Proceedings of Machine Learning Research vol 65:1–23, 2017

Sampling from a log-concave distribution with compact support with proximal Langevin Monte Carlo

Nicolas Brosse NICOLAS.BROSSE@POLYTECHNIQUE.EDU
Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.

Alain Durmus ALAIN.DURMUS@TELECOM-PARISTECH.FR
LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France.

Éric Moulines ERIC.MOULINES@POLYTECHNIQUE.EDU
Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.

Marcelo Pereyra M.PEREYRA@HW.AC.UK
School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, U.K.

Abstract

This paper presents a detailed theoretical analysis of the Langevin Monte Carlo sampling algorithm recently introduced in Durmus et al. (2016) when applied to log-concave probability distributions that are restricted to a convex body K. This method relies on a regularisation procedure involving the Moreau-Yosida envelope of the indicator function associated with K. Explicit convergence bounds in total variation norm and in Wasserstein distance of order 1 are established. In particular, we show that the complexity of this algorithm given a first order oracle is polynomial in the dimension of the state space. Finally, some numerical experiments are presented to compare our method with competing MCMC approaches from the literature.

Keywords: Markov chain Monte Carlo methods; Langevin Algorithm; Bayesian inference; convex body

1. Introduction

Many statistical inference problems involve estimating parameters subject to constraints on the parameter space. In a Bayesian setting, these constraints define a posterior distribution π with bounded support. Some examples include truncated data problems which arise naturally in failure and survival time studies Klein and Moeschberger (2005), ordinal data models Johnson and Albert (2006), constrained lasso and ridge regressions Celeux et al. (2012), Latent Dirichlet Allocation Blei et al. (2003), and non-negative matrix factorization Paisley et al. (2014). Drawing samples from such constrained distributions is a challenging problem that has been investigated in many papers; see Gelfand et al. (1992), Pakman and Paninski (2014), Lan and Shahbaba (2015), Bubeck et al. (2015). All these works are based on efficient Markov Chain Monte Carlo methods to approximate the posterior distribution; however, with the exception of the recent work Bubeck et al. (2015), these methods are not theoretically well understood and do not provide any theoretical guarantees on the estimations delivered.

Recently a new MCMC method has been proposed in Durmus et al. (2016) to sample from a non-smooth log-concave probability distribution on $\mathbb{R}^d$. This method is mainly based on a carefully designed regularised version of the target distribution π that enjoys a number of favourable properties that are useful for MCMC simulation. In this study, we analyse the complexity of this algorithm when applied to log-concave distributions constrained to a convex set, with a focus on complexity as the dimension of the state space increases. More precisely, we establish explicit bounds in total variation norm and in Wasserstein distance of order 1 between the iterates of the Markov kernel defined by the algorithm and the target density π.

The paper is organised as follows. Section 2.1 introduces the MCMC method of Durmus et al. (2016). The main complexity result is stated in Section 2.2 and compared to previous works on the subject. The proof of this result is presented in Section 3 and Section 4. The methodology is then illustrated and compared to other approaches via experiments in Section 5. Proofs are finally reported in Section 6.

2. The Moreau-Yosida Unadjusted Langevin Algorithm (MYULA)

2.1. Presentation of MYULA

Let π be a probability measure on $\mathbb{R}^d$ with density w.r.t. the Lebesgue measure given for all $x \in \mathbb{R}^d$ by $\pi(x) = \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$, where $U : \mathbb{R}^d \to (-\infty, +\infty]$ is a measurable function. In the sequel, U will be referred to as the potential associated with π. Assume for the moment that U is continuously differentiable. Then, the unadjusted Langevin algorithm (ULA) introduced in Parisi (1981) (see also Roberts and Tweedie (1996)) can be used to sample from π. This algorithm is based on the overdamped Langevin stochastic differential equation (SDE) associated with U,

$$\mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \ , \tag{1}$$

where $(B_t)_{t \geq 0}$ is a d-dimensional Brownian motion. Under mild assumptions on ∇U, this SDE has a unique strong solution $(Y_t)_{t \geq 0}$ and defines a strong Markovian semigroup $(P_t)_{t \geq 0}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ which is ergodic with respect to π, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-field on $\mathbb{R}^d$. Since simulating exact solutions of (1) is in general computationally impossible or very hard, ULA considers the Euler-Maruyama discretization associated with (1) to approximate samples from π. Precisely, ULA constructs the discrete-time Markov chain $(X_k)_{k \geq 0}$, started at X_0, given for $k \in \mathbb{N}$ by:

$$X_{k+1} = X_k - \gamma \nabla U(X_k) + \sqrt{2\gamma} Z_{k+1} \ ,$$

where $\gamma > 0$ is the stepsize and $(Z_k)_{k \in \mathbb{N}}$ is a sequence of i.i.d. standard Gaussian d-dimensional vectors; the process $(X_k)_{k \geq 0}$ is used as approximate samples from π. However, the ULA algorithm cannot be directly applied to a distribution π restricted to a compact convex set. Let $\mathsf{K} \subset \mathbb{R}^d$ be a convex body, i.e. a compact convex set with non-empty interior and $\iota_{\mathsf{K}} : \mathbb{R}^d \to \{0, +\infty\}$ be the (convex) indicator function of K, defined for $x \in \mathbb{R}^d$ by,

$$\iota_{\mathsf{K}}(x) = \begin{cases} +\infty & \text{if } x \notin \mathsf{K}, \\ 0 & \text{if } x \in \mathsf{K} \ . \end{cases}$$

Let $f : \mathbb{R}^d \to \mathbb{R}$. In this paper we consider any probability density π associated to a potential $U : \mathbb{R}^d \to (-\infty, +\infty]$ of the form

$$U = f + \iota_{\mathsf{K}} \ , \tag{2}$$

and assume that the function f and the convex body K satisfy the following assumptions. For $x \in \mathbb{R}^d$ and $r > 0$, denote by $\mathrm{B}(x, r)$ the closed ball of center x and radius r: $\mathrm{B}(x, r) = \{y \in \mathbb{R}^d : \|y - x\| \leq r\}$.

H1

(i) *f is convex.*

(ii) *f is continuously differentiable on $\mathbb{R}^d$ and gradient Lipschitz with Lipschitz constant L_f, i.e. for all $x, y \in \mathbb{R}^d$*

$$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\| \ . \tag{3}$$

H2 *There exist $r, R > 0$, $r \leq R$, such that,*

$$\mathrm{B}(0, r) \subset \mathsf{K} \subset \mathrm{B}(0, R) \ .$$

To apply ULA, Durmus et al. (2016) suggested to carefully regularize U in such a way that 1) the convexity of U is preserved (this property is key to the theoretical analysis of the algorithm), 2) the regularisation of U is continuously differentiable and gradient Lipschitz (this regularity property is key to the algorithm's stability), and 3) the resulting approximation is close to π (e.g. in total variation norm). The tool used to construct such an approximation is the Moreau-Yosida envelope of ι_{K}, $\iota_{\mathsf{K}}^{\lambda} : \mathbb{R}^d \to \mathbb{R}_+$ defined for $x \in \mathbb{R}^d$ (see e.g. (Rockafellar and Wets, 1998, Chapter 1 Section G)) by,

$$\iota_{\mathsf{K}}^{\lambda}(x) = \inf_{y \in \mathbb{R}^d} \left(\iota_{\mathsf{K}}(y) + (2\lambda)^{-1} \|x - y\|^2 \right) = (2\lambda)^{-1} \|x - \operatorname{proj}_{\mathsf{K}}(x)\|^2 \ , \tag{4}$$

where $\lambda > 0$ is a regularization parameter and $\operatorname{proj}_{\mathsf{K}}$ is the projection onto K. By (Rockafellar and Wets, 1998, Example 10.32, Theorem 9.18), the function $\iota_{\mathsf{K}}^{\lambda}$ is convex and continuously differentiable with gradient given for all $x \in \mathbb{R}^d$ by:

$$\nabla \iota_{\mathsf{K}}^{\lambda}(x) = \lambda^{-1}(x - \operatorname{proj}_{\mathsf{K}}(x)) \ . \tag{5}$$

Moreover, (Rockafellar and Wets, 1998, Proposition 12.19) implies that $\iota_{\mathsf{K}}^{\lambda}$ is λ^{-1}-gradient Lipschitz: for all $x, y \in \mathbb{R}^d$,

$$\left\| \nabla \iota_{\mathsf{K}}^{\lambda}(x) - \nabla \iota_{\mathsf{K}}^{\lambda}(y) \right\| \leq \lambda^{-1} \|x - y\| \ . \tag{6}$$

Adding f to $\iota_{\mathsf{K}}^{\lambda}$ under **H1** leads to the regularization $U^{\lambda} : \mathbb{R}^d \to \mathbb{R}$ of the potential U defined for all $x \in \mathbb{R}^d$ by

$$U^{\lambda}(x) = f(x) + \iota_{\mathsf{K}}^{\lambda}(x) \ . \tag{7}$$

The following lemma shows that the probability measure π^{λ} on $\mathbb{R}^d$, with density with respect to the Lebesgue measure, also denoted by π^{λ} and given for all $x \in \mathbb{R}^d$ by

$$\pi^{\lambda}(x) = \frac{\mathrm{e}^{-U^{\lambda}(x)}}{\int_{\mathbb{R}^d} \mathrm{e}^{-U^{\lambda}(s)} \mathrm{d}s} \ , \tag{8}$$

is well defined. It also shows that U^{λ} has a minimizer $x^{\star} \in \mathbb{R}^d$, a fact that will be used in Section 4.

Lemma 1 *Assume **H1**-(i) and **H2**. For all $\lambda > 0$,*

a) U^{λ} has a minimizer $x^{\star} \in \mathbb{R}^d$, i.e. for all $x \in \mathbb{R}^d$, $U^{\lambda}(x) \geq U^{\lambda}(x^{\star})$.

b) e^{-U^λ} *defines a proper density of a probability measure on* $\mathbb{R}^d$*, i.e.*

$$0 < \int_{\mathbb{R}^d} \mathrm{e}^{-U^\lambda(y)} \mathrm{d}y < +\infty \, .$$

Proof Note that (Durmus et al., 2016, Proposition 1) provides a proof in a more general case. Given the specific form of U^λ, a short and self-contained proof can be found in Section 6.1. ■

Under **H**1, for all $\lambda > 0$, π^λ is log-concave and U^λ is continuously differentiable by (5), with ∇U^λ given for all $x \in \mathbb{R}^d$ by

$$\nabla U^\lambda(x) = -\nabla \log \pi^\lambda(x) = \nabla f(x) + \lambda^{-1}(x - \mathrm{proj}_{\mathsf{K}}(x)) \, . \tag{9}$$

In addition, by (6), ∇U^λ is Lipschitz with constant $L \leq L_f + \lambda^{-1}$. Since U^λ is continuously differentiable, ULA is well defined. The algorithm proposed in Durmus et al. (2016) then proceeds by using the Euler-Maruyama discretization of the Langevin equation associated with U^λ, with π^λ as proxy, to generate approximate samples from π. Precisely, it uses the Markov chain $(X_k)_{k\in\mathbb{N}}$, started at X_0, given for all $k \in \mathbb{N}$ by

$$X_{k+1} = (1 - \tfrac{\gamma}{\lambda})X_k - \gamma \nabla f(X_k) + \tfrac{\gamma}{\lambda}\, \mathrm{proj}_{\mathsf{K}}(X_k) + \sqrt{2\gamma} Z_{k+1} \, , \tag{10}$$

where $(Z_k)_{k\in\mathbb{N}}$ is a sequence of i.i.d. standard Gaussian d-dimensional vectors and $\gamma > 0$ is the stepsize. Note that one iteration (10) requires a projection onto the convex body K and the evaluation of ∇f. The kernel of the homogeneous Markov chain defined by (10) is given for $x \in \mathbb{R}^d$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by,

$$R_\gamma(x, \mathsf{A}) = (4\pi\gamma)^{-d/2} \int_{\mathsf{A}} \exp\left(-(4\gamma)^{-1} \left\| y - x + \gamma \nabla U^\lambda(x) \right\|^2 \right) \mathrm{d}y \, , \tag{11}$$

where U^λ is defined in (7). Since the target density for the Markov chain (10) is the regularized measure π^λ and not π, the algorithm is named the Moreau-Yosida regularized Unadjusted Langevin Algorithm (MYULA).

2.2. Context and contributions

The total variation distance between two probability measures μ and ν is defined by $\|\mu - \nu\|_{\mathrm{TV}} = 2\sup_{\mathsf{A}\in\mathcal{B}(\mathbb{R}^d)} |\mu(\mathsf{A}) - \nu(\mathsf{A})|$. Let $\phi, \psi : \mathbb{R}_+ \to \mathbb{R}_+$. Denote by $\phi = \tilde{\mathcal{O}}(\psi)$ or $\phi = \tilde{\Omega}(\psi)$ if there exist $C, c \geq 0$ such that for all $t \in \mathbb{R}_+$ $\phi(t) \leq C\psi(t)(\log t)^c$ or $\phi(t) \geq C\psi(t)(\log t)^c$ respectively. Our main result is the following:

Theorem 2 *Assume* ***H****1 and* ***H****2. For all* $\varepsilon > 0$ *and* $x \in \mathbb{R}^d$*, there exist* $\lambda > 0$ *and* $\gamma \in \left(0, \lambda(1 + L_f^2\lambda^2)^{-1}\right)$ *such that,*

$$\|\delta_x R_\gamma^n - \pi\|_{\mathrm{TV}} \leq \varepsilon \quad \textit{for} \quad n = \tilde{\Omega}(d^5) \, ,$$

where R_γ *is defined in* (11)*.*

The proof of Theorem 2 follows from combining Proposition 6 and Proposition 4 below. Note that these two results imply explicit bounds between R_γ^n and π for all $n \in \mathbb{N}$ and $\gamma > 0$.

The problem of sampling from a probability measure restricted to a convex compact support has been investigated in several works, mainly in the fields of theoretical computer science and Bayesian statistics. In computer science, a line of works starting with Dyer and Frieze (1991) has studied the convergence of the ball walk and the hit-and-run algorithm towards the uniform density on a convex body K, or more generally to a log-concave density. The best complexity result is achieved by (Lovász and Vempala, 2007, Theorem 2.1) who establishes a mixing time for these two algorithms of order $\tilde{\mathcal{O}}(d^4)$. However, observe that contrary to Theorem 2, this result assumes that π is in near-isotropic position, i.e. there exists $C \in \mathbb{R}_+^*$ such that for all $u \in \mathbb{R}^d$, $\|u\| = 1$,

$$C^{-1} \leq \int_{\mathbb{R}^d} \langle u, x \rangle^2 \, \pi(\mathrm{d}x) \leq C \; . \tag{12}$$

Note that (Lovász and Vempala, 2007, Section 2.5) gives also an algorithm of complexity $\tilde{\mathcal{O}}(d^5)$ which provides an invertible linear map T of $\mathbb{R}^d$ such that the measure π_T defined for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$\pi_T(\mathsf{A}) = \pi(T^{-1}(\mathsf{A})) \; ,$$

is log-concave and near-isotropic. Also note that, unlike our method, each iteration of the ball walk or the hit-and-run algorithm requires a call to a zero-order oracle, which given $x \in \mathbb{R}^d$, returns the value $U(x)$. MYULA does not require to fulfill the condition (12) and is thus dispensed of preprocessing step. However, MYULA needs a first-order oracle which returns the value $\nabla f(x)$ for $x \in \mathbb{R}^d$.

As emphasized in the introdution, probability distributions with convex compact supports or more generally with constrained parameters arise naturally in Bayesian statistics. Gelfand et al. (1992) includes many examples of such problems and suggests to use a Gibbs sampler, see also Rodriguez-Yam et al. (2004). (Chen et al., 2012, Chapter 6) addresses the subject with the additional difficulty of computing normalizing constants. Recently, Pakman and Paninski (2014) adapted the Hamiltonian Monte Carlo method to sample from a truncated multivariate gaussian, and Lan and Shahbaba (2015) suggested a new approach which consists in mapping the constrained domain to a sphere in an augmented space. However, these methods are not well understood from a theoretical viewpoint, and do not provide any theoretical guarantees for the estimations delivered.

Concerning the ULA algorithm, when U is continuously differentiable, the first explicit convergence bounds have been obtained by Dalalyan (2016), Durmus and Moulines (2015), Durmus and Moulines (2016). In the constrained case $U = f + \iota_{\mathsf{K}}$, Bubeck et al. (2015) suggests a projection step in ULA i.e. to consider the Markov chain $(\tilde{X}_k)_{k \geq 0}$, defined for all $k \in \mathbb{N}$ by

$$\tilde{X}_{k+1} = \mathrm{proj}_{\mathsf{K}} \left(\tilde{X}_k - \gamma \nabla U(\tilde{X}_k) + \sqrt{2\gamma} Z_{k+1} \right) \; . \tag{13}$$

with $\tilde{X}_0 = 0$. This method is referred to as the Projected Langevin Monte Carlo (PLMC) algorithm. As in MYULA, one iteration of PLMC requires a projection onto K and an evaluation of ∇f. Let $\tilde{R}_\gamma$ be the Markov kernel defined by (13). Bubeck et al. (2015) proved that for all $\varepsilon > 0$, $\|\delta_0 \tilde{R}_\gamma^n - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $n = \tilde{\Omega}(d^7)$ if π is the uniform density on K and $n = \tilde{\Omega}(d^{12})$ if π is a log-concave density. Theorem 2 improves these bounds for the MYULA algorithm. Note however that the iterations of PLMC stay within the constraint set K and this property can be useful in some specific problems. Nevertheless, there is a wide range of settings where this property is not particularly beneficial, for example in the case of the computation of volumes discussed in Section 5, or in Bayesian model selection where it is necessary to estimate marginal likelihoods.

3. Distance between π and π^λ

In this section, we derive bounds between π and π^λ in total variation and in Wasserstein distance (recall that π is associated with a potential of the form (2) and π^λ is given by (8)). It is shown that the approximation error in both distances can be made arbitrarily small by adjusting the regularisation parameter λ.

The main quantity of interest to analyze the distance between π and π^λ will appear to be the integral of $x \mapsto \mathrm{e}^{-(2\lambda)^{-1}\|x-\mathrm{proj}_{\mathsf{K}}(x)\|^2}$ over $\mathbb{R}^d$. This constant is linked to useful notions borrowed from the field of convex geometry (Kampf, 2009, Proposition 3). Indeed, Fubini's theorem gives the following equality:

$$\begin{aligned}\int_{\mathbb{R}^d} \mathrm{e}^{-(2\lambda)^{-1}\|x-\mathrm{proj}_{\mathsf{K}}(x)\|^2}\mathrm{d}x &= \int_{\mathbb{R}_+}\int_{\mathbb{R}^d} \mathbb{1}_{[\|x-\mathrm{proj}_{\mathsf{K}}(x)\|,+\infty)}(t)\lambda^{-1}t\mathrm{e}^{-t^2/(2\lambda)}\mathrm{d}x\mathrm{d}t \;, \\ &= \int_{\mathbb{R}_+} \mathrm{Vol}\left(\mathsf{K}+\mathrm{B}(0,t)\right)\lambda^{-1}t\mathrm{e}^{-t^2/(2\lambda)}\mathrm{d}t \;, \end{aligned} \tag{14}$$

where $\mathsf{A}+\mathsf{B}$ is the Minkowski sum of $\mathsf{A},\mathsf{B} \subset \mathbb{R}^d$, i.e. $\mathsf{A}+\mathsf{B} = \{x+y : x \in \mathsf{A}, y \in \mathsf{B}\}$, and we have used in the last line that for all $t \in \mathbb{R}_+$, $\mathsf{K}+\mathrm{B}(0,t) = \{x \in \mathbb{R}^d \,:\, \|x-\mathrm{proj}_{\mathsf{K}}(x)\| \leq t\}$. It turns out that $t \mapsto \mathrm{Vol}(\mathsf{K}+\mathrm{B}(0,t))$ on $\mathbb{R}_+$ is a polynomial. More precisely, Steiner's formula states that for all $t \geq 0$,

$$\mathrm{Vol}(\mathsf{K}+\mathrm{B}(0,t)) = \sum_{i=0}^{d} t^i \kappa_i \mathscr{V}_{d-i}(\mathsf{K}) \;, \tag{15}$$

where $\{\mathscr{V}_i(\mathsf{K})\}_{0\leq i\leq d}$ are the intrinsic volumes of K, κ_i denotes the volume of the unit ball in $\mathbb{R}^i$, i.e.

$$\kappa_i = \pi^{i/2}/\mathbf{\Gamma}(1+i/2) \;, \tag{16}$$

and $\mathbf{\Gamma} : \mathbb{R}_+^* \to \mathbb{R}_+^*$ is the Gamma function. We refer to (Schneider, 2013, Chapter 4.2) for this result and an introduction to this topic. Combining (14) and (15) gives:

$$\int_{\mathbb{R}^d} \mathrm{e}^{-(2\lambda)^{-1}\|x-\mathrm{proj}_{\mathsf{K}}(x)\|^2}\mathrm{d}x = \sum_{i=0}^{d} \mathscr{V}_i(\mathsf{K})(2\pi\lambda)^{(d-i)/2} \;. \tag{17}$$

This expression will provide a precise analysis of the distance in total variation and Wasserstein distance between π and π^λ, in particular when π is the uniform density on K. However, in more general cases, an additional assumption on the relation between f and K is necessary to bound the distance between π and π^λ. Under **H**1-(i) and **H**2, f has a minimum x_{K} on K. Define

$$\widetilde{\mathsf{K}} = \{x \in \mathsf{K} \mid \mathrm{B}(x,r) \subset \mathsf{K}\} \;. \tag{18}$$

$\widetilde{\mathsf{K}}$ has the following property.

Lemma 3 *Assume* ***H*2**. *$\widetilde{\mathsf{K}}$ is a non-empty convex compact set.*

Proof The proof is postponed to Section 6.2. ■

H3

(i) There exists $\Delta_1 > 0$ *such that* $\exp\left(\inf_{\mathsf{K}^c}(f) - \max_{\mathsf{K}}(f)\right) \geq \Delta_1$.

(ii) There exists $\Delta_2 \geq 0$ *such that* $0 \leq f(\operatorname{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})) - f(x_{\mathsf{K}}) \leq \Delta_2$.

Under **H**3-(i), the application of Steiner's formula is possible and reveals the precise dependence of the bounds with respect to the intrinsic volumes of K. A complementary view is possible under **H**3-(ii). The obtained bounds are less precise regarding K but more robust with respect to f. Note that if $x_{\mathsf{K}} \in \widetilde{\mathsf{K}}$, Δ_2 can be chosen equal to 0. On the other hand, if f is assumed to be ℓ-Lipschitz inside K, Δ_2 is less than ℓR.

Proposition 4 *Assume* ***H****1-(i) and* ***H****2.*

a) Assume ***H****3-(i). For all* $\lambda > 0$,

$$\|\pi^\lambda - \pi\|_{\mathrm{TV}} \leq 2\left(1 + \Delta_1 \mathsf{D}(\mathsf{K},\lambda)^{-1}\right)^{-1} , \tag{19}$$

where,

$$\mathsf{D}(\mathsf{K},\lambda) = (\operatorname{Vol}\mathsf{K})^{-1}\sum_{i=0}^{d-1}(2\uppi\lambda)^{(d-i)/2}\mathscr{V}_i(\mathsf{K}) , \tag{20}$$

and $\mathscr{V}_i(\mathsf{K})$ *are defined in (15).*

b) In addition, assuming ***H****3-(i), for all* $\lambda \in \left(0, (2\uppi)^{-1}(r/d)^2\right)$,

$$\|\pi^\lambda - \pi\|_{\mathrm{TV}} \leq 2^{3/2}\Delta_1^{-1}(\uppi\lambda)^{1/2} d r^{-1} . \tag{21}$$

c) Assume ***H****3-(ii). For all* $\lambda \in \left(0, 16^{-1}(r/d)^2\right]$,

$$\|\pi^\lambda - \pi\|_{\mathrm{TV}} \leq (4/r)\exp\left(4\lambda\left(\Delta_2/r\right)^2\right)\left\{\sqrt{\lambda}(d+\Delta_2) + (2\lambda\Delta_2)/r\right\} . \tag{22}$$

Proof The proof is postponed to Section 6.3. ■

In the particular case where $f = 0$ and π is the uniform density on K, Δ_1 equals 1 and the inequality (19) is in fact an equality. The dependence of the upper bound in (19) w.r.t. to λ, d, r is sharp. Indeed, for the cube C of side c, $\mathsf{D}(\mathsf{C},\lambda)$ can be explicitly computed. (Klain and Rota, 1997, Theorem 4.2.1) gives for $i \in \{0,\ldots,d\}$, $\mathscr{V}_i(\mathsf{C}) = \binom{d}{i}c^i$, which implies:

$$\mathsf{D}(\mathsf{C},\lambda) = \left(1 + c^{-1}\sqrt{2\uppi\lambda}\right)^d - 1 ,$$

$$\|\pi^\lambda - \pi\|_{\mathrm{TV}} = 2\left\{1 - \left(1 + c^{-1}\sqrt{2\uppi\lambda}\right)^{-d}\right\} , \text{ for } U = \iota_{\mathsf{C}} .$$

For two probability measures μ and ν on $\mathcal{B}(\mathbb{R}^d)$, the Wasserstein distance of order $p \in \mathbb{N}^*$ between μ and ν is defined by

$$W_p(\mu, \nu) = \left(\inf_{\zeta \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \, \mathrm{d}\zeta(x, y) \right)^{1/p} ,$$

where $\Pi(\mu, \nu)$ is the set of transference plans of μ and ν. ζ is a transference plan of μ and ν if it is a probability measure on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ such that for all $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$, $\zeta(\mathsf{A} \times \mathbb{R}^d) = \mu(\mathsf{A})$ and $\zeta(\mathbb{R}^d \times \mathsf{A}) = \nu(\mathsf{A})$.

Proposition 5 *Assume **H**1-(i) and **H**2.*

*a) Assume **H**3-(i). For all $\lambda > 0$,*

$$W_1(\pi, \pi^\lambda) \leq \Delta_1^{-1} \mathsf{E}(\mathsf{K}, \lambda, R) ,$$

where

$$\mathsf{E}(\mathsf{K}, \lambda, R) = (\mathrm{Vol}(\mathsf{K}))^{-1} \sum_{i=0}^{d-1} \mathscr{V}_i(\mathsf{K}) \, (2\pi\lambda)^{(d-i)/2} \left\{ 2R + [\lambda(d - i + 2)]^{1/2} \right\} ,$$

and $\mathscr{V}_i(\mathsf{K})$ are defined in (15).

*b) In addition, assuming **H**3-(i), for all $\lambda \in \left(0, (2\pi)^{-1} d^{-2} r^2\right)$,*

$$W_1(\pi, \pi^\lambda) \leq \Delta_1^{-1} (2\pi\lambda)^{1/2} d r^{-1} \left(2R + r \left(3/(2d\pi)\right)^{1/2} \right) .$$

*c) Assume **H**3-(ii). For all $\lambda \in \left(0, 16^{-1}(r/d)^2\right]$,*

$$W_1(\pi, \pi^\lambda) \leq 4 \exp\left(4\lambda \left(\Delta_2/r\right)^2 \right) \left\{ \sqrt{\lambda}(d + \Delta_2)(R/r) + (2\lambda\Delta_2 R)/r^2 + \sqrt{\pi\lambda} \right\} .$$

Proof The proof is postponed to Section 6.4. ■

Note that the bounds in Wasserstein distance between π and π^λ are roughly similar to those obtained in total variation norm.

4. Convergence analysis of MYULA

We now analyse the convergence of the Markov kernel R_γ, given by (11), to the target density π^λ defined in (8). For $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$, explicit bounds in total variation norm and in Wasserstein distance between $\delta_x R_\gamma^n$ and π^λ are provided in Proposition 6 and Proposition 7. Because of the regularisation procedure performed in Section 2.1, the convergence analysis of MYULA (10) is an application of results of Durmus and Moulines (2015) and Durmus and Moulines (2016).

4.1. Convergence in total variation norm

Define $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ for all $r \geq 0$ by

$$\omega(r) = r^2 / \left\{2\mathbf{\Phi}^{-1}(3/4)\right\}^2 \ , \tag{23}$$

where $\mathbf{\Phi}(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2} \mathrm{d}t$.

Proposition 6 *Assume **H**1 and **H**2. Let $\lambda > 0$, L be the Lipschitz constant of ∇U^λ defined in (7) and $\bar{\gamma} \in \left(0, \lambda^{-1} L^{-2}\right)$. Then for all $\varepsilon > 0$ and $x \in \mathbb{R}^d$, we get:*

$$\|\delta_x R_\gamma^n - \pi^\lambda\|_{\mathrm{TV}} \leq \varepsilon \ , \tag{24}$$

provided that $n > T\gamma^{-1}$ with

$$T = (\log\{A_2(x)\} - \log(\varepsilon/2)) \Big/ (-\log(\kappa)) \ , \tag{25a}$$

$$\gamma \leq \frac{-d + \sqrt{d^2 + (2/3) A_1(x) \varepsilon^2 (L^2 T)^{-1}}}{2A_1(x)/3} \wedge \bar{\gamma} \ , \tag{25b}$$

where

$$A_1(x) = L^2 \left(\|x - x^\star\|^2 + 2(d + 8\lambda^{-1} R^2) \mathrm{e}^{\gamma(\lambda^{-1} - \bar{\gamma} L^2)} (\lambda^{-1} - \bar{\gamma} L^2)^{-1} \right) \ ,$$
$$\log(\kappa) = -\log(2)(4\lambda)^{-1} \left[\log\left\{ \left(1 + \mathrm{e}^{(8\lambda)^{-1} \omega\{\max(1, 4R)\}}\right) (1 + \max(1, 4R)) \right\} + \log(2) \right]^{-1} \ ,$$
$$A_2(x) = 6 + 2^{3/2} \left(d\lambda + 8R^2\right)^{1/2} + 2(A_1(x)/L^2)^{1/2} \ ,$$

and $x^\star$ is a minimizer of U^λ.

Proof To apply (Durmus and Moulines, 2015, Theorem 21), it is sufficient to check the assumption (Durmus and Moulines, 2015, H3), i.e. there exist $\tilde{R} \geq 0$ and $m > 0$ such that for all $x, y \in \mathbb{R}^d$, $\|x - y\| \geq \tilde{R}$,

$$\left\langle \nabla U^\lambda(x) - \nabla U^\lambda(y), x - y \right\rangle \geq m \|x - y\|^2 \ . \tag{26}$$

By (5) and the Cauchy-Schwarz inequality, we have:

$$\left\langle \nabla \iota_{\mathsf{K}}^\lambda(x) - \nabla \iota_{\mathsf{K}}^\lambda(y), x - y \right\rangle \geq \lambda^{-1} \left(\|x - y\|^2 - 2\left\{ \sup_{z \in \mathsf{K}} \|z\| \right\} \|x - y\| \right) \ ,$$

which implies under **H**1-(i) and **H**2 that (26) holds for $\tilde{R} = 4R$ and $m = (2\lambda)^{-1}$. ■

Combining Proposition 4 and Proposition 6 determines the stepsize γ and the number of samples n to get $\|\delta_{x^\star} R_\gamma^n - \pi\|_{\mathrm{TV}} \leq \varepsilon$. λ is chosen of order $\varepsilon^2 r^2 d^{-2} \Delta_1^2$ under **H**3-(i) and $\varepsilon^2 r^2 \min(d^{-2}, \Delta_2^{-2})$ under **H**3-(ii). The orders of magnitude of n in d, ε, R, r are reported in Table 1, along with the results of Bubeck et al. (2015). The dependency of n towards Δ_1, Δ_2 is presented in Table 2. A detailed table is provided in Appendix A.

Upper bound on n to get $\|\delta_{x^\star} R_\gamma^n - \pi\|_{\mathrm{TV}} \leq \varepsilon$	$d \to +\infty$	$\varepsilon \to 0$	$R \to +\infty$	$r \to 0$
Proposition 4 and Proposition 6	$\tilde{\mathcal{O}}(d^5)$	$\tilde{\mathcal{O}}(\varepsilon^{-6})$	$\tilde{\mathcal{O}}(R^4)$	$\tilde{\mathcal{O}}(r^{-4})$
(Bubeck et al., 2015, Theorem 1) π uniform on K	$\tilde{\mathcal{O}}(d^7)$	$\tilde{\mathcal{O}}(\varepsilon^{-8})$	$\tilde{\mathcal{O}}(R^6)$	$\tilde{\mathcal{O}}(r^{-6})$
(Bubeck et al., 2015, Theorem 1) π log concave	$\tilde{\mathcal{O}}(d^{12})$	$\tilde{\mathcal{O}}(\varepsilon^{-12})$	$\tilde{\mathcal{O}}(R^{18})$	$\tilde{\mathcal{O}}(r^{-18})$

Table 1: dependency of n on d, ε, R and r to get $\|\delta_{x^\star} R_\gamma^n - \pi\|_{\mathrm{TV}} \leq \varepsilon$

Upper bound on n to get $\|\delta_{x^\star} R_\gamma^n - \pi\|_{\mathrm{TV}} \leq \varepsilon$	$\Delta_1 \to 0$	$\Delta_2 \to +\infty$
Proposition 4 and Proposition 6	$\tilde{\mathcal{O}}(\Delta_1^{-4})$	$\tilde{\mathcal{O}}(\Delta_2^{4})$

Table 2: dependency of n on Δ_1 and Δ_2 to get $\|\delta_{x^\star} R_\gamma^n - \pi\|_{\mathrm{TV}} \leq \varepsilon$

4.2. Convergence in Wasserstein distance for strongly convex f

In this section, f is assumed to satisfy an additional assumption.

H4 $f : \mathbb{R}^d \mapsto \mathbb{R}$ *is m-strongly convex , i.e. there exists $m > 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + (m/2) \|x - y\|^2 \ . \tag{27}$$

Note that under **H**4, U^λ defined in (7) is m-strongly convex as well. The following Proposition 7 relies on the convergence analysis in Wasserstein distance done in Durmus and Moulines (2016), which assumes that f is strongly convex. It may be possible to extend the range of validity of these results but this work goes beyond the scope of this paper.

Proposition 7 *Assume **H**1 and **H**4. Let $\lambda > 0$, L be the Lipschitz constant of ∇U^λ defined in (7) and $\kappa = (2mL)(m+L)^{-1}$. Let $\varepsilon > 0$ and $x \in \mathbb{R}^d$. We have,*

$$W_2(\delta_x R_\gamma^n, \pi^\lambda) \leq \varepsilon \ ,$$

provided that,

$$\gamma \leq \frac{m}{L^2} \left\{ -\frac{13}{12} + \left[\left(\frac{13}{12} \right)^2 + \frac{\varepsilon^2 \kappa^2}{8md} \right]^{1/2} \right\} \wedge \frac{1}{m+L} \ ,$$
$$n \geq 2(\kappa\gamma)^{-1} \left\{ -\log(\varepsilon^2/4) + \log\left(\|x - x^\star\|^2 + d/m \right) \right\} \ .$$

Proof The proof is postponed to Section 6.5. ■

Combining Proposition 5 and Proposition 7 determines the stepsize γ and the number of samples n to get $W_1(\delta_{x^\star} R_\gamma^n, \pi) \leq \varepsilon$. λ is chosen of order $\varepsilon^2 \Delta_1^2 r^2 d^{-2} R^{-2}$ under **H**3-(i) and $\varepsilon^2 r^2 R^{-2} \min(d^{-2}, \Delta_2^{-2})$ under **H**3-(ii). The orders of magnitude of n in $d, \varepsilon, R, r, \Delta_1, \Delta_2$ are reported in Tables 3 and 4.

Upper bound on n to get $W_1(\delta_{x^\star} R_\gamma^n, \pi) \leq \varepsilon$	$d \to +\infty$	$\varepsilon \to 0$	$R \to +\infty$	$r \to 0$
Proposition 5-c) and Proposition 7	$\tilde{\mathcal{O}}(d^5)$	$\tilde{\mathcal{O}}(\varepsilon^{-6})$	$\tilde{\mathcal{O}}(R^4)$	$\tilde{\mathcal{O}}(r^{-4})$

Table 3: dependency of n on d, ε, R and r to get $W_1(\delta_{x^\star} R_\gamma^n, \pi) \leq \varepsilon$

Upper bound on n to get $W_1(\delta_{x^\star} R_\gamma^n, \pi) \leq \varepsilon$	$\Delta_1 \to 0$	$\Delta_2 \to +\infty$
Proposition 5-c) and Proposition 7	$\tilde{\mathcal{O}}(\Delta_1^{-4})$	$\tilde{\mathcal{O}}(\Delta_2^4)$

Table 4: dependency of n on Δ_1 and Δ_2 to get $W_1(\delta_{x^\star} R_\gamma^n, \pi) \leq \varepsilon$

5. Numerical experiments

In this section we illustrate MYULA with the following three numerical experiments: computation of the volume of a high-dimensional convex set, sampling from a truncated multivariate Gaussian distribution, and Bayesian inference with the constrained LASSO model. We benchmark our results with model-specific specialised algorithms, namely the hit-and-run algorithm Lovász and Vempala (2006) for set volume computation, the wall HMC (WHMC) Pakman and Paninski (2014) for truncated Gaussian models, and the auxiliary-variable Gibbs sampler for the Bayesian lasso model Park and Casella (2008). Where relevant we also compare with the Random Walk Metropolis Hastings (RWM) algorithm.

First we consider the computation of the volume of a high-dimensional hypercube. In a manner akin to Cousins and Vempala (2015), to apply MYULA to this problem we use an annealing strategy involving truncated Gaussian distributions whose variance is gradually increased at each step $i \in \mathbb{N}$ of the annealing process. Precisely, for $M \in \mathbb{N}^\star$ and $i \in \{0, \ldots, M-1\}$, the potential U_i (2) of the phase i is given for all $x \in \mathbb{R}^d$ by, $U_i(x) = (2\sigma_i^2)^{-1} \|x\|^2 + \iota_{\mathsf{K}}$ where $\mathsf{K} = [-1,1]^d$. Observing that,

$$\frac{\int_{\mathbb{R}^d} \mathrm{e}^{-U_{i+1}(x)} \mathrm{d}x}{\int_{\mathbb{R}^d} \mathrm{e}^{-U_i(x)} \mathrm{d}x} = \pi_i(g_i) \ , \quad g_i(x) = \mathrm{e}^{2^{-1}(\sigma_i^{-2} - \sigma_{i+1}^{-2})\|x\|^2} \ , \tag{28}$$

where π_i is the probability measure associated with U_i, the volume of K is

$$\mathrm{Vol}(\mathsf{K}) = \prod_{i=0}^{M-1} \pi_i(g_i) \int_{\mathbb{R}^d} \mathrm{e}^{-U_0(x)} \ ,$$

where $U_M = \iota_{\mathsf{K}}$. To use MYULA we consider for all $i \in \{0, \ldots, M-1\}$ the potential $U_i^{\lambda_i}$ defined for all $x \in \mathbb{R}^d$ by $U_i^{\lambda_i}(x) = (2\sigma_i^2)^{-1} \|x\|^2 + \iota_{\mathsf{K}}^{\lambda_i}$ where $\iota_{\mathsf{K}}^{\lambda_i}$ is given by (4). We choose the step-size γ_i proportional to $1/\{d \max(d, \sigma_i^{-1})\}$ and the regularization parameter λ_i is set equal to $2\gamma_i$. The counterpart of (28) is then

$$\frac{\int_{\mathbb{R}^d} \mathrm{e}^{-U_{i+1}^{\lambda_{i+1}}(x)} \mathrm{d}x}{\int_{\mathbb{R}^d} \mathrm{e}^{-U_i^{\lambda_i}(x)} \mathrm{d}x} = \pi_i^{\lambda_i}\left(g_i^{\lambda_i}\right) \ , \quad g_i^{\lambda_i}(x) = \mathrm{e}^{2^{-1}(\sigma_i^{-2} - \sigma_{i+1}^{-2})\|x\|^2 + \iota_{\mathsf{K}}^{\lambda_i} - \iota_{\mathsf{K}}^{\lambda_{i+1}}} \ ,$$

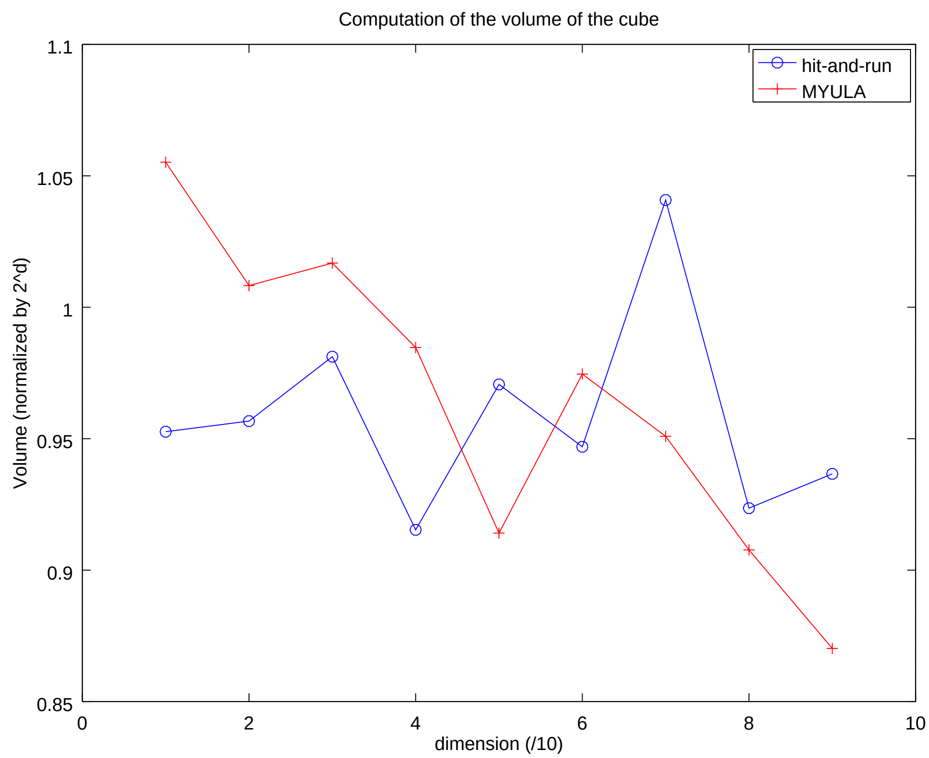

Figure 1: Computation of the volume of the cube with MYULA and hit-and-run algorithm.

where $\pi_i^{\lambda_i}$ is the probability measure associated with $U_i^{\lambda_i}$, and the volume of K is

$$\mathrm{Vol}(\mathsf{K}) = \prod_{i=0}^{M-1} \pi_i^{\lambda_i}(g_i^{\lambda_i}) \int_{\mathbb{R}^d} \mathrm{e}^{-U_0^{\lambda_0}(x)} ,$$

where $U_M^{\lambda_M} = U_M = \iota_{\mathsf{K}}$.

Figure 1 shows the volume estimates (over 10 experiments) obtained with MYULA and the hit-and-run algorithm for a unit hypercube of dimension d ranging from $d = 10$ to $d = 90$ (to simplify visual comparison the estimates are normalised w.r.t. the true volume). Observe that the estimates of MYULA are in agreement with the results of the hit-and-run algorithm, which serves as a benchmark for this problem. The outputs of both algorithms are at similar distances with respect to the true value 1.

Moreover, the second experiment we consider is the simulation from a d-dimensional truncated Gaussian distribution restricted on a convex set K_d, with mode zero at the boundary of the set, and covariance matrix Σ with (i,j)th element given by $(\Sigma)_{i,j} = 1/(1+|i-j|)$. Let $\boldsymbol{\beta} \in \mathbb{R}^d$. The potential U, given by (2) and associated with the density $\pi(\boldsymbol{\beta})$, is given by $U(\boldsymbol{\beta}) = (1/2)\left\langle \boldsymbol{\beta}, \Sigma^{-1}\boldsymbol{\beta} \right\rangle + \iota_{\mathsf{K}_d}(\boldsymbol{\beta})$. We consider three scenarios of increasing dimension: $d = 2$ with

$\mathsf{K}_2 = [0,5] \times [0,1]$, $d = 10$ with $\mathsf{K}_{10} = [0,5] \times [0,0.5]^9$, and $d = 100$ with $\mathsf{K}_{100} = [0,5] \times [0,0.5]^{99}$. We generate 10^6 samples for MYULA, 10^5 samples for WHMC, and 10^6 samples for RWM (in all cases the initial 10% is discarded as burn-in period). Regarding algorithm parameters, we set $\gamma = 1/1000$ and $\lambda = 2\gamma$ for MYULA, and adjust the parameters of RWM and WHMC such that their acceptance rates are approximately 25% and 70%.

Table 5 shows the results obtained with each method for the model $d = 2$, and by performing 100 repetitions to obtain 95% confidence intervals. For this model we also report a solution by a cubature integration Narasimhan and Johnson (2016) which provides a ground truth. Moreover, Figure 2 and Figure 3 show the results for the first three coordinates of $\boldsymbol{\beta}$ (i.e., $\beta_1, \beta_2, \beta_3$) for $d = 10$ and $d = 100$ respectively. Observe the good performance of MYULA as dimensionality increases, particularly in the challenging case $d = 100$ where it performs comparably to the specialised algorithm WHMC.

Method	Mean	Covariance
Truth	$\begin{bmatrix} 0.790 \\ 0.488 \end{bmatrix}$	$\begin{bmatrix} 0.326 & 0.017 \\ 0.017 & 0.080 \end{bmatrix}$
RWM	$\begin{bmatrix} 0.791 \pm 0.013 \\ 0.486 \pm 0.002 \end{bmatrix}$	$\begin{bmatrix} 0.330 \pm 0.011 & 0.017 \pm 0.002 \\ 0.017 \pm 0.002 & 0.080 \pm 0.0003 \end{bmatrix}$
WHMC	$\begin{bmatrix} 0.789 \pm 0.005 \\ 0.490 \pm 0.005 \end{bmatrix}$	$\begin{bmatrix} 0.324 \pm 0.008 & 0.017 \pm 0.002 \\ 0.017 \pm 0.002 & 0.079 \pm 0.0007 \end{bmatrix}$
MYULA	$\begin{bmatrix} 0.758 \pm 0.052 \\ 0.484 \pm 0.016 \end{bmatrix}$	$\begin{bmatrix} 0.309 \pm 0.038 & 0.017 \pm 0.009 \\ 0.017 \pm 0.009 & 0.088 \pm 0.002 \end{bmatrix}$

Table 5: Mean and covariance of $\boldsymbol{\beta}$ in dimension 2 obtained by RWM, WHMC and MYULA.

Finally, we also report an experiment involving the analysis of a real dataset with an ℓ_1-norm constrained Bayesian LASSO model (i.e. least squares regression subject to an ℓ_1-ball constraint). Precisely, the observations $Y = \{Y_1, \ldots, Y_n\} \in \mathbb{R}^n$, for $n \geq 1$, are assumed to be distributed from the Gaussian distribution with mean $X\boldsymbol{\beta}$ and covariance matrix $\sigma^2 \, \mathrm{I}_n$, where $X \in \mathbb{R}^{n \times d}$ is the design matrix, $\boldsymbol{\beta} \in \mathbb{R}^d$ is the regression parameter, $\sigma^2 > 0$ and I_n is the identity matrix of dimension n. The prior on $\boldsymbol{\beta}$ is the uniform distribution over the ℓ_1 ball, $\mathrm{B_o}(0, s) = \{\boldsymbol{\beta} \in \mathbb{R}^d \quad \|\boldsymbol{\beta}\|_1 \leq s\}$, for $s > 0$, where $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^{d} |\boldsymbol{\beta}_i|$, $\boldsymbol{\beta}_i$ is the i-th component of $\boldsymbol{\beta}$. The potential U^s, for $s > 0$, associated with the posterior distribution is given for all $\boldsymbol{\beta} \in \mathbb{R}^d$ by $U^s(\boldsymbol{\beta}) = \|Y - X\boldsymbol{\beta}\|^2 + \iota_{\mathrm{B_o}(0,s)}(\boldsymbol{\beta})$. We consider in our experiment the diabetes data set[1], which consists in $n = 442$ observations and $d = 10$ explanatory variables.

Figure 4 shows the "LASSO paths" obtained using MYULA, the WHMC algorithm, and with the specialised Gibbs sampler of Park and Casella (2008) (these paths are the posterior marginal medians associated with π^s for $s = t \left\|\boldsymbol{\beta}^{\mathrm{OLS}}\right\|_1$, $t \in [0,1]$, and where $\boldsymbol{\beta}^{\mathrm{OLS}}$ is the estimate obtained by the ordinary least square regression). The dot lines represent the confidence interval at level 95%, obtained by performing 100 repetitions. MYULA estimates were obtained by using 10^5 samples (with the initial 10^4 samples discarded as burn-in period) and stepsize $s^{3/2} \times 10^{-5}$. WHMC

1. http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

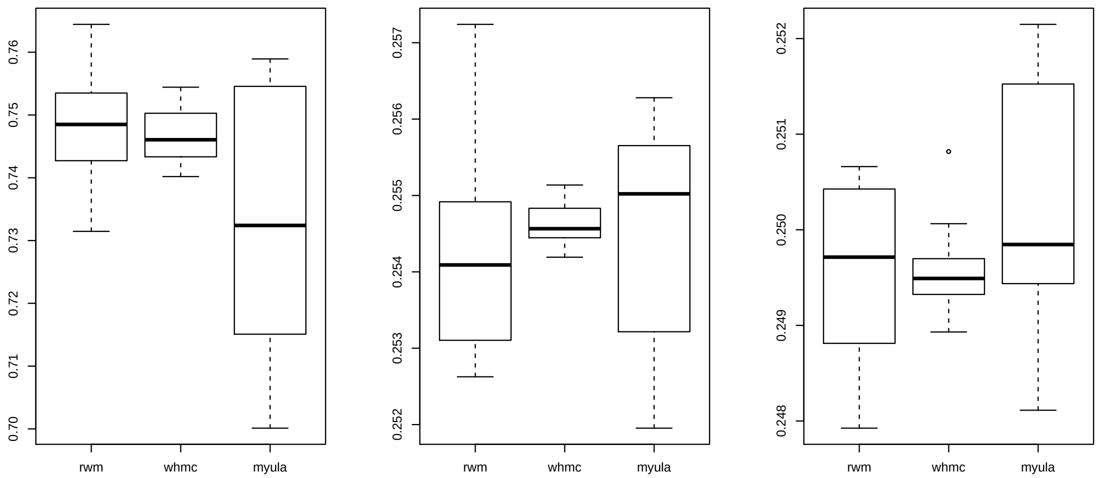

Figure 2: Boxplots of $\beta_1, \beta_2, \beta_3$ for the truncated Gaussian variable in dimension 10.

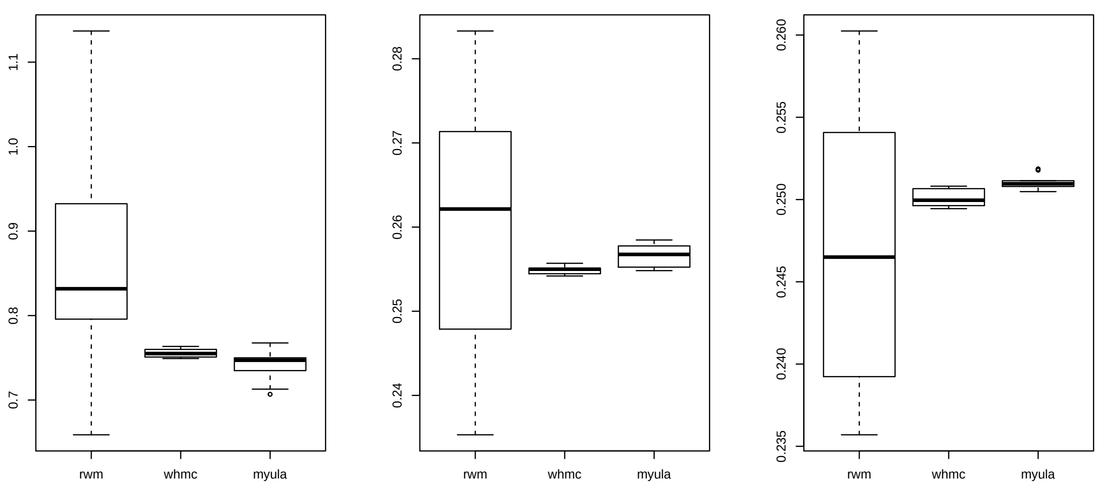

Figure 3: Boxplots of $\beta_1, \beta_2, \beta_3$ for the truncated Gaussian variable in dimension 100.

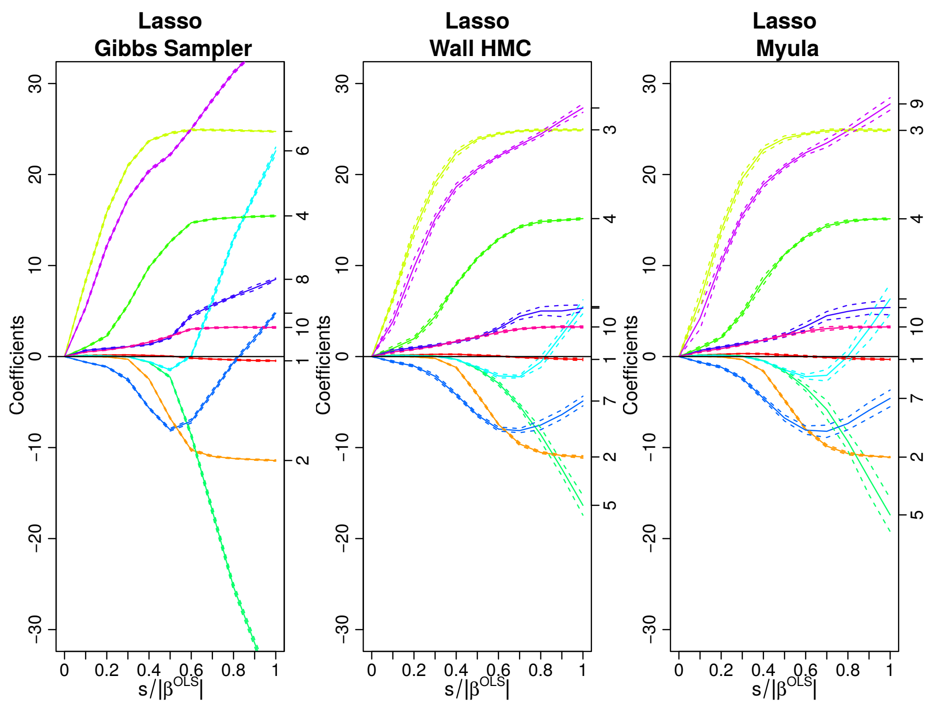

Figure 4: Lasso path for the Gibbs sampler, Wall HMC and MYULA algorithms.

estimates were obtained by using 10^4 samples (with the initial 10^3 samples discarded as burn-in period), and by adjusting parameters to achieve an acceptance rate of approximately 90%. Finally, the Gibbs sampler is targeting an unconstrained LASSO model with prior $\boldsymbol{\beta} \mapsto (2s)^{-d}\mathrm{e}^{-\|\boldsymbol{\beta}\|_1/s}$, for $s > 0$.

6. Proofs

6.1. Proof of Lemma 1

Since f is a (proper) convex function, there exist $a \in \mathbb{R}$, $b \in \mathbb{R}^d$ such that $f(x) \geq a + \langle b, x \rangle$ (Rockafellar, 2015, Theorem 23.4). By **H**2 and a straightforward calculation, for $\|x\| \geq R + 4\lambda \|b\| + 2\left\{\lambda(|a| + R\|b\|)\right\}^{1/2}$, we have,

$$U^{\lambda}(x) \geq (4\lambda)^{-1}(\|x\| - R)^2 \;,$$

which concludes the proof.

6.2. Proof of Lemma 3

Under **H**2, $0 \in \widetilde{\mathsf{K}}$. Let $x_1, x_2 \in \widetilde{\mathsf{K}}$ and $t \in [0,1]$. We have by definition of $\widetilde{\mathsf{K}}$ (18) that $\mathrm{B}(tx_1 + (1-t)x_2, r) \subset t\mathrm{B}(x_1, r) + (1-t)\mathrm{B}(x_2, r) \subset \mathsf{K}$, which implies that $\widetilde{\mathsf{K}}$ is convex.

To show that $\widetilde{\mathsf{K}}$ is close, it is enough to show that $\widetilde{\mathsf{K}} = \{x \in \mathsf{K} \mid \mathrm{dist}(x, \mathsf{K}^c) \geq r\}$ where $\mathrm{dist}(x, \mathsf{K}^c) = \inf_{y \in \mathsf{K}^c} \|x - y\|$ since $x \mapsto \mathrm{dist}(x, \mathsf{K}^c)$ is Lipschitz continuous. First by definition, we have $\widetilde{\mathsf{K}} \subset \{x \in \mathsf{K} \mid \operatorname{dist}(x, \mathsf{K}^c) \geq r\}$. To show the converse, let $x \in \{y \in \mathsf{K} \mid \operatorname{dist}(y, \mathsf{K}^c) \geq r\}$. Then, $\mathrm{B_o}(x, r) \subset \mathsf{K}$, where $\mathrm{B_o}(x, r) = \left\{y \in \mathbb{R}^d \mid \|y - x\| < r\right\}$, which yields $\mathrm{B}(x, r) \subset \mathsf{K}$ since K is assumed to be close. This result then concludes the proof by definition of $\widetilde{\mathsf{K}}$.

6.3. Proof of Proposition 4

a) By a direct calculation, we have:

$$\|\pi^\lambda - \pi\|_{\mathrm{TV}} = \int_{\mathbb{R}^d} \left|\pi(x) - \pi^\lambda(x)\right| \mathrm{d}x = 2\left(1 + \left\{\int_{\mathsf{K}^c} \mathrm{e}^{-U^\lambda(x)} \mathrm{d}x\right\}^{-1} \int_{\mathsf{K}} \mathrm{e}^{-f(x)} \mathrm{d}x\right)^{-1} \tag{29}$$

$$\leq 2\left(1 + \exp\left(\min_{\mathsf{K}^c}(f) - \max_{\mathsf{K}}(f)\right) A\right)^{-1} . \tag{30}$$

where

$$A = \operatorname{Vol}(\mathsf{K}) \Big/ \int_{\mathsf{K}^c} \mathrm{e}^{-(2\lambda)^{-1}\|x - \operatorname{proj}_{\mathsf{K}}(x)\|^2} \mathrm{d}x \ . \tag{31}$$

The conclusion follows then from (17) and **H**3-(i).

b) We give two proofs for this result, which both consist in lower bounding A. The obtained bounds are identical up to an universal constant. The first one is simpler and was suggested by a referee. The second one is more involved ; however, it has the benefit of establishing the relation between the intrinsic volumes of K and the bound on the total variation norm.

Under **H**2, we have $\mathsf{K} + \mathrm{B}(0, t) \subset (1 + t/r)\mathsf{K}$ and using (14),

$$\begin{aligned}
\int_{\mathsf{K}^c} \mathrm{e}^{-(1/2\lambda)\|x - \operatorname{proj}_{\mathsf{K}}(x)\|^2} \mathrm{d}x &\leq \left\{\int_{\mathbb{R}_+} \operatorname{Vol}(\mathsf{K}(1 + t/r))\lambda^{-1} t \mathrm{e}^{-t^2/(2\lambda)} \mathrm{d}t - \operatorname{Vol}(\mathsf{K})\right\} \\
&= \operatorname{Vol}(\mathsf{K}) \left\{\int_{\mathbb{R}_+} (1 + t/r)^d \lambda^{-1} t \mathrm{e}^{-t^2/(2\lambda)} \mathrm{d}t - 1\right\} \\
&= \operatorname{Vol}(\mathsf{K}) \sum_{i=1}^d \binom{d}{i} \left(\frac{\sqrt{2\lambda}}{r}\right)^i \mathbf{\Gamma}(1 + i/2) \\
&\leq \operatorname{Vol}(\mathsf{K}) \sum_{i=1}^d \left(\frac{\sqrt{2\lambda} d}{r}\right)^i ,
\end{aligned}$$

where the second equality follows from developping $(1 + t/r)^d$, making the change of variable $t \mapsto t^2/(2\lambda)$ and using the Gamma function and the last inequality from $\binom{d}{i}\mathbf{\Gamma}(1 + i/2) \leq d^i$ for $i \in \{1, \ldots, d\}$. For $\lambda \in \left(0, r^2 d^{-2}/8\right]$, we get

$$A^{-1} \leq \sum_{i=1}^d \left(\frac{\sqrt{2\lambda} d}{r}\right)^i \leq \frac{2\sqrt{2\lambda} d}{r} .$$

Combining it with (30) and **H**3-(i) concludes the proof.

For the second proof, it is necessary to introduce first a generalized notion of the intrinsic volumes (15), the mixed volumes. Let $\mathcal{K}$ be the class of convex bodies of $\mathbb{R}^d$, $\mathsf{K}_1, \ldots, \mathsf{K}_m \in \mathcal{K}$ and

$\lambda_1, \ldots, \lambda_m \geq 0$. By (Schneider, 2013, Theorem 5.1.7), there is a nonnegative symmetric function $\mathcal{V} : (\mathcal{K})^d \to \mathbb{R}_+$, the mixed volume, such that,

$$\mathrm{Vol}(\lambda_1 \mathsf{K}_1 + \ldots + \lambda_m \mathsf{K}_m) = \sum_{i_1, \ldots, i_d = 1}^{m} \lambda_{i_1} \ldots \lambda_{i_d} \mathcal{V}(\mathsf{K}_{i_1}, \ldots, \mathsf{K}_{i_d}) \ . \tag{32}$$

Let $m > 1$, $a_1, \ldots, a_m \geq 0$ and $\mathsf{K}_1, \ldots, \mathsf{K}_m, \mathsf{L}$ be $(m+1)$ convex bodies in $\mathbb{R}^d$ such that $\mathsf{K}_1 \subset \mathsf{L}$. By unicity of the coefficients of the polynomial in $\lambda_1, \ldots, \lambda_m$ (32) and (Schneider, 2013, p.282), we have:

$$\mathcal{V}(a_1 \mathsf{K}_1, \ldots, a_m \mathsf{K}_m) = \left(\prod_{i=1}^{m} a_i \right) \mathcal{V}(\mathsf{K}_1, \ldots, \mathsf{K}_m) \ , \tag{33}$$

$$\mathcal{V}(\mathsf{K}_1, \mathsf{K}_2, \ldots, \mathsf{K}_m) \leq \mathcal{V}(\mathsf{L}, \mathsf{K}_2, \ldots, \mathsf{K}_m) \ . \tag{34}$$

Denote by B the unity ball of $\mathbb{R}^d$, $\mathsf{B} = \mathrm{B}(0,1)$. Taking $m = 2, \mathsf{K}_1 = \mathsf{K}, \mathsf{K}_2 = \mathsf{B}, \lambda_1 = 1, \lambda_2 = t$ in (32), we get:

$$\mathrm{Vol}(\mathsf{K} + \mathrm{B}(0,t)) = \sum_{i=0}^{d} t^i \binom{d}{i} \mathcal{V}(\mathsf{K}[d-i], \mathsf{B}[i]) \ , \tag{35}$$

where for a set $A \subset \mathbb{R}^d$, the notation $A[i]$ means A repeated i times: $A[i] = A, \ldots, A$ i times. The quermassintegrals of K are defined for $i \in \{0, \ldots, d\}$ by $\mathscr{W}_i(\mathsf{K}) = \mathcal{V}(\mathsf{K}[d-i], \mathsf{B}[i])$ (Schneider, 2013, equation 5.31). We get then by (35) and (15),

$$\binom{d}{i} \mathscr{W}_i(\mathsf{K}) = \kappa_i \mathscr{V}_{d-i}(\mathsf{K}) \ , \tag{36}$$

where κ_i is given by (16).

The proof consists then in identifying an upper bound on $\mathscr{V}_i(\mathsf{K})(\mathrm{Vol}\,\mathsf{K})^{-1}$ for $i \in \{0, \ldots, d\}$. First, the sequence $\{i! \mathscr{V}_i(\mathsf{K})\}_{0 \leq i \leq d}$ is shown to be log-concave, i.e. for $i \in \{1, \ldots, d-1\}$

$$(i! \mathscr{V}_i(\mathsf{K}))^2 \geq (i+1)! \mathscr{V}_{i+1}(\mathsf{K}) (i-1)! \mathscr{V}_{i-1}(\mathsf{K}) \ . \tag{37}$$

The Aleksandrov-Fenchel inequality (Schneider, 2013, equation 7.66) states, for $i \in \{1, \ldots, d-1\}$,

$$\mathscr{W}_i(\mathsf{K})^2 \geq \mathscr{W}_{i-1}(\mathsf{K}) \mathscr{W}_{i+1}(\mathsf{K}) \ . \tag{38}$$

By (16), $\kappa_i / \kappa_{i-2} = (2\pi)/i$ and the log convexity of the gamma function, we get for $i \in \{1, \ldots, d-1\}$:

$$\frac{1}{i+1} \frac{\kappa_i}{\kappa_{i+1}} = \frac{1}{i} \frac{\kappa_{i-2}}{\kappa_{i-1}} \leq \frac{1}{i} \frac{\kappa_{i-1}}{\kappa_i} \ . \tag{39}$$

Combining (39), (38) and (36) shows (37).

The log-concavity of $\{i! \mathscr{V}_i(\mathsf{K})\}_{0 \leq i \leq d}$ gives for $i \in \{0, \ldots, d-1\}$,

$$\frac{\mathscr{V}_i(\mathsf{K})}{\mathscr{V}_{i+1}(\mathsf{K})} \leq \frac{\mathscr{V}_{d-1}(\mathsf{K})}{\mathrm{Vol}(\mathsf{K})} = \frac{d}{2} \frac{\mathscr{W}_1(\mathsf{K})}{\mathscr{W}_0(\mathsf{K})} \ . \tag{40}$$

Combining the definition of the quermassintegrals, (33), (34) and **H**2 give:

$$r\mathscr{W}_1(\mathsf{K}) = \mathcal{V}(\mathsf{K},\ldots,\mathsf{K},\mathrm{B}(0,r)) \leq \mathcal{V}(\mathsf{K},\ldots,\mathsf{K},\mathsf{K}) = \mathscr{W}_0(\mathsf{K}) \; . \tag{41}$$

By (41) and (40), we get:

$$\mathsf{D}(\mathsf{K},\lambda) \leq \sum_{i=1}^{d} \left\{ dr^{-1}(\pi\lambda/2)^{1/2} \right\}^{i} \; , \tag{42}$$

where $\mathsf{D}(\mathsf{K},\lambda)$ is defined in (20). For all $\lambda \in \left(0, 2\pi^{-1}(r/d)^2\right)$, (19) gives then,

$$\|\pi^{\lambda} - \pi\|_{\mathrm{TV}} \leq 2\left\{ 1 + \exp\left(\min_{\mathsf{K}^c}(f) - \max_{\mathsf{K}}(f) \right) \left(\left\{ dr^{-1}(\pi\lambda/2)^{1/2} \right\}^{-1} - 1 \right) \right\}^{-1} \; .$$

Using that for all $a, b \in \mathbb{R}_+^*$, $b \geq 2$, $(1 + a(b-1))^{-1} \leq b^{-1}/(b^{-1} + a/2)$ and **H**3-(i), we get for $\lambda \in \left(0, 2\pi^{-1}(r/d)^2\right)$

$$\|\pi^{\lambda} - \pi\|_{\mathrm{TV}} \leq 2^{3/2}(\pi\lambda)^{1/2} dr^{-1} \left\{ (2\pi\lambda)^{1/2} dr^{-1} + \Delta_1 \right\}^{-1} \; .$$

c) The proof consists in using (29) to bound $\|\pi^{\lambda} - \pi\|_{\mathrm{TV}}$. In the first step we give an upper bound on $\int_{\mathbb{R}^d} \mathrm{e}^{-U^{\lambda}(x)} \mathrm{d}x / \int_{\mathsf{K}} \mathrm{e}^{-f(x)} \mathrm{d}x$. By Fubini's theorem, similarly to (14) we have

$$\int_{\mathbb{R}^d} \mathrm{e}^{-U^{\lambda}(x)} \mathrm{d}x \leq \int_{\mathbb{R}_+} \int_{\mathsf{K}+\mathrm{B}(0,t)} \mathrm{e}^{-f(x)} \lambda^{-1} t \mathrm{e}^{-t^2/(2\lambda)} \mathrm{d}x \mathrm{d}t \; . \tag{43}$$

Let $t \geq 0$. By definition of $\widetilde{\mathsf{K}}$, using Lemma 3 and $\mathsf{K} - \mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}}) + \mathrm{B}(0,t) \subset (1 + t/r)(\mathsf{K} - \mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}}))$, we have

$$\begin{aligned} \int_{\mathsf{K}+\mathrm{B}(0,t)} \mathrm{e}^{-f(x)} \mathrm{d}x &= \int_{\mathsf{K}-\mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})+\mathrm{B}(0,t)} \mathrm{e}^{-f(x+\mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}}))} \mathrm{d}x \\ &\leq \int_{(1+t/r)(\mathsf{K}-\mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}}))} \mathrm{e}^{-f(x+\mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}}))} \mathrm{d}x \\ &= (1+t/r)^d \int_{\mathsf{K}-\mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})} \mathrm{e}^{-f((1+t/r)x+\mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}}))} \mathrm{d}x \; . \end{aligned} \tag{44}$$

By **H**1-(i) f is convex and therefore for all $x \in \mathsf{K} - \mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})$,

$$\begin{aligned} f((1+t/r)x + \mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})) &\geq (t/r)\left\{ f(x + \mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})) - f(\mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})) \right\} + f(x + \mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})) \\ &\geq -(\Delta_2 t)/r + f(x + \mathrm{proj}_{\widetilde{\mathsf{K}}}(x_{\mathsf{K}})) \; . \end{aligned}$$

Combining it with (43) and (44), we get

$$\int_{\mathbb{R}^d} \mathrm{e}^{-U^{\lambda}(x)} \mathrm{d}x \leq \left(\int_{\mathsf{K}} \mathrm{e}^{-f(x)} \mathrm{d}x \right) \int_{\mathbb{R}_+} (1+t/r)^d \mathrm{e}^{(\Delta_2 t)/r} \lambda^{-1} t \mathrm{e}^{-t^2/(2\lambda)} \mathrm{d}t \; . \tag{45}$$

We now bound $B = \int_{\mathsf{K}^c} \mathrm{e}^{-U^\lambda(x)}\mathrm{d}x / \int_{\mathsf{K}} \mathrm{e}^{-f(x)}\mathrm{d}x$. Using (45) and an integration by parts, we have

$$\begin{aligned} B &\le \int_{\mathbb{R}_+} \left\{(1+t/r)^d \mathrm{e}^{(\Delta_2 t)/r} - 1\right\} \lambda^{-1} t \mathrm{e}^{-t^2/(2\lambda)}\mathrm{d}t \\ &\le \int_{\mathbb{R}_+} (1+t/r)^{d-1}\mathrm{e}^{(\Delta_2 t)/r} r^{-1}\left(d+\Delta_2+(\Delta_2 t)/r\right)\mathrm{e}^{-t^2/(2\lambda)}\mathrm{d}t \;. \end{aligned}$$

Since for all $t \ge 0$, $(\Delta_2 t)/r - t^2/(2\lambda) \le -t^2/(4\lambda) + 4\lambda(\Delta_2/r)^2$, it holds

$$B \le \frac{1}{r}\exp\left(4\lambda\left(\frac{\Delta_2}{r}\right)^2\right)\int_{\mathbb{R}_+}(1+t/r)^{d-1}\left(d+\Delta_2+(\Delta_2 t)/r\right)\mathrm{e}^{-t^2/(4\lambda)}\mathrm{d}t \;.$$

By developping $(1+t/r)^{d-1}$, using the change of variable $t \mapsto t^2/(4\lambda)$ and the definition of the Gamma function, we have

$$B \le \frac{2\lambda}{r}\exp\left(4\lambda\left(\frac{\Delta_2}{r}\right)^2\right)\sum_{i=0}^{d-1}\binom{d-1}{i}\left(\frac{2\sqrt{\lambda}}{r}\right)^i\left\{\frac{d+\Delta_2}{2\sqrt{\lambda}}\mathbf{\Gamma}\left(\frac{1+i}{2}\right)+\frac{\Delta_2}{r}\mathbf{\Gamma}\left(1+\frac{i}{2}\right)\right\} \;.$$

Using that for all $i \in \{0,\ldots,d-1\}$, $\binom{d-1}{i}\mathbf{\Gamma}(1+i/2) \le d^i$, we get for $\lambda \in \left(0, 16^{-1}r^2d^{-2}\right]$

$$B \le \frac{2}{r}\exp\left(4\lambda\left(\frac{\Delta_2}{r}\right)^2\right)\left\{\sqrt{\lambda}(d+\Delta_2)+\frac{2\lambda\Delta_2}{r}\right\} \;,$$

which combined with (29) concludes the proof.

6.4. Proof of Proposition 5

a) The proof relies on a control of the Wasserstein distance by a weighted total variation. The arguments are similar to those of Proposition 4. (Villani, 2009, Theorem 6.15) implies:

$$W_1(\pi, \pi^\lambda) \le \int_{\mathbb{R}^d} \|x\| \left|\pi(x) - \pi^\lambda(x)\right|\mathrm{d}x = C + D \;, \tag{46}$$

where

$$C = \int_{\mathsf{K}^c}\|x\|\,\pi^\lambda(x)\mathrm{d}x \;, \quad D = \left\{1 - \frac{\int_{\mathsf{K}}\mathrm{e}^{-f}}{\int_{\mathbb{R}^d}\mathrm{e}^{-U^\lambda}}\right\}\int_{\mathsf{K}}\|x\|\,\pi(x)\mathrm{d}x \;. \tag{47}$$

We bound these two terms separately. First using the same decomposition as in (14), $\|x\| \le R + \|x - \mathrm{proj}_{\mathsf{K}}(x)\|$ and that for all $t \in \mathbb{R}_+$, $\mathsf{K} + \mathrm{B}(0,t) = \{x \in \mathbb{R}^d \,:\, \|x - \mathrm{proj}_{\mathsf{K}}(x)\| \le t\}$, we get

$$\begin{aligned} C &= \left(\int_{\mathbb{R}^d}\mathrm{e}^{-U^\lambda}\right)^{-1}\int_0^{+\infty}\int_{\mathsf{K}^c}\mathrm{e}^{-f(x)}\,\|x\|\,t\lambda^{-1}\mathrm{e}^{-t^2/(2\lambda)}\mathbb{1}_{[\|x-\mathrm{proj}_{\mathsf{K}}(x)\|,+\infty)}(t)\,\mathrm{d}x\,\mathrm{d}t \qquad (48)\\ &\le \mathrm{e}^{\max_{\mathsf{K}}(f)-\min_{\mathsf{K}^c}(f)}\int_0^{+\infty}(R+t)t\lambda^{-1}\mathrm{e}^{-t^2/(2\lambda)}\left(\frac{\mathrm{Vol}(\mathsf{K}+\mathrm{B}(0,t))-\mathrm{Vol}(\mathsf{K})}{\mathrm{Vol}(\mathsf{K})}\right)\mathrm{d}t \;. \qquad (49) \end{aligned}$$

Combining (15)-(49), **H**3-(i) and using $\mathscr{V}_d(\mathsf{K}) = \mathrm{Vol}(\mathsf{K})$ give

$$C \le \Delta_1^{-1}\sum_{i=0}^{d-1}\kappa_{d-i}\frac{\mathscr{V}_i(\mathsf{K})}{\mathrm{Vol}(\mathsf{K})}\int_0^{+\infty}(Rt^{d-i+1}+t^{d-i+2})\lambda^{-1}\mathrm{e}^{-t^2/(2\lambda)}\mathrm{d}t \;. \tag{50}$$

Using (16), for all $k \geq 0$, $\int_{\mathbb{R}_+} t^k \mathrm{e}^{t^2/(2\lambda)}\mathrm{d}t = (2\lambda)^{(k+1)/2}\boldsymbol{\Gamma}((k+1)/2)$ and for all $a > 1$, $\boldsymbol{\Gamma}(a+1/2) \leq a^{1/2}\boldsymbol{\Gamma}(a)$ (by log-convexity of the Gamma function), we have

$$C \leq \Delta_1^{-1} \sum_{i=0}^{d-1} \frac{\mathscr{V}_i(\mathsf{K})}{\mathrm{Vol}(\mathsf{K})} (2\pi\lambda)^{(d-i)/2} \left\{ R + [\lambda(d-i+2)]^{1/2} \right\} . \tag{51}$$

Regarding D defined in (47), by **H**2, **H**3-(i), (30) and (17), we get:

$$D \leq R\Delta_1^{-1}\mathsf{D}(\mathsf{K}, \lambda) , \tag{52}$$

where $\mathsf{D}(\mathsf{K}, \lambda)$ is defined in (20). Combining (51) and (52) in (46) concludes the proof.

b) Using (40) and (41) in (51) gives for all $\lambda \in \left(0, (2\pi)^{-1}r^2d^{-2}\right)$

$$\begin{aligned} C &\leq \Delta_1^{-1} \sum_{i=0}^{d-1} \left(\frac{d}{r}\sqrt{\frac{\pi\lambda}{2}} \right)^{d-i} \left\{ R + [\lambda(d-i+2)]^{1/2} \right\} \\ &\leq \Delta_1^{-1}(2\pi\lambda)^{1/2} d r^{-1} \left(R + r\left(\frac{3}{2d\pi}\right)^{1/2} \right) . \end{aligned}$$

Finally this bound, (52), (42) and (46) conclude the proof.

c) The proof still relies on the decomposition (46), where C and D are defined in (47). Eq. (48) gives

$$C \leq \int_0^{+\infty} (R+t)t\lambda^{-1}\mathrm{e}^{-t^2/(2\lambda)} \left(\frac{\int_{\mathsf{K}+\mathrm{B}(0,t)} \mathrm{e}^{-f(x)}\mathrm{d}x}{\int_{\mathsf{K}} \mathrm{e}^{-f(x)}\mathrm{d}x} - 1 \right) \mathrm{d}t .$$

Under **H**3-(ii), following the steps of Section 6.3-c) to upper bound the term $\int_{\mathsf{K}+\mathrm{B}(0,t)} \mathrm{e}^{-f(x)}\mathrm{d}x / \int_{\mathsf{K}} \mathrm{e}^{-f(x)}\mathrm{d}x$, we have

$$\begin{aligned} C &\leq \int_0^{+\infty} (R+t)t\lambda^{-1}\mathrm{e}^{-t^2/(2\lambda)} \left((1+t/r)^d \mathrm{e}^{(t\Delta_2)/r} - 1 \right) \mathrm{d}t \\ &= C_1 + C_2 , \end{aligned}$$

where

$$\begin{aligned} C_1 &= R \int_0^{+\infty} t\lambda^{-1}\mathrm{e}^{-t^2/(2\lambda)} \left((1+t/r)^d \mathrm{e}^{(t\Delta_2)/r} - 1 \right) \mathrm{d}t , \\ C_2 &= \int_0^{+\infty} t^2\lambda^{-1}\mathrm{e}^{-t^2/(2\lambda)} \left((1+t/r)^d \mathrm{e}^{(t\Delta_2)/r} - 1 \right) \mathrm{d}t . \end{aligned}$$

C_1 is upper bounded in the same way as B in Section 6.3-c). Regarding C_2, since for all $t \geq 0$, $(\Delta_2 t)/r - t^2/(2\lambda) \leq -t^2/(4\lambda) + 4\lambda(\Delta_2/r)^2$, developping $(1+t/r)^d$ and using the change of variable $t \mapsto t^2/(4\lambda)$ we get

$$\begin{aligned} C_2 &\leq \mathrm{e}^{4\lambda(\Delta_2/r)^2} \sum_{i=0}^{d} \binom{d}{i} r^{-i} \int_{\mathbb{R}_+} t^{i+2}\lambda^{-1}\mathrm{e}^{-t^2/(4\lambda)}\mathrm{d}t \\ &\leq 4\sqrt{\lambda}\mathrm{e}^{4\lambda(\Delta_2/r)^2} \frac{\sqrt{\pi}}{2} \sum_{i=0}^{d} \binom{d}{i} \left(\frac{2\sqrt{\lambda}}{r} \right)^i \boldsymbol{\Gamma}\left(\frac{3}{2} + \frac{i}{2} \right) . \end{aligned}$$

Using $\binom{d}{i}\mathbf{\Gamma}((3+i)/2) \leq (\sqrt{\pi}/2)d^i$ for $i \in \{0, \dots, d\}$, we have for $\lambda \in \left(0, 16^{-1}r^2d^{-2}\right]$,

$$\begin{aligned} C_2 &\leq 2\sqrt{\pi\lambda}\mathrm{e}^{4\lambda(\Delta_2/r)^2} \sum_{i=0}^{d} \left(\frac{2\sqrt{\lambda}d}{r}\right)^i \\ &\leq 4\sqrt{\pi\lambda}\mathrm{e}^{4\lambda(\Delta_2/r)^2} . \end{aligned}$$

D defined in (47) is upper bounded by RB where B is defined in Section 6.3-c). Combining the bounds on C_1, C_2, D gives the result.

6.5. Proof of Proposition 7

Assume that $\gamma \in \left(0, (m+L)^{-1}\right)$. (Durmus and Moulines, 2016, Theorem 5) gives for all $n \in \mathbb{N}^\star$:

$$W_2^2(\delta_x R_\gamma^n, \pi^\lambda) \leq 2\left(1 - (\kappa\gamma)/2\right)^n \left\{\|x - x^\star\|^2 + d/m\right\} + u(\gamma) ,$$

where,

$$u(\gamma) = 2\kappa^{-1}L^2d\gamma(\kappa^{-1} + \gamma)\left(2 + \frac{L^2\gamma}{m} + \frac{L^2\gamma^2}{6}\right) .$$

Noting that $\kappa\gamma \leq 1$ and $L^2\gamma^2 \leq 1$, it is then sufficient for γ, n to satisfy,

$$\begin{aligned} 4\kappa^{-2}L^2d\gamma\left(2 + \frac{1}{6} + \frac{L^2\gamma}{m}\right) &\leq \varepsilon^2/2 , \\ 2\left(1 - (\kappa\gamma)/2\right)^n \left\{\|x - x^\star\|^2 + d/m\right\} &\leq \varepsilon^2/2 , \end{aligned}$$

which concludes the proof.

Acknowledgments

The authors wish to express their thanks to the anonymous referees for several helpful remarks, in particular concerning a simplified proof of Proposition 4.

References

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

Sebastien Bubeck, Ronen Eldan, and Joseph Lehec. Finite-time analysis of projected langevin monte carlo. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, NIPS'15, pages 1243–1251, Cambridge, MA, USA, 2015. MIT Press. URL http://dl.acm.org/citation.cfm?id=2969239.2969378.

Gilles Celeux, Mohammed El Anbari, Jean-Michel Marin, and Christian P. Robert. Regularization in regression: Comparing bayesian and frequentist methods in a poorly informative situation. *Bayesian Anal.*, 7(2):477–502, 06 2012. doi: 10.1214/12-BA716. URL http://dx.doi.org/10.1214/12-BA716.

Ming-Hui Chen, Qi-Man Shao, and Joseph G Ibrahim. *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media, 2012.

Ben Cousins and Santosh Vempala. Computation of the volume of convex bodies, Jun 2015. URL http://fr.mathworks.com/matlabcentral/fileexchange/43596-volume-computation-of-convex-bodies.

Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016.

A. Durmus and E. Moulines. Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm. *ArXiv e-prints*, July 2015. Accepted for publication in Ann. Appl. Probab.

A. Durmus and E. Moulines. High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm. *ArXiv e-prints*, May 2016.

A. Durmus, E. Moulines, and M. Pereyra. Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau. *ArXiv e-prints*, December 2016. Accepted for publication in SIAM J. Imaging Sciences.

Martin Dyer and Alan Frieze. Computing the volume of convex bodies: a case where randomness provably helps. *Probabilistic combinatorics and its applications*, 44:123–170, 1991.

A. E. Gelfand, A. F. Smith, and T.-M. Lee. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87(418): 523–532, 1992.

Valen E Johnson and James H Albert. *Ordinal data modeling*. Springer Science & Business Media, 2006.

Jürgen Kampf. On weighted parallel volumes. *Beiträge Algebra Geom*, 50(2):495–519, 2009.

Daniel A Klain and Gian-Carlo Rota. *Introduction to geometric probability*. Cambridge University Press, 1997.

John P Klein and Melvin L Moeschberger. *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media, 2005.

S. Lan and B. Shahbaba. Sampling constrained probability distributions using Spherical Augmentation. *ArXiv e-prints*, June 2015.

László Lovász and Santosh Vempala. Hit-and-run from a corner. *SIAM Journal on Computing*, 35 (4):985–1005, 2006. doi: 10.1137/S009753970544727X. URL http://dx.doi.org/10.1137/S009753970544727X.

László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307–358, May 2007. ISSN 1042-9832. doi: 10.1002/rsa.v30:3. URL http://dx.doi.org/10.1002/rsa.v30:3.

Balasubramanian Narasimhan and Steven G. Johnson. *cubature: Adaptive Multivariate Integration over Hypercubes*, 2016. URL https://CRAN.R-project.org/package=cubature. R package version 1.3-6.

John Paisley, David M Blei, and Michael I Jordan. Bayesian nonnegative matrix factorization with stochastic variational inference. In *Handbook of Mixed Membership Models and Their Applications*, pages 205–224. Chapman and Hall/CRC, 2014.

Ari Pakman and Liam Paninski. Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542, 2014.

G. Parisi. Correlation functions and computer simulations. *Nuclear Physics B*, 180:378–384, 1981.

T. Park and G. Casella. The Bayesian lasso. *J. Amer. Statist. Assoc.*, 103(482):681–686, 2008. ISSN 0162-1459. doi: 10.1198/016214508000000337. URL http://dx.doi.org/10.1198/016214508000000337.

G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. ISSN 1350-7265. doi: 10.2307/3318418. URL http://dx.doi.org/10.2307/3318418.

R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998. ISBN 3-540-62772-3. doi: 10.1007/978-3-642-02431-3. URL http://dx.doi.org/10.1007/978-3-642-02431-3.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

Gabriel Rodriguez-Yam, Richard A Davis, and Louis L Scharf. Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. *Unpublished manuscript*, 2004.

Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*. Number 151. Cambridge University Press, 2013.

C. Villani. *Optimal transport : old and new*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin, 2009. ISBN 978-3-540-71049-3. URL http://opac.inria.fr/record=b1129524.

Appendix A. Details of the orders of magnitude for Table 1 and Table 2

	$d \to +\infty$	$\varepsilon \to 0$	$R \to +\infty$	$r \to 0$	$\Delta_1 \to 0$	$\Delta_2 \to +\infty$
L, λ^{-1}	d^2	ε^{-2}	1	r^{-2}	Δ_1^{-2}	Δ_2^2
$A_1(x)$	d^4	ε^{-4}	R^2	r^{-4}	Δ_1^{-4}	Δ_2^4
$-\log(\kappa)$	1	1	R^{-2}	1	1	1
$A_2(x)$	1	ε^{-1}	R	r^{-1}	Δ_1^{-1}	Δ_2
T	1	$\log(\varepsilon^{-1})$	R^2	$\log(r^{-1})$	$\log(\Delta_1^{-1})$	$\log(\Delta_2)$
γ	d^{-5}	ε^6	R^{-2}	r^{-4}	Δ_1^4	Δ_2^{-4}

Table 6: dependency of $L, A_1(x), -\log(\kappa), A_2(x), T, \gamma$ on d, ε, R, r, Δ_1 and Δ_2.