Bayesian Methods in Imaging Sciences

Marcelo Pereyra[(1)] and Jean-Yves Tourneret[(2)]

[(1)] Heriot-Watt University

Maxwell Institute for Mathematical Sciences & School of Mathematical and Computer Sciences
Edinburgh, UK, m.pereyra@hw.ac.uk

[(2)] University of Toulouse

ENSEEIHT-IRIT-TéSA
Toulouse, France, jean-yves.tourneret@enseeiht.fr

Wednesday, February 6, 2019

Outline

Bayesian Inference

Posterior Distribution

$$\pi(\boldsymbol{x}) \triangleq p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}) = \frac{p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})p(\boldsymbol{x};\boldsymbol{\theta})}{p(\boldsymbol{y};\boldsymbol{\theta})}$$

Notations

- $\boldsymbol{x} = [x_1, \ldots, x_N]^T$: unknown vector of interest
- $\boldsymbol{y} = [y_1, \ldots, y_M]^T$: observation vector associated with $\boldsymbol{x}$
- $\boldsymbol{\theta}$: vector gathering the deterministic parameters and hyperparameters of the statistical model

Vocabulary

- $p(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$: likelihood of the statistical model
- $p(\boldsymbol{x};\boldsymbol{\theta})$: prior distribution assigned to the vector $\boldsymbol{x}$
- $p(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta})$: posterior distribution of interest

Bayesian Inference

Many interesting properties

- Possibility of computing uncertainty measures such as confidence intervals
- Multiple estimators of $\boldsymbol{x}$: maximum a posteriori (MAP), minimum mean square error (MMSE), posterior median (robustness), ...
- Model selection: determine the model order, the number of unknown parameters, ...

Denoising

Problem of interest

$$\arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y}-\boldsymbol{x}\|^2 + \lambda\phi(\boldsymbol{x})$$

- Various regularizations: TV, ℓ_1, ℓ_p, ...
- Other data fidelity terms might be considered

Noisy

TV Denoised

Deconvolution

Problem of interest

$$\arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \lambda\phi(\boldsymbol{x})$$

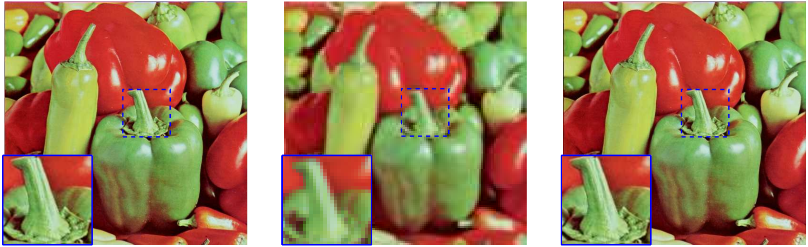
- $\boldsymbol{H}$ is a blurring operator
- Possibility of considering various regularizations: TV, ℓ_1, ℓ_p, ...

Other applications

Super-resolution, compressed sensing

$$\arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{H}\boldsymbol{x}\|^2 + \lambda\phi(\boldsymbol{x})$$

where $\boldsymbol{S}$ is a decimation matrix, a sensing matrix, ...



Ground truth (left), Observed image (middle), Reconstruction (right).

Outline

- Part 1: Inverse Problems for Image Processing
- Part 2: The Gibbs Sampler: Blocking, Moving, Collapsing
- Part 3: Langevin and Hamiltonian MCMC
- Part 4: Proximal MCMC Algorithms
- Part 5: Conclusion

The Gibbs Sampler

General Principle

To sample according to a distribution $\pi(\boldsymbol{x})$ with $\boldsymbol{x} = (x_1, ..., x_N)$, one can use the following idea

- Initialization: generate a vector $\boldsymbol{x} = (x_1, ..., x_N)$ according to an initial proposal π_0
- Sample according to the full conditional distributions of the target distribution π

$$\pi_i(x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)$$

  for $i = 1, 2, \ldots, N$.

Remarks

- Asymptotic convergence to the distribution of interest $\pi(\boldsymbol{x})$
- Requires to know the conditional distributions of π
- Acceptance rate of each draw equal to 1.

The Gibbs Sampler

Limitations

- Variables x_i strongly correlated
- High-dimensional vector $\boldsymbol{x}$
- The conditional distributions can be known but difficult to sample
- Difficulties to escape from local minima of $\pi(\boldsymbol{x})$

References

- C. Y. Chi, J. M. Mendel, Improved maximum likelihood detection and estimation of Bernoulli-Gaussian processes, IEEE Trans. Inf. Theory, vol. 30, pp. 429-434, March 1984.
- M. Lavielle, Optimal segmentation of random processes, IEEE Trans. Signal Process., vol. 46, no 5, May 1998.
- S. Bourguignon, H. Carfantan, Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data, in Proc. SSP, Bordeaux, France, 2005.
- T. Veit, J. Idier, Rééchantillonnage de l'échelle dans les algorithmes MCMC pour les problèmes inverses bilinéaires, in Proc. GRETSI, Troyes, 2007.
- G. Kail, J.-Y. Tourneret, N. Dobigeon and F. Hlawatsch, "Blind Deconvolution of Sparse Pulse Sequences under a Minimum Distance Constraint: A Partially Collapsed Gibbs Sampler Method," IEEE Trans. Sig. Process., vol. 60, no. 6, pp. 2727-2743, June 2012.

The Gibbs Sampler

Simple tricks

- Block Gibbs sampler
- Use appropriate moves to accelerate the convergence

Metropolis-within-Gibbs sampler

Given $\boldsymbol{x}^{(t)}$,

1. Sample according to the proposal $\boldsymbol{z}_t \sim q(\boldsymbol{z}|\boldsymbol{x}^{(t)})$.
2. Acceptance-Rejection

$$\boldsymbol{x}^{(t+1)} = \begin{cases} \boldsymbol{z}_t & \text{with prob.} \quad \rho(\boldsymbol{x}^{(t)}, \boldsymbol{z}_t) \\ \boldsymbol{x}^{(t)} & \text{with prob.} \quad 1 - \rho(\boldsymbol{x}^{(t)}, \boldsymbol{z}_t) \end{cases}$$

with

$$\rho(\boldsymbol{x}, \boldsymbol{z}) = \min \left\{ \frac{\pi(\boldsymbol{z})}{\pi(\boldsymbol{x})} \, \frac{q(\boldsymbol{x}|\boldsymbol{z})}{q(\boldsymbol{z}|\boldsymbol{x})}, 1 \right\} .$$

Example: Spectral Analysis of Astrophysical Data

Reference

- S. Bourguignon, H. Carfantan, Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data, in Proc. SSP, Bordeaux, France, 2005.

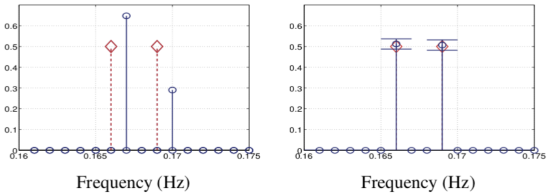


Fig. 5. Simulation results with 2 close spectral lines ($\diamond$). Left: SMLR solution. Right: $\widehat{\boldsymbol{X}} \pm \sigma_{\widehat{\boldsymbol{X}}}$.

Partially Collapsed Gibbs Sampler (PCGS)

General Principles

Three operations that do not change the asymptotic distribution

- **Marginalization**: replace a conditional distribution of π by sampling a variable that was conditioned, e.g.,

$$\text{replace } \pi(A|B,C) \text{ by } \pi(A,B|C)$$

- **Permutation**
- **Trimming**: remove some consecutive draws of variables when these variables are not conditioned

Références

- D. A. Van Dyk and T. Park, "Partially Collapsed Gibb Samplers: Theory and Methods," J. American Statistical Association, vol. 103, pp. 70-796, 2008.
- T. Park and D. A. Van Dyk, "Partially Collapsed Gibb Samplers: Illustrations and Applications," J. Computational Graphical Statistics, vol. 18, pp. 283-305, 2009.

Partially Collapsed Gibbs Sampler (PCGS)

Standard Gibbs Sampler

- $\pi(A|B,C)$
- $\pi(B|A,C)$
- $\pi(C|A,B)$

Marginalization

- $\pi(A,C|B)$
- $\pi(B|A,C)$
- $\pi(C|A,B)$

Permutation

- $\pi(A,C|B)$
- $\pi(C|A,B)$
- $\pi(B|A,C)$

Partially Collapsed Gibbs Sampler (PCGS)

Trimming and permutation

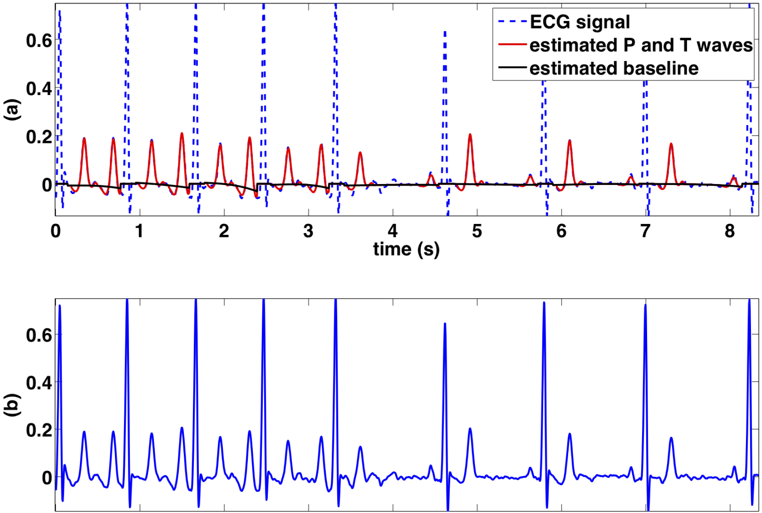
- $\pi(A|B)$
- $\pi(B|A, C)$
- $\pi(C|A, B)$

Remarks

- The variable C has disappeared in the first simulation, which can accelerate convergence
- Necessity of being able to marginalize with respect to the variable C
- Example of application

  C. Lin, C. Mailhes and J.-Y. Tourneret, "P- and T-Wave Delineation in ECG Signals Using a Bayesian Approach and a Partially Collapsed Gibbs Sampler," IEEE Trans. Biomed. Eng., vol. 57, no. 12, pp. 2840 - 2849, Dec. 2010.

ECG Delineation



Typical example

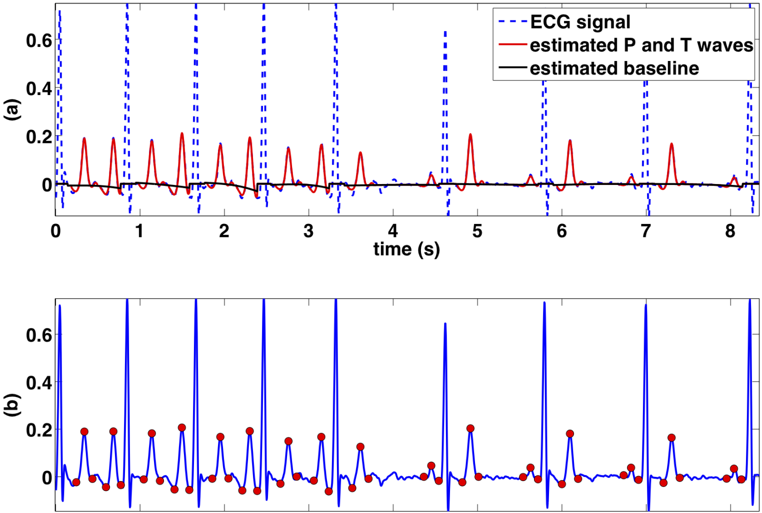


Illustration of improved convergence for the PCGS

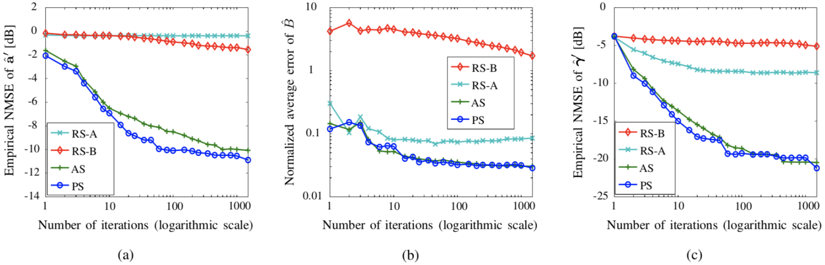


Fig. 3. Detection/estimation performance versus the number of iterations: (a) Empirical NMSE of $\hat{\mathbf{a}}'$, (b) normalized average error of $\hat{B} = \|\hat{\mathbf{b}}\|^2$, (c) empirical NMSE of $\hat{\boldsymbol{\gamma}}'$.

Alternatives

Other ideas

- Simulated Tempering: introduce a "temperature" as in simulated annealing, i.e., consider a sequence of distributions

$$\pi_i(\boldsymbol{x}) = \frac{1}{Z_i} \exp\left(-\frac{\pi(\boldsymbol{x})}{T_i}\right)$$

- Exchange some information from several chains generated in parallel Population Markov Chain Monte Carlo, Metropolis Coupled Markov Chain Monte Carlo (MCMCMC), ...
- Population Monte Carlo
- ...

Alternatives

References

- O. Cappé, A. Guillin, J-M. Marin, and C. P. Robert. Population Monte Carlo. J. Comput. Graph. Statist., vol. 13, no. 4, pp. 907-929, 2004.
- Radford M. Neal, "Sampling from Multimodal Distributions Using Tempered Transitions," Statist. Comput., vol. 6, no. 4, pp. 353-366, Dec. 1992.
- P. Chen, James D. B. Nelson and J.-Y. Tourneret, "Toward a Sparse Bayesian Markov Random Field Approach to Hyperspectral Unmixing and Classification," IEEE Trans. Image Process., vol. 26, no. 1, pp. 426-438, Jan. 2017.
- C. J. Geyer and E. A. Thompson, "Annealing Markov chain Monte Carlo with applications to ancestral inference," J. Amer. Stat. Soc., vol. 90, no. 431, pp. 909-920, 1995.
- K. B. Laskey and J. W. Myers, "Population Markov Chain Monte Carlo," Mach. Learn., vol. 50, pp. 175-196, 2003.
- F. Costa, H. Batatia, T. Oberlin, C. D'Giano and J.-Y. Tourneret, "Bayesian EEG Source Localization Using a Structured Sparsity Prior," Neuroimage, vol. 144, Part A, pp. 142-152, Jan. 2017.

Outline

Special Issue "Stochastic Simulation and Optimization in Signal Processing" (S. McLaughlin, M. Pereyra, A. O. Hero, J.-Y. Tourneret and J.-C. Pesquet)

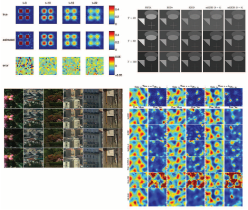
IEEE JOURNAL OF
SELECTED TOPICS IN SIGNAL PROCESSING

IEEE Signal Processing Society

www.ieee.org/sp/index.html

FEBRUARY 2016 VOLUME 10 NUMBER 1 IJSTGY (ISSN 1932-4553)
MARCH 2016 VOLUME 10 NUMBER 2

ISSUE ON STOCHASTIC SIMULATION AND OPTIMIZATIONS IN SIGNAL PROCESSING



Clockwise from top left: Fig. 7 from the paper by J. Murphy and S. J. Godsill, p. 328; Fig. 7 from the paper by J Konečný *et al.*, p. 242; Fig. 7 from the paper by F. Septier, and G. W. Peters., p. 312; Fig. 9 from the paper by J. Tan *et al*, p. 389.

For the February 2016 issue, see p. 1 for Table of Contents.
For the March 2016 issue, see p. 219 for Table of Contents.

Outline

Monte Carlo Methods Based on the Langevin Diffusion

Langevin diffusion on $\mathbb{R}^N$

$$dX(t) = \frac{1}{2}\nabla \log \pi \left[X(t)\right] dt + dW(t), \quad X(0) = \boldsymbol{x}_0 \in \mathbb{R}^N, \tag{1}$$

where W is a Brownian motion on $\mathbb{R}^N$.

Under appropriate conditions, $X(t)$ converges in distribution to π when $t \to \infty$, and can thus lead to an interesting sampling strategy for π.

Remark 1: Good convergence properties when $-\log \pi$ is strongly convex, even in very high dimension.

Remark 2: Slow convergence when π is heavy-tailed (e.g., if $X(t)$ is assigned an ℓ_q prior with $q < 1$).

Monte Carlo Methods Based on the Langevin Diffusion

Unfortunately, sampling $X(t)$ according to the previous differential equation is generally difficult.

We can consider a discrete approximation, e.g., Euler-Maruyama

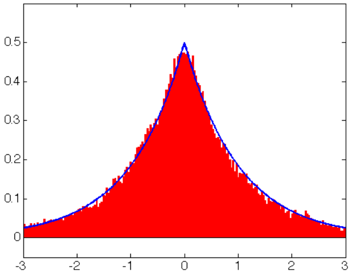
$$X^{(t+1)} = X^{(t)} + \frac{\delta}{2}\nabla \log \pi \left(X^{(t)}\right) + \sqrt{\delta} Z_{m+1}, \quad Z_{m+1} \sim \mathcal{N}(0, \mathbb{I}_N) \qquad (2)$$

where δ is a discretization parameter.

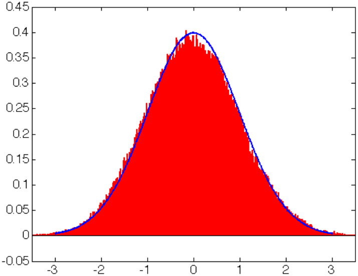
Assuming some regularity conditions for π and δ, fast convergence of (2) to a distribution close to π [Durmus and Moulines, 2015].

Numerical illustrations

Histograms obtained for a sample size equal to $10\,000$ generated by ULA.



$\pi(x) \propto \exp(-|x|)$



$\pi(x) \propto \exp(-x^2)$

Metropolis Adjusted Langevin Algorithm (MALA)

In MALA, the approximation error is corrected by an MH step ensuring that $\pi(\boldsymbol{x})$ is the invariant distribution of the Markov chain.

This acceptance step reduces the asymptotic bias and increases the variance of the generated sample. Thus there is a possible increase of the mean square error at a given time instant.

Good convergence properties are obtained for an acceptance rate $\rho(\delta) \approx 0.6$.

To adjust δ automatically, one can introduce in MALA a stochastic optimization method to minimize the energy $(\rho(\delta) - 0.6)^2$, leading to

$$X^{(t+1)} \sim K_{\delta_t}\left(\cdot | X^{(t)}\right)$$

$$\delta_{t+1} = \delta_t + \gamma_{t+1}[\delta_t - (\rho_{\mathrm{MH}}(t+1) - 0.6)]$$

where K_δ is the MALA kernel with a stepsize δ, $\rho_{\mathrm{MH}}(t)$ is the acceptance ratio of the MH step at iteration t, and $\{\gamma_t\}_{t=1}^{\infty}$ is a decreasing sequence.

Riemannian MALA

Improve the convergence speed of MALA by replacing δ by a matrix $\boldsymbol{\Sigma}(\boldsymbol{x})$ leading to the following update

$$X^{(t+1)} = X^{(t)} + \boldsymbol{\Sigma}\left(X^{(t)}\right)\nabla \log \pi\left(X^{(t)}\right) + \sqrt{2\boldsymbol{\Sigma}\left(X^{(t)}\right)}Z_{m+1} \quad (3)$$

$$Z_{m+1} \sim \mathcal{N}(0, \mathbb{I}_N)$$

This update can be obtained by a Langevin diffusion on a Riemannian Manifold with a metric defined by the matrix $\boldsymbol{\Sigma}(\boldsymbol{x})$ [Girolami and Calderhead, 2011].

Riemannian and Euclidean gradients are related by $\tilde{\nabla} g(\boldsymbol{x}) = \boldsymbol{\Sigma}(\boldsymbol{x})\nabla g(\boldsymbol{x})$. Idea close to gradient preconditioning in optimization.

Riemannian and Adaptive MALA

Standard choices of matrices $\boldsymbol{\Sigma}$

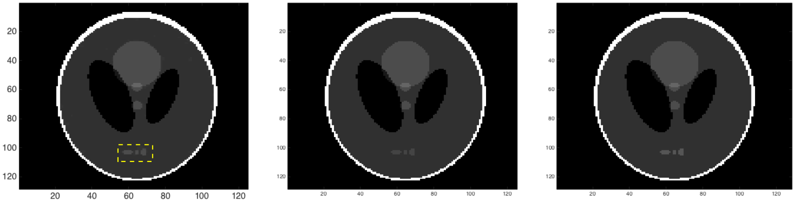
1. Inverse Fisher information matrix ("natural" metric) $\Longleftrightarrow$ optimization by natural gradient [Girolami and Calderhead, 2011].
2. Positive semidefinite version of the inverse Hessian matrix [Zhang and Sutton, 2011] [Betancourt, 2013] $\Longleftrightarrow$ Newton optimization.
3. Inverse curvature of a quadratic majorant [Marnissi et al., 2014] $\Longleftrightarrow$ Optimization by majoration-minimization.
4. Optimise $\boldsymbol{\Sigma}$ online to learn the covariance matrix associated with $\pi(x)$ [Atchadé, 2006].

Simulation results

2D tomographic inversion - robust total variation prior

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left[-\|\boldsymbol{y} - \Phi\mathcal{F}\boldsymbol{x}\|^2/2\sigma^2 - \beta\rho_H(\|\nabla_d \boldsymbol{x}\|_2)\right]$$

An adaptive MALA algorithm is used to compute the confidence region $C_\alpha^* = \{\boldsymbol{x} : p(\boldsymbol{x}|\boldsymbol{y}) \geq \gamma_\alpha\}$ such that $\mathrm{P}\left[\boldsymbol{x} \in C_\alpha|\boldsymbol{y}\right] = 1 - \alpha$, which can be used as a measure of uncertainty for some specific parts of the image.



A posteriori mean (tumor intensity: 0.30) | lower bound (tumor intensity: 0.27) | upper bound (tumor intensity: 0.33)

Hamiltonian Monte Carlo (HMC) Method

Auxiliary Gaussian vector $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ defined in $\mathbb{R}^N$.

Augmented distribution $\pi(\boldsymbol{x}, \boldsymbol{w}) \propto \pi(\boldsymbol{x}) \exp(-\frac{1}{2}\boldsymbol{w}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{w})$, whose marginal distribution is the target distribution $\pi(\boldsymbol{x})$.

The HMC method is based on the property according to which the trajectories defined by "Hamiltonian dynamics" preserve the level sets of $\pi(\boldsymbol{x}, \boldsymbol{w})$.

Hamiltonian Monte Carlo Method

An initial point $(\boldsymbol{x}_0, \boldsymbol{w}_0) \in \mathbb{R}^{2N}$ for the differential equations

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= -\nabla_{\boldsymbol{w}} \log \pi(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{\Sigma}^{-1}\boldsymbol{w} \\ \frac{\mathrm{d}\boldsymbol{w}}{\mathrm{d}t} &= \nabla_{\boldsymbol{x}} \log \pi(\boldsymbol{x}, \boldsymbol{w}) = \nabla_{\boldsymbol{x}} \log \pi(\boldsymbol{x}) \end{aligned} \tag{4}$$

generates a point $(\boldsymbol{x}_t, \boldsymbol{w}_t)$ such that $\pi(\boldsymbol{x}_t, \boldsymbol{w}_t) = \pi(\boldsymbol{x}_0, \boldsymbol{w}_0)$. In other words, the deterministic Hamiltonian proposal admits $\pi(\boldsymbol{x}, \boldsymbol{w})$ as invariant distribution.

Combining (4) with the sampling step $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, whose invariant distribution is $\pi(\boldsymbol{x}, \boldsymbol{w})$, produces an ergodic Markov chain.

To obtain vectors distributed according to $\pi(\boldsymbol{x})$, the augmented state $(\boldsymbol{x}^{(t)}, \boldsymbol{w}^{(t)})$ can be projected onto the original space by removing $\boldsymbol{w}^{(t)}$.

Hamiltonian equations cannot be solved analytically.

Leap-frog approximation [Neal, 2013]

$$\begin{aligned} \boldsymbol{w}^{(t+\delta/2)} &= \boldsymbol{w}^{(t)} + \frac{\delta}{2}\nabla_{\boldsymbol{x}} \log \pi\left(\boldsymbol{x}^{(t)}\right) \\ \boldsymbol{x}^{(t+\delta)} &= \boldsymbol{x}^{(t)} + \delta \boldsymbol{\Sigma}^{-1}\boldsymbol{w}^{(t+\delta/2)} \\ \boldsymbol{w}^{(t+\delta)} &= \boldsymbol{w}^{(t+\delta/2)} + \frac{\delta}{2}\nabla_{\boldsymbol{x}} \log \pi\left(\boldsymbol{x}^{(t+\delta)}\right) \end{aligned} \tag{5}$$

where the parameter δ is used to control the discretization stepsize.

The approximation error is corrected by an MH step ensuring that $\pi(\boldsymbol{x}, \boldsymbol{w})$ is the invariant distribution of the Markov chain.

Remark: if $\delta = t$, HMC and MALA algorithms are equivalent.

Example: Image Restoration with Poisson Noise

Scaling properties of several samplers

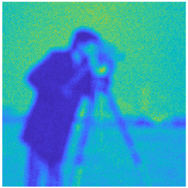
- Unadjusted Langevin algorithm (ULA)
- Metropolis adjusted Langevin algorithm (MALA)
- Hamiltonian Monte Carlo (HMC)
- No U-turn Hamiltonian Monte Carlo (NUTS)
- Bouncy particle sampler (BPS)
- Non-reversible rejection-free strategy

Reference

- J. Tachella *et al.*, Bayesian Restoration of High-Dimensional Photon-Starved Images, in Proc. Eusipco, Roma, Italy, 2019.

Image Restoration with Poisson Noise

Ground Truth.



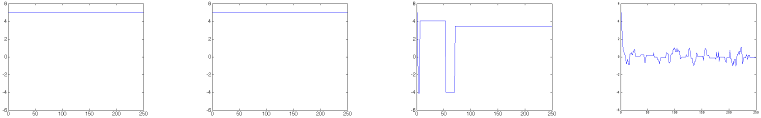
Noisy Image.

Restored Image.

Outline

Limitations of Langevin and Hamiltonian MCMC Algorithms

- Geometric convergence of ULA, MALA and HMC is only guaranteed when $\nabla \log \pi$ is Lipchitz continuous with a Lipchitz constant $L > 2\delta^{-1}$.
- For example, MALA and HMC can fail, e.g., when $\pi(x) \propto \exp\left(-\gamma |x|^q\right)$ with $q > 2$, or $q = 2$ and $\delta > 2\gamma^{-1}$.



MALA HMC MALTA S-MMALA

Generation according to $\pi(x) \propto \exp\{-x^4\}$ with MALA, HMC, truncated MALA [Roberts and Tweedie, 1996], and Riemannian MALA (S-MMALA) [Girolami and Calderhead, 2011].

Proximal Langevin Algorithms

Proximal Langevin Algorithms use a regularized version of Langevin diffusion [Pereyra, 2015, Durmus et al., 2016]

$$\boldsymbol{X}^{\lambda}: \quad \mathrm{d}\boldsymbol{X}_t^{\lambda} = \frac{1}{2}\nabla \log \pi_{\lambda}\left(\boldsymbol{X}_t^{\lambda}\right)\mathrm{d}t + \mathrm{d}W_t, \quad 0 \leq t \leq T, \quad \boldsymbol{X}^{\lambda}(0) = x_0,$$

where $\log \pi_{\lambda}$ is the concave Moreau envelop of $\log \pi$

$$\log \pi_{\lambda}(\boldsymbol{x}) = \sup_{\boldsymbol{u}\in\mathbb{R}^d} \left[\log \pi(\boldsymbol{u}) - (2\lambda)^{-1}\|\boldsymbol{u}-\boldsymbol{x}\|_2^2\right] .$$

Remark 1: if $\log \pi$ is concave, then $\log \pi_{\lambda}(\boldsymbol{x})$ is λ-Lipchitz differentiable.

Remark 2: $\boldsymbol{X}^{\lambda} \to \boldsymbol{X}$ when $\lambda \to 0$, which provides an interesting strategy to sample approximately according to π.

Proximal Langevin Algorithms

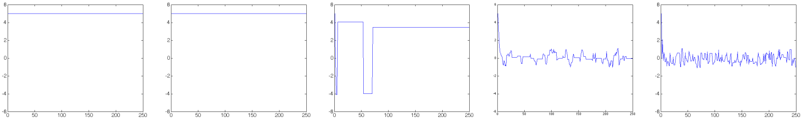
The proximal ULA algorithm is defined from this discrete approximation of $\boldsymbol{X}^{\lambda}$

$$X_{m+1}^{\lambda} = (1 - \tfrac{\delta}{\lambda}) X_m^{\lambda} + \tfrac{\delta}{\lambda} \operatorname{prox}_{\log \pi}^{\lambda}\{X_m^{\lambda}\} + \sqrt{2\delta} Z_{m+1}$$

based on the equality $\nabla \log \pi_{\lambda}(\boldsymbol{x}) = [\boldsymbol{x} - \operatorname{prox}_{\log \pi}^{\lambda}(\boldsymbol{x})]/\lambda$, where

$$\operatorname{prox}_{\log \pi}^{\lambda} = \arg\max_{\boldsymbol{u} \in \mathbb{R}^d} \left[\log \pi(\boldsymbol{u}) - (2\lambda)^{-1} \|\boldsymbol{u} - \boldsymbol{x}\|_2^2 \right] .$$

In the proximal MALA algorithm, the approximation error is corrected at each MH step with the target distribution π.



MALA HMC MALTA S-MMALA Prox. MALA

Generation according to $\pi(x) \propto \exp\{-x^4\}$ avec MALA, HMC, truncated MALA [Roberts and Tweedie, 1996], Riemannian MALA (S-MMALA) [Girolami and Calderhead, 2011], and proximal MALA [Pereyra, 2015].

Outline

- Part 1: Inverse Problems for Image Processing
- Part 2: The Gibbs Sampler: Blocking, Moving, Collapsing
- Part 3: Langevin and Hamiltonian MCMC
- Part 4: Proximal MCMC Algorithms
- Part 5: Conclusion

Conclusion

The main stochastic simulation methods piloted by optimization include

- Langevin MCMC
- Hamiltoninan MCMC
- Proximal MCMC

Optimization will be clearly important in the near future to build new MCMC methods adapted to high-dimensional problems.

Thanks for your attention!

Assistant Professor Position in Medical Imaging in the University of Toulouse (Oct. 2019). Please contact me!

Bibliography:

Atchadé, Y. (2006).
An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift.
Methodology and Computing in Applied Probability, 8(2):235–254.

Betancourt, M. (2013).
A general metric for Riemannian manifold Hamiltonian Monte Carlo.
In Nielsen, F. and Barbaresco, F., editors, *National Conference on the Geometric Science of Information*, Lecture Notes in Computer Science 8085, pages 327–334. Springer.

Durmus, A. and Moulines, E. (2015).
Non-asymptotic convergence analysis for the unadjusted langevin algorithm.
Accepted for publication in Ann. Appl. Probab. 1507.05021, arXiv.

Durmus, A., Moulines, E., and Pereyra, M. (2016).
Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau.
ArXiv e-prints.

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.

Marnissi, Y., Benazza-Benyahia, A., Chouzenoux, E., and Pesquet, J.-C. (2014).
Majorize-minimize adapted Metropolis Hastings algorithm. application to multichannel image recovery.
In *Proc. European Signal Process. Conf. (EUSIPCO 2014)*, pages 1332–1336, Lisbon, Portugal.

Neal, R. (2013).
MCMC using Hamiltonian dynamics.
In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, pages 113–162. Chapman & Hall/CRC Press.

Pereyra, M. (2015).
Proximal Markov chain Monte Carlo algorithms.
Statistics and Computing.
open access paper, http://dx.doi.org/10.1007/s11222-015-9567-4.

Roberts, G. O. and Tweedie, R. L. (1996).
Exponential convergence of Langevin distributions and their discrete approximations.
Bernulli, 2(4):341–363.

Zhang, Y. and Sutton, C. A. (2011).
Quasi-Newton methods for Markov chain Monte Carlo.
In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 24 (NIPS 2011)*, pages 2393–2401.