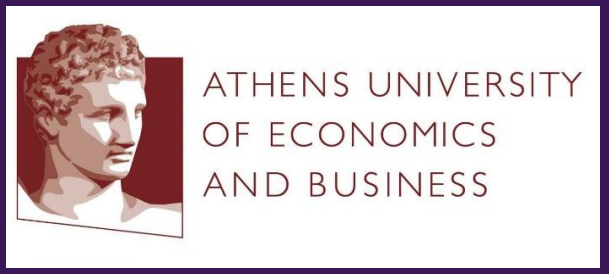
Gaussian invariant Markov chain Monte Carlo

Petros Dellaportas (UCL and AUEB)

Joint work with: Michalis Titsias (Google, Deepmind), Angelos Alexopoulos (AUEB) and Siran Liu (UCL)



CRETE, July 2026

Markov chain Monte Carlo (MCMC)

MCMC draws samples from an ergodic Markov chain $\{X_0, X_1, \dots, X_{n-1}\}_{n\geq 0}$ with state space $\mathcal{X} \subseteq \mathbb{R}^{\mathrm{d}}$.

$\mathbb{E}_\pi[F] \coloneqq \pi(F) = \int F(x)\,\pi(x)\,dx$ estimated by

$$\mu_n(F) = \frac{1}{n}\sum_{i=0}^{n-1} F(X_i), \qquad X_i \in \mathcal{X}.$$

Central Limit Theorem:

$$\mu_n(F) \xrightarrow{\mathrm{D}} N\left(\pi(F), \frac{\sigma_F^2}{n}\right).$$

Asymptotic variance:

$$\sigma_F^2 \coloneqq \mathrm{Var}_\pi[F(X_0)] + 2\sum_{i=0}^{\infty} \mathrm{Cov}_\pi\{F(X_0), F(X_{i+1})\}.$$

Metropolis-Hastings Algorithm

Draw proposal $y \sim q(y|x)$; accept with probability

$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)\, q(x|y)}{\pi(x)\, q(y|x)}\right\}.$$

Resulting density of the transition kernel:

$$P(y|x) = \alpha(x, y)q(y|x) + \left(1 - \int \alpha(x, z)q(z|x)dz\right)\delta_x(y).$$

If $\pi(x) = \mathcal{N}(x|\mu, \Sigma)$, neither achieves i.i.d. sampling.

Random Walk Metropolis (RWM):

$$q(y|x) = \mathcal{N}(y \mid x, 2\gamma\Sigma).$$

Metropolis Adjusted Langevin Algorithm (MALA):

$$q(y|x) = \mathcal{N}(y|x + \gamma\Sigma\nabla \log \pi(x)\,, 2\gamma\Sigma).$$

Gaussian Invariance: A key property

For $\pi(x) = \mathcal{N}(x|\mu, \Sigma)$, a MH sampler is Gaussian invariant if

$$\mathcal{N}(y|\mu, \Sigma)q(x|y) = \mathcal{N}(x|\mu, \Sigma)q(y|x).$$

Under Gaussian invariance:

$\alpha(x, y) = 1$ ⇒ all proposed samples are accepted.

Gaussian Invariant RWM (GI-RWM)

Suppose we start with a Gaussian approximation to the target:

$$\hat{\pi}(x) = \mathcal{N}(x|\mu, \Sigma).$$

GI-RWM: an **autoregressive RWM-like proposal**:

$$q(y|x) = \mathcal{N}(y|(1-\gamma)x + \gamma\mu, (2\gamma - \gamma^2)\Sigma), \quad \gamma \in (0,2).$$

Properties:

- $\gamma = 1$: reduces to independent Metropolis.
- $\gamma \approx 0$: local RWM-like moves.
- Reversible $\Longrightarrow$ invariant w.r.t. $\hat{\pi}(x)$.
- **But**: requires specifying both μ and Σ

Hessian-based GI Methods

Avoid explicit covariance Σ by using the Hessian of $\log \pi(x)$:

$$A_x = -(\nabla^2 \log \pi(x))^{-1}.$$

Proposal distribution:

$$q(y|x) = \mathcal{N}(y|x + \gamma A_x \nabla \log \pi(x), (2\gamma - \gamma^2) A_x).$$

Allows local adaptation to curvature of target distribution.

Gaussian Invariant MALA (GI-MALA)

GI-RWM: an **autoregressive RWM-like proposal**:

$$q(y|x) = \mathcal{N}(y|(1-\gamma)x + \gamma\mu, (2\gamma - \gamma^2)\Sigma), \quad \gamma \in (0,2).$$

- But: requires specifying both μ and Σ

We avoid this by using gradient information, GI-MALA:

$$q(y|x) = \mathcal{N}(y|x + \gamma\Sigma\nabla \log \pi(x), (2\gamma - \gamma^2)\Sigma).$$

Depends only on Σ, not on μ.

Gaussian invariance: reversible for the entire family of Gaussians with covariance Σ and arbitrary mean μ.

Computational complexity

- GI-RWM and GI-MALA: **similar cost** to standard RWM/MALA.
- Hessian-based methods: higher cost (due to Hessian), but **better local adaptation**.

Efficiency

- GI-RWM and GI-MALA: **Better efficiency** than the state of the art MCMC algorithms in terms of effective sample size for a series of challenging MCMC problems: Large logistic regression with correlated covariates, Gaussian process regression, Binary Gaussian process classification, log-Gaussian Cox processes
- Illustration of results later…

Poisson equation for control variates

A function $\hat{F}$ satisfies the Poisson equation for F if

$$P\hat{F} - \hat{F} = -F + \pi(F)$$

Observation: extremely hard to solve for continuous state-space Markov chains, even if $\pi(F)$ is known.

But if you can solve it, then an interesting **control variate** is

$$\mathrm{U} = \hat{F} - \mathrm{P}\hat{F} \qquad \mathrm{E}_\pi[\mathrm{U}(\mathrm{X})] = 0$$

Then

$$F - U = F - \hat{F} + P\hat{F} = \pi(F) \qquad \mathrm{Var}_\pi[F - U] = 0$$

Replace $\mu_n(F) := \frac{1}{n}\sum_{i=0}^{n-1} F(X_i)$

by $\mu_{n,\hat{F}}(F) = \frac{1}{n}\sum_{i=0}^{n-1}\{F(X_i) + P\hat{F}(X_i) - \hat{F}(X_i)\}$

Poisson equation in Gaussian invariant MCMC

General **solution** of the Poisson equation:

$$\hat{F}(x) = F(x) - \pi(F) + \sum_{n=1}^{\infty}\left(P^n F(x) - \pi(F)\right).$$

In standard HM:

$$PF(x) = \int \alpha(x,y)q(y|x)F(y)dy + \left(1 - \int \alpha(x,z)q(z|x)dz\right)F(x)$$
$$= F(x) + \int \alpha(x,y)\big(F(y) - F(x)\big)q(y|x)dy$$

r **Gaussian invariant MCMC with Gaussian target:**

$$\alpha(x,y) = 1 \quad \Rightarrow \quad PF(x) = \int F(y)q(y|x)dy\,,$$

i.e. expectation under a Gaussian transition.

In this case Poisson equation admits **analytic** solutions.

- For non-Gaussian targets, these Gaussian-based solutions can still act as powerful control variates, reducing variance.

D+Kontoyiannis, RSSB, 2012; Alexopoulos, Titsias and D, Stas+comp, 2023; Liu, D and Titsias, arxiv, 2024

Variance reduction via Poisson equation

Replace integral with unbiased estimator:

$$\alpha(X_i, Y_i)\big(G(Y_i) - G(X_i)\big)$$

- Note that this is a Monte Carlo estimator of sample size one!

- But we can reduce its variance too by adding another control variate of the form $G(Y_i) - \mathbb{E}_{q(\cdot|X_i)}[G]$

Final estimator:

$$\mu_{n,G}(F) = \frac{1}{n}\sum_{i=0}^{n-1}\left\{F(X_i) + \widehat{\beta_1}H_1(X_i, Y_i) + \widehat{\beta_2}H_2(X_i, Y_i)\right\}$$

- $H_1(X_i, Y_i) = \alpha(X_i, Y_i)\big(G(Y_i) - G(X_i)\big)$
- $H_2(X_i, Y_i) = G(Y_i) - \mathbb{E}_{q(\cdot|X_i)}[G]$
- $(\widehat{\beta_1}, \widehat{\beta_2})$ chosen to **minimize asymptotic variance** (D & Kontoyiannis, RSSB, 2012).
- $(\widehat{\beta_1}, \widehat{\beta_2})$ are estimated from the MCMC samples X_i.

$$\mu_{n,\hat{F}}(F) = \frac{1}{n}\sum_{i=0}^{n-1}\{F(X_i) + P\hat{F}(X_i) - \hat{F}(X_i)\}$$

$$PF(x) = F(x) + \int \alpha(x, y)\big(F(y) - F(x)\big)q(y|x)dy$$

- For non-Gaussian targets, these Gaussian-based solutions can still act as powerful control variates, reducing variance.

➢ We do not have the exact solution $\hat{F}$ but we have $G \approx \hat{F}$

The estimator with a control variate is

$$\mu_{n,G}(F) = \frac{1}{n}\sum_{i=0}^{n-1}\left\{F(X_i) + \int \alpha(X_i, y)\big(G(y) - G(X_i)\big)q(y|X_i)dy\right\}$$

Optimal estimator involves this intractable integral because $\alpha(X_i, y)$ depends on $\pi(y)$.

But in MH we already have one proposal $Y_i \sim q(\cdot\,|X_i)$ and $\alpha(X_i, Y_i)$.

Experiments on Bayesian logistic regression

Effective Sample Size (ESS) (10,000 samples, 5,000 burn-in).

Dataset	Method	Min	Median	Max
Heart ($N = 270, d = 13$)	GI-MALA	2787	3400	3982
	MALA	2029	2270	2439
Australian ($N = 690, d = 14$)	GI-MALA	3550	4621	5225
	MALA	1947	2194	2446
German ($N = 1000, d = 24$)	GI-MALA	3288	5433	5999
	MALA	1573	1818	1975

GI-MALA achieves **1. 5– 3x larger ESS** than standard MALA.

Variance Reduction with Control Variates (GI-MALA)

Ratios: $\text{variance}\left(\mu_{\text{n}}(F)\right)/\text{variance}\left(\mu_{\text{n,G}}(F)\right.$ for $F(x) = x)$

Dataset	$n = 1{,}000$	$n = 10{,}000$	$n = 50{,}000$	$n = 200{,}000$
Heart ($d = 13$)	3.2– 7.4	3.6– 7.0	3.2– 6.7	3.1– 6.1
Australian ($d = 14$)	1.7– 7.8	3.3– 7.7	1.1– 4.8	3.2– 8.1

Latent Gaussian models

Latent Gaussian models are widely used in Bayesian statistics, particularly in high-dimensional settings.

The target distribution in these models often takes the form:

$$\pi(x) \propto \exp\{ g(x)\}\mathcal{N}(x|0,\Sigma_0)$$

- $g(x)$ represents the log-likelihood, and $\mathcal{N}(x|0,\Sigma_0)$ is a Gaussian prior.
- Examples include Gaussian process regression, classification, and Bayesian inverse problems.

GI-MALA for latent Gaussian models

We apply GI-MALA by allowing A_x in the proposal

$$q(y|x) = \mathcal{N}(y|x + \gamma A_x \nabla \log \pi(x)\,, (2\gamma - \gamma^2)A_x)$$

to adapt to the local curvature of the target

$$A_x = (\Sigma_0^{-1} + L_x)^{-1}, \qquad (L_x)_{ii} = -\frac{\partial^2 g(x)}{\partial x^{(i)} \partial x^{(i)}}.$$

- Computational cost per iteration: $\mathcal{O}(d^2) \Longrightarrow$ feasible for high dimensions.

High dimensional experiments

Gaussian process **regression**:

$$\pi(x) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2\right\}\mathcal{N}(x|0,\Sigma_0)$$

Gaussian process **classification**:

$$\pi(x) \propto \prod_{i=1}^{n} P(y_i|x_i)\mathcal{N}(x|0,\Sigma_0)$$

Log-Gaussian Cox Process:

$$\pi(Z) \propto \exp\left(\sum_{i=1}^{n} Z(x_i) - \int_D \exp\big(Z(x)\big)\,dx\right)\cdot \mathcal{N}(Z|0,\Sigma_0),$$

where $Z(x)$ is the log-intensity modelled as a Gaussian process $\mathcal{N}(Z|0,\Sigma_0)$.

Gaussian process regression results

Comparison of sampling methods in simulated GP regression dataset with $\sigma^2 = 0.01$ and $d = 1000$.
All numbers are averages across ten repeats.

Method	Time (s)	ESS (Min, Med, Max)	Min ESS/s
GI-MALA	13.1	(4258.1, 4905.1, 5000.0)	325.06
MALA	13.2	(649.4, 880.1, 1072.2)	49.30
mGrad	21.2	(882.7, 1097.5, 1282.0)	41.59
Ellipt	7.0	(10.7, 43.6, 149.5)	1.54
pCN	6.5	(7.9, 36.7, 129.7)	1.22
pCNL	13.3	(12.8, 52.8, 230.7)	0.97

mGrad: Titsias and papaspiliopoulos, 2018, JRSSB; **Ellipt**: Elliptical slice sampling, Nishihara, Murray, Adams 2014, JMLR; **pCN,pCNL**: Cotter, Roberts, Stuart, White, 2013, St. Science.

Binary GP classification

Heart Dataset ($d = 270$)

Method	Time (s)	ESS (Min, Med, Max)	Min ESS/s
GI-MALA	0.7	(343, 868, 1443)	517
MALA	0.6	(210, 525, 898)	379
mGrad	0.9	(198, 587, 1083)	215
Ellipt	0.6	(14, 41, 97)	23
pCN	0.2	(8, 28, 79)	33
pCNL	0.5	(25, 70, 151)	55

Australian Credit Dataset ($d = 690$)

Method	Time (s)	ESS (Min, Med, Max)	Min ESS/s
GI-MALA	2.9	(283, 667, 1487)	98
MALA	3.5	(182, 469, 1018)	53
mGrad	11.1	(191, 488, 1167)	17
Ellipt	2.4	(6, 24, 93)	2
pCN	1.2	(4, 17, 76)	3
pCNL	2.3	(5, 21, 95)	2

GI-MALA dominates in efficiency (ESS/s).

Variance reduction with control variates.

Variance Reduction Factors for $F(x) = x$

	Heart ($d = 270$)	Australian ($d = 690$)
$n = 1k$	3.2– 20.2	3.3– 15.5
$n = 10k$	2.8– 17.6	3.6– 18.7
$n = 50k$	3.6– 18.9	1.0– 6.9
$n = 200k$	3.1– 18.9	1.0– 13.0

Log-Gaussian Cox model $(d = 4096)$

Comparison of Sampling Methods

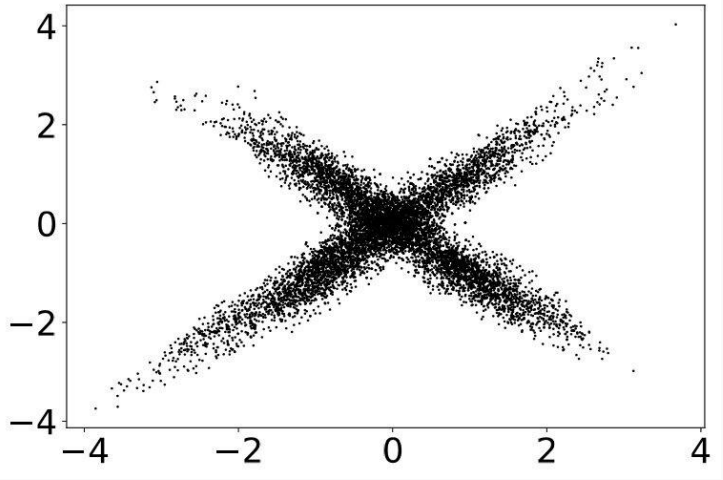
Method	Time (s)	ESS (Min, Med, Max)	Min ESS/s
GI-MALA	52.4	$(189, 651, 1141)$	3.61
MALA	54.2	$(147, 512, 890)$	2.71
mGrad	163.2	$(191, 821, 1699)$	1.17
Ellipt	88.3	$(4, 17, 70)$	0.05
pCN	83.3	$(3, 12, 62)$	0.04
pCNL	167.5	$(5, 5, 5)$	0.03

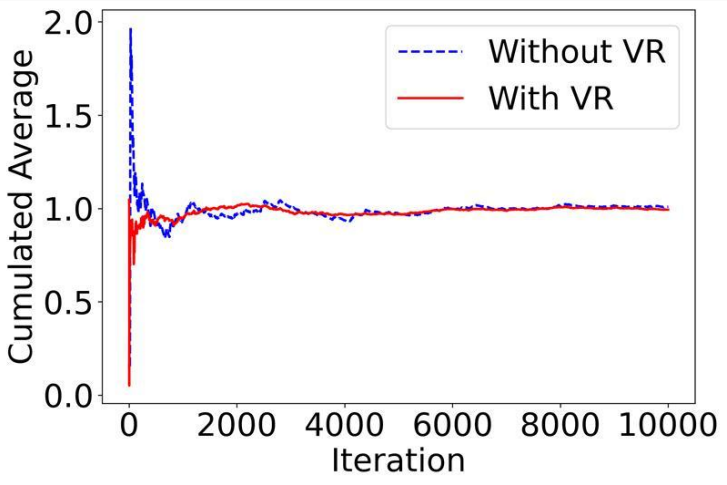
Variance Reduction Factors for $F\ (x)\ =\ x$

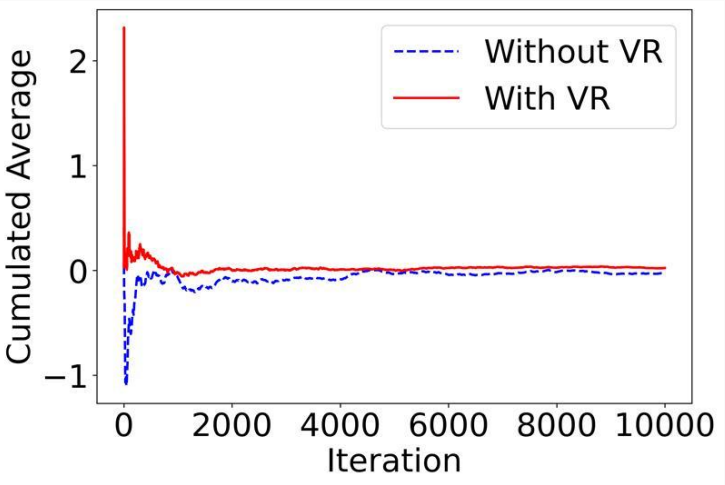
$n\ =\ 1{,}000$	$n\ =\ 10{,}000$
$2.38 – 127.62$	$1.40 – 121.42$

Gaussian mixture example

Target density and ergodic averages







Variance Reduction Factors

$F(x)$	$x^{(1)}$	$x^{(2)}$	$\left(x^{(1)}\right)^2$	$\left(x^{(2)}\right)^2$	$x^{(1)}x^{(2)}$
	3.983	2.5	7.25	6.34	2.18

Theory

- Precise conditions for Geometric ergodicity of our algorithms.

- Optimal acceptance rate and choice of γ for finite dimension d: theory differs from RWM/MALA since acceptance rates depend on **Gaussianity of target**.

- LLN, CLT for our estimators.

Discussion

- We introduced **Gaussian invariant MCMC proposals**

- Key property: analytical Poisson equation solutions $\Longrightarrow$ **control variates for variance reduction**.

- Applications to latent Gaussian models: **high-dimensional targets**, large ESS, low-cost variance reduction.