University of BRISTOL

Contractivity of Markov Processes

Sam Power

University of Bristol

4 April, 2023

Contractivity of Markov Processes

Lecture 1

Sam Power

University of Bristol

4 April, 2023

Introduction

- If a Markov chain satisfies some form of 'curvature' or 'contractivity' condition, then the chain can be very nice, and we can draw many interesting conclusions about the chain and its invariant measure.
- This condition is naturally framed in the language of optimal transport theory:

$$\mathcal{T}_{1,\mathsf{d}}\left(P\left(x,\cdot\right),P\left(y,\cdot\right)\right) \leqslant \left(1-\kappa\right)\cdot \mathsf{d}\left(x,y\right).$$

- In this Section, I will explain the meaning and consequences of this condition.

Metric Spaces

- A metric is a symmetric mapping $\mathrm{d} : E \times E \to \mathbb{R}_+$ satisfying
  1. $\mathrm{d}(x, y) \geqslant 0$ (with equality holding if and only if $x = y$), and
  2. The triangle inequality: $\mathrm{d}(x, y) \leqslant \mathrm{d}(x, z) + \mathrm{d}(z, y)$.
- Given a space E and a metric d on E, we call (E, d) a metric space.

Examples of Metrics (1/2)

1. The discrete metric on arbitrary E, where $\mathrm{d}(x, y) = \mathrm{I}[x \neq y]$.
2. The Euclidean metric on $E = \mathrm{R}^d$, where $\mathrm{d}(x, y) = \|x - y\|_2$.
3. The Hamming metric on $E = \{\pm 1\}^d$, where $\mathrm{d}(x, y) = \sum_{i \in [d]} \mathrm{I}[x_i \neq y_i]$.
4. The ℓ_p metric on $E = \mathrm{R}^d$, where $\mathrm{d}(x, y) = \|x - y\|_p$.
   - $p \in \{1, 2, \infty\}$ will (as usual) typically be the interesting cases.

Examples of Metrics (2/2)

1. The V-weighted discrete metric on arbitrary E, where

$$\mathsf{d}(x, y) = \mathrm{I}[x \neq y] \cdot (V(x) + V(y))$$

   for some $V : E \to \mathrm{R}_+$ which satisfies $V(x) > 0$ for all $x \in E$.

2. The f-transformed Euclidean metric on $E = \mathrm{R}^d$, where

$$\mathsf{d}(x, y) = f(\|x - y\|_2)$$

   for some $f : \mathrm{R}_+ \to \mathrm{R}_+$ which is increasing, concave, and satisfies $f(0) = 0$.

Lipschitz Functions

- One often thinks about metric spaces in terms of points and distances.
- One can dually think about functions on those spaces.
- Say that a function $f : E \to \mathrm{R}$ is Lipschitz with respect to d if

$$|f|_{\mathrm{Lip}(\mathrm{d})} := \sup \left\{ \frac{|f(y) - f(x)|}{\mathrm{d}(x, y)} : x, y \in E, x \neq y \right\}$$

  is finite, and write $\mathrm{Lip}(E, \mathrm{d})$ for the set of all such functions.
- Note that $|\cdot|_{\mathrm{Lip}(\mathrm{d})}$ is a seminorm.
- Exercise: Given the Lipschitz seminorm, can you recover the underlying metric?

Examples of Lipschitz Seminorms

- For d the discrete metric, it holds that

$$|f|_{\mathrm{Lip}(\mathsf{d})} = |f|_{\mathrm{osc}} := \sup\{|f(y) - f(x)| : x, y \in E, x \neq y\}.$$

- For d the Euclidean metric on $E = \mathrm{R}^d$, it holds (for f smooth enough) that

$$|f|_{\mathrm{Lip}(\mathsf{d})} = \|\nabla f\|_\infty := \sup\{\|\nabla f(x)\|_2 : x \in E\}.$$

- <u>Exercise</u>: Describe $|\cdot|_{\mathrm{Lip}(\mathsf{d})}$ when
  1. d is a V-weighted discrete metric on E.
  2. d is the ℓ_p metric on $E = \mathrm{R}^d$ for $p \in \{1, \infty\}$

Probability Measures and Couplings

- Given nice (E, d), write $\mathcal{P}(E)$ for the space of probability measures on E.
- Given $\alpha, \beta \in \mathcal{P}(E)$, write $\Gamma(\alpha, \beta)$ for the set of couplings of (α, β), i.e. the set of probability distributions $\gamma \in \mathcal{P}(E \times E)$ whose marginals are given by (α, β).
- Couplings are useful for comparing probability measures.
  - "Can we jointly construct draws from (α, β) which are similar in certain ways?"

Transport Distances

- For $p \geqslant 1$, define the (p, d)-transport distance on $\mathcal{P}(E)$ by

$$\begin{aligned} \mathcal{T}_{p,\mathsf{d}} : \mathcal{P}(E) \times \mathcal{P}(E) &\to \mathrm{R}_+ \\ (\alpha, \beta) &\mapsto \inf \left\{ \left(\int \gamma(\mathrm{d}x, \mathrm{d}y) \cdot \mathsf{d}(x, y)^p \right)^{1/p} : \gamma \in \Gamma(\alpha, \beta) \right\} \end{aligned}$$

- It can be shown that $\mathcal{T}_{p,\mathsf{d}}$ is itself a metric over the space $\mathcal{P}(E)$.
- In the special case $p = 1$, we have the 'Kantorovich' dual representation

$$\mathcal{T}_{1,\mathsf{d}}(\alpha, \beta) = \sup \left\{ \alpha(f) - \beta(f) : |f|_{\mathrm{Lip}(\mathsf{d})} \leqslant 1 \right\}.$$

Curvature Definitions

- Define ‘curvature’ $\kappa(x, y) \in (-\infty, 1]$ so that

$$\mathcal{T}_{1,\mathsf{d}}\left(P\left(x,\cdot\right),P\left(y,\cdot\right)\right)=\left(1-\kappa\left(x,y\right)\right)\cdot\mathsf{d}\left(x,y\right).$$

- Define also

$$\begin{aligned}\kappa\left(x\right) &:= \inf\left\{\kappa\left(x,y\right): y \in E\setminus\{x\}\right\}\\ \kappa &:= \inf\left\{\kappa\left(x\right): x \in E\right\}\\ &= \inf\left\{\kappa\left(x,y\right): x \neq y \in E\right\}.\end{aligned}$$

- Our (major) standing assumption will be that $\kappa > 0$.
- If d is ”ϵ-geodesic”, then it suffices to take infima over $\{x \neq y \in E : \mathsf{d}(x, y) \leqslant \epsilon\}$.

Why Use Curvature?

- Curvature is a strong, *uniform* assumption (c.f. convexity).
- It is thus well-adapted to non-asymptotic analysis.
- It is also uncommonly well-suited to ‘non-equilibrium’ Markov processes.
- It is quite ‘concrete’ (via couplings of dynamics).
- It is quite robust to details (reversibility, absolute continuity, perturbations, etc.).
- If curvature isn’t enough to make a problem easy, then it may be too hard.

Will Curvature Hold?

1. For a given kernel P and metric d, curvature need not hold.
2. This can even be the case for nice kernels which mix very well.
3. If P is nice, then there is often *some* nice metric d for which curvature holds.
4. κ looks at the worst-case curvature; can miss things (e.g. 'near-curvature').
5. We get an 'honest' picture if the worst-case is close to the average-case.

Composition Laws (1)

- Exercise: Suppose that $\{P_i : i \in [N]\}$ is a collection of contractive kernels with respect to the same metric d, with P_i having curvature $\kappa_i \in (0, 1)$. Consider the mixture kernel

$$P^c(x, \cdot) = \sum_{i \in [N]} c_i \cdot P_i(x, \cdot)$$

for some probability vector $c \in \Delta^{N-1}$. Is P^c contractive? If so, what is its curvature?

Composition Laws (2)

- Exercise: Suppose that $\{P_i : i \in [N]\}$ is a collection of contractive kernels with respect to the same metric d, with P_i having curvature $\kappa_i \in (0, 1)$. Consider the composite kernel

$$P^{\text{composite}} = P_N P_{N-1} \cdots P_2 P_1.$$

Is $P^{\text{composite}}$ contractive? If so, what is its curvature?

Composition Laws (3)

- Exercise: Let $\{P_i : i \in [N]\}$ be a collection of contractive kernels each acting on different metric spaces, i.e. for $i \in [N]$, P_i is a Markov kernel on (E_i, d_i) with curvature $\kappa_i \in (0, 1)$. Define

$$E = E_1 \times E_2 \times \cdots \times E_N$$

$$\mathsf{d}(x, y) = \sum_{i \in [N]} \mathsf{d}_i(x_i, y_i),$$

and define a Markov kernel $P^{\text{tensorised}}$ on (E, d) by

$$P^{\text{tensorised}}(x, \mathrm{d}y) = \sum_{i \in [N]} c_i \cdot \bigotimes_{j \in [N] \setminus \{i\}} \delta(x_j, \mathrm{d}y_j) \cdot P_i(x_i, \mathrm{d}y_i),$$

i.e. with probability c_i, update only the ith coordinate of x, using the kernel P_i. Is $P^{\text{tensorised}}$ contractive? If so, what is its curvature?

Operator Norms (1)

- Let $f \in \mathrm{Lip}(E, \mathsf{d})$. Then $\|Pf\|_{\mathrm{Lip}(\mathsf{d})} \leqslant (1-\kappa) \cdot \|f\|_{\mathrm{Lip}(\mathsf{d})}$.
- For $\alpha, \beta \in \mathcal{P}(E)$, it holds that $\mathcal{T}_{1,\mathsf{d}}(\alpha P, \beta P) \leqslant (1-\kappa) \cdot \mathcal{T}_{1,\mathsf{d}}(\alpha, \beta)$.
- It thus follows that P admits a unique invariant measure π, to which one has convergence in $\mathcal{T}_{1,\mathsf{d}}$ at rate $(1-\kappa)^t$.

Operator Norms (2)

- Assume that P is reversible with respect to its invariant measure π. Then as an operator on $\mathrm{L}_0^2(\pi)$, it holds that

$$\|P\|_{\mathrm{op}} \leqslant 1 - \kappa.$$

  In particular, all eigenvalues of P (except for 1) lie in the interval $[-(1-\kappa), 1-\kappa]$, and the spectral gap of P is at least κ.

Equilibrium Properties

- Under a 'locality' condition on P, it holds that

$$\|f\|_{\mathrm{L}^2(\pi)}^2 \lesssim \kappa^{-1} \cdot \|f\|_{\mathrm{Lip}(\mathrm{d})}^2 .$$

  - Can control the variance of Lipschitz functions.

- Under an additional 'bounded moves' condition on P, for $\lambda \cdot \|f\|_{\mathrm{Lip}(\mathrm{d})} \lesssim \min\left\{\sigma_\infty^{-1}, \kappa\right\}$, it holds that

$$\log \pi\left(\exp\left(\lambda \cdot \left(f - \pi\left(f\right)\right)\right)\right) \lesssim \frac{1}{2}\lambda^2 \cdot \kappa^{-1} \cdot \|f\|_{\mathrm{Lip}(\mathrm{d})}^2$$

  - Can control the fluctuations of Lipschitz functions.

Further Results

- Can control non-asymptotic $\{$ bias, variance, fluctuations, ... $\}$ of ergodic averages of Markov chains.
- Can control regularity of solutions to the Poisson equation.
- Can extend to continuous-time processes.
- Can establish (generalised) gradient. contractivity