AACIS Seminar 10/02/23 13:00

Preconditioning for MCMC

Max Hird (UCL)

Joint work with Sam Livingstone (UCL)

Outline

Introductory Material

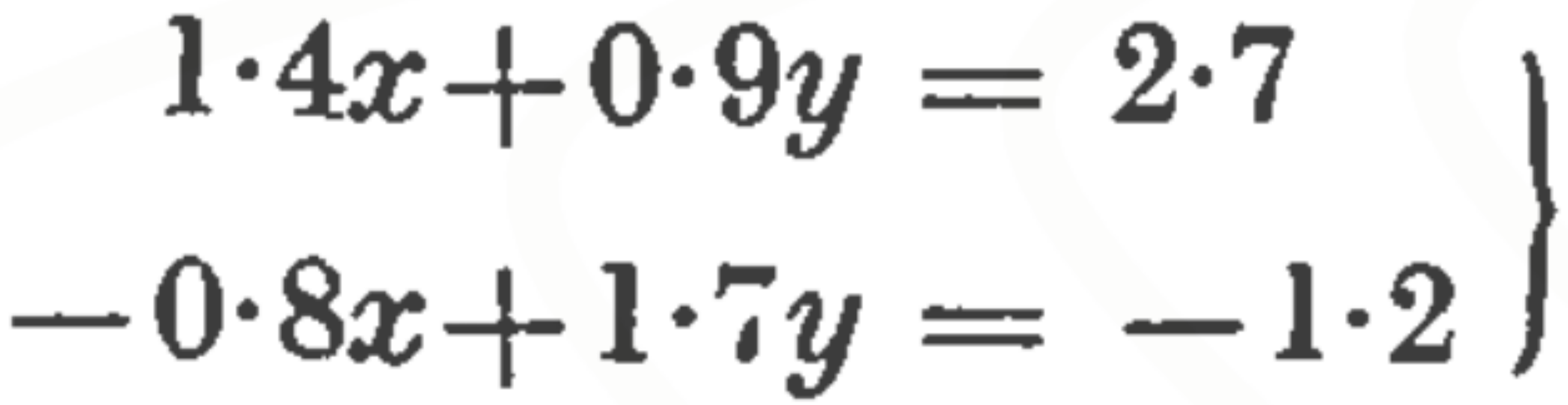
- Intro to Conditioning
- Intro to Markov Chain Monte Carlo (MCMC)
- Preconditioning in MCMC
    - Condition Number

Our Contribution

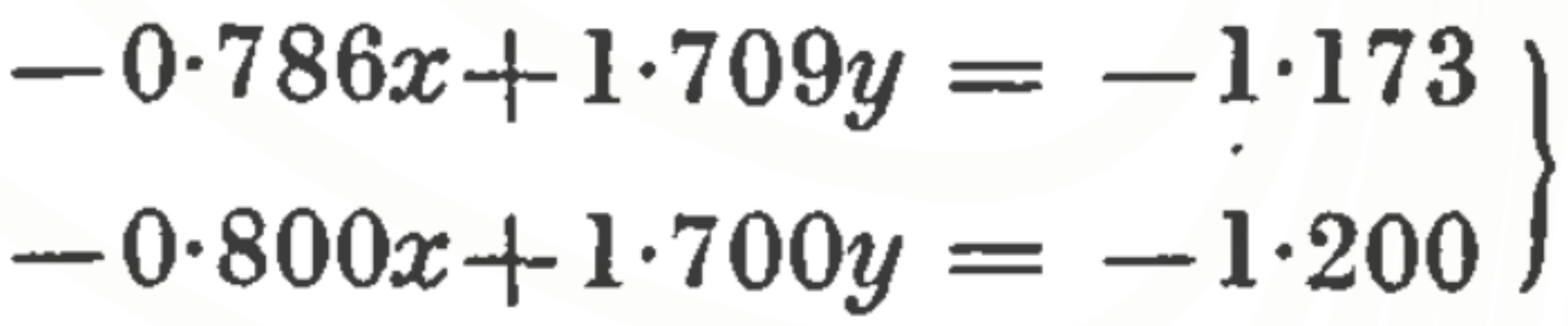
-
    - Linear Preconditioning
    - Nonlinear Preconditioning
- Summary
- References

Preconditioning

20th C Maths starts being concerned with *computability* and not simply *conceivability*:



$$\left.\begin{array}{ll} e_1 & 1{\cdot}4x+0{\cdot}9y = 2{\cdot}7 \\ e_2 & -0{\cdot}8x+1{\cdot}7y = -1{\cdot}2 \end{array}\right\} \Longleftrightarrow \left.\begin{array}{ll} 0.01 \times e_1 + e_2 & -0{\cdot}786x+1{\cdot}709y = -1{\cdot}173 \\ e_2 & -0{\cdot}800x+1{\cdot}700y = -1{\cdot}200 \end{array}\right\}$$



well-conditioned ill-conditioned

`It is certainly true that a trivial modification improves the conditioning'

Turing coins the *condition number* and defines it in multiple ways:

- N-condition number: $\|A\|_F\|A^{-1}\|_F$ where $\|A\|_F := \sqrt{\text{Tr}(A^*A)}$
- M-condition number: $M(A)M(A^{-1})$ where $M(A) := \max_{ij}|m_{ij}|$

The condition number ≥ 1, and 1 is the best possible value

Preconditioning: applying a transformation to reduce the condition number Turing [1948]

Markov Chain Monte Carlo

Sample $X_1, \ldots, X_n$ from a π-stationary Markov Chain, initial dist$^{\underline{n}}$ μ_0, form the estimator

$$\hat{f}_n := \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

Markov Chain CLT gives us that

$$\sqrt{n} \left(\hat{f}_n - E_\pi(f) \right) \xrightarrow{d} N\left(0, \sigma_f^2\right)$$

where

$$\sigma_f^2 := Var_\pi(f(X)) + 2 \sum_{i=1}^{\infty} \text{Cov}\left(f(X_1), f(X_{1+k})\right)$$

Ideal MCMC is *quick to equilibrate* and has *low autocorrelation in equilibrium*

Want to estimate $E_\pi(f(X))$

Sample iid $X_1, \ldots, X_n \sim \pi$, form the estimator:

$$\bar{f}_n := \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

Bias is 0, Variance is $n^{-1} Var_\pi(f(X))$

Unnormalised π is no (theoretical) barrier

Sampling is impossible for interesting π

MCMC: algorithms

Generic structure of an MCMC algorithm: given an initial state $X_0 \sim \mu_0$ and a *proposal density* $q_\theta(x \to .\,)$ with parameters $\theta \in \Theta$

1. Propose a new state $Y_{i+1} \sim q_\theta(X_i \to .\,)$
2. Set $X_{i+1} = Y_{i+1}$ with probability $\alpha(X_i, Y_{i+1})$, otherwise set $X_{i+1} = X_i$

Step 2. is the *Metropolis-Hastings* accept/reject step - ensures π-stationarity

Step 1. defines the algorithm:

- $q_\theta(x \to .\,) = p_\theta(\,.\,)$: Independent Metropolis-Hastings
- $q_\theta(x \to .\,) = N(x,\ \sigma^2 \mathbf{I}_d)$: Random Walk Metropolis
- $q_\theta(x \to .\,) = N\left(x + \sigma^2 \nabla_x \log \pi,\ 2\sigma^2 \mathbf{I}_d\right)$: Metropolis Adjusted Langevin Algorithm
- $q_\theta(x \to .\,) =$ the distribution of the position of a particle after T seconds, with initial position x and initial momentum $p \sim N(0, \mathbf{I}_d)$, evolving according to Hamiltonian dynamics: HMC

Metropolis et al. [1953]
Hastings [1970]

MCMC: quantities of interest

Recall: Ideal MCMC is *quick to equilibrate* and has *low autocorrelation in equilibrium* (low autocorrelation $\Longrightarrow$ low asymptotic variance, modulo f)

Time to equilibrium of a particular algorithm is measured by the ϵ*-mixing time*:

$$\tau(\epsilon, \mu_0) := \inf\left\{n : d\left(\mathscr{L}(X_n | X_0 \sim \mu_0), \pi\right) \leq \epsilon\right\}$$

Asymptotic variance *and* time to equilibrium strongly depend on the *spectral gap*: defining the *operator* P of the Markov chain, which acts on $L^2(\pi)$

$Pf(x) := E(f(Y))$ where Y is the first state in the Markov chain, started at x.

P has an eigenvalue at 1 (Pconst.=const.) and spectrum$(P) \subset [-1,1]$

The *spectral gap* ρ is the distance between 1 and the nearest point in the spectrum λ_{max} (bigger is better)

$$\sigma_f^2 = \frac{1 + \lambda_{\text{max}}}{\rho} Var_\pi(f)$$

Condition number in MCMC

Target in the form $\pi \propto \exp(-U(x))$ on $\mathbb{R}^d$ such that $m\mathbf{I}_d \leq \nabla_x^2 U(x) \leq M\mathbf{I}_d$ for all $x \in \mathbb{R}^d$:
$U : \mathbb{R}^d \to \mathbb{R}$ is m-*strongly convex* and M-*smooth*

m-strong convexity:

Unimodal

m measures the curvature of $U(x)$

e.g. posterior with concave log-likelihood, Gaussian prior

M-smoothness

$\nabla_x U(x)$ is M-Lipschitz

Discretisations work nicely

Convex quadratic upper and lower bound on $U(x)$

The condition number associated with *sampling from* π is

$$\kappa := \sup_{x\in\mathbb{R}^d} \|\nabla_x^2 U(x)\|_2 \sup_{x\in\mathbb{R}^d} \|\nabla_x^2 U(x)^{-1}\|_2$$

If $m\mathbf{I}_d \leq \nabla_x^2 U(x) \leq M\mathbf{I}_d$ is tight $\kappa = M/m$

As $\kappa \to 1$, the eigenvalues of $\nabla_x^2 U(x)$ get squeezed together, and π starts to look more like an isotropic Gaussian

Importance of the condition number

	Spectral Gap	ϵ-Mixing Time
Upper bounds	∃ $O(\frac{\sqrt{\log d}}{\kappa\sqrt{d}})$ (on a Gaussian) Lee et al. [2021] $O(\frac{\log d}{\kappa d})$ Lee et al. [2021] $O(\frac{\sqrt{\log d}}{\kappa\sqrt{d}})$ (on a Gaussian) Lee et al. [2021]	∀ $O(d\kappa^2 \log\frac{1}{\epsilon})$ Dwivedi et al. [2019] $O(d\kappa \log\frac{1}{\epsilon})$ Dwivedi et al. [2019] $O(d^{\frac{2}{3}}\kappa \log\frac{1}{\epsilon})$ Chen et al. [2019]
Lower bounds	∀ $O(\frac{1}{\kappa d})$ Andrieu et al. [2022]	∃ $O(\frac{\kappa d}{\log^2 d})$ $(\epsilon = e^{-1})$ Lee et al. [2021]

Key: ● - RWM ● - MALA ● - HMC

All bounds up to logarithmic factors, mixing times in TV

Preconditioning in MCMC

Preconditioning involves a process $\{X_i\}$ in $\mathscr{X}$, a process $\{Y_i\}$ in $\mathscr{Y}$, and a transformation $g : \mathscr{X} \to \mathscr{Y}$

We sample Y from a well-conditioned distribution and apply a Metropolis-Hastings accept/reject to $X = g^{-1}(Y)$ such that $\{X_i\}$ forms our samples to use in $\hat{f}_n$

Encapsulates much of adaptive MCMC and therefore generative models: learning a complex distribution is seen as equivalent to learning parameters θ of a map g_θ^{-1} which we apply to samples from a simple distribution

Adaptive MCMC: access to π (unnormalised)

- Sampling via measure transport, Marzouk et al. [2016]
- HMC with Inverse Autoregressive Flows, Hoffman et al. [2019]

Generative Models: access to samples from π

- GANs, Goodfellow et al. [2014]
- Normalizing flows, Papamakarios [2021]

Linear Preconditioning

When $Y = g(X) = LX$ for $L \in GL_d(\mathbb{R})$ the condition number of the distribution of Y is

$$\kappa_L := \sup_{y\in\mathbf{R}^d} \|\nabla_y^2 \tilde{U}(y)\|_2 \sup_{y\in\mathbf{R}^d} \|\nabla_y^2 \tilde{U}(y)^{-1}\|_2 = \sup_{x\in\mathbf{R}^d} \|L^{-T}\nabla_x^2 U(x) L^{-1}\|_2 \sup_{x\in\mathbf{R}^d} \|L\nabla_x^2 U(x)^{-1} L^T\|_2$$

Used in all major MCMC software packages (Stan, Tensorflow, Pyro etc.) even though theory is lacking.

Intuition: set L to be the square root of some *representative* of $\nabla_x^2 U(x)$ i.e. Precision, $\nabla_x^2 U(x*)$ for $x*$ the mode, hope that $\kappa_L \ll \kappa$, doesn't always work:

Diagonal Preconditioning: $L = \text{diag}(\Sigma_\pi)^{-\frac{1}{2}}$

Gaussian target:

$\nabla_x^2 U(x) = \Sigma_\pi^{-1}$ so $\kappa_L = \|\text{diag}(\Sigma_\pi)^{\frac{1}{2}}\Sigma_\pi^{-1}\text{diag}(\Sigma_\pi)^{\frac{1}{2}}\|_2 \|\text{diag}(\Sigma_\pi)^{-\frac{1}{2}}\Sigma_\pi\text{diag}(\Sigma_\pi)^{-\frac{1}{2}}\|_2 = \|C_\pi^{-1}\|_2\|C_\pi\|_2$

There exist Gaussian targets for which $L = \text{diag}(\Sigma_\pi)^{-\frac{1}{2}}$ *increases the condition number*

$$\Sigma_\pi = \begin{pmatrix} 4.07, -3.90, 1.66 \\ -3.90, 3.73, -1.59 \\ 1.66, -1.59, 0.72 \end{pmatrix} \implies \kappa = 23{,}000, \kappa_L = 31{,}000$$

Linear Preconditioning: Bounding κ_L

SVD on L: $L = U\Sigma V^T$, $\Sigma = \text{diag}(\sigma_i : i \in [d])$, $\{v_i : \|v_i\| = 1, i \in [d]\}$ the right singular vectors

$\{(\lambda_i(x), v_i(x)) : \|v_i(x)\| = 1, i \in [d]\}$ the eigenvalue/vector pairs of $\nabla_x^2 U(x)$

Condition 1 (C1): There exists an $\epsilon > 0$ s.t. for all $i \in [d]$ and $x \in \mathbb{R}^d$

$$(1+\epsilon)^{-\frac{1}{2}} \leq \frac{\lambda_i(x)}{\sigma_i^2} \leq (1+\epsilon)^{\frac{1}{2}}$$

Condition 2 (C2): There exists a $\delta > 0$ s.t. for all $i, j \in [d]$ and $x \in \mathbb{R}^d$

$$\|v_i(x) - v_i\| \leq \sqrt{2\delta} \text{ and } \|v_i(x) - v_j\| \geq \sqrt{2(1-\delta)} \text{ for } i \neq j$$

Theorem 1: Assuming C1 and C2 we are able to make the following upper bound

$$\kappa_L \leq (1+\epsilon)\left(1 + \delta\sqrt{\sum_{i=1}^{d} \sigma_i^2 \sum_{i=1}^{d} \sigma_i^{-2}}\right)^4$$

There exist conditions C1', C2' which *only involve* $\lambda_i(x), v_i(x)$ that *imply* C1 and C2

Bounds inform decisions at each stage of the process: pre-check, constructive, verification

Nonlinear Preconditioning

Call κ_g the condition number after general transform $g : \mathcal{X} \to \mathcal{Y}$

Proposition: It is impossible to use linear preconditioning to achieve optimality ($\kappa_g = 1$) when π is not a Gaussian

Proof Sketch: The only distribution with $\kappa = 1$ is an isotropic Gaussian. Assume, seeking a contradiction, that we can linearly transform the state variable of a non-Gaussian to reach a Gaussian. Then we could simply take the inverse of the transform to reach a non-Gaussian from a Gaussian, which is impossible due to closure of Gaussians under linear transformations

Proposition: There exist targets with *arbitrarily high condition number* that *gets worse* under *any linear preconditioning whatsoever* (excluding $L = \mathbf{I}_d$)

Change of variables: $\tilde{U}(g(x)) = U(x) + \log|\det J(g(x))|$ so we need $\frac{1}{2}\|g(x)\|^2 = U(x) + \log|\det J(g(x))|$

which is a particular form of the *Monge-Ampère equation*.

Nonlinear Preconditioning the Langevin Diffusion

Our Contribution

$$dY_t = \frac{1}{2}\nabla_y \log \tilde{\pi}(Y_t)dt + dB_t$$

Defining $f := g^{-1}$ such that $X = f(Y)$, Itô's Lemma gives:

$$dX_t = \frac{1}{2}(J(f(Y_t))\nabla_y \log \tilde{\pi}(Y_t) + L(f(Y_t)))dt + J(f(Y_t))dB_t$$

where $L_i(f(Y_t)) = \Delta_y f_i(Y_t)$. Changing variables, calculus:

$$dX_t = \frac{1}{2}G(X_t)^{-1}\nabla_x \log \pi(X_t)dt + \Gamma(X_t)dt + G(X_t)^{-\frac{1}{2}}dB_t$$

$$\Gamma_i(X_t) = \frac{1}{2}\sum_{j=1}^{d}\frac{\partial}{\partial x_j}\left(G(X_t)^{-1}_{ij}\right)$$

where $G(X_t)^{-1} = J(f(Y_t))J(f(Y_t))^T = (J(g(X_t))^T J(g(X_t))^{-1}$. This is exactly the diffusion on a manifold with contravariant metric $G(X_t)^{-1}$.

Diffusion forms the basis of *Riemannian Manifold* MALA: parameter space as a manifold with *Expected Fisher Information* as metric

Betancourt [2013]: Use $G(X_t)^{-1} = \nabla_x^2 U(X_t)^{-1}$

Xifara et al. [2014]

Livingstone and Girolami [2014]

Girolami and Calderhead [2011]

Rao [1945]

Our Contribution

Nonlinear Preconditioning the Hamiltonian

Recall $p \sim N(0, \mathbf{I}_d)$ so $\nu(p) \propto \left(-\frac{1}{2}p^T p\right)$. Make the transformation $p \to \tilde{p} := f(p)$.

$$\tilde{\nu}(\tilde{p}) \propto \nu(f^{-1}(\tilde{p}))\,|\det J(f^{-1}(\tilde{p})|$$

$$= \exp\left(-\frac{1}{2}f^{-1}(\tilde{p})^T f^{-1}(\tilde{p})\right)|\det J(f(p))|^{-1}$$

$$= \exp\left(-\frac{1}{2}f^{-1}(\tilde{p})^T f^{-1}(\tilde{p}) - \log|\det J(f(p))|\right)$$

In particular $f(p) = \sqrt{G(x)}p$ has a Jacobian $J(f(p)) = \sqrt{G(x)}$ so

$$\tilde{\nu}(\tilde{p}) = \exp\left(-\frac{1}{2}\tilde{p}^T G(x)^{-1}\tilde{p} - \frac{1}{2}\log|\det G(x)|\right)$$

The joint dist$^{\underline{n}}$ targeted by HMC is

$$\pi(x, \tilde{p}) \propto \pi(x)\tilde{\nu}(\tilde{p}\,|\,x) = \exp\left(-U(x) - \frac{1}{2}\tilde{p}^T G(x)^{-1}\tilde{p} - \frac{1}{2}\log|\det G(x)|\right)$$

which has Hamiltonian

$$H(x, \tilde{p}) = U(x) + \frac{1}{2}\tilde{p}^T G(x)^{-1}\tilde{p} + \frac{1}{2}\log|\det G(x)|$$

Girolami and Calderhead [2011]

Unification via Nonlinear Preconditioning

Our Contribution

Recent algorithms inspired by `mirror descent' technique use heuristic in the last slide: simulate process using the Langevin diffusion, and transport to samples using a `mirror map':

Zhang et al. [2020]:

Well-conditioned process: $dY_t = \frac{1}{2}\nabla_x \log \pi(X_t)dt + \nabla_x^2 h(X_t)^{\frac{1}{2}} dB_t$ f map: $X_t = \nabla_y h^*(Y_t)$ (no MH accept/reject)

$h : \mathbb{R}^d \to \mathbb{R}$ is convex, h^* its convex conjugate, $\nabla_y h^* = (\nabla_x h)^{-1}$

Dynamics can be shown to be equivalent to Langevin on a *Hessian Manifold* i.e. a manifold with Hessian metric: $G(X_t)^{-1} = \nabla_x^2 h(X_t)^{-1}$

Chewi et al. [2020] propose using $h = U$, matching the metric proposed in Betancourt [2013]: $G(X_t)^{-1} = \nabla_x^2 U(X_t)^{-1}$

Therefore use a transformation such that $J(g(X)) = \nabla_x^2 U(X)^{\frac{1}{2}}$ (since recall: $G(X_t)^{-1} = J(f(Y_t))J(f(Y_t))^T = (J(g(X_t))^T J(g(X_t))^{-1})$

Nemirovskii and Yudin [1979]
Hsieh and Cevher [2018]
Chewi et al. [2020]

Hessian Based Transformation

Our Contribution

g s.t. $J(g(X)) = \nabla_x^2 U(X)^{\frac{1}{2}}$ makes sense:

$$\nabla_y^2 \tilde{U}(y) = J(g)^{-T} \nabla_x^2 U(x) J(g)^{-1} + J(g)^{-T} \nabla_x^2 \log|\det J(g)| J(g)^{-1} + R$$

$$= \mathbf{I}_d + \nabla_x^2 U(x)^{-\frac{1}{2}} \nabla_x^2 \log|\det J(g)| \nabla_x^2 U(x)^{-\frac{1}{2}} + R$$

R is a remainder involving derivatives of $\nabla_x^2 U(x)^{\frac{1}{2}}$ and $U(x)$.

Go from conditions on $\nabla_x^2 U(x)$ being *global* in the case of linear preconditioning to *local*

Make the guess: $g(X) = \nabla_x^2 U(X)^{\frac{1}{2}} X - c(X)$. Jacobian is $J(g) = \nabla_x^2 U(X)^{\frac{1}{2}} + \partial\left(\nabla_x^2 U(X)^{\frac{1}{2}}, X\right) - J(c)$ where

$\partial\left(\nabla_x^2 U(X)^{\frac{1}{2}}, X\right) \in \mathbb{R}^{d \times d}$ has jth column $\left(\dfrac{\partial}{\partial x_j} \nabla_x^2 U(X)^{\frac{1}{2}}\right) X$

In 1 dimension:

$$c(X) = \sum_{k=2}^{\infty} \frac{(-1)^k x^k}{k!} \frac{\partial^{k-1}}{\partial x^{k-1}} \left(\sqrt{\frac{\partial^2}{\partial x^2} U(x)}\right)$$

$$= x\sqrt{\frac{\partial^2}{\partial x^2} U(x)} - \int_0^x \sqrt{\frac{\partial^2}{\partial t^2} U(t)dt}$$

In d dimensions: need to solve $\partial\left(\nabla_x^2 U(X)^{\frac{1}{2}}, X\right) = J(c)$

Compromise: let $g(X) = \nabla_x^2 U(X)^{\frac{1}{2}} X$

Summary

Introductory Material

- Intro to Preconditioning
  - Condition Number
- Intro to Markov Chain Monte Carlo (MCMC)
  - Defined quantities of interest: spectral gap, ϵ-mixing time
  - Recently introduced bounds on the quantities, polynomial in dimension and condition
- Preconditioning in MCMC
    - Condition Number

Our Contribution

-
    - Linear Preconditioning
      - Global conditions on the Hessian of the potential characterise the effectiveness
      - Bound can be used: as a pre-check, constructively, or for verification
    - Nonlinear Preconditioning
      - Derive Riemannian manifold techniques as an instance of nonlinear preconditioning
      - Identify Mirror Langevin techniques as the same
      - Use these classes to identify nonlinear transformations

Thanks! 🐒

References I

Turing, A. M., (1948). ROUNDING-OFF ERRORS IN MATRIX PROCESSES. The Quarterly Journal of Mechanics and Applied Mathematics, 1(1), 287-308. https://doi.org/10.1093/qjmam/1.1.287

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equation of state calculations by fast computing machines. Journal of Chemical Physics, 21, 1087-1092.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 57(1), 97–109. https://doi.org/10.2307/2334940

Lee, Y. T., & Shen, R., & Tian, K. (2021). Lower Bounds on Metropolized Sampling Methods for Well-Conditioned Distributions. ArXiv, abs/2106.05480.

Chen, Y., & Dwivedi, R., & Wainwright, M. J., & Yu, B. (2019). Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients. ArXiv, abs/1905.12247.

Dwivedi, R., & Chen, Y., & Wainwright, M. J., & Yu, B. (2019). Log-concave sampling: Metropolis-Hastings algorithms are fast. Journal of Machine Learning Research. 20(183). 1-42

Andrieu, C., & Lee, A., & Power, S., & Wang, A. Q. (2022). Explicit convergence bounds for Metropolis Markov chains: isoperimetry, spectral gaps and profiles. ArXiv, abs/2211.08959

References II

Marzouk, Y., & Moselhy, T., & Parno, M., & Spantini, A. (2016). Sampling via Measure Transport: An Introduction. 10.1007/978-3-319-11259-6_23-1.

Hoffman, M., & Sountsov, P., & Dillon, J. V., & Langmore, I., & Tran, D., Vasudevan, S. (2019). NeuTra-lizing Bad Geometry in Hamiltonian Monte Carlo Using Neural Transport. ArXiv, abs/1903.03704

Goodfellow, I., & Pouget-Abadie, J., & Mirza, M., & Xu, B., & Warde-Farley, D., & Ozair, S., & Courville, A. & Bengio, Y.. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems. 3. 10.1145/3422622.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan. B. (2022). Normalizing flows for probabilistic modeling and inference. J. Mach. Learn. Res. 22, 1, Article 57

Livingstone, S., & Girolami, M. (2014). Information-Geometric Markov Chain Monte Carlo Methods Using Diffusions. Entropy. 16. 10.3390/e16063074.

Xifara, T., Sherlock, C., Livingstone, S., Byrne, S., & Girolami, M.A. (2014). Langevin diffusions and the Metropolis-adjusted Langevin algorithm. Statistics & Probability Letters, 91, 14-19.

References III

Girolami, M. and Calderhead, B. (2011), Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73: 123-214. https://doi.org/10.1111/j.1467-9868.2010.00765.x

Rao, C.R. (1945) Information and the Accuracy Attainable in the Estimation of Statistical Parameters. Bulletin of Calcutta Mathematical Society, 37, 81-91.

Betancourt, M. (2012). A General Metric for Riemannian Manifold Hamiltonian Monte Carlo. International Conference on Geometric Science of Information.

Hsieh, Y., & Cevher, V. (2018). Mirrored Langevin Dynamics. ArXiv, abs/1802.10174.

Nemirovskii, A. S. & Yudin, D. B. (1979). Complexity of Problems and Efficiency of Optimization Methods.

Chewi, S., Le Gouic, T., Lu, C., Maunu, T., Rigollet, P. and Stromme, A. (2020). Exponential ergodicity of mirror-Langevin diffusions. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20). Curran Associates Inc., Red Hook, NY, USA, Article 1642, 19573–19585.

Zhang, K. S., Peyré G., Fadili, J. M., Pereyra, M. (2020). Wasserstein Control of Mirror Langevin Monte Carlo. Proceedings of Machine Learning Research, Conference on Learning Theory (COLT). pp. 1-28