Discussion of Eric Moulines talk ”Some results on adaptive MCMC Algorithms”

Krzysztof Latuszynski

University of Warwick, UK

January 3, 2011

the Results

- I would like to congratulate the authors for this very impressive and subtle paper!
- it develops theory for establishing ergodicity and SLLN in very general scenarios
- the theory is extremely powerful and allows to address the two most important and challenging Adaptive/Interacting algorithms in the field.

the Results

- I would like to congratulate the authors for this very impressive and subtle paper!
- it develops theory for establishing ergodicity and SLLN in very general scenarios
- the theory is extremely powerful and allows to address the two most important and challenging Adaptive/Interacting algorithms in the field.

the Results

- I would like to congratulate the authors for this very impressive and subtle paper!
- it develops theory for establishing ergodicity and SLLN in very general scenarios
- the theory is extremely powerful and allows to address the two most important and challenging Adaptive/Interacting algorithms in the field.

the Results

- In case of the original Adaptive Metropolis algorithm of Haario et al.,
- the theory implies convergence of marginal distributions to π.
- These conclusions have been verified for $\mathcal{X} = \mathbb{R}^d$
  in the geometrically ergodic setting,
  for super-exponentially decaying targets,
  reproving the recent result of Saksmann and Vihola (2010).

- The analysis of the Interacting Tempering Algorithm of Kou at al is even more impressive,
- Interacting Tempering has been analyzed for the first time in this generality,
- In particular, in the geometrically ergodic setting
- the SLLN for unbounded functionals has been established.

the Results

- In case of the original Adaptive Metropolis algorithm of Haario et al.,
- the theory implies convergence of marginal distributions to π.
- These conclusions have been verified for $\mathcal{X} = \mathbb{R}^d$ in the geometrically ergodic setting, for super-exponentially decaying targets, reproving the recent result of Saksmann and Vihola (2010).
- The analysis of the Interacting Tempering Algorithm of Kou at al is even more impressive,
- Interacting Tempering has been analyzed for the first time in this generality,
- In particular, in the geometrically ergodic setting
- the SLLN for unbounded functionals has been established.

the Results

- In case of the original Adaptive Metropolis algorithm of Haario et al.,
- the theory implies convergence of marginal distributions to π.
- These conclusions have been verified for $\mathcal{X} = \mathbb{R}^d$
  in the geometrically ergodic setting,
  for super-exponentially decaying targets,
  reproving the recent result of Saksmann and Vihola (2010).

- The analysis of the Interacting Tempering Algorithm of Kou at al is even more impressive,
- Interacting Tempering has been analyzed for the first time in this generality,
- In particular, in the geometrically ergodic setting
- the SLLN for unbounded functionals has been established.

the Results

- In case of the original Adaptive Metropolis algorithm of Haario et al.,
- the theory implies convergence of marginal distributions to π.
- These conclusions have been verified for $\mathcal{X} = \mathbb{R}^d$
  in the geometrically ergodic setting,
  for super-exponentially decaying targets,
  reproving the recent result of Saksmann and Vihola (2010).

- The analysis of the Interacting Tempering Algorithm of Kou at al is even more impressive,
- Interacting Tempering has been analyzed for the first time in this generality,
- In particular, in the geometrically ergodic setting
- the SLLN for unbounded functionals has been established.

the Results

- In case of the original Adaptive Metropolis algorithm of Haario et al.,
- the theory implies convergence of marginal distributions to π.
- These conclusions have been verified for $\mathcal{X} = \mathbb{R}^d$
  in the geometrically ergodic setting,
  for super-exponentially decaying targets,
  reproving the recent result of Saksmann and Vihola (2010).

- The analysis of the Interacting Tempering Algorithm of Kou at al is even more impressive,
- Interacting Tempering has been analyzed for the first time in this generality,
- In particular, in the geometrically ergodic setting
- the SLLN for unbounded functionals has been established.

the Results

- In case of the original Adaptive Metropolis algorithm of Haario et al.,
- the theory implies convergence of marginal distributions to π.
- These conclusions have been verified for $\mathcal{X} = \mathbb{R}^d$
  in the geometrically ergodic setting,
  for super-exponentially decaying targets,
  reproving the recent result of Saksmann and Vihola (2010).

- The analysis of the Interacting Tempering Algorithm of Kou at al is even more impressive,
- Interacting Tempering has been analyzed for the first time in this generality,
- In particular, in the geometrically ergodic setting
- the SLLN for unbounded functionals has been established.

the Results

- In case of the original Adaptive Metropolis algorithm of Haario et al.,
- the theory implies convergence of marginal distributions to π.
- These conclusions have been verified for $\mathcal{X} = \mathbb{R}^d$
  in the geometrically ergodic setting,
  for super-exponentially decaying targets,
  reproving the recent result of Saksmann and Vihola (2010).

- The analysis of the Interacting Tempering Algorithm of Kou at al is even more impressive,
- Interacting Tempering has been analyzed for the first time in this generality,
- In particular, in the geometrically ergodic setting
- the SLLN for unbounded functionals has been established.

Technical Conditions for ergodicity

- The Theory is very delicate and is building on the following crucial conditions.
- A1: For any $\theta \in \Theta$, there exists π_θ, s.t. $\pi_\theta = P_\theta \pi_\theta$.
- A2(a): For any $\epsilon > 0$, there exists a non-decreasing sequence $r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)}, \cdot) - \pi_{\theta_{n-r_\epsilon(n)}}\|_{TV}\right] \le \epsilon.$$

- A2(b): For any $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$$

- the dependence on θ in π_θ above, is crucial for Interacting Tempering, however I will drop it for clarity in subsequent slides.

Technical Conditions for ergodicity

- The Theory is very delicate and is building on the following crucial conditions.
- A1: For any $\theta \in \Theta$, there exists π_θ, s.t. $\pi_\theta = P_\theta \pi_\theta$.
- A2(a): For any $\epsilon > 0$, there exists a non-decreasing sequence $r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)}, \cdot) - \pi_{\theta_{n-r_\epsilon(n)}}\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$$

- the dependence on θ in π_θ above, is crucial for Interacting Tempering, however I will drop it for clarity in subsequent slides.

Technical Conditions for ergodicity

- The Theory is very delicate and is building on the following crucial conditions.
- A1: For any $\theta \in \Theta$, there exists π_θ, s.t. $\pi_\theta = P_\theta \pi_\theta$.
- A2(a): For any $\epsilon > 0$, there exists a non-decreasing sequence $r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)}, \cdot) - \pi_{\theta_{n-r_\epsilon(n)}}\|_{TV}\right] \le \epsilon.$$

- A2(b): For any $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$$

- the dependence on θ in π_θ above, is crucial for Interacting Tempering, however I will drop it for clarity in subsequent slides.

Technical Conditions for ergodicity

- The Theory is very delicate and is building on the following crucial conditions.
- A1: For any $\theta \in \Theta$, there exists π_θ, s.t. $\pi_\theta = P_\theta \pi_\theta$.
- A2(a): For any $\epsilon > 0$, there exists a non-decreasing sequence $r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)}, \cdot) - \pi_{\theta_{n-r_\epsilon(n)}}\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$$

- the dependence on θ in π_θ above, is crucial for Interacting Tempering, however I will drop it for clarity in subsequent slides.

Technical Conditions for ergodicity

- The Theory is very delicate and is building on the following crucial conditions.
- A1: For any $\theta \in \Theta$, there exists π_θ, s.t. $\pi_\theta = P_\theta \pi_\theta$.
- A2(a): For any $\epsilon > 0$, there exists a non-decreasing sequence $r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)}, \cdot) - \pi_{\theta_{n-r_\epsilon(n)}}\|_{TV}\right] \le \epsilon.$$

- A2(b): For any $\epsilon > 0$,

$$\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$$

- the dependence on θ in π_θ above, is crucial for Interacting Tempering, however I will drop it for clarity in subsequent slides.

Comparison to containment - I

- A2(a): For any $\epsilon > 0$, $\exists\, r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)},\cdot) - \pi\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0$, $\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$
- Containment C(a): define $M_\epsilon(x,\theta) := \inf_n\{\|P^n_\theta(x,\cdot) - \pi\|_{TV} \leq \epsilon\}$, and assume

$$\forall \delta > 0, \epsilon > 0, \ \exists\, M_{\epsilon,\delta} \quad \text{s.t.} \quad \forall n \ \ P(M_\epsilon(X_n,\theta_n) \leq M_{\epsilon,\delta}) \geq 1-\delta.$$

- Diminishing Adaptation C(b): $\lim_{n\to\infty} \mathbb{E}\left[D(\theta_{n-1},\theta_n)\right] = 0.$
- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon$, then
  A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.

Comparison to containment - I

- A2(a): For any $\epsilon > 0$, $\exists\ r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)},\cdot) - \pi\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0$, $\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$

- Containment C(a): define $M_\epsilon(x,\theta) := \inf_n\{\|P^n_\theta(x,\cdot) - \pi\|_{TV} \leq \epsilon\}$, and assume

$$\forall\delta > 0, \epsilon > 0, \ \exists\ M_{\epsilon,\delta} \quad \text{s.t.} \quad \forall n\ \ P(M_\epsilon(X_n,\theta_n) \leq M_{\epsilon,\delta}) \geq 1-\delta.$$

- Diminishing Adaptation C(b): $\lim_{n\to\infty} \mathbb{E}\left[D(\theta_{n-1},\theta_n)\right] = 0.$
- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon$, then A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.

Comparison to containment - I

- A2(a): For any $\epsilon > 0$, $\exists\ r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)},\cdot) - \pi\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0$, $\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$
- Containment C(a): define $M_\epsilon(x,\theta) := \inf_n\{\|P^n_\theta(x,\cdot) - \pi\|_{TV} \leq \epsilon\}$, and assume

$$\forall\delta > 0, \epsilon > 0, \ \exists\ M_{\epsilon,\delta} \quad \text{s.t.} \quad \forall n\ \ P(M_\epsilon(X_n,\theta_n) \leq M_{\epsilon,\delta}) \geq 1-\delta.$$

- Diminishing Adaptation C(b): $\lim_{n\to\infty} \mathbb{E}\left[D(\theta_{n-1},\theta_n)\right] = 0.$
- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon$, then
  A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.

Comparison to containment - I

- A2(a): For any $\epsilon > 0, \ \exists \ r_\epsilon(n),$ s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)}, \cdot) - \pi\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0,$ $\quad \lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$
- Containment C(a): define $M_\epsilon(x, \theta) := \inf_n\{\|P^n_\theta(x, \cdot) - \pi\|_{TV} \leq \epsilon\}$, and assume

$$\forall \delta > 0, \epsilon > 0, \ \exists \ M_{\epsilon,\delta} \quad \text{s.t.} \quad \forall n \ \ P(M_\epsilon(X_n, \theta_n) \leq M_{\epsilon,\delta}) \geq 1 - \delta.$$

- Diminishing Adaptation C(b): $\lim_{n\to\infty} \mathbb{E}\left[D(\theta_{n-1}, \theta_n)\right] = 0.$
- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}.$
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon,$ then A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}.$

Comparison to containment - I

- A2(a): For any $\epsilon > 0$, $\exists\ r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)},\cdot) - \pi\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0$, $\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$
- Containment C(a): define $M_\epsilon(x,\theta) := \inf_n\{\|P^n_\theta(x,\cdot) - \pi\|_{TV} \leq \epsilon\}$, and assume

$$\forall \delta > 0, \epsilon > 0, \ \exists\ M_{\epsilon,\delta} \quad \text{s.t.} \quad \forall n \ \ P(M_\epsilon(X_n,\theta_n) \leq M_{\epsilon,\delta}) \geq 1-\delta.$$

- Diminishing Adaptation C(b): $\lim_{n\to\infty} \mathbb{E}\left[D(\theta_{n-1},\theta_n)\right] = 0.$
- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon$, then A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.

Comparison to containment - I

- A2(a): For any $\epsilon > 0$, $\exists\, r_\epsilon(n)$, s.t. $\limsup_{n\to\infty} r_\epsilon(n)/n = 0$ and

$$\limsup_{n\to\infty} \mathbb{E}\left[\|P^{r_\epsilon(n)}_{\theta_{n-r_\epsilon(n)}}(X_{n-r_\epsilon(n)}, \cdot) - \pi\|_{TV}\right] \leq \epsilon.$$

- A2(b): For any $\epsilon > 0$, $\lim_{n\to\infty} \sum_{j=0}^{r_\epsilon(n)-1} \mathbb{E}\left[D(\theta_{n-r_\epsilon(n)+j}, \theta_{n-r_\epsilon(n)})\right] = 0.$
- Containment C(a): define $M_\epsilon(x,\theta) := \inf_n\{\|P^n_\theta(x,\cdot) - \pi\|_{TV} \leq \epsilon\}$, and assume

$$\forall \delta > 0, \epsilon > 0, \ \exists\, M_{\epsilon,\delta} \quad \text{s.t.} \quad \forall n \ \ P(M_\epsilon(X_n, \theta_n) \leq M_{\epsilon,\delta}) \geq 1 - \delta.$$

- Diminishing Adaptation C(b): $\lim_{n\to\infty} \mathbb{E}\left[D(\theta_{n-1}, \theta_n)\right] = 0.$
- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon$, then
  A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.

Comparison to containment - I

- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon$, then A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.
- Therefore A2(a), A2(b) generalize C(a), C(b) (rather then weaken) and the generalization is in settings where $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.
- We shall **try to investigate**, what happens if $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.

Comparison to containment - I

- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon$, then A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.
- Therefore A2(a), A2(b) generalize C(a), C(b) (rather then weaken) and the generalization is in settings where $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.
- We shall **try to investigate**, what happens if $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.

Comparison to containment - I

- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \text{const}(\epsilon) = r_\epsilon,$ then
  A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.
- Therefore A2(a), A2(b) generalize C(a), C(b) (rather then weaken) and the generalization is in settings where $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.
- We shall **try to investigate**, what happens if $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.

Comparison to containment - I

- C(a), C(b) $\Rightarrow$ A2(a), A2(b) by taking e.g. $r_\epsilon(n) = M_{\epsilon/2,\epsilon/2}$.
- if $r_\epsilon(n) = \mathrm{const}(\epsilon) = r_\epsilon$, then
  A2(a), A2(b) $\Rightarrow$ C(a), C(b) by taking e.g. $M_{\epsilon,\delta} := r_{\epsilon\delta}$.
- Therefore A2(a), A2(b) generalize C(a), C(b) (rather then weaken) and the generalization is in settings where $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.
- We shall **try to investigate**, what happens if $r_\epsilon(n)$ needs to grow to ∞ as $n \to \infty$.

What Adaptive MCMC is designed for?

- In a typical Adaptive MCMC setting the parameter space Θ is large
- there is an optimal $\theta_* \in \Theta$ s.t. P_{θ_*} converges quickly.
- there are arbitrary bad values in Θ, say if $\theta \in \bar{\Theta} - \Theta$ then P_θ is not ergodic.
- if $\theta \in \Theta_* :=$ a region close to θ_*, then P_θ shall inherit good convergence properties of P_{θ_*}.
- When using adaptive MCMC we hope θ_n will eventually find the region Θ_* and stay there essentially forever. And that the adaptive algorithm $\mathcal{A}$ will inherit the good convergence properties of Θ_* in the limit.
- We are looking for a Theorem:
  You can actually run your Adaptive MCMC algorithm $\mathcal{A}$, and it will do what it is supposed to do! (under verifiable conditions)

What Adaptive MCMC is designed for?

- In a typical Adaptive MCMC setting the parameter space Θ is large
- there is an optimal $\theta_* \in \Theta$ s.t. P_{θ_*} converges quickly.
- there are arbitrary bad values in Θ, say if $\theta \in \bar{\Theta} - \Theta$ then P_θ is not ergodic.
- if $\theta \in \Theta_* :=$ a region close to θ_*, then P_θ shall inherit good convergence properties of P_{θ_*}.
- When using adaptive MCMC we hope θ_n will eventually find the region Θ_* and stay there essentially forever. And that the adaptive algorithm $\mathcal{A}$ will inherit the good convergence properties of Θ_* in the limit.
- We are looking for a Theorem:
  You can actually run your Adaptive MCMC algorithm $\mathcal{A}$, and it will do what it is supposed to do! (under verifiable conditions)

What Adaptive MCMC is designed for?

- In a typical Adaptive MCMC setting the parameter space Θ is large
- there is an optimal $\theta_* \in \Theta$ s.t. P_{θ_*} converges quickly.
- there are arbitrary bad values in Θ, say if $\theta \in \bar{\Theta} - \Theta$ then P_θ is not ergodic.
- if $\theta \in \Theta_* :=$ a region close to θ_*, then P_θ shall inherit good convergence properties of P_{θ_*}.
- When using adaptive MCMC we hope θ_n will eventually find the region Θ_* and stay there essentially forever. And that the adaptive algorithm $\mathcal{A}$ will inherit the good convergence properties of Θ_* in the limit.
- We are looking for a Theorem:
  You can actually run your Adaptive MCMC algorithm $\mathcal{A}$, and it will do what it is supposed to do! (under verifiable conditions)

What Adaptive MCMC is designed for?

- In a typical Adaptive MCMC setting the parameter space Θ is large
- there is an optimal $\theta_* \in \Theta$ s.t. P_{θ_*} converges quickly.
- there are arbitrary bad values in Θ, say if $\theta \in \bar{\Theta} - \Theta$ then P_θ is not ergodic.
- if $\theta \in \Theta_* :=$ a region close to θ_*, then P_θ shall inherit good convergence properties of P_{θ_*}.
- When using adaptive MCMC we hope θ_n will eventually find the region Θ_* and stay there essentially forever. And that the adaptive algorithm $\mathcal{A}$ will inherit the good convergence properties of Θ_* in the limit.
- We are looking for a Theorem:
  You can actually run your Adaptive MCMC algorithm $\mathcal{A}$, and it will do what it is supposed to do! (under verifiable conditions)

What Adaptive MCMC is designed for?

- In a typical Adaptive MCMC setting the parameter space Θ is large
- there is an optimal $\theta_* \in \Theta$ s.t. P_{θ_*} converges quickly.
- there are arbitrary bad values in Θ, say if $\theta \in \bar{\Theta} - \Theta$ then P_θ is not ergodic.
- if $\theta \in \Theta_* :=$ a region close to θ_*, then P_θ shall inherit good convergence properties of P_{θ_*}.
- When using adaptive MCMC we hope θ_n will eventually find the region Θ_* and stay there essentially forever. And that the adaptive algorithm $\mathcal{A}$ will inherit the good convergence properties of Θ_* in the limit.
- We are looking for a Theorem: *You can actually run your Adaptive MCMC algorithm $\mathcal{A}$, and it will do what it is supposed to do! (under verifiable conditions)*

What Adaptive MCMC is designed for?

- In a typical Adaptive MCMC setting the parameter space Θ is large
- there is an optimal $\theta_* \in \Theta$ s.t. P_{θ_*} converges quickly.
- there are arbitrary bad values in Θ, say if $\theta \in \bar{\Theta} - \Theta$ then P_θ is not ergodic.
- if $\theta \in \Theta_* :=$ a region close to θ_*, then P_θ shall inherit good convergence properties of P_{θ_*}.
- When using adaptive MCMC we hope θ_n will eventually find the region Θ_* and stay there essentially forever. And that the adaptive algorithm $\mathcal{A}$ will inherit the good convergence properties of Θ_* in the limit.
- We are looking for a Theorem:
You can actually run your Adaptive MCMC algorithm $\mathcal{A}$, and it will do what it is supposed to do! (under verifiable conditions)

What Adaptive MCMC is designed for?

- In a typical Adaptive MCMC setting the parameter space Θ is large
- there is an optimal $\theta_* \in \Theta$ s.t. P_{θ_*} converges quickly.
- there are arbitrary bad values in Θ, say if $\theta \in \bar{\Theta} - \Theta$ then P_θ is not ergodic.
- if $\theta \in \Theta_* :=$ a region close to θ_*, then P_θ shall inherit good convergence properties of P_{θ_*}.
- When using adaptive MCMC we hope θ_n will eventually find the region Θ_* and stay there essentially forever. And that the adaptive algorithm $\mathcal{A}$ will inherit the good convergence properties of Θ_* in the limit.
- We are looking for a Theorem:
  You can actually run your Adaptive MCMC algorithm $\mathcal{A}$, and it will do what it is supposed to do! (under verifiable conditions)

a new class: `AdapFail` Algorithms

- an adaptive algorithm $\mathcal{A} \in$ `AdapFail`, if with positive probability, it is asymptotically less efficient then ANY MCMC algorithm with fixed θ.
- more formally, `AdapFail` can be defined e.g. as follows: $\mathcal{A} \in$ `AdapFail`, if

$$\forall_{\epsilon_* > 0}, \quad \exists_{0<\epsilon<\epsilon_*}, \quad \exists_{\delta>0}, \quad \text{s.t.} \quad \inf_{\theta \in \Theta} \lim_{n \to \infty} P\Big(M_\epsilon(X_n, \theta_n) > M_\epsilon(\tilde{X}_n, \theta)\Big) > \delta.$$

- QuasiLemma: Assume the geometrically ergodic setting and assume Θ is big enough to contain arbitrary slowly converging kernels.
  If containment doesn't hold for $\mathcal{A}$ then $\mathcal{A} \in$ `AdapFail`.
- If A2(a), A2(b) hold but C(a), C(b) do not hold, then $\mathcal{A} \in$ `AdapFail`, but it deteriorates slowly enough (due to more restrictive A2(b)), so that marginal distributions (still) converge, and SLLN (still) holds.

a new class: AdapFail Algorithms

- an adaptive algorithm $\mathcal{A} \in$ AdapFail, if with positive probability, it is asymptotically less efficient then ANY MCMC algorithm with fixed θ.
- more formally, AdapFail can be defined e.g. as follows: $\mathcal{A} \in$ AdapFail, if

$$\forall_{\epsilon_* > 0}, \quad \exists_{0 < \epsilon < \epsilon_*}, \quad \exists_{\delta > 0}, \quad \text{s.t.} \quad \inf_{\theta \in \Theta} \lim_{n \to \infty} P\Big(M_\epsilon(X_n, \theta_n) > M_\epsilon(\tilde{X}_n, \theta)\Big) > \delta.$$

- QuasiLemma: Assume the geometrically ergodic setting and assume Θ is big enough to contain arbitrary slowly converging kernels.
  If containment doesn't hold for $\mathcal{A}$ then $\mathcal{A} \in$ AdapFail.
- If A2(a), A2(b) hold but C(a), C(b) do not hold, then $\mathcal{A} \in$ AdapFail, but it deteriorates slowly enough (due to more restrictive A2(b)), so that marginal distributions (still) converge, and SLLN (still) holds.

a new class: AdapFail Algorithms

- an adaptive algorithm $\mathcal{A} \in$ AdapFail, if with positive probability, it is asymptotically less efficient then ANY MCMC algorithm with fixed θ.
- more formally, AdapFail can be defined e.g. as follows: $\mathcal{A} \in$ AdapFail, if

$$\forall_{\epsilon_* > 0}, \quad \exists_{0<\epsilon<\epsilon_*}, \quad \exists_{\delta>0}, \quad \text{s.t.} \quad \inf_{\theta \in \Theta} \lim_{n \to \infty} P\Big(M_\epsilon(X_n, \theta_n) > M_\epsilon(\tilde{X}_n, \theta)\Big) > \delta.$$

- QuasiLemma: Assume the geometrically ergodic setting and assume Θ is big enough to contain arbitrary slowly converging kernels.
  If containment doesn't hold for $\mathcal{A}$ then $\mathcal{A} \in$ AdapFail.
- If A2(a), A2(b) hold but C(a), C(b) do not hold, then $\mathcal{A} \in$ AdapFail, but it deteriorates slowly enough (due to more restrictive A2(b)), so that marginal distributions (still) converge, and SLLN (still) holds.

a new class: AdapFail Algorithms

- an adaptive algorithm $\mathcal{A} \in$ AdapFail, if with positive probability, it is asymptotically less efficient then ANY MCMC algorithm with fixed θ.
- more formally, AdapFail can be defined e.g. as follows: $\mathcal{A} \in$ AdapFail, if

$$\forall_{\epsilon_*>0}, \quad \exists_{0<\epsilon<\epsilon_*}, \quad \exists_{\delta>0}, \quad \text{s.t.} \quad \inf_{\theta\in\Theta} \lim_{n\to\infty} P\Big(M_\epsilon(X_n, \theta_n) > M_\epsilon(\tilde{X}_n, \theta)\Big) > \delta.$$

- QuasiLemma: Assume the geometrically ergodic setting and assume Θ is big enough to contain arbitrary slowly converging kernels.
If containment doesn't hold for $\mathcal{A}$ then $\mathcal{A} \in$ AdapFail.
- If A2(a), A2(b) hold but C(a), C(b) do not hold, then $\mathcal{A} \in$ AdapFail, but it deteriorates slowly enough (due to more restrictive A2(b)), so that marginal distributions (still) converge, and SLLN (still) holds.

a new class: AdapFail Algorithms

- an adaptive algorithm $\mathcal{A} \in$ AdapFail, if with positive probability, it is asymptotically less efficient then ANY MCMC algorithm with fixed θ.
- more formally, AdapFail can be defined e.g. as follows: $\mathcal{A} \in$ AdapFail, if

$$\forall_{\epsilon_* > 0}, \quad \exists_{0<\epsilon<\epsilon_*}, \quad \exists_{\delta>0}, \quad \text{s.t.} \quad \inf_{\theta \in \Theta} \lim_{n \to \infty} P\Big(M_\epsilon(X_n, \theta_n) > M_\epsilon(\tilde{X}_n, \theta)\Big) > \delta.$$

- QuasiLemma: Assume the geometrically ergodic setting and assume Θ is big enough to contain arbitrary slowly converging kernels.
  If containment doesn't hold for $\mathcal{A}$ then $\mathcal{A} \in$ AdapFail.
- If A2(a), A2(b) hold but C(a), C(b) do not hold, then $\mathcal{A} \in$ AdapFail, but it deteriorates slowly enough (due to more restrictive A2(b)), so that marginal distributions (still) converge, and SLLN (still) holds.

some QuasiConjectures

- If containment does not hold for $\mathcal{A}$ then the CLT does not hold for $\mathcal{A}$. (Can one prove that $\sigma_{\mathrm{as}} > K$ for arbitrary large K?)
- the Adaptive Metropolis $\notin$ `AdapFail`, (looking forward to talks by Matti and Yves that may be related)
- the Interactive Tempering $\notin$ `AdapFail`, (the uniformly ergodic case in Bercu, Del Moral, Doucet 2010)
- i.e. Containment holds for both.

some QuasiConjectures

- If containment does not hold for $\mathcal{A}$ then the CLT does not hold for $\mathcal{A}$. (Can one prove that $\sigma_{\text{as}} > K$ for arbitrary large K?)
- the Adaptive Metropolis $\notin$ `AdapFail`, (looking forward to talks by Matti and Yves that may be related)
- the Interactive Tempering $\notin$ `AdapFail`, (the uniformly ergodic case in Bercu, Del Moral, Doucet 2010)
- i.e. Containment holds for both.

some QuasiConjectures

- If containment does not hold for $\mathcal{A}$ then the CLT does not hold for $\mathcal{A}$. (Can one prove that $\sigma_{\text{as}} > K$ for arbitrary large K?)
- the Adaptive Metropolis $\notin$ `AdapFail`, (looking forward to talks by Matti and Yves that may be related)
- the Interactive Tempering $\notin$ `AdapFail`, (the uniformly ergodic case in Bercu, Del Moral, Doucet 2010)
- i.e. Containment holds for both.

some QuasiConjectures

- If containment does not hold for $\mathcal{A}$ then the CLT does not hold for $\mathcal{A}$.
  (Can one prove that $\sigma_{\mathrm{as}} > K$ for arbitrary large K?)
- the Adaptive Metropolis $\notin$ `AdapFail`,
  (looking forward to talks by Matti and Yves that may be related)
- the Interactive Tempering $\notin$ `AdapFail`,
  (the uniformly ergodic case in Bercu, Del Moral, Doucet 2010)
- i.e. Containment holds for both.

Design of Algorithms

- We have substantial insight into ergodicity of adaptive MCMC
- is it possible to use this to design algorithms that are robust and easier to analyze?
- Ex. for the Adaptive Metropolis, consider the following slight modification [Roberts, Rosenthal (2009)] with proposal distribution

$$Q_n(x, \cdot) = (1-\beta)\mathcal{N}(x, (2.38)^2 \Sigma_n/d) + \beta\mathcal{N}(x, (0.1/d)Id).$$

- the above modification appears more tractable and C(a), C(b) are known to hold for both, exponentially and super-exponentially decaying tails (Bai et al 2009).
- is it possible to modify other challenging adaptive algorithms to make them easier to analyze without destroying their empirical properties?

Design of Algorithms

- We have substantial insight into ergodicity of adaptive MCMC
- is it possible to use this to design algorithms that are robust and easier to analyze?
- Ex. for the Adaptive Metropolis, consider the following slight modification [Roberts, Rosenthal (2009)] with proposal distribution

$$Q_n(x, \cdot) \;=\; (1-\beta)\mathcal{N}(x, (2.38)^2\Sigma_n/d) \;+\; \beta\mathcal{N}(x, (0.1/d)Id).$$

- the above modification appears more tractable and C(a), C(b) are known to hold for both, exponentially and super-exponentially decaying tails (Bai et al 2009).
- is it possible to modify other challenging adaptive algorithms to make them easier to analyze without destroying their empirical properties?

Design of Algorithms

- We have substantial insight into ergodicity of adaptive MCMC
- is it possible to use this to design algorithms that are robust and easier to analyze?
- Ex. for the Adaptive Metropolis, consider the following slight modification [Roberts, Rosenthal (2009)] with proposal distribution

$$Q_n(x, \cdot) = (1-\beta)\mathcal{N}(x, (2.38)^2 \Sigma_n / d) + \beta \mathcal{N}(x, (0.1/d) Id).$$

- the above modification appears more tractable and C(a), C(b) are known to hold for both, exponentially and super-exponentially decaying tails (Bai et al 2009).
- is it possible to modify other challenging adaptive algorithms to make them easier to analyze without destroying their empirical properties?

Design of Algorithms

- We have substantial insight into ergodicity of adaptive MCMC
- is it possible to use this to design algorithms that are robust and easier to analyze?
- Ex. for the Adaptive Metropolis, consider the following slight modification [Roberts, Rosenthal (2009)] with proposal distribution

$$Q_n(x, \cdot) = (1-\beta)\mathcal{N}(x, (2.38)^2\Sigma_n/d) + \beta\mathcal{N}(x, (0.1/d)Id).$$

- the above modification appears more tractable and C(a), C(b) are known to hold for both, exponentially and super-exponentially decaying tails (Bai et al 2009).
- is it possible to modify other challenging adaptive algorithms to make them easier to analyze without destroying their empirical properties?

Design of Algorithms

- We have substantial insight into ergodicity of adaptive MCMC
- is it possible to use this to design algorithms that are robust and easier to analyze?
- Ex. for the Adaptive Metropolis, consider the following slight modification [Roberts, Rosenthal (2009)] with proposal distribution

$$Q_n(x, \cdot) \;=\; (1-\beta)\mathcal{N}(x, (2.38)^2\Sigma_n/d) \;+\; \beta\mathcal{N}(x, (0.1/d)Id).$$

- the above modification appears more tractable and C(a), C(b) are known to hold for both, exponentially and super-exponentially decaying tails (Bai et al 2009).
- is it possible to modify other challenging adaptive algorithms to make them easier to analyze without destroying their empirical properties?