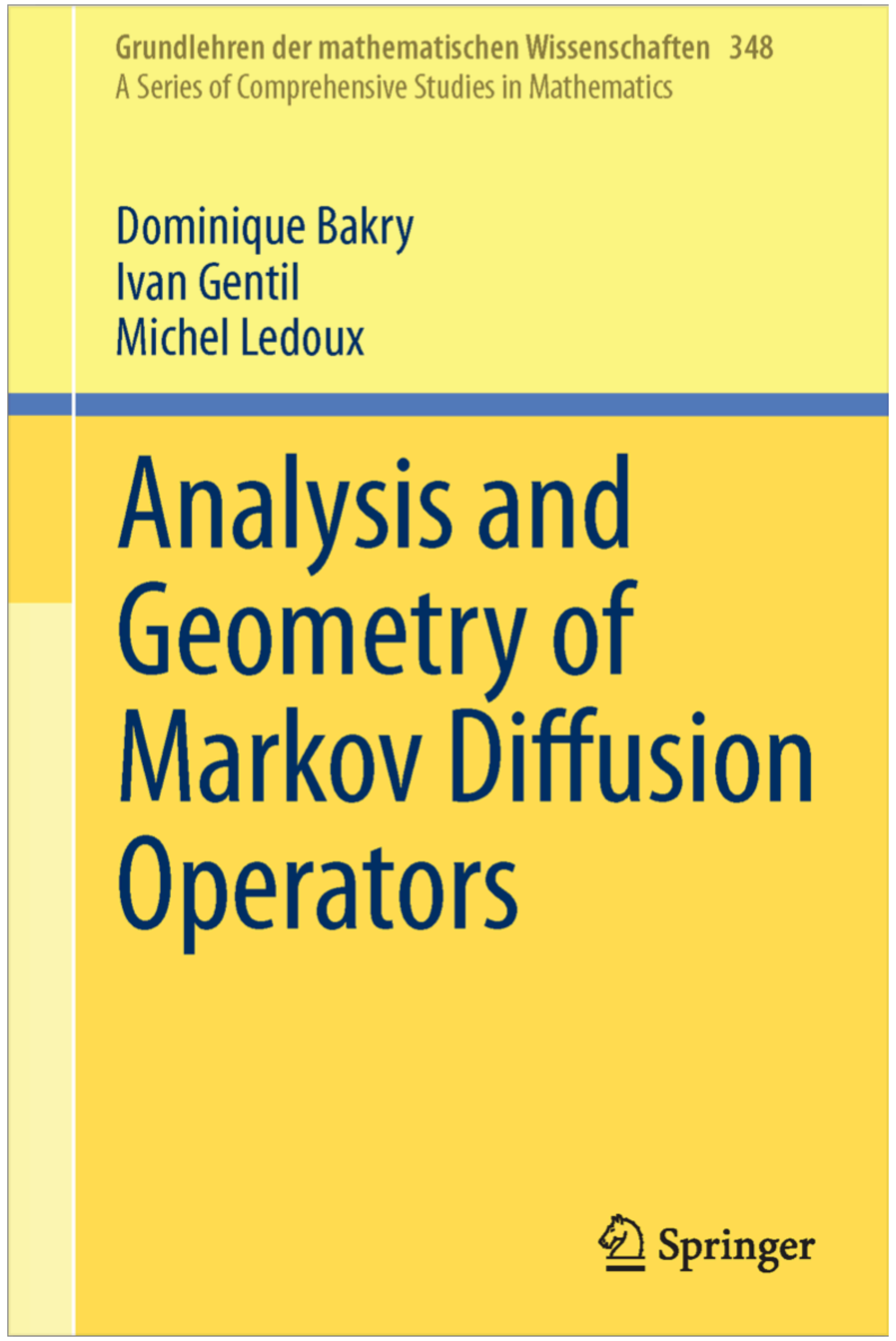


Some Key References (1)

- Bakry, Gentil, Ledoux - "Analysis and Geometry of Markov Diffusion Operators"

- Bonnefont - "Poincaré Inequality with Explicit Constant in Dimension $d \geq 1$", Lecture Notes

- Chafaï, Lehec - "Logarithmic Sobolev Inequalities Essentials", Lecture Notes



(back)

Some Key References (2)

- Chen - “Geometric Flows for Applied Mathematicians”, Lecture Notes
- Chewi - “Log-Concave Sampling”, Book-in-Progress
- van Handel - “Probability in High Dimension”, Lecture Notes

Geometric Flows for Applied Mathematicians*

Xiaohui Chen

This version: December 24, 2020

Some Additional References

- Chen - "Eigenvalues, Inequalities, and Ergodic Theory"

- Montenegro, Tetali - "Mathematical Aspects of Mixing Times in Markov Chains"

- Wang - "Functional Inequalities, Markov Semigroups and Spectral Theory"