Sticky proposal densities for adaptive MCMC methods

L. Martino[†],R. Casarin[‡], F. Leisen[§], D. Luengo[¶],

[†]University of Helsinki, [‡]Universitá Ca' Foscari,
[§]University of Kent, [¶]Universidad Politecnica de Madrid.

MCQMC, 2014

Introduction

- Markov Chain Monte Carlo (MCMC) methods convert samples from a proposal pdf $\tilde{q}(x) \propto q(x)$, into correlated samples from a target pdf $\tilde{\pi}(x) \propto \pi(x)$, generating a chain.

$$x_0 \Longrightarrow x_1 \Longrightarrow \ldots x_t \underbrace{\Longrightarrow}_{K(x_t|x_{t-1})} x_{t+1} \Longrightarrow \ldots x_{t+\tau} \sim \tilde{\pi}(x)$$

Introduction

- Markov Chain Monte Carlo (MCMC) methods convert samples from a proposal pdf $\tilde{q}(x) \propto q(x)$, into correlated samples from a target pdf $\tilde{\pi}(x) \propto \pi(x)$, generating a chain.

$$x_0 \Longrightarrow x_1 \Longrightarrow \ldots x_t \underbrace{\Longrightarrow}_{K(x_t|x_{t-1})} x_{t+1} \Longrightarrow \ldots x_{t+\tau} \sim \tilde{\pi}(x)$$

- Within the Monte Carlo (MC) techniques:
  - **[Gilks et al. (1992)]**: *adaptive rejection sampling* (**ARS**),
  - **[Gilks et al. (1995)]**: *adaptive rejection Metropolis sampling* (**ARMS**),

  are samplers from **univariate** pdfs.

INTRODUCTION

- Markov Chain Monte Carlo (MCMC) methods convert samples from a proposal pdf $\tilde{q}(x) \propto q(x)$, into correlated samples from a target pdf $\tilde{\pi}(x) \propto \pi(x)$, generating a chain.

$$x_0 \Longrightarrow x_1 \Longrightarrow \dots x_t \underbrace{\Longrightarrow}_{K(x_t|x_{t-1})} x_{t+1} \Longrightarrow \dots x_{t+\tau} \sim \tilde{\pi}(x)$$

- Within the Monte Carlo (MC) techniques:
  - **[Gilks et al. (1992)]**: *adaptive rejection sampling* (**ARS**),
  - **[Gilks et al. (1995)]**: *adaptive rejection Metropolis sampling* (**ARMS**),

  are samplers from **univariate** pdfs.
  - They are often used within Gibbs sampling.
  - Both techniques present different limitations.

Introduction

- Markov Chain Monte Carlo (MCMC) methods convert samples from a proposal pdf $\tilde{q}(x) \propto q(x)$, into correlated samples from a target pdf $\tilde{\pi}(x) \propto \pi(x)$, generating a chain.

$$x_0 \Longrightarrow x_1 \Longrightarrow \ldots x_t \underbrace{\Longrightarrow}_{K(x_t|x_{t-1})} x_{t+1} \Longrightarrow \ldots x_{t+\tau} \sim \tilde{\pi}(x)$$

- Within the Monte Carlo (MC) techniques:
  - **[Gilks et al. (1992)]**: *adaptive rejection sampling* (**ARS**),
  - **[Gilks et al. (1995)]**: *adaptive rejection Metropolis sampling* (**ARMS**),

  are samplers from **univariate** pdfs.
  - They are often used within Gibbs sampling.
  - Both techniques present different limitations.
- **GOAL:** Overcoming these drawbacks by proposing a more general and efficient class of adaptive samplers.

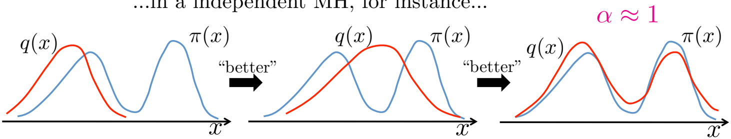
Performance

- The performance of an MCMC method depends strictly on the discrepancy between proposal, q and target, π.

Performance

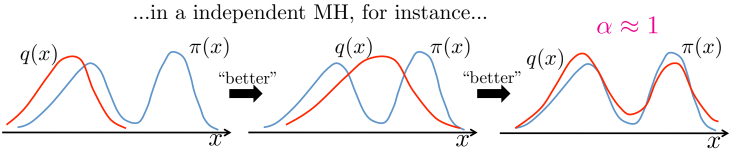
- The performance of an MCMC method depends strictly on the discrepancy between proposal, q and target, π.
- If proposal=target, we have an exact sampler.

...in a independent MH, for instance...



Performance

- The performance of an MCMC method depends strictly on the discrepancy between proposal, q and target, π.
- If proposal=target, we have an exact sampler.



- Need of *adapting* the proposal density, while ensuring ergodicity.

Adaptive procedures

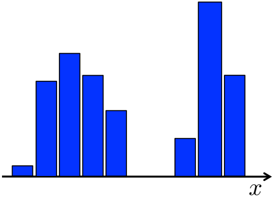
- **Parametric:** Learn parameters of the proposal (location and/or scale parameter).
- **Non-parametric:** Approximate the target via non-parametric procedures (as in **kernel density estimation**).

Adaptive procedures

- **Parametric:** Learn parameters of the proposal (location and/or scale parameter).
- **Non-parametric:** Approximate the target via non-parametric procedures (as in **kernel density estimation**).
  - Simple idea: Update the proposal taking into account the histogram of the generated samples (after "burn-in"):

$$x_1, \ldots, x_t, \ldots, x_{t+\tau} \ldots$$

$$\text{proposal} = \beta_t \times \text{random walk} + (1 - \beta_t) \times$$



OTHER USEFUL INFORMATION

- We have several evaluations of the target pdf available (at least at each state of the chain).

$$x_1, \ldots, x_t, \ldots, x_{t+\tau},$$

$$\pi(x_1), \ldots, \pi(x_t), \ldots, \pi(x_{t+\tau}).$$

- Can we incorporate all this information (or a subset) in the learning procedure?

OTHER USEFUL INFORMATION

- We have several evaluations of the target pdf available (at least at each state of the chain).

$$x_1, \dots, x_t, \dots, x_{t+\tau},$$

$$\pi(x_1), \dots, \pi(x_t), \dots, \pi(x_{t+\tau}).$$

- Can we incorporate all this information (or a subset) in the learning procedure?
- AIM: Interpolative construction of a proposal q which depends on a subset $\mathcal{S}_t \subset \{x_1, \dots, x_t\}$,

$$\tilde{q}(x) = \tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t).$$

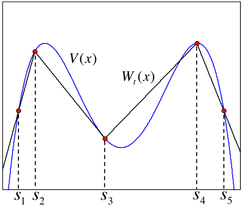
- Adaptive proposal $\Longrightarrow$ adaptive MCMC.

Interpolation procedures

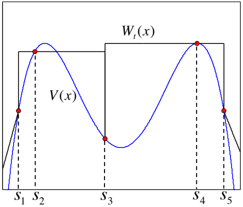
- Consider a set of support points $\mathcal{S}_t = \{s_1, \ldots, s_{m_t}\}$, and

$$V(x) = \log[\pi(x)], \qquad W_t(x) = \log[q_t(x|\mathcal{S}_t)].$$

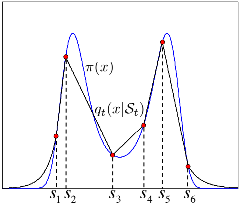
- Interpolation procedure:



(a) **P2**: log-domain



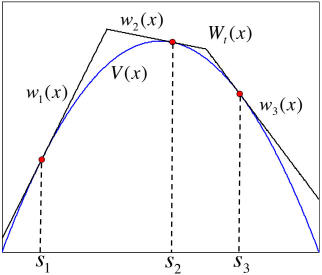
(b) **P3**: log-domain



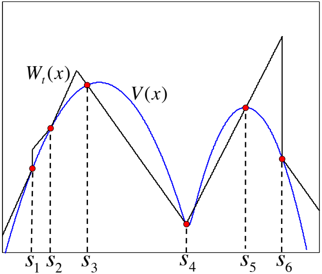
(c) **P4**: pdf-domain

Interpolation procedures

- Similar to the constructions in the *adaptive rejection sampling* (ARS) **[Gilks et al., 1992]** and *adaptive rejection Metropolis sampling* (ARMS) methods **[Gilks et al., 1995]**.



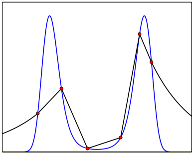
(d) log-domain (ARS)



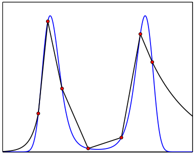
(e) **P1**: log-domain (ARMS)

- **ARS:** only for log-concave pdfs.
- **ARMS:** sometimes incomplete adaptation.

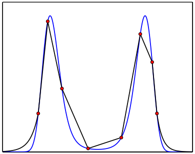
Interpolation procedures



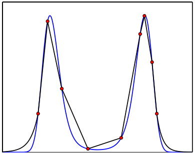
(f) **P4**: $|\mathcal{S}_t| = 6$



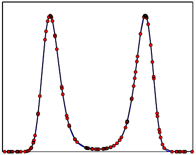
(g) **P4**: $|\mathcal{S}_t| = 7$



(h) **P4**: $|\mathcal{S}_t| = 8$



(i) **P4**: $|\mathcal{S}_t| = 9$



(j) **P4**: $|\mathcal{S}_t| > 100$

- Here the points are not adaptively chosen.

Drawing from q_t

1. Calculate analytically the area below each piece, i.e.,

$$\int_{s_j}^{s_{j+1}} q_t(x|\mathcal{S}_t)dx = A_j, \quad j = 0, \dots, m_t,$$

   denoting $s_0 = -\infty$ and $s_{m_t+1} = +\infty$.

2. Choose a j^*-th piece according to

$$\omega_j = \frac{A_j}{\sum_{j=1}^{n} A_j}, \quad j = 0, \dots, m_t.$$

3. Draw a sample x' from $q_t(x|\mathcal{S}_t)$ with $x \in (s_{j^*}, s_{j^*+1})$.

P2 $\longrightarrow$ exponential pieces
P3 $\longrightarrow$ uniform pieces
P4 $\longrightarrow$ linear pieces

Computational cost - efficiency

- More points: better approximation of the target $\Rightarrow$ more efficiency (i.e., less correlation $\Leftrightarrow$ faster convergence).
- More points: to draw from q_t is more costly.

$m_t \uparrow \quad \Longrightarrow$ efficiency $\uparrow$ + computational cost $\uparrow$

- **Desired adaptive strategy:** manage the set $\mathcal{S}_t$ in order to build a "good" proposal with a small number m_t of points, keeping the ergodicity of the sampler.

Adaptive Sticky Metropolis (ASM)

1. **Construction of the proposal:** Build a proposal $q_t(x|\mathcal{S}_t)$, using the set $\mathcal{S}_t = \{s_1, \ldots, s_{m_t}\}$ (e.g., using **P1**, **P2**, **P3** and **P4**).

Adaptive Sticky Metropolis (ASM)

1. **Construction of the proposal:** Build a proposal $q_t(x|\mathcal{S}_t)$, using the set $\mathcal{S}_t = \{s_1, \ldots, s_{m_t}\}$ (e.g., using **P1**, **P2**, **P3** and **P4**).
2. **MH step:**
   2.1 Draw x' from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$.
   2.2 Set $x_{t+1} = x'$ and $z = x_t$ with probability

$$\alpha = 1 \wedge \frac{\pi(x')q_t(x_t|\mathcal{S}_t)}{\pi(x_t)q_t(x'|\mathcal{S}_t)},$$

   and set $x_{t+1} = x_t$ and $z = x'$, with probability $1 - \alpha$.

Adaptive Sticky Metropolis (ASM)

1. **Construction of the proposal:** Build a proposal $q_t(x|\mathcal{S}_t)$, using the set $\mathcal{S}_t = \{s_1, \ldots, s_{m_t}\}$ (e.g., using **P1**, **P2**, **P3** and **P4**).
2. **MH step:**
   2.1 Draw x' from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$.
   2.2 Set $x_{t+1} = x'$ and $z = x_t$ with probability

$$\alpha = 1 \wedge \frac{\pi(x')q_t(x_t|\mathcal{S}_t)}{\pi(x_t)q_t(x'|\mathcal{S}_t)},$$

   and set $x_{t+1} = x_t$ and $z = x'$, with probability $1 - \alpha$.
3. **Test to update $\mathcal{S}_t$:** Set

$$\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{z\} \quad \text{with prob.} \quad P_a = \eta(d_t(z)),$$

otherwise $\mathcal{S}_{t+1} = \mathcal{S}_t$.

Adaptive Sticky Metropolis (ASM)

1. **Construction of the proposal:** Build a proposal $q_t(x|\mathcal{S}_t)$, using the set $\mathcal{S}_t = \{s_1, \ldots, s_{m_t}\}$ (e.g., using **P1**, **P2**, **P3** and **P4**).
2. **MH step:**
   2.1 Draw x' from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$.
   2.2 Set $x_{t+1} = x'$ and $z = x_t$ with probability
   $$\alpha = 1 \wedge \frac{\pi(x')q_t(x_t|\mathcal{S}_t)}{\pi(x_t)q_t(x'|\mathcal{S}_t)},$$
   and set $x_{t+1} = x_t$ and $z = x'$, with probability $1 - \alpha$.
3. **Test to update $\mathcal{S}_t$:** Set
   $$\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{z\} \quad \text{with prob.} \quad P_a = \eta(d_t(z)),$$
   otherwise $\mathcal{S}_{t+1} = \mathcal{S}_t$.
   - $d_t(z) \Rightarrow$ a positive measure of the distance in z between the q_t and π.

Adaptive Sticky Metropolis (ASM)

1. **Construction of the proposal:** Build a proposal $q_t(x|\mathcal{S}_t)$, using the set $\mathcal{S}_t = \{s_1, \ldots, s_{m_t}\}$ (e.g., using **P1**, **P2**, **P3** and **P4**).
2. **MH step:**
   2.1 Draw x' from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$.
   2.2 Set $x_{t+1} = x'$ and $z = x_t$ with probability

$$\alpha = 1 \wedge \frac{\pi(x')q_t(x_t|\mathcal{S}_t)}{\pi(x_t)q_t(x'|\mathcal{S}_t)},$$

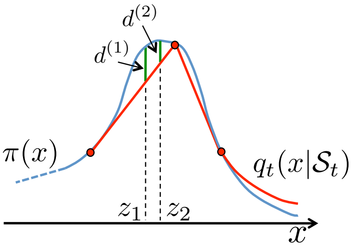
   and set $x_{t+1} = x_t$ and $z = x'$, with probability $1 - \alpha$.

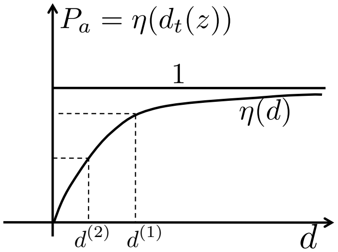
3. **Test to update $\mathcal{S}_t$:** Set

$$\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{z\} \quad \text{with prob.} \quad P_a = \eta(d_t(z)),$$

   otherwise $\mathcal{S}_{t+1} = \mathcal{S}_t$.
   - $d_t(z) \Rightarrow$ a positive measure of the distance in z between the q_t and π.
   - $\eta : \mathbb{R}^+ \to [0, 1] \Rightarrow$ increasing, with $\eta(0) = 0$, $\eta(\infty) = 1$.

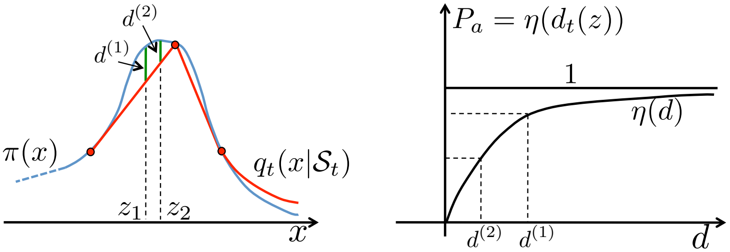
Control test: Update of $\mathcal{S}_t$





$$\boxed{\int_{\mathcal{X}} |\pi(x) - q_t(x|\mathcal{S}_t)|dx \to 0 \quad \Longrightarrow \quad P_a \to 0}$$

Control test: Update of $\mathcal{S}_t$

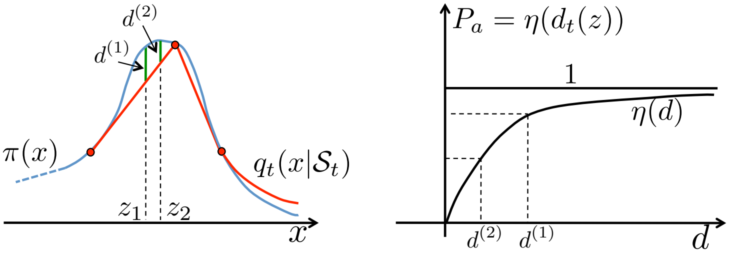


$$\boxed{\int_{\mathcal{X}} |\pi(x) - q_t(x|\mathcal{S}_t)|dx \rightarrow 0 \quad \Longrightarrow \quad P_a \rightarrow 0}$$

We obtain, at the same time, both:

- **Efficiency:** we add points where (and when) exactly needed.
- **Bounded computational cost:** since $P_a \rightarrow 0$, m_T is controlled.

Control test: Update of $\mathcal{S}_t$



$$\boxed{\int_{\mathcal{X}} |\pi(x) - q_t(x|\mathcal{S}_t)|dx \to 0 \quad \Longrightarrow \quad P_a \to 0}$$

We obtain, at the same time, both:

- **Efficiency:** we add points where (and when) exactly needed.
- **Bounded computational cost:** since $P_a \to 0$, m_T is controlled.

Exactly as in the ARS [Gilks et al., 1992].

An example of ASM

1. Build $q_t(x|\mathcal{S}_t)$.
2. Draw $x' \sim \tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$.
3. Set $x_{t+1} = x'$ and $z = x_t$ with probability

$$\alpha = 1 \wedge \frac{\pi(x')q_t(x_t|\mathcal{S}_t)}{\pi(x_t)q_t(x'|\mathcal{S}_t)},$$

   otherwise set $x_{t+1} = x_t$ and $z = x'$.
4. Draw $u' \sim \mathcal{U}([0,1])$. If

$$u' \geq \frac{\min[\pi(z), q_t(z|\mathcal{S}_t)]}{\max[\pi(z), q_t(z|\mathcal{S}_t)]},$$

   set $\mathcal{S}_{t+1} = \mathcal{S}_t \cup \{z\}$, otherwise set $\mathcal{S}_{t+1} = \mathcal{S}_t$.

Other possible tests

$\mathbf{d_t(z)}$	$\boldsymbol{\eta}(\mathbf{d})$	**Type**
$d_t(z) = 1 - \frac{\min[\pi(z), q_t(z\|\mathcal{S}_t)]}{\max[\pi(z), q_t(z\|\mathcal{S}_t)]}$	$\eta(d) = d$, with $d \in [0, 1]$	random (similar to ARS, ARMS)
$d_t(z) = \|\pi(x) - q_t(x\|\mathcal{S}_t)\|$	$\eta(d) = 1 - \exp(-d)$, with $d \in \mathbb{R}^+$	random
$d_t(z) = \|\pi(x) - q_t(x\|\mathcal{S}_t)\|$	$\eta(d) = 1$ if $d_t(z) > \varepsilon$ $\eta(d) = 0$ if $d_t(z) \leq \varepsilon$	deterministic

- With the deterministic test, at some $t^* < \infty$, the adaptation could be stopped, depending on ε.

Ergodicity

- Based on a result in **[Holden et at., 2009]**:
    1. The adaptation procedure must use z instead of x_{t+1}.

Ergodicity

- Based on a result in **[Holden et at., 2009]**:
  1. The adaptation procedure must use z instead of x_{t+1}.
  2. The proposal must satisfy the *strong Doeblin's condition*, i.e., there exists a value $a_t \in (0,1]$, $\forall t \in \mathbb{N}$, such that

$$\frac{1}{a_t}\tilde{q}_t(x|\mathcal{S}_t) \geq \tilde{\pi}(x), \quad \forall x \in \mathcal{X}.$$

Ergodicity

- Based on a result in **[Holden et at., 2009]**:
  1. The adaptation procedure must use z instead of x_{t+1}.
  2. The proposal must satisfy the *strong Doeblin's condition*, i.e., there exists a value $a_t \in (0,1]$, $\forall t \in \mathbb{N}$, such that

$$\frac{1}{a_t}\tilde{q}_t(x|\mathcal{S}_t) \geq \tilde{\pi}(x), \quad \forall x \in \mathcal{X}.$$

- **Fulfilled by ASM** (note that we can always change the type of tails used in the proposal construction).
- For more details, see

[Holden09]: L. Holden, R. Hauge, and M. Holden. "Adaptive Independent Metropolis-Hastings." The Annals of Applied Probability, 19(1): 395-413, 2009.

Adaptive Sticky Multiple Try Metropolis

1. **Construction of the proposal:** Build $q_t(x|\mathcal{S}_t)$ using the set $\mathcal{S}_t$.

Adaptive Sticky Multiple Try Metropolis

1. **Construction of the proposal:** Build $q_t(x|\mathcal{S}_t)$ using the set $\mathcal{S}_t$.
2. **MTM step:**
   2.1 Draw $x'_1, \dots, x'_M$ from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$ and compute the weights $w_t(x'_i) = \frac{\pi(x'_i)}{q_t(x'_i|\mathcal{S}_t)}$.

Adaptive Sticky Multiple Try Metropolis

1. **Construction of the proposal:** Build $q_t(x|\mathcal{S}_t)$ using the set $\mathcal{S}_t$.
2. **MTM step:**
   2.1 Draw $x'_1, \ldots, x'_M$ from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$ and compute the weights $w_t(x'_i) = \frac{\pi(x'_i)}{q_t(x'_i|\mathcal{S}_t)}$.
   2.2 Select $x' = x'_j \in \{x'_1, ..., x'_M\}$ with probability $\frac{w_t(x'_j)}{\sum_{i=1}^{M} w_t(x'_i)}$.

Adaptive Sticky Multiple Try Metropolis

1. **Construction of the proposal:** Build $q_t(x|\mathcal{S}_t)$ using the set $\mathcal{S}_t$.
2. **MTM step:**
   2.1 Draw $x'_1, \ldots, x'_M$ from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$ and compute the weights $w_t(x'_i) = \frac{\pi(x'_i)}{q_t(x'_i|\mathcal{S}_t)}$.
   2.2 Select $x' = x'_j \in \{x'_1, ..., x'_M\}$ with probability $\frac{w_t(x'_j)}{\sum_{i=1}^{M} w_t(x'_i)}$.
   2.3 Set the auxiliary points $x^*_i = x'_i$ and $z_i = x'_i$, $i \neq j$ and $x^*_j = x_t$.

Adaptive Sticky Multiple Try Metropolis

1. **Construction of the proposal:** Build $q_t(x|\mathcal{S}_t)$ using the set $\mathcal{S}_t$.
2. **MTM step:**
   - 2.1 Draw $x'_1, \dots, x'_M$ from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$ and compute the weights $w_t(x'_i) = \frac{\pi(x'_i)}{q_t(x'_i|\mathcal{S}_t)}$.
   - 2.2 Select $x' = x'_j \in \{x'_1, ..., x'_M\}$ with probability $\frac{w_t(x'_j)}{\sum_{i=1}^{M} w_t(x'_i)}$.
   - 2.3 Set the auxiliary points $x^*_i = x'_i$ and $z_i = x'_i$, $i \neq j$ and $x^*_j = x_t$.
   - 2.4 Set $x_{t+1} = x'$ and $z_j = x_t$ with probability

$$\alpha = \min\left[1, \frac{w_t(x'_1) + \cdots + w_t(x'_M)}{w_t(x^*_1) + \cdots + w_t(x^*_M)}\right],$$

     and set $x_{t+1} = x_t$ and $z_j = x'_j$, with probability $1 - \alpha$.

Adaptive Sticky Multiple Try Metropolis

1. **Construction of the proposal:** Build $q_t(x|\mathcal{S}_t)$ using the set $\mathcal{S}_t$.
2. **MTM step:**
   2.1 Draw $x'_1, \dots, x'_M$ from $\tilde{q}_t(x) \propto q_t(x|\mathcal{S}_t)$ and compute the weights $w_t(x'_i) = \frac{\pi(x'_i)}{q_t(x'_i|\mathcal{S}_t)}$.
   2.2 Select $x' = x'_j \in \{x'_1, ..., x'_M\}$ with probability $\frac{w_t(x'_j)}{\sum_{i=1}^M w_t(x'_i)}$.
   2.3 Set the auxiliary points $x^*_i = x'_i$ and $z_i = x'_i$, $i \neq j$ and $x^*_j = x_t$.
   2.4 Set $x_{t+1} = x'$ and $z_j = x_t$ with probability

$$\alpha = \min\left[1, \frac{w_t(x'_1) + \cdots + w_t(x'_M)}{w_t(x^*_1) + \cdots + w_t(x^*_M)}\right],$$

   and set $x_{t+1} = x_t$ and $z_j = x'_j$, with probability $1 - \alpha$.
3. **Test to update $\mathcal{S}_t$:** Set

$$\mathcal{S}_t = \begin{cases} \mathcal{S}_{t-1} \cup \{z_i\} & \text{with prob.} \quad \eta_i(d_t(z_i)), i = 1, \dots, M \\ \mathcal{S}_{t-1} & \text{with prob.} \quad 1 - \sum_{i=1}^M \eta_i(d_t(z_i)). \end{cases}$$

Ergodicity

- The proof of ASMTM is an extension of the results in **[Holden09]** . See

  L. Martino, R. Casarin, F. Leisen, D. Luengo, "Adaptive Sticky Generalized Metropolis", arXiv:1308.3779, 2013.

- The proof is valid for ASM and ASMTM for a generic construction of the proposal (not only univariate).
- The proposal must fulfill the Doeblin's condition.

Higher dimensions: ASM within Gibbs

- This approach is not confined only to the one-dimensional case. It can be used to the multidimensional setting via a suitable interpolation procedure (still an open problem).

Higher dimensions: ASM within Gibbs

- This approach is not confined only to the one-dimensional case. It can be used to the multidimensional setting via a suitable interpolation procedure (still an open problem).
- Sticky proposals: easy to be implemented in one-dimension.

Higher dimensions: ASM within Gibbs

- This approach is not confined only to the one-dimensional case. It can be used to the multidimensional setting via a suitable interpolation procedure (still an open problem).
- Sticky proposals: easy to be implemented in one-dimension.
- Within Gibbs: we need efficient samplers to draw from the full-conditional pdfs (as close as possible to an exact sampler).

Numerical results

- **Target pdf:**

$$\tilde{\pi}(x) \propto \pi(x) = 0.5\mathcal{N}(7,1) + 0.5\mathcal{N}(-7,0.1), \qquad (1)$$

- **Goal:** Estimating the mean of $X \sim \tilde{\pi}(x)$ ($E[X] = 0$).
- **Experimental Setting:**
  - Use all the generated samples ($T = 5000$) without removing any "burn-in" period.
  - Perform 2000 runs using an initial $\mathcal{S}_0 = \{-10, -8, 5, 10\}$.
- We compare with the Standard ARMS method [Gilks et al., 1995] which corresponds to the first row of Table 1.
- ARMS is often used within Gibbs.

Numerical results

Algorithm	MSE	ACF(1)	ACF(10)	ACF(50)	m_T	Time
ARMS-P1 (Gilks)	10.0395	0.4076	0.3250	0.2328	118.1912	1.0000
ARMS-P2	15.6756	0.8955	0.7210	0.4639	7.6126	0.1195
ARMS-P3	0.2398	0.8753	0.4410	0.0296	131.3360	0.3589
ARMS-P4	0.2874	0.8882	0.4758	0.0418	42.8872	0.2291
ASM-P1	3.0277	0.1284	0.1099	0.0934	152.6301	1.2274
ASM-P2	2.9952	0.1306	0.1125	0.0929	71.1478	0.2757
ASM-P3	0.0290	0.0535	0.0165	0.0077	279.6570	0.6494
ASM-P4	0.0354	0.0354	0.0195	0.0086	84.8742	0.3297
ASMTM-P1 ($M = 10$)	0.6720	0.0726	0.0696	0.0624	159.0060	2.3547
ASMTM-P1 ($M = 50$)	0.1666	0.0430	0.0395	0.0316	160.7579	6.4518
ASMTM-P2 ($M = 10$)	0.5632	0.0588	0.0525	0.0443	72.1628	1.1291
ASMTM-P2 ($M = 50$)	0.1156	0.0345	0.0303	0.0231	72.5270	4.3802
ASMTM-P3 ($M = 10$)	0.0105	0.0045	0.0001	0.0001	315.7808	2.6022
ASMTM-P3 ($M = 50$)	0.0099	0.0063	0.0001	0.0001	360.7323	10.5935
ASMTM-P4 ($M = 10$)	0.0108	0.0036	0.0011	0.0014	92.6660	1.8618
ASMTM-P4 ($M = 50$)	0.0098	0.0001	0.0001	0.0001	101.7775	7.2475

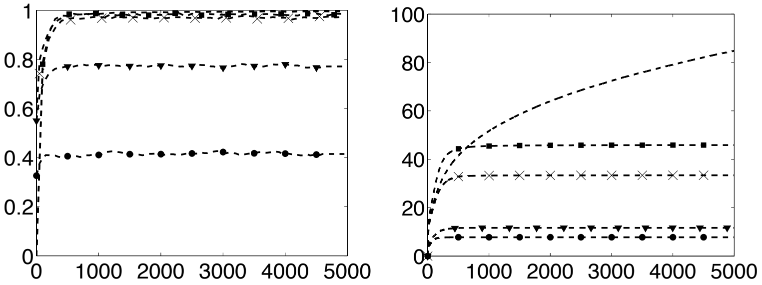
Table: Different columns: the mean square error (MSE), the autocorrelation function (ACF(k)) at different lags, $k = 1, 10, 50$, the final number of support points (m_T), the computing times normalized w.r.t. ARMS [Gilks et al., 95] (Time).

Numerical results

- **ASM schemes provide better results than the standard ARMS in all cases, regardless of the scheme used to build the proposal.**
- ASM-P4 is also faster then ARMS (-P1, [Gilks95]), providing better results.
- ASM is also quite robust w.r.t. the choice of the initial set $\mathcal{S}_0$.
- Good results are also obtained with other kinds of distributions; see

L. Martino, R. Casarin, F. Leisen, D. Luengo, "Adaptive Sticky Generalized Metropolis", arXiv:1308.3779, 2013.

Numerical results



(k) α vs t (ASM-P4) (l) m_t vs t (ASM-P4)

Figure: Averaged α and number of support points m_t over the ASM chain iterations. In each plot the results of the ASM-P4 with random test Ex-3, $\beta = 1$, (line without symbol) is compared with the results of a deterministic test with $\varepsilon = 0.005$ (square), $\varepsilon = 0.01$ (cross), $\varepsilon = 0.1$ (triangle) and $\varepsilon = 0.2$ (circle).

CONCLUSIONS

Advantages:

- ASM is a valid alternative for ARS and ARMS.
- Good performance $\Rightarrow$ ASM is an asymptotically exact sampler.
- Really useful within Gibbs.

Limitations:

- Difficult to build the proposal in higher-dimension.

Future:

- **Can we use a Gaussian Process (GP) as proposal pdf?** this can solve the previous limitation . . . (work in progress)

- Thank you very much!
- Any questions?

Main references:

[Gilks92]: W. R. Gilks and P. Wild. “Adaptive Rejection Sampling for Gibbs Sampling.” Applied Statistics, 41(2): 337-348, 1992.

[Gilks95]: W. R. Gilks, N. G. Best and K. K. C. Tan. “Adaptive Rejection Metropolis Sampling within Gibbs Sampling.” Applied Statistics, 44(4): 455-472, 1995.

[Holden09]: L. Holden, R. Hauge, and M. Holden. “Adaptive Independent Metropolis-Hastings.” The Annals of Applied Probability, 19(1): 395-413, 2009.

Further info:

L. Martino, R. Casarin, F. Leisen, D. Luengo, "Adaptive Sticky Generalized Metropolis", arXiv:1308.3779, 2013.