Auxiliary gradient-based sampling algorithms

Sam Power

Cambridge Centre for Analysis
Cantab Capital Institute for the Mathematics of Information

sp825@cam.ac.uk

August 22, 2018

Summary of Paper

- Comparison between auxiliary and marginal samplers
- Dominance result
  - Marginal Sampler $\succcurlyeq$ Auxiliary Sampler
    - Ordered by asymptotic variance (Peskun)
- Scheme for devising both auxiliary and marginal samplers
  - Context for existing algorithms (MALA variants, pCN variants, $\cdots$)
  - Introduces new algorithms (aGrad, mGrad, $\cdots$)
- Several examples and experiments

Notation

- Throughout:
  - Targeting a measure with density $\pi(x) \propto \exp(-\Phi(x))$
  - Current location: x
  - Intermediate / Auxiliary location: $u \sim q^{\textbf{Int}}(u|x)$
  - Proposed new location: $y \sim q^{\textbf{Aux}}(y|x, u)$ or $\sim q^{\textbf{Marg}}(y|x)$
  - Σ will denote the covariance matrix of some Gaussian random variable
  - δ will be a step size or similar
  - The gradient, Hessian of Φ will be denoted g, H respectively

Auxiliary Samplers - Core Idea

- Sample $u|x \sim q^{\textbf{Int}}(u|x)$ to be close to x
- Target the extended measure $\tilde{\pi}(x, u) = \pi(x) q^{\textbf{Int}}(u|x)$
- Sampling $x|u$ is typically *intractable* $\cdots$
- $\cdots$ so, use an approximation: for $y \approx x$

$$\pi(y) \approx \pi^{A}(y; x) \tag{1}$$

$$\tilde{\pi}(y|u) \approx \pi^{A}(y; x) \times q^{\textbf{Int}}(u|y) =: q^{\textbf{Aux}}(y|x, u) \tag{2}$$

- If $(\pi^{A}, q^{\textbf{Int}})$ are chosen compatibly, we can use $q^{\textbf{Aux}}$ as a proposal
- Then, Metropolise:

$$\mathrm{acc}(x \to y|u) = 1 \wedge \frac{\tilde{\pi}(y|u) \cdot q^{\textbf{Aux}}(x|y, u)}{\tilde{\pi}(x|u) \cdot q^{\textbf{Aux}}(y|x, u)} \tag{3}$$

Marginal Samplers

- We don't actually care about u - can we get rid of it?
- The y we propose is drawn from

$$q^{\textbf{Marg}}(y|x) = \int q^{\textbf{Aux}}(y|x,u)q^{\textbf{Int}}(u|x)du \tag{4}$$

- If this is tractable, we can forget u and Metropolise with respect to this instead

$$\mathrm{acc}(x \to y) = 1 \wedge \frac{\pi(y) \cdot q^{\textbf{Marg}}(x|y)}{\pi(y) \cdot q^{\textbf{Marg}}(y|x)} \tag{5}$$

- $\rightsquigarrow$ Marginal samplers!
- Theorem: Marginal sampler $\succcurlyeq$ Auxiliary sampler.

Algorithm Design

- How do we choose $\pi^A, q^{\textbf{Int}}$?
- Simplest π^A: log-linear approximation

$$\begin{aligned} \pi(y) &= \exp(-\Phi(y)) && (6) \\ &\approx \exp(-\Phi(x) - \langle g, y - x \rangle) && (7) \end{aligned}$$

- Choose $q^{\textbf{Int}}$ so that $\pi^A(y; x) \times q^{\textbf{Int}}(u|y)$ can be sampled easily (in y)
- Let $q^{\textbf{Int}}(u|x) = \mathcal{N}(x, \delta\Sigma)$, then

$$q^{\textbf{Aux}}(y|x, u) = \mathcal{N}(u - \delta\Sigma g, \delta\Sigma) \tag{8}$$

- Marginal sampler is then

$$q^{\textbf{Marg}}(y|x) = \mathcal{N}(x - \delta\Sigma g, 2\delta\Sigma) \tag{9}$$

a.k.a. preconditioned MALA

General Constructions

- When π has the form

$$\pi(x) \propto \mathcal{N}(x|0, \Sigma) \times \exp(-\Phi(x)) \tag{10}$$

then we can get tighter quadratic approximations:

$$\pi^A(y; x) \propto \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y - \langle g, y - x \rangle\right) \tag{11}$$

and better proposals.

- $\rightsquigarrow$ MALA, pCN, pCNL, ESS (sort of)
- Variety of different algorithms from different choices of π^A, $q^{\textbf{Int}}$
  - $\cdots$ mostly use linear approximations to Φ, Gaussian $q^{\textbf{Int}}$
  - $\Phi \approx$ **Quadratic** mentioned in appendix; not pursued (due to cost)

Auxiliary MALA, General Target

- Define the auxiliary variable by a Gaussian perturbation of x

$$q^{\textbf{Int}}(u|x) = \mathcal{N}(x, \delta I) \tag{12}$$

- Define the local approximation by

$$\pi^{A}(y; x) \propto \exp\left(-\langle g, y - x\rangle\right) \tag{13}$$

- This leads to the auxiliary sampler

$$q^{\textbf{Aux}}(y|x, u) = \mathcal{N}(u - \delta g, \delta I) \tag{14}$$

- which marginalises to

$$q^{\textbf{Marg}}(y|x) = \mathcal{N}(x - \delta g, 2\delta I) \tag{15}$$

Auxiliary MALA, Gaussian change-of-measure

- Define the auxiliary variable by a Gaussian perturbation of x

$$q^{\textbf{Int}}(u|x) = \mathcal{N}(x, \delta I) \tag{16}$$

- Define the local approximation by

$$\pi^A(y; x) \propto \exp\left(-\frac{1}{2}y^T C^{-1} y - \langle g, y - x\rangle\right) \tag{17}$$

- This leads to the auxiliary sampler

$$A \triangleq C(C + \delta I)^{-1} \tag{18}$$

$$q^{\textbf{Aux}}(y|x, u) = \mathcal{N}\left(A(u - \delta g), \delta A\right) \tag{19}$$

- which marginalises to

$$q^{\textbf{Marg}}(y|x) = \mathcal{N}\left(A(x - \delta g), \delta\left[A + A^2\right]\right) \tag{20}$$

Auxiliary pMALA, Gaussian change-of-measure

- Define the auxiliary variable by a Gaussian perturbation of x

$$q^{\mathbf{Int}}(u|x) = \mathcal{N}(x, \delta\Sigma) \tag{21}$$

- Define the local approximation by

$$\pi^{A}(y; x) \propto \exp\left(-\frac{1}{2}y^{T}C^{-1}y - \langle g, y - x\rangle\right) \tag{22}$$

- This leads to the auxiliary sampler

$$J \triangleq I + \delta\Sigma^{1/2}C^{-1}\Sigma^{1/2} \tag{23}$$

$$B \triangleq \Sigma^{1/2}J^{-1}\Sigma^{-1/2} \tag{24}$$

$$q^{\mathbf{Aux}}(y|x, u) = \mathcal{N}\left(B(u - \delta\Sigma g), \delta B\Sigma\right) \tag{25}$$

- which marginalises to

$$q^{\mathbf{Marg}}(y|x) = \mathcal{N}\left(B(x - \Sigma\delta g), \delta\left[B + B^{2}\right]\Sigma\right) \tag{26}$$

Auxiliary Autoregressive, Gaussian change-of-measure

- Define the auxiliary variable by a Gaussian perturbation of x

$$q^{\mathbf{Int}}(u|x) = \mathcal{N}(Fx, (I - F^2)C) \tag{27}$$

where F is symmetric, commutes with C

- Define the local approximation by

$$\pi^A(y;x) \propto \exp\left(-\frac{1}{2}y^T C^{-1} y - \langle g, y - x\rangle\right) \tag{28}$$

- This leads to the auxiliary sampler

$$q^{\mathbf{Aux}}(y|x,u) = \mathcal{N}\left(Fu - (I - F^2)Cg, (I - F^2)C\right) \tag{29}$$

- which marginalises to

$$q^{\mathbf{Marg}}(y|x) = \mathcal{N}\left(F^2 u - (I - F^2)Cg, (I - F^4)C\right) \tag{30}$$

Second-Order Construction: Auxiliary

- If we choose

$$\pi^{A}(y;x) \propto \exp\left(-\langle g, y-x\rangle - \frac{1}{2}\|y-x\|_{H}^{2}\right) \tag{31}$$

$$q^{\mathbf{Int}}(u|x) = \mathcal{N}(x, \delta\Sigma) \tag{32}$$

and write

$$J \triangleq I + \delta\Sigma^{1/2}H\Sigma^{1/2} \tag{33}$$

$$B \triangleq \Sigma^{1/2}J^{-1}\Sigma^{-1/2} \tag{34}$$

$$\mu = B(u + \delta\Sigma\{Hx - g\}) \tag{35}$$

then we obtain the auxiliary sampler

$$q^{\mathbf{Aux}}(y|x,u) = \mathcal{N}(\mu, \delta B\Sigma) \tag{36}$$

Second-Order Construction: Marginal

- This leads to the marginal sampler

$$q^{\textbf{Marg}}(y|x) = \mathcal{N}(x - \delta B \Sigma g, \delta \left[B + B^2\right] \Sigma) \tag{37}$$

- When $\Sigma = I, \delta = \lambda^{-1}$, proposal $\approx$ Levenberg-Marquardt
- Still high per-iteration cost.
- Should lead to dimension-free (fast) mixing when

$$\pi(x) = \mathcal{N}(x|0, \Sigma) \times \exp(-\Phi(x)). \tag{38}$$