Convergence and well-definedness of ellitpical slice sampling

Daniel Rudolf

Universität Passau

June 2023

Joint work with

M. Hasenpflug and V. Natarovskii

Motivation/Goal

For $\rho\colon \mathbb{R}^d \to (0,\infty)$ and Gaussian reference measure $\mathcal{N}(0,C)$ with covariance matrix C sample (approximately) w.r.t.

$$\mu(A) = \frac{\int_A \rho(x)\,\mathcal{N}(0,C)(\mathrm{d}x)}{\int_{\mathbb{R}^d} \rho(x)\,\mathcal{N}(0,C)(\mathrm{d}x)}, \qquad A \in \mathcal{B}(\mathbb{R}^d).$$

Idea Markov chain Monte Carlo:

- Run a Markov chain $X_1, X_2, \ldots$ with limit distribution μ;
- After sufficiently many steps return X_n.

Ideal slice sampling

For $t \in (0, \|\rho\|_\infty)$ define the **level set** of ρ (of level t) by

$$G(t) := \left\{ x \in \mathbb{R}^d \mid \rho(x) > t \right\}, \quad \text{and let} \quad \mu_{0,t}(A) := \frac{\mathcal{N}(0,C)(A \cap G(t))}{\mathcal{N}(0,C)(G(t))}, \quad A \in \mathcal{B}(\mathbb{R}^d).$$

Transition of an **ideal slice sampling** Markov chain $(X_n)_{n \in \mathbb{N}}$:

Assume that $X_n = x$, then the next state X_{n+1} is generated by:

1. Draw $T \sim \mathrm{Unif}(0, \rho(x))$, call the realization t.
2. Draw $X_{n+1} \sim \mu_{0,t}$.

Slice sampling

The corresponding transition kernel is given by

$$S(x,A) = \frac{1}{\rho(x)} \int_0^{\rho(x)} \mu_{0,t}(A)\,\mathrm{d}t, \quad x \in \mathbb{R}^d, \quad A \in \mathcal{B}(\mathbb{R}^d)$$

Contra: Difficult to sample $\mu_{0,t}$ on $G(t)$.

Idea: Run a suitable Markov chain on the level set $G(t)$.

(see e.g. Neal 2003, MacKay 2003, Murray et al. 2010, ...)

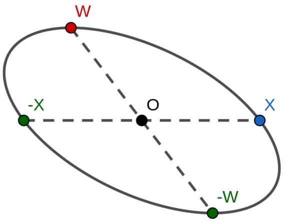
Elliptical slice sampling

For $x, w \in \mathbb{R}^d$ let $p_{x,w} : [0, 2\pi) \to \mathbb{R}^d$ be

$$p_{x,w}(\theta) := \cos(\theta)\ x + \sin(\theta)\ w.$$

For $\theta_1, \theta_2 \in [0, 2\pi)$ intervals on the circle are

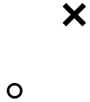
$$I(\theta_1, \theta_2) = \begin{cases} [\theta_1, \theta_2) & \theta_1 < \theta_2 \\ [0, \theta_2] \cup [\theta_1, 2\pi) & \theta_1 \geq \theta_2. \end{cases}$$



Elliptical slice sampling

input: current state $X_n = x \in \mathbb{R}^d$
output: next state X_{n+1}

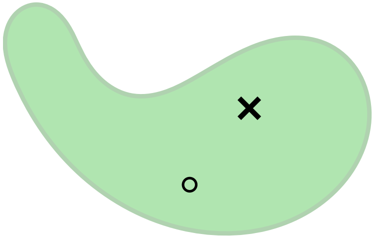
1. draw $t \sim \mathrm{Unif}(0, \rho(x))$
2. draw $w \sim \mathcal{N}(0, C)$
3. set $\theta_{\text{out}} := \mathrm{shrink}(0, p_{x,w}^{-1}(G(t)))$
4. $X_{n+1} := p_{x,w}(\theta_{\text{out}})$



Elliptical slice sampling

input: current state $X_n = x \in \mathbb{R}^d$
output: next state X_{n+1}

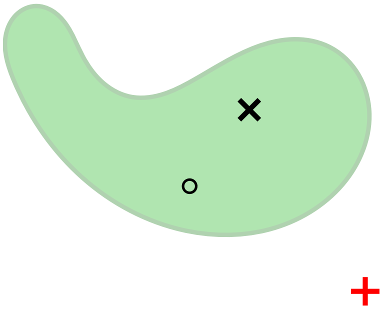
1. draw $t \sim \mathrm{Unif}(0, \rho(x))$
2. draw $w \sim \mathcal{N}(0, C)$
3. set $\theta_{\text{out}} := \mathrm{shrink}(0, p_{x,w}^{-1}(G(t)))$
4. $X_{n+1} := p_{x,w}(\theta_{\text{out}})$



Elliptical slice sampling

input: current state $X_n = x \in \mathbb{R}^d$
output: next state X_{n+1}

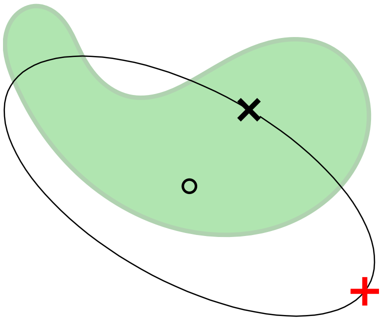
1. draw $t \sim \mathrm{Unif}(0, \rho(x))$
2. draw $w \sim \mathcal{N}(0, C)$
3. set $\theta_{\text{out}} := \mathrm{shrink}(0, p_{x,w}^{-1}(G(t)))$
4. $X_{n+1} := p_{x,w}(\theta_{\text{out}})$



Elliptical slice sampling

input: current state $X_n = x \in \mathbb{R}^d$
output: next state X_{n+1}

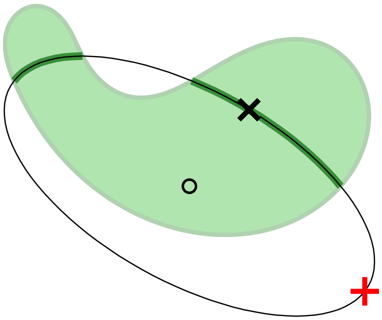
1. draw $t \sim \mathrm{Unif}(0, \rho(x))$
2. draw $w \sim \mathcal{N}(0, C)$
3. set $\theta_{\text{out}} := \mathrm{shrink}(0, p_{x,w}^{-1}(G(t)))$
4. $X_{n+1} := p_{x,w}(\theta_{\text{out}})$



Elliptical slice sampling

input: current state $X_n = x \in \mathbb{R}^d$
output: next state X_{n+1}

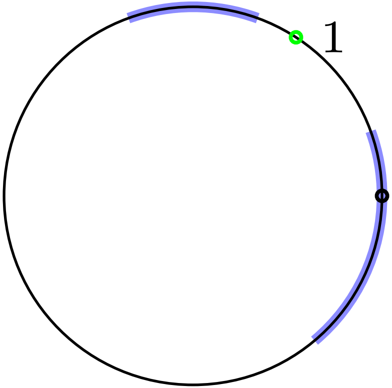
1. draw $t \sim \mathrm{Unif}(0, \rho(x))$
2. draw $w \sim \mathcal{N}(0, C)$
3. set $\theta_{\text{out}} := \mathrm{shrink}(0, p_{x,w}^{-1}(G(t)))$
4. $X_{n+1} := p_{x,w}(\theta_{\text{out}})$



shrink procedure within elliptical slice sampling

input: $S \in \mathcal{B}([0, 2\pi))$, $\theta_{\text{in}} \in S$;
output: $\theta_{\text{out}} \in S$

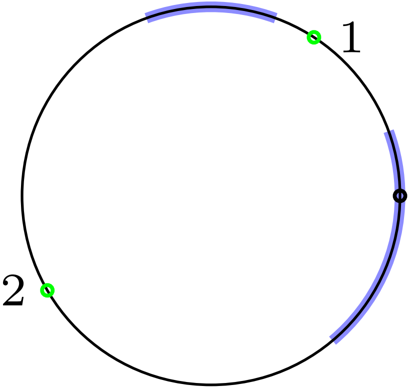
set $i := 1$, draw $\gamma_1 \sim \mathrm{Unif}[0, 2\pi)$;
set $\gamma_1^{\min} := \gamma_1$, $\gamma_1^{\max} := \gamma_1$;
while $\gamma_i \notin S$:
    if $\gamma_i \in I(\gamma_i^{\min}, \theta_{\text{in}})$:
        set $\gamma_{i+1}^{\min} := \gamma_i$ and $\gamma_{i+1}^{\max} := \gamma_i^{\max}$;
    else
        set $\gamma_{i+1}^{\min} := \gamma_i^{\min}$ and $\gamma_{i+1}^{\max} := \gamma_i$;
    draw $\gamma_{i+1} \sim \mathrm{Unif}(I(\gamma_{i+1}^{\min}, \gamma_{i+1}^{\max}))$;
    set $i := i + 1$;
return $\theta_{\text{out}} := \gamma_i$.



shrink procedure within elliptical slice sampling

input: $S \in \mathcal{B}([0, 2\pi))$, $\theta_{\text{in}} \in S$;
output: $\theta_{\text{out}} \in S$

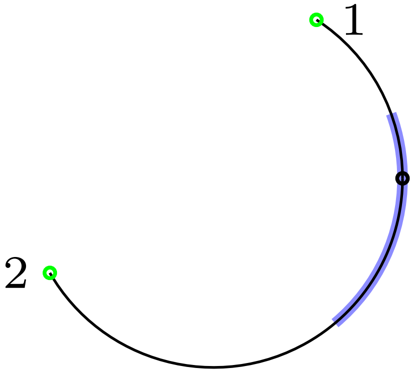
```
set $i := 1$, draw $\gamma_1 \sim \mathrm{Unif}[0, 2\pi)$;
set $\gamma_1^{\min} := \gamma_1$, $\gamma_1^{\max} := \gamma_1$;
while $\gamma_i \notin S$:
    if $\gamma_i \in I(\gamma_i^{\min}, \theta_{\text{in}})$:
        set $\gamma_{i+1}^{\min} := \gamma_i$ and $\gamma_{i+1}^{\max} := \gamma_i^{\max}$;
    else
        set $\gamma_{i+1}^{\min} := \gamma_i^{\min}$ and $\gamma_{i+1}^{\max} := \gamma_i$;
    draw $\gamma_{i+1} \sim \mathrm{Unif}(I(\gamma_{i+1}^{\min}, \gamma_{i+1}^{\max}))$;
    set $i := i + 1$;
return $\theta_{\text{out}} := \gamma_i$.
```



shrink procedure within elliptical slice sampling

input: $S \in \mathcal{B}([0, 2\pi))$, $\theta_{\text{in}} \in S$;
output: $\theta_{\text{out}} \in S$

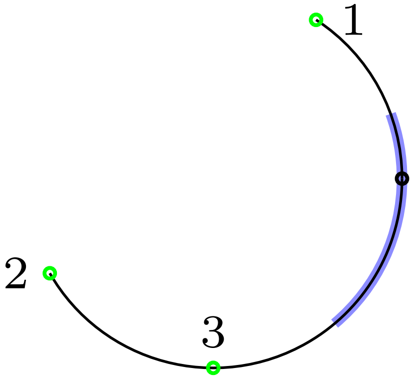
set $i := 1$, draw $\gamma_1 \sim \mathrm{Unif}[0, 2\pi)$;
set $\gamma_1^{\min} := \gamma_1$, $\gamma_1^{\max} := \gamma_1$;
while $\gamma_i \notin S$:
  if $\gamma_i \in I(\gamma_i^{\min}, \theta_{\text{in}})$:
    set $\gamma_{i+1}^{\min} := \gamma_i$ and $\gamma_{i+1}^{\max} := \gamma_i^{\max}$;
  else
    set $\gamma_{i+1}^{\min} := \gamma_i^{\min}$ and $\gamma_{i+1}^{\max} := \gamma_i$;
  draw $\gamma_{i+1} \sim \mathrm{Unif}(I(\gamma_{i+1}^{\min}, \gamma_{i+1}^{\max}))$;
  set $i := i + 1$;
return $\theta_{\text{out}} := \gamma_i$.



shrink procedure within elliptical slice sampling

input: $S \in \mathcal{B}([0, 2\pi))$, $\theta_{\text{in}} \in S$;
output: $\theta_{\text{out}} \in S$

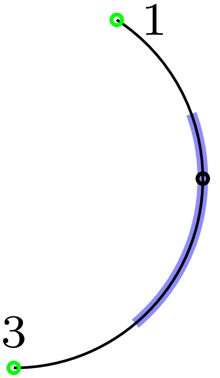
set $i := 1$, draw $\gamma_1 \sim \mathrm{Unif}[0, 2\pi)$;
set $\gamma_1^{\min} := \gamma_1$, $\gamma_1^{\max} := \gamma_1$;
while $\gamma_i \notin S$:
  if $\gamma_i \in I(\gamma_i^{\min}, \theta_{\text{in}})$:
    set $\gamma_{i+1}^{\min} := \gamma_i$ and $\gamma_{i+1}^{\max} := \gamma_i^{\max}$;
  else
    set $\gamma_{i+1}^{\min} := \gamma_i^{\min}$ and $\gamma_{i+1}^{\max} := \gamma_i$;
  draw $\gamma_{i+1} \sim \mathrm{Unif}(I(\gamma_{i+1}^{\min}, \gamma_{i+1}^{\max}))$;
  set $i := i + 1$;
return $\theta_{\text{out}} := \gamma_i$.



shrink procedure within elliptical slice sampling

input: $S \in \mathcal{B}([0, 2\pi))$, $\theta_{\text{in}} \in S$;
output: $\theta_{\text{out}} \in S$

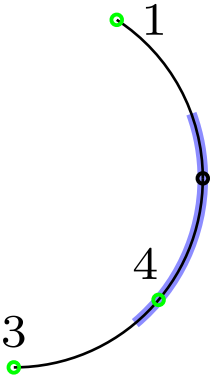
set $i := 1$, draw $\gamma_1 \sim \mathrm{Unif}[0, 2\pi)$;
set $\gamma_1^{\min} := \gamma_1$, $\gamma_1^{\max} := \gamma_1$;
while $\gamma_i \notin S$:
    if $\gamma_i \in I(\gamma_i^{\min}, \theta_{\text{in}})$:
        set $\gamma_{i+1}^{\min} := \gamma_i$ and $\gamma_{i+1}^{\max} := \gamma_i^{\max}$;
    else
        set $\gamma_{i+1}^{\min} := \gamma_i^{\min}$ and $\gamma_{i+1}^{\max} := \gamma_i$;
    draw $\gamma_{i+1} \sim \mathrm{Unif}(I(\gamma_{i+1}^{\min}, \gamma_{i+1}^{\max}))$;
    set $i := i + 1$;
return $\theta_{\text{out}} := \gamma_i$.



shrink procedure within elliptical slice sampling

input: $S \in \mathcal{B}([0, 2\pi))$, $\theta_{\text{in}} \in S$;
output: $\theta_{\text{out}} \in S$

set $i := 1$, draw $\gamma_1 \sim \mathrm{Unif}[0, 2\pi)$;
set $\gamma_1^{\min} := \gamma_1$, $\gamma_1^{\max} := \gamma_1$;
while $\gamma_i \notin S$:
  if $\gamma_i \in I(\gamma_i^{\min}, \theta_{\text{in}})$:
    set $\gamma_{i+1}^{\min} := \gamma_i$ and $\gamma_{i+1}^{\max} := \gamma_i^{\max}$;
  else
    set $\gamma_{i+1}^{\min} := \gamma_i^{\min}$ and $\gamma_{i+1}^{\max} := \gamma_i$;
  draw $\gamma_{i+1} \sim \mathrm{Unif}(I(\gamma_{i+1}^{\min}, \gamma_{i+1}^{\max}))$;
  set $i := i + 1$;
return $\theta_{\text{out}} := \gamma_i$.



Shrinkage procedure within elliptical slice sampling

For $i \in \mathbb{N}$ let $(\gamma_i, \gamma_i^{\text{min}}, \gamma_i^{\text{max}}) \in [0, 2\pi)^3$ be realization of random vector $(\Gamma_i, \Gamma_i^{\text{min}}, \Gamma_i^{\text{max}})$.

Define stopping time

$$\tau_S := \inf \left\{ n \in \mathbb{N} : (\Gamma_n, \Gamma_n^{\text{min}}, \Gamma_n^{\text{max}}) \in S \times [0, 2\pi)^2 \right\}$$

Lemma (Natarovskii, Hasenplug, R.)

For ρ lower semi-continuous, i.e., all level sets $G(t)$ are open, we have

$$\mathbb{P}(\tau_{p_{x,w}^{-1}(G(t))} < \infty \mid \theta_{\text{in}} = 0) = 1, \qquad \forall x, w \in \mathbb{R}^d \quad \text{and} \quad t \in (0, \rho(x)).$$

Shrinkage procedure within elliptical slice sampling

Lemma (Natarovskii, Hasenplug, R.)

For open $S \in \mathcal{B}([0, 2\pi))$ and $\theta \in S$, $A \in \mathcal{B}(S)$ define transition kernel

$$Q(\theta, A) := \mathbb{P}(\Gamma_{\tau_S} \in A, \tau_S < \infty \mid \theta_{\text{in}} = \theta),$$

corresponding to the shrinkage procedure.

Then, Q is reversible w.r.t. Unif(S).

For ρ lower semi-cont., this implies reversibility of elliptical slice sampling w.r.t. μ.

Convergence result

Theorem (Natarovskii, R., Sprungk)

Assume

- ρ lower semi-continuous;
- ρ is bounded away from 0 and ∞ on any compact set of $\mathbb{R}^d$;
- $\exists r > 0, R < \infty$, such that $\forall x \in \mathbb{R}^d$ with $\|x\| > R$ holds

$$\inf_{y:\ \|y\| \leq r\|x\|} \rho(y) \geq \rho(x).$$

Then, $\exists K > 0, \exists \alpha \in (0, 1)$, such that

$$\|\mathbb{P}(X_n \in \cdot \mid X_1 = x) - \mu\|_{\text{tv}} \leq K\alpha^n(1 + \|x\|), \qquad \forall x \in \mathbb{R}^d.$$

(Assumption e.g. satisfied in Gaussian setting, for multivariate t-distribution, logistic regression)

References

- M. Hasenpflug, V. Natarovskii, D. Rudolf, *Reversibility of elliptical slice sampling revisited*, 2023, arxiv
- D. MacKay, *Information theory, inference and learning algorithms*, Cambridge University press, (2003)
- I. Murray, R. Adams, and D. MacKay, *Elliptical slice sampling*, Journal of Machine Learning Research: W&CP, (2010)
- V. Natarovskii, D. Rudolf, B. Sprungk, *Geometric convergence of elliptical slice sampling*, Proc. 38th Internat. Conf. Mach. Learn. (2021)
- R. Neal, *Slice sampling*, Ann. Statist. (2003)

Motivation: $\rho(x) = \exp(\|x\|)$ and reference measure $\mathcal{N}(0, I)$