Adaptive MCMC

Mikhail Savelov

December 6, 2024

Contents

1 Introduction **2**

2 Preliminaries **3**
2.1 Martingales 3
2.2 Markov processes and transition kernels 3
2.3 Wasserstein ergodicity 3

3 Adaptive MCMC **5**
3.1 Preliminaries 5
3.2 Martingale decomposition 5
3.3 The Poisson equation 6
3.4 Convergence Analysis 8

4 Numerical Experiments **13**
4.1 Benchmarks 13
4.2 Adaptive Metropolis 13
4.3 AIR 13

Chapter 1

Introduction

Chapter 2

Preliminaries

2.1 Martingales

See Kallenberg 2021

2.2 Markov processes and transition kernels

Notation

Given a measurable space $(\mathcal{X}, \mathcal{F}_{\mathcal{X}})$, let $\mathbb{M}_1(\mathcal{X})$ denote the set of all probability measures on it. Let P be a transition kernel on $(\mathcal{X}, \mathcal{F}_{\mathcal{X}})$ which defines a linear operator $P \colon \mathbb{M}_1(\mathcal{X}) \to \mathbb{M}_1(\mathcal{X})$ given by

$$\mu P(B) = \int_{\mathcal{X}} P(x, B)\mu(dx), \quad \mu \in \mathbb{M}_1(\mathcal{X}), B \in \mathcal{F}_{\mathcal{X}}.$$

Further, for a measurable function $f \colon \mathcal{X} \to \mathbb{R}$ and $\mu \in \mathbb{M}_1(\mathcal{X})$ we denote

$$\mu(f) = \int_{\mathcal{X}} f(x)\mu(dx),$$
$$Pf(x) = \int_{\mathcal{X}} f(y)P(x, dy)$$

whenever well-defined.

2.3 Wasserstein ergodicity

See Douc et al. 2018 and Rudolf and Schweizer 2017

Let d be a metric which is assumed to be lower semi-continuous with respect to the product topology of $\mathcal{X}$. For two probability measures $\nu, \mu \in \mathbb{M}_1(\mathcal{X})$ we define the Wasserstein distance by

$$\mathcal{W}(\nu, \mu) := \inf_{\xi \in \mathcal{C}(\nu,\mu)} \int_{\mathcal{X}^2} d(x, y)\xi(dx, dy),$$

where $\mathcal{C}(\mu, \nu)$ is the set of all couplings of ν, μ, that is, all probability measures on product σ-algebra $\mathcal{F}_{\mathcal{X}} \otimes \mathcal{F}_{\mathcal{X}}$ with marginals ν and μ.

For a measurable function $f\colon \mathcal{X} \to \mathbb{R}$ we denote Lipschitz semi-norm as

$$\|f\|_d = \sup_{\substack{x,y\in\mathcal{X}\\ x\neq y}} \frac{|f(x)-f(y)|}{d(x,y)}$$

Theorem 1 (Kantorovich-Rubenstein duality)**.**

$$\mathcal{W}(\nu, \mu) = \sup_{\|f\|_d \le 1} |\nu(f) - \mu(f)| \,.$$

Further, we define the following quantities:

1) The eccentricity (see Joulin and Ollivier 2010) is defined as

$$E(x) = \int_{\mathcal{X}} d(x,y)\pi(dy).$$

2) Coarse diffusion coefficient (see Joulin and Ollivier 2010)

$$\mathrm{diff}(x, P) = \int_{\mathcal{X}} \int_{\mathcal{X}} d(x', x'')^2 P(x, dx') P(x, dx'').$$

3) The Wasserstein contraction coefficient for transition kernel P is

$$\tau(P) = \sup_{\substack{x,y\in\mathcal{X}\\ x\neq y}} \frac{\mathcal{W}(\delta_x P, \delta_y P)}{d(x,y)}.$$

Proposition 2. *For the Wasserstein contraction coefficient one has the properties of*

i) submultiplicativity, that is $\tau(P\tilde{P}) \le \tau(P)\tau(\tilde{P})$*, and,*

ii) contractivity, that is $\mathcal{W}(\nu P, \mu P) \le \tau(P)\mathcal{W}(\nu, \mu)$*,*

for any transition kernels P*,* $\tilde{P}$ *and any probability measures* $\nu, \mu \in \mathbb{M}_1(\mathcal{X})$*.*

Chapter 3

Adaptive MCMC

3.1 Preliminaries

Taken from Hofstadler et al. 2024 and Laitinen and Vihola 2024

We consider a state space $\mathcal{X}$ and a parameter space $\mathcal{I}$, which are assumed to be Polish with countable generated σ-algebras $\mathcal{F}_\mathcal{X}$ and $\mathcal{F}_\mathcal{I}$. Given a family of transition kernels $\{P_\gamma\}_{\gamma\in\mathcal{I}}$ on $(\mathcal{X}, \mathcal{F}_\mathcal{X})$, let $(X_n, \Gamma_n)_{n\in\mathbb{N}_0}$ be an adaptive MCMC chain evolving in the space $\mathcal{X} \times \mathcal{I}$, where $(\Gamma_n)_{n\in\mathbb{N}_0}$ denotes the sequence of (possibly random) adaptation parameters. The family $\{P_\gamma\}_{\gamma\in\mathcal{I}}$ must satisfy the following non-restrictive regularity condition:

Assumption (A1) (Regularity)**.** *$(\gamma, x) \mapsto P_\gamma(x, A)$ is $\mathcal{F}_\mathcal{I} \otimes \mathcal{F}_\mathcal{X}$-measurable for all $A \in \mathcal{F}_\mathcal{X}$.*

This ensures that $\big((\gamma, x), A\big) \mapsto P_\gamma(x, A)$ defines a Markov transition kernel from $(\mathcal{I} \times \mathcal{X}, \mathcal{F}_\mathcal{I} \otimes \mathcal{F}_\mathcal{X})$ to $(\mathcal{X}, \mathcal{F}_\mathcal{X})$.

We assume that π is the invariant distribution of $P_\gamma(x, \cdot)$ for any $x \in \mathcal{X}$ and $\gamma \in \mathcal{I}$. Moreover, we assume that any kernel P_γ is ψ-irreducible and aperiodic.

For a fixed test function $\varphi \in L^1(\pi)$, we consider an estimator

$$\widehat{\pi}_n(\varphi) := \frac{1}{n} \sum_{k=1}^{n} \varphi(X_k) \tag{3.1}$$

and investigate the convergence properties

$$\widehat{\pi}_n(\varphi) \xrightarrow{n\to\infty} \pi(\varphi), \tag{3.2}$$

3.2 Martingale decomposition

Following Laitinen and Vihola 2024, we make the following two general assumptions:

Assumption (A2) (Iterative structure)**.** *There is a filtration $(\mathcal{F}_k)_{k\geq 0}$ such that $(X_k, \Gamma_k)_{k\in\mathbb{N}_0}$ is $(\mathcal{F}_k)_{k\in\mathbb{N}_0}$-adapted, and the following holds for all $k \geq 0$ and all measurable $B \in \mathcal{F}_{\mathcal{X}}$:*

$$\mathbb{P}\left[X_{k+1} \in B \mid \mathcal{F}_k\right] = P_{\Gamma_k}(X_k, B) \qquad a.s.$$

Natural filtration

Assumption (A3) (Solutions of the Poisson equation)**.** *There exists a measurable mapping $(x, \gamma) \mapsto u_\gamma(x)$ from $\mathcal{X} \times \mathcal{I}$ to $\mathbb{R}$ which satisfies the following:*

$$u_\gamma(x) - (P_\gamma u_\gamma)(x) = \varphi(x) - \pi(\varphi) \qquad \text{for all } y \in \mathcal{X} \text{ and } \gamma \in \mathcal{I}. \tag{3.3}$$

Our desired convergence (3.2) can be then analysed using the following decomposition:

$$\begin{aligned}
\sum_{k=1}^{n} \varphi(X_k) - \pi(\varphi) &= \sum_{k=1}^{n} u_{\Gamma_k}(X_k) - P_{\Gamma_k} u_{\Gamma_k}(X_k) && (3.4)\\
&= \sum_{k=1}^{n} u_{\Gamma_{k-1}}(X_k) - P_{\Gamma_{k-1}} u_{\Gamma_{k-1}}(X_{k-1}) \qquad \big(=: M_n\big)\\
&+ \sum_{k=1}^{n} u_{\Gamma_k}(X_k) - u_{\Gamma_{k-1}}(X_k) \qquad \big(=: A_n\big)\\
&+ \sum_{k=1}^{n} P_{\Gamma_{k-1}} u_{\Gamma_{k-1}}(X_{k-1}) - P_{\Gamma_k} u_{\Gamma_k}(X_k) \qquad \big(=: R_n\big),
\end{aligned}$$

3.3 The Poisson equation

Look into Douc et al. 2018 for Poisson equation.

As shown in Hofstadler et al. 2024, (A3) holds under following assumption on Wasserstein contraction:

Assumption (A4). *For each $\gamma \in \mathcal{I}$ we assume that π is the invariant distribution of P_γ and that for all $k \in \mathbb{N}$*

$$\tau(P_\gamma^k) \leq C\tau^k,$$

with $C \in [1, \infty)$ and $\tau \in [0, 1)$ independent of γ.

Proposition 3. *Let $\gamma \in \mathcal{I}$, $\varphi\colon \mathcal{X} \to \mathbb{R}$ be π integrable and Lipschitz with constant $L \in (0, \infty)$ and let $E(x) < \infty$ for any $x \in \mathcal{X}$. If (A4) holds, then the function $u_\gamma\colon \mathcal{X} \to \mathbb{R}$ given via*

$$u_\gamma(x) = \sum_{k=0}^{\infty} P_\gamma^k \left(\varphi - \pi(\varphi)\right)(x) \tag{3.5}$$

is well defined and we have

i) $|u_\gamma(x)| \le \frac{C\|\varphi\|_d}{1-\tau} E(x)$ *for any* $x \in \mathcal{X}$,

ii) $\|u_\gamma\|_d \le \frac{C\|\varphi\|_d}{1-\tau}$,

iii) $u_\gamma(x) - P_\gamma u_\gamma(x) = \varphi(x) - \pi(\varphi)$ *for any* $x \in \mathcal{X}$.

We now make an assumption on smoothness w.r.t. parameter γ. For this purpose, we denote a metric on $\mathcal{I}$ as $\tilde{d}$. We aim to have Lipschitz-smoothness of mappings $\gamma \mapsto u_\gamma(x)$ for any $x \in \mathcal{X}$. The following would suffice to have desired smoothness:

$$\sum_{n=1}^{\infty} \mathcal{W}(\delta_x P_\gamma^n, \delta_x P_{\gamma'}^n) \le \tilde{d}(\gamma, \gamma') \cdot g(x)$$

for all $\gamma, \gamma' \in \mathcal{I}$, $x \in \mathcal{X}$ and some measurable $g\colon \mathcal{X} \to \mathbb{R}$.

Assumption (A5). *There exists* $\tilde{L} < \infty$*, s.t. for all* $\gamma, \gamma' \in \mathcal{I}$ *holds*

$$\sup_{\|f\|_d \le 1} \|(P_\gamma - P_{\gamma'})f\|_d \le \tilde{L} \cdot \tilde{d}(\gamma, \gamma').$$

Lemma 4. *Under (A4) and (A5) for any* $x \in \mathcal{X}$*,* $\gamma, \gamma' \in \mathcal{I}$ *and* $n \in \mathbb{N}$ *it holds*

$$\mathcal{W}(\delta_x P_\gamma^n, \delta_x P_{\gamma'}^n) \le n\tau^{n-1} \cdot C^2 \tilde{L} \tilde{d}(\gamma, \gamma') E(x).$$

Consequently, for u_γ *defined in* (3.5) *we have for any* $x \in \mathcal{X}$*:*

$$\sup_{\substack{\gamma, \gamma' \in \mathcal{I} \\ \gamma \ne \gamma'}} \frac{|u_\gamma(x) - u_{\gamma'}(x)|}{\tilde{d}(\gamma, \gamma')} \le \frac{\tilde{L} C^2 \|\varphi\|_d}{(1-\tau)^2} E(x).$$

Proof. We observe that for any $n \in \mathbb{N}_0$ and $x \in \mathcal{X}$ it holds

$$\begin{aligned} P_\gamma^n \varphi(x) - P_{\gamma'}^n \varphi(x) &= \sum_{i=0}^{n-1} \left(P_\gamma^{i+1} P_{\gamma'}^{n-(i+1)} \varphi(x) - P_\gamma^i P_{\gamma'}^{n-i} \varphi(x) \right) \\ &= \sum_{i=0}^{n-1} P_\gamma^i (P_\gamma - P_{\gamma'}) P_{\gamma'}^{n-i-1} \varphi(x). \end{aligned}$$

Denote $\psi(x) := (P_\gamma - P_{\gamma'}) P_{\gamma'}^{n-i-1} \varphi(x)$. Observing that $\pi(\psi) = \pi P_\gamma P_{\gamma'}^{n-i-1}(\varphi) - \pi P_{\gamma'}^{n-i}(\varphi) = 0$ yields

$$\begin{aligned} \left| P_\gamma^i \psi(x) \right| &= \left| \delta_x P_\gamma^i(\psi) - \pi(\psi) \right| \\ &\le \|\psi\|_d \cdot \mathcal{W}(\delta_x P_\gamma^i, \pi) \\ &\le \|\psi\|_d \cdot C\tau^i \mathcal{W}(\delta_x, \pi) \\ &= \|\psi\|_d \cdot C\tau^i E(x). \end{aligned}$$

Assumption (A5) allows to write

$$\begin{aligned}
\|\psi\|_d &= \|(P_\gamma - P_{\gamma'})P_{\gamma'}^{n-i-1}\varphi\|_d \\
&\leq \|P_{\gamma'}^{n-i-1}\varphi\|_d \cdot \tilde{L}\tilde{d}(\gamma,\gamma') \\
&\leq \|\varphi\|_d \cdot C\tau^{n-i-1} \cdot \tilde{L}\tilde{d}(\gamma,\gamma').
\end{aligned}$$

Hence, we can bound

$$\left|P_\gamma^n\varphi(x) - P_{\gamma'}^n\varphi(x)\right| \leq n\tau^{n-1} \cdot \|\varphi\|_d C^2 \tilde{L}\tilde{d}(\gamma,\gamma')E(x).$$

It directly follows that

$$\begin{aligned}
\mathcal{W}(\delta_x P_\gamma^n, \delta_x P_{\gamma'}^n) &= \sup_{\|f\|_d \leq 1} \left|P_\gamma^n f(x) - P_{\gamma'}^n f(x)\right| \\
&\leq n\tau^{n-1} \cdot C^2 \tilde{L}\tilde{d}(\gamma,\gamma')E(x).
\end{aligned}$$

Finally,

$$\begin{aligned}
|u_\gamma(x) - u_{\gamma'}(x)| &\leq \sum_{n=1}^{\infty} \left|P_\gamma^n\varphi(x) - P_{\gamma'}^n\varphi(x)\right| \\
&\leq \left(\sum_{n=1}^{\infty} n\tau^{n-1}\right) \cdot \|\varphi\|_d C^2 E(x) \cdot \tilde{L}\tilde{d}(\gamma,\gamma') \\
&= \tilde{d}(\gamma,\gamma') \cdot \frac{\tilde{L}C^2\|\varphi\|_d}{(1-\tau)^2} E(x).
\end{aligned}$$

□

3.4 Convergence Analysis

Here we analyze the terms M_n, A_n, R_n as defined in (3.4). We initially assume that (A2), (A4) hold, and u_γ is given by (3.5).

Assumption (A6).

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\mathrm{diff}(X_n, P_{\Gamma_n})\right] = \Lambda < \infty.$$

Lemma 5. *Under (A6), $M_n = \sum_{k=1}^n u_{\Gamma_{k-1}}(X_k) - P_{\Gamma_{k-1}}u_{\Gamma_{k-1}}(X_{k-1})$ defined in (3.4) is a martingale w.r.t. $(\mathcal{F}_n)_{n\in\mathbb{N}}$.*

Proof. Denote $\Delta_k := u_{\Gamma_{k-1}}(X_k) - P_{\Gamma_{k-1}} u_{\Gamma_{k-1}}(X_{k-1})$ for $k \in \mathbb{N}$.

$$\begin{aligned}
\mathbb{E}\left[|\Delta_k|\right] &= \mathbb{E}\left[|u_{\Gamma_{k-1}}(X_k) - P_{\Gamma_{k-1}} u_{\Gamma_{k-1}}(X_{k-1})|\right] \\
&\leq \mathbb{E}\left[\int_{\mathcal{X}} |u_{\Gamma_{k-1}}(X_k) - u_{\Gamma_{k-1}}(x)| P_{\Gamma_{k-1}}(X_{k-1}, dx)\right] \\
&\leq \frac{C\|\varphi\|_d}{1-\tau} \mathbb{E}\left[\int_{\mathcal{X}} d(X_k, x) P_{\Gamma_{k-1}}(X_{k-1}, dx)\right] \\
&= \frac{C\|\varphi\|_d}{1-\tau} \mathbb{E}\left[\int_{\mathcal{X}} \int_{\mathcal{X}} d(x', x) P_{\Gamma_{k-1}}(X_{k-1}, dx') P_{\Gamma_{k-1}}(X_{k-1}, dx)\right] \\
&\leq \frac{C\|\varphi\|_d}{1-\tau} \sqrt{\mathbb{E}\left[\int_{\mathcal{X}} \int_{\mathcal{X}} d(x', x)^2 P_{\Gamma_{k-1}}(X_{k-1}, dx') P_{\Gamma_{k-1}}(X_{k-1}, dx)\right]} \\
&\leq \frac{C\|\varphi\|_d}{1-\tau} \sqrt{\Lambda} < \infty,
\end{aligned}$$

where Jensen's inequality is used in the end

Disintegration Thm (See Thm. 8.5 in Kallenberg 2021) yields

$$\begin{aligned}
\mathbb{E}\left[\Delta_k \mid \mathcal{F}_{k-1}\right] &\overset{\text{a.s}}{=} \mathbb{E}\left[u_{\Gamma_{k-1}}(X_k) \mid \mathcal{F}_{k-1}\right] - P_{\Gamma_{k-1}} u_{\Gamma_{k-1}}(X_{k-1}) \\
&\overset{\text{a.s}}{=} P_{\Gamma_{k-1}} u_{\Gamma_{k-1}}(X_{k-1}) - P_{\Gamma_{k-1}} u_{\Gamma_{k-1}}(X_{k-1}) \\
&= 0.
\end{aligned}$$

□

Lemma 6. *Under (A6) for M_n defined in* (3.4) *it holds*

$$\frac{M_n}{n} \xrightarrow{a.s.} 0.$$

Proof. Consider the martingale $V_n = \sum_{k=1}^{n} \frac{\Delta_k}{k}$. The differences Δ_k are uniformly bounded in L^2:

$$\begin{aligned}
\mathbb{E}\left[\Delta_k^2\right] &\leq \frac{C^2\|\varphi\|_d^2}{(1-\tau)^2} \mathbb{E}\left[\int_{\mathcal{X}} \int_{\mathcal{X}} d(x', x)^2 P_{\Gamma_{k-1}}(X_{k-1}, dx') P_{\Gamma_{k-1}}(X_{k-1}, dx)\right] \\
&\leq \frac{C^2\|\varphi\|_d^2}{(1-\tau)^2} \Lambda.
\end{aligned}$$

Because martingale differences are orthogonal, we have

$$\begin{aligned}
\mathbb{E}\left[V_n^2\right] &= \sum_{k=1}^{n} \frac{\mathbb{E}\left[\Delta_k^2\right]}{k^2} \\
&\leq \frac{C^2\|\varphi\|_d^2}{(1-\tau)^2} \Lambda \cdot \sum_{k=1}^{\infty} \frac{1}{k^2} \\
&= \frac{C^2\|\varphi\|_d^2}{(1-\tau)^2} \Lambda \cdot \frac{\pi^2}{6}.
\end{aligned}$$

That is, V_n is a L^2-bounded martingale, which converges to an a.s. finite $V_\infty = \sum_{k=1}^{\infty} \frac{\Delta_k}{k}$ almost surely (see Corollary E.3.5 in Douc et al. 2018). Whenever V_∞ is finite, Kronecker's lemma implies that

$$\frac{M_n}{n} = \frac{1}{n}\sum_{k=1}^{n} k\frac{\Delta_k}{k} \xrightarrow{\text{a.s.}} 0.$$

□

Assumption (A7).

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[E(X_n)^2\right] = K < \infty.$$

Lemma 7. *Under (A7) for* R_n *defined in* (3.4) *it holds*

$$\frac{R_n}{n} \xrightarrow{a.s.} 0, \qquad \frac{R_n}{\sqrt{n}} \xrightarrow{P} 0.$$

Proof. For $p \in \{1, \frac{1}{2}\}$ we have

$$\begin{aligned}\left|\frac{R_n}{n^p}\right| &= \frac{1}{n^p}\left|P_{\Gamma_n}u_{\Gamma_n}(X_n) - P_{\Gamma_0}u_{\Gamma_0}(X_0)\right| \\ &\le \frac{1}{n^p}\left|P_{\Gamma_n}u_{\Gamma_n}(X_n)\right| + \frac{1}{n^p}\left|P_{\Gamma_0}u_{\Gamma_0}(X_0)\right|\end{aligned}$$

Under (A1), $P_{\Gamma_0}u_{\Gamma_0}(X_0)$ is finite for each $\omega \in \Omega$ and hence the second term converges a.s. and, hence, in probability.

For the first term we bound

$$\begin{aligned}\mathbb{P}\left[\frac{1}{n^p}\left|P_{\Gamma_n}u_{\Gamma_n}(X_n)\right| > \varepsilon\right] &\le \frac{1}{n^{2p}\varepsilon^2}\mathbb{E}\left[P_{\Gamma_n}u_{\Gamma_n}(X_n)^2\right] \\ &= \frac{1}{n^{2p}\varepsilon^2}\mathbb{E}\left[\mathbb{E}\left[u_{\Gamma_n}(X_{n+1}) \mid \mathcal{F}_n\right]^2\right] \\ &\le \frac{1}{n^{2p}\varepsilon^2}\mathbb{E}\left[\mathbb{E}\left[u_{\Gamma_n}(X_{n+1})^2 \mid \mathcal{F}_n\right]\right] \\ &\le \left(\frac{C\|\varphi\|_d}{1-\tau}\right)^2 \cdot \frac{1}{n^{2p}\varepsilon^2}\mathbb{E}\left[\ E(X_{n+1})^2\right] \\ &\le \left(\frac{C\|\varphi\|_d}{1-\tau}\right)^2 \cdot \frac{K}{n^{2p}\varepsilon^2}.\end{aligned}$$

This shows $\sum_{n=1}^{\infty} \mathbb{P}\left[\frac{1}{n}\left|P_{\Gamma_n}u_{\Gamma_n}(X_n)\right| > \varepsilon\right] < \infty$. By Borel-Cantelli lemma it follows that

$$\frac{1}{n}\left|P_{\Gamma_n}u_{\Gamma_n}(X_n)\right| \xrightarrow{\text{a.s.}} 0.$$

We also have $\mathbb{P}\left[\frac{1}{\sqrt{n}}\left|P_{\Gamma_n}u_{\Gamma_n}(X_n)\right| > \varepsilon\right] \xrightarrow{n\to\infty} 0$, yielding

$$\frac{1}{\sqrt{n}}\left|P_{\Gamma_n}u_{\Gamma_n}(X_n)\right| \xrightarrow{P} 0.$$

□

For $k \in \mathbb{N}$, denote $D_k := \|\tilde{d}(\Gamma_{k-1}, \Gamma_k)\|_2 = \sqrt{\mathbb{E}\left[\tilde{d}(\Gamma_{k-1}, \Gamma_k)^2\right]}$.

Assumption (A8). *Suppose the following series converges:*

$$\sum_{k=1}^{\infty} \frac{1}{k} D_k =: D < \infty.$$

Assumption (A9). *Suppose for some $p > 0$ it holds*

$$\frac{1}{n^p} \sum_{k=1}^{n} D_k \xrightarrow{n\to\infty} 0.$$

Lemma 8. *Under (A5), (A7) and (A8) it holds*

$$\frac{A_n}{n} \xrightarrow{a.s.} 0.$$

Proof. (A5) yields

$$\begin{aligned}
\left|\frac{A_n}{n}\right| &\le \frac{1}{n} \sum_{k=1}^{n} \left|u_{\Gamma_k}(X_k) - u_{\Gamma_{k-1}}(X_k)\right| \\
&\le \frac{\tilde{L}C^2 \|\varphi\|_d}{(1-\tau)^2} \cdot \frac{1}{n} \sum_{k=1}^{n} \tilde{d}(\Gamma_{k-1}, \Gamma_k) E(X_k).
\end{aligned}$$

Consider $Y_n := \sum_{k=1}^{n} \frac{1}{k} \tilde{d}(\Gamma_{k-1}, \Gamma_k) E(X_k)$. Since all terms are nonnegative, for each $\omega \in \Omega$ there exists a pointwise limit $Y_\infty(\omega) := \lim_{n\to\infty} Y_n(\omega) \in [0, \infty]$.

We apply Cauchy–Schwarz and use (A7) and (A8):

$$\begin{aligned}
\mathbb{E}\left[Y_n\right] &\le \sum_{k=1}^{n} \frac{1}{k} \sqrt{\mathbb{E}\left[\tilde{d}(\Gamma_{k-1}, \Gamma_k)^2\right] \mathbb{E}\left[E(X_k)^2\right]} \\
&\le \sqrt{K} \sum_{k=1}^{n} \frac{1}{k} \sqrt{\mathbb{E}\left[\tilde{d}(\Gamma_{k-1}, \Gamma_k)^2\right]} \\
&= \sqrt{K} \sum_{k=1}^{n} \frac{1}{k} D_k \\
&\le \sqrt{K} \cdot D < \infty.
\end{aligned}$$

By monotone convergence, Y_∞ satisfies $\mathbb{E}\left[Y_\infty\right] \le \sqrt{K} \cdot D$ and hence it is a.s. finite. Finally, Kronecker's lemma implies that

$$\frac{1}{n} \sum_{k=1}^{n} \tilde{d}(\Gamma_{k-1}, \Gamma_k) E(X_k) = \frac{1}{n} \sum_{k=1}^{n} k \cdot \frac{1}{k} \tilde{d}(\Gamma_{k-1}, \Gamma_k) E(X_k) \xrightarrow{\text{a.s.}} 0.$$

□

Lemma 9. *Under (A5), (A7) and (A9) (for some $p > 0$) it holds:*

$$\frac{A_n}{n^p} \xrightarrow{L_1} 0.$$

As consequence, convergence in probability also holds.

Proof.

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{A_n}{n^p}\right|\right] &\leq \frac{\tilde{L}C^2\|\varphi\|_d}{(1-\tau)^2} \cdot \frac{1}{n^p} \sum_{k=1}^{n} \mathbb{E}\left[\tilde{d}(\Gamma_{k-1}, \Gamma_k) E(X_k)\right] \\
&\leq \frac{\tilde{L}C^2\|\varphi\|_d}{(1-\tau)^2} \cdot \frac{1}{n^p} \sum_{k=1}^{n} \sqrt{\mathbb{E}\left[\tilde{d}(\Gamma_{k-1}, \Gamma_k)^2\right] \mathbb{E}\left[E(X_k)^2\right]} \\
&\leq \frac{\tilde{L}C^2\|\varphi\|_d}{(1-\tau)^2} \sqrt{K} \cdot \frac{1}{n^p} \sum_{k=1}^{n} D_k \xrightarrow{n\to\infty} 0.
\end{aligned}$$

□

Chapter 4

Numerical Experiments

4.1 Benchmarks

See Magnusson et al. 2024

4.2 Adaptive Metropolis

4.3 AIR

Bibliography

Joulin, Aldéric and Yann Ollivier (2010). "Curvature, concentration and error estimates for Markov chain Monte Carlo." In: *The Annals of Probability* 38.6. ISSN: 0091-1798. DOI: 10.1214/10-aop541. URL: http://dx.doi.org/10.1214/10-AOP541.

Rudolf, Daniel and Nikolaus Schweizer (2017). *Perturbation theory for Markov chains via Wasserstein distance*. arXiv: 1503.04123 [stat.CO]. URL: https://arxiv.org/abs/1503.04123.

Douc, Randal et al. (2018). *Markov chains*. Operations Research and Financial Engineering. Cham, Switzerland: Springer.

Kallenberg, Olav (2021). *Foundations of modern probability*. 3rd ed. Probability Theory and Stochastic Modelling. Cham, Switzerland: Springer.

Hofstadler, Julian et al. (2024). *Almost sure convergence rates of adaptive increasingly rare Markov chain Monte Carlo*. arXiv: 2402.12122 [math.NA]. URL: https://arxiv.org/abs/2402.12122.

Laitinen, Pietari and Matti Vihola (2024). *An invitation to adaptive Markov chain Monte Carlo convergence theory*. arXiv: 2408.14903 [math.PR]. URL: https://arxiv.org/abs/2408.14903.

Magnusson, Måns et al. (2024). *posteriordb: Testing, Benchmarking and Developing Bayesian Inference Algorithms*. arXiv: 2407.04967 [stat.CO]. URL: https://arxiv.org/abs/2407.04967.