Simulation and monte carlo methods

The Coupled Rejection Sampler

Omar El Mamoune, Piero Pelosi, Sarakpy NY

ENSAE Paris

April 29, 2025

Overview

1. General definition of coupling
2. Coupled Rejection-Sampling Method
3. Multidimensional Gaussian Case
4. Validity of Thorisson's algorithm
5. Comparison between the two methods
6. Application: Coupled Random Walk Metropolis

I. General definition of coupling

Definition

Using the standard formalism of probability theory, let X_1 and X_2 be two random variables defined on the probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$. A *coupling* of X_1 and X_2 is a new probability space $(\Omega, \mathcal{F}, P)$ on which there exist two random variables Y_1 and Y_2 such that Y_1 has the same distribution as X_1 while Y_2 has the same distribution as X_2.

In the article, we consider a class of couplings, which preserve mass over the diagonal of the joint distribution, that is, couplings (X, Y) such that $\mathbb{P}(X = Y) > 0$. Such couplings are ubiquitous in proving a number of inequalities used, for example, to study the convergence of Monte Carlo algorithms.

Coupling Inequality

Theorem: Coupling Inequality

Let μ and ν be probability measures on a measurable space $(S, \mathcal{S})$. For any coupling (X, Y) of μ and ν,

$$\|\mu - \nu\|_{\mathrm{TV}} \ \le \ \mathbb{P}[X \neq Y].$$

Useful for proving convergence and limit theorems

II. Coupled Rejection-Sampling Method

- Let p and q be probability densities defined on a space χ

Definition: Diagonal Coupling

A pair of random variables (X, Y), defined on the same probability space and taking values in χ, is a diagonal coupling of p and q if:

$$\mathbb{P}(X \in A) = \int_A p(x)\, dx, \quad \mathbb{P}(Y \in A) = \int_A q(x)\, dx$$

for all measurable sets $A \subset \chi$, and

$$\mathbb{P}(X = Y) > 0.$$

- A **maximal coupling** maximizes $\mathbb{P}(X = Y)$
- Useful for rejection sampling and dependent proposal schemes

Coupled Rejection-Sampling Method

Definition: Dominating Pair

Let $(\hat{p}, \hat{q})$ be a pair of probability densities. We say it dominates (p, q) if there exist constants $M(p, \hat{p}) < \infty$ and $M(q, \hat{q}) < \infty$ such that:

$$p(x) \leq M(p, \hat{p})\, \hat{p}(x), \quad q(x) \leq M(q, \hat{q})\, \hat{q}(x), \quad \forall x \in \mathbb{R}^d$$

- Construct a diagonal coupling $\hat{\Gamma}$ of $\hat{p}, \hat{q}$ that dominates the independent product $p \otimes q$
- Use $\hat{\Gamma}$ as a proposal in an acceptance-rejection scheme to sample from a diagonal coupling with marginals p and q

Algorithm 1: Rejection Coupling of (p, q)

Algorithm 1: Rejection-coupling of (p, q)

Function `RejectionCoupling`($\hat{\Gamma}$, p, q):

    // Supposing $\hat{\Gamma} \succeq p \otimes q$ is a coupling of $\hat{p}$ and $\hat{q}$.

    Set $A_X = 0$ and $A_Y = 0$ // Acceptance flags

    while $A_X = 0$ *and* $A_Y = 0$ **do**

        Sample $X_1, Y_1 \sim \hat{\Gamma}$, $U \sim \mathcal{U}(0, 1)$

        if $U < \frac{p(X_1)}{M(p,\hat{p})\hat{p}(X_1)}$ **then** set $A_X = 1$

        if $U < \frac{q(Y_1)}{M(q,\hat{q})\hat{q}(Y_1)}$ **then** set $A_Y = 1$

    Sample X_2, Y_2 from $p \otimes q$

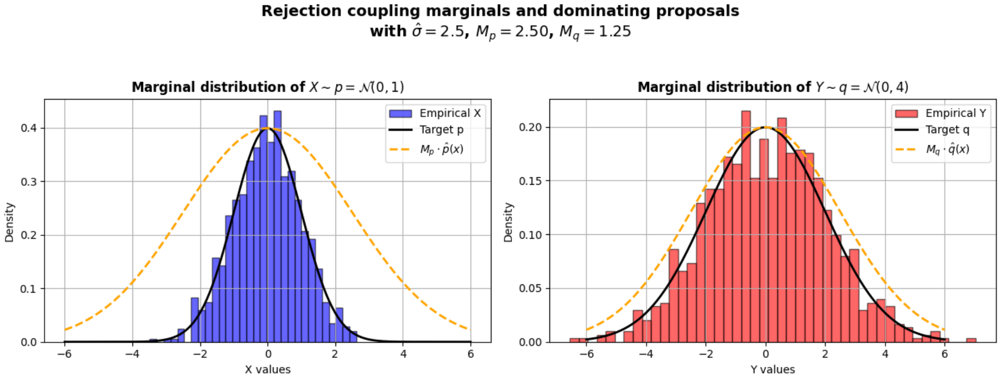
    Return: $X = A_X\, X_1 + (1 - A_X)\, X_2$, $Y = A_Y\, Y_1 + (1 - A_Y)\, Y_2$

- The algorithm samples $(X_1, Y_1) \sim \hat{\Gamma}$ and uses a shared $U \sim \mathcal{U}(0, 1)$
- Accepts X_1 or Y_1 based on rejection conditions with respect to p, $\hat{p}$, q, $\hat{q}$
- If not accepted, samples from the product $p \otimes q$
- **Output:** a sample (X, Y) from a valid diagonal coupling of p and q

Visualization

- We use Gaussian targets: $p = \mathcal{N}(0, 1)$, $q = \mathcal{N}(0, 4)$
- Dominating proposal: $\hat{p}, \hat{q} \sim \mathcal{N}(0, \hat{\sigma}^2)$, with $\hat{\sigma} = 2.5$
- Constants: $M_p = 2.5$, $M_q = 1.25$
- Samples are drawn using a placeholder $\texttt{sample\_pq} \sim p \otimes q$
- Metric: match rate $\mathbb{P}(|X - Y| < \varepsilon) \approx 2.9\%$

Visualization



- Histogram of empirical marginals vs targets
- Dominating bounds $M_p\hat{p}(x)$, $M_q\hat{q}(x)$ shown as dashed lines
- Visual validation of dominance and marginal accuracy

III. Multidimensional Gaussian Case

- Goal: Build a Coupled Rejection Sampler (CRS) from $\mathcal{N}(\mu_p, \Sigma_p)$ and $\mathcal{N}(\mu_q, \Sigma_q)$
- We use a common proposal $\hat{\Gamma}$ based on a shared covariance $\hat{\Sigma}$
- Method: **Reflection-Maximal Coupling** from Bou-Rabee et al. (2020)

Diagonal Dominating Coupling Condition

Let $\hat{\Sigma}$ satisfy:

$$\hat{\Sigma}^{-1} \preceq \Sigma_p^{-1}, \quad \hat{\Sigma}^{-1} \preceq \Sigma_q^{-1}$$

(Loewner ordering)

- This guarantees domination:

$$\mathcal{N}(x; \mu_p, \Sigma_p) \leq \frac{\det(2\pi\hat{\Sigma})^{1/2}}{\det(2\pi\Sigma_p)^{1/2}} \mathcal{N}(x; \mu_p, \hat{\Sigma})$$

- And similarly for $\mathcal{N}(x; \mu_q, \Sigma_q)$

Multidimensional Gaussian Case: Dominating Proposals

Proposition 5: Gaussian Diagonal Coupling via a Dominating Covariance

Let $\hat{\Sigma}$ satisfy:

$$\hat{\Sigma}^{-1} \preceq \Sigma_p^{-1}, \quad \hat{\Sigma}^{-1} \preceq \Sigma_q^{-1}$$

Then $\mathcal{N}(\mu_p, \hat{\Sigma})$ and $\mathcal{N}(\mu_q, \hat{\Sigma})$ dominate $\mathcal{N}(\mu_p, \Sigma_p)$ and $\mathcal{N}(\mu_q, \Sigma_q)$

Set:

$$M(p, \hat{p}) = \frac{\det(2\pi\hat{\Sigma})^{1/2}}{\det(2\pi\Sigma_p)^{1/2}}, \quad M(q, \hat{q}) = \frac{\det(2\pi\hat{\Sigma})^{1/2}}{\det(2\pi\Sigma_q)^{1/2}}$$

Algorithm 1 with reflection-maximal coupling $\hat{\Gamma}$ yields a diagonal coupling of $\mathcal{N}(\mu_p, \Sigma_p)$ and $\mathcal{N}(\mu_q, \Sigma_q)$.

Optimizing Coupling via Dominating Covariance

- Goal: Choose $\hat{\Sigma}$ to maximize coupling probability.
- Criterion: Maximize $\mathbb{P}(A_X = 1)\mathbb{P}(A_Y = 1) \Rightarrow \log\det(\hat{\Sigma}^{-1})$

$$\max_{\hat{\Sigma}^{-1}} \log\det(\hat{\Sigma}^{-1}) \quad \text{s.t.} \quad \hat{\Sigma}^{-1} \preceq \Sigma_p^{-1}, \quad \hat{\Sigma}^{-1} \preceq \Sigma_q^{-1}, \quad \hat{\Sigma}^{-1} \succeq 0$$

Proposition 7: Optimal Solution

$$\hat{\Sigma}_{\text{opt}} = CVUV^\top C^\top, \quad C = \Sigma_q^{1/2}, \quad VDV^\top = C^\top \Sigma_p^{-1} C$$

where V is orthonormal, D diagonal, and $U_{ii} = \frac{1}{\min(1, D_{ii})}$

Computing $\Sigma_q^{1/2}$*:* via Cholesky decomposition:

$$\Sigma_q = CC^\top$$

where C is lower triangular with positive diagonal entries.

Reflection-Maximal Coupling Algorithm

Algorithm 3: Reflection-maximal coupling

Function `ReflectionCoupling`$(a,\ b,\ \Sigma)$:

- $z = \Sigma^{-1/2}(a - b)$
- $e = z / \|z\|$
- Sample $\dot{X} \sim \mathcal{N}(0, I)$ and $U \sim \mathcal{U}(0,1)$
- **if** $\mathcal{N}(\dot{X}; 0, I)\, U < \mathcal{N}(\dot{X} + z; 0, I)$ **then**
  - Set $\dot{Y} = \dot{X} + z$
- **else**
  - Set $\dot{Y} = \dot{X} - 2\left\langle e, \dot{X} \right\rangle e$
- Set $X = a + \Sigma^{1/2}\dot{X}$ and $Y = b + \Sigma^{1/2}\dot{Y}$
- **Return:** X, Y

- To construct a diagonal coupling of $\mathcal{N}(\mu_p, \Sigma_p)$ and $\mathcal{N}(\mu_q, \Sigma_q)$, we compute the diagonal dominating proposal $\hat{\Sigma}$
- The final step is to apply **Algorithm 1** using $\hat{\Sigma}$, resulting in a reflection-maximal coupling
- This method enables exact coupling of Gaussian marginals

Empirical Test of Rejection Coupling

- We evaluate `Rejection Coupling` on two Gaussian test cases:

Test Case 1: Close Gaussians

- $\mu_p = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma_p = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$
- $\mu_q = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}, \quad \Sigma_q = \begin{bmatrix} 1.05 & 0.25 \\ 0.25 & 1.1 \end{bmatrix}$
- Coupling success rate: $\mathbb{P}(X = Y) \approx 0.8908$

Test Case 2: Separated Gaussians

- $\mu_p = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \Sigma_p = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- $\mu_q = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \quad \Sigma_q = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
- Coupling success rate: $\mathbb{P}(X = Y) \approx 0.0010$

IV. Validity of Thorisson's algorithm

Algorithm 5: Modified Thorisson algorithm

Function `ThorissonCoupling`(p, q, C):
- Sample $X \sim p$
- Sample $U \sim \mathcal{U}(0,1)$
- **if** $U < \min(\frac{q(X)}{p(X)}, C)$ **then**
  - Set $Y = X$
- **else**
  - Set $A = 0$
  - **while** $A \neq 1$ **do**
    - Sample $U \sim \mathcal{U}(0,1)$
    - Sample $Z \sim q$
    - **if** $U > \min\left(1, C\frac{p(Z)}{q(Z)}\right)$ **then**
      - Set $A = 1$
    - Set $Y = Z$
- **Return:** X, Y

Validity of Thorisson's algorithm

Conditions: we will check if we have :

1. $X \sim p, \; Y \sim q$
2. $\mathbb{P}(X = Y) > 0$
3. A rejection-sampling step is present

Proof $Y \sim q$

Since in the first step $X \sim p$, let us show that $Y \sim q$. Suppose $A \subseteq \mathcal{X}$ is a measurable subset. We have

$$
\begin{aligned}
P(Y \in A) &= P\big(Y \in A, \text{step1}\big) + P\big(Y \in A, \text{step2}\big) \\
P\big(Y \in A, \text{step1}\big) &= \mathbb{E}\big[\mathbf{1}\{Y \in A, \text{step1}\}\big] \\
&= \int_A \int_0^1 \mathbf{1}\Big(u < \min\big(\tfrac{q(x)}{p(x)}, C\big)\Big)\, p(x)\, du\, dx \\
&= \int_A \min\big(q(x), C\, p(x)\big)\, dx
\end{aligned}
$$

$$
P(\text{step1}) \;=\; \int_{\mathcal{X}} \min\big(q(x), C\, p(x)\big)\, dx
$$

Continuation of the proof

$$P\big(Y \in A, \text{step2}\big) = \int_A \Big[q(x) - \min\{q(x), Cp(x)\}\Big]\, dx \tag{1}$$

For (1) to hold it is necessary that

$$\int_A q(x) - \min\{q(x), Cp(x)\}\, dx \;=\; P\big(Y \in A \mid \text{step2}\big)\, P(\text{step2}).$$

We know that

$$P(\text{step2}) \;=\; 1 - P(\text{step1}) \;=\; 1 - \int_{\mathcal{X}} \min\{q(x), Cp(x)\}\, dx.$$

Thus, given step2, Y has density

$$\tilde{q}(x) = \frac{q(x) - \min\{q(x), Cp(x)\}}{1 - \int_{\mathcal{X}} \min\{q(s), Cp(s)\}\, ds}.$$

It is then sufficient to note that step 2 is a standard acceptance–rejection procedure, where one simulates $\tilde{q}(x)$ using the proposal law $q(x)$, to conclude that $Y \sim q$.

Rewriting of step 2 of Algorithm 5

Note that

$$\frac{\tilde{q}(z)}{q(z)} = \frac{1 - \min\left\{1, \dfrac{Cp(z)}{q(z)}\right\}}{1 - \int_{\mathcal{X}} \min\{q(s), Cp(s)\}\, ds} \leq \frac{1}{1 - \int_{\mathcal{X}} \min\{q(s), Cp(s)\}\, ds} = M.$$

Thus,

$$1 - \min\left\{1, \frac{Cp(z)}{q(z)}\right\} = \frac{\tilde{q}(z)}{M\, q(z)}.$$

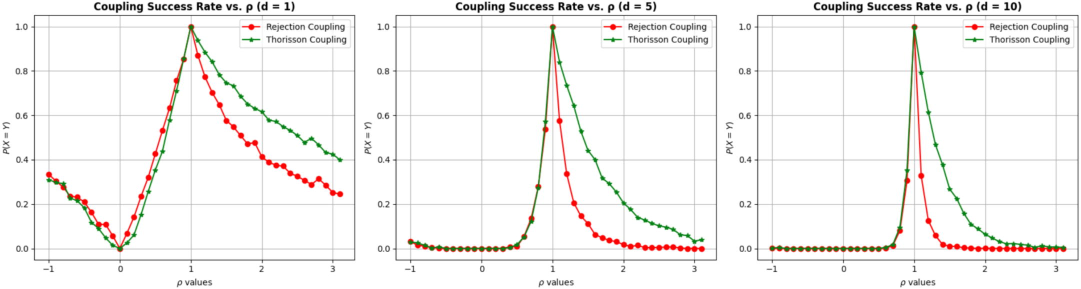
Hence instructions 9–13 can be rewritten:

- Sample $Z \sim q$
- Sample $V \sim \mathrm{Unif}(0,1)$ with $V = 1 - U$ (where U is the uniform draw in the original algorithm)
- If $V < 1 - \min\left\{1, \dfrac{Cp(z)}{q(z)}\right\} = \dfrac{\tilde{q}(z)}{M\, q(z)}$ then
    - Set $A = 1$
    - Set $Y = Z$

One can thus clearly recognise the acceptance–rejection mechanism. Finally, note that

$$P(X = Y) = P(\text{step1}) = \int_{\mathcal{X}} \min\{q(x), Cp(x)\}\, dx > 0.$$

Coupling success rate vs. ρ



Time comparaison

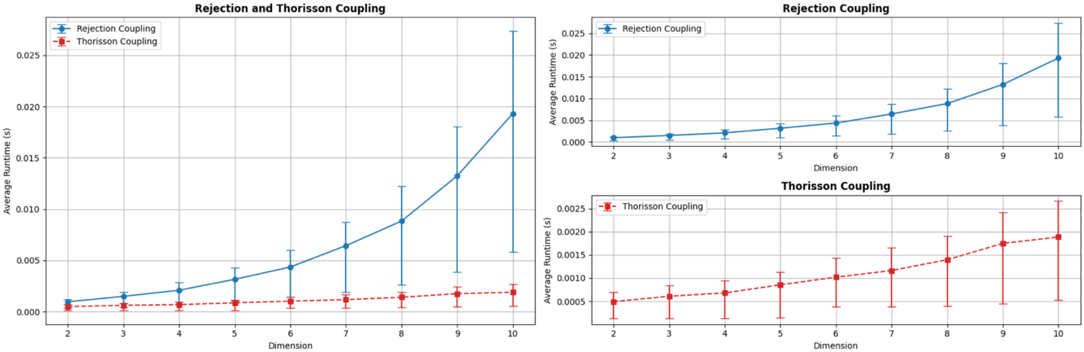


Figure: Average Runtime

VI. Coupled Random Walk Metropolis (RWM)

Goal: Sample from a target distribution $\pi(x)$ using MCMC.

Algorithm Steps RWM

Given current state x_t:

1. **Proposal:** Sample $x' = x_t + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \Sigma)$
2. **Acceptance:** Compute acceptance ratio:

$$\alpha = \min\left(1, \frac{\pi(x')}{\pi(x_t)}\right)$$

3. **Update:** With probability α, set $x_{t+1} = x'$, otherwise, set $x_{t+1} = x_t$

Application: Use a standard multivariate normal:

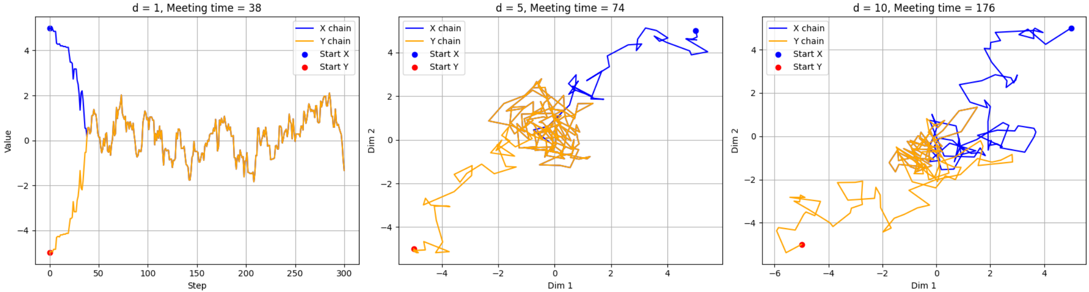
$$\pi(x) \propto \exp\left(-\frac{1}{2}\|x\|^2\right)$$

and couple initial chains using a **reflection-maximal Gaussian** to study *coupling time*.

Coupled RWM: Methodology and Setup

- Run **two RWM chains** from initial points to study convergence
- Use **reflection-maximal coupling** at each step:
  - Proposals: $\mathcal{N}(x_t, \Sigma)$, $\mathcal{N}(y_t, \Sigma)$ with $\Sigma = \sigma^2 I$
  - Shared $u \sim \mathcal{U}(0, 1)$ synchronizes acceptance
- **Meeting time τ:** first step where $\|x_t - y_t\| < 10^{-10}$
- After meeting: chains evolve identically
- **Experimental goal:** analyze impact of dimensionality $d = 1, 5, 10$ on coupling
- **Setup:**
  - $\sigma = 0.5$, $\Sigma = \sigma^2 I_d$, 300 iterations
  - $x_0 = (5, \ldots, 5) \in \mathbb{R}^d$, $\quad y_0 = (-5, \ldots, -5) \in \mathbb{R}^d$

Results Across Dimensions



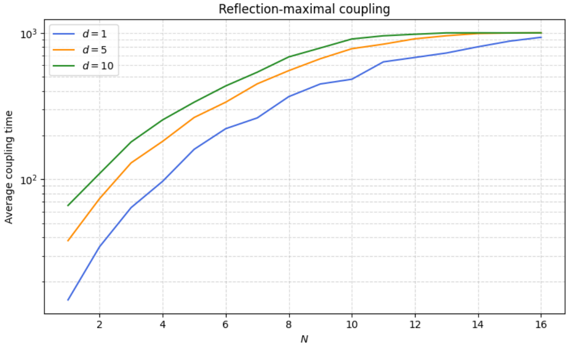
- $d = 1$: chains meet quickly with smooth convergence
- $d = 5, 10$: longer coupling times and more erratic paths
- Meeting time τ shows how dimensionality affects MCMC convergence

Average Coupling Time vs Proposal Scale

- Investigate how **proposal scale** affects coupling in RWM:

$$\Sigma = \left(\frac{1}{N}\right)^2 I$$

- Dimensions tested: $d = 1, 5, 10$, with $N \in \{1, \ldots, 16\}$
- For each (d, N): 50 repetitions, 1000 max iterations



Questions?

Thank you for your attention!

Annexe (1/2)

La densité d'une loi $\mathcal{N}(0, \sigma^2)$ est donnée par :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Ainsi, dans notre cas :

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad q(y) = \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{y^2}{8}\right),$$

et pour la proposition commune $\hat{p}(x) = \hat{q}(y)$ avec $\hat{\sigma} = 2.5$:

$$\hat{p}(x) = \hat{q}(y) = \frac{1}{\sqrt{2\pi(2.5)^2}} \exp\left(-\frac{x^2}{2(2.5)^2}\right).$$

On cherche les plus petits M_p et M_q tels que :

$$p(x) \leq M_p \hat{p}(x), \quad q(y) \leq M_q \hat{q}(y), \quad \text{pour tout } x, y.$$

Annexe (2/2)

On calcule :

$$\frac{p(x)}{\hat{p}(x)} = 2.5 \exp\left(-\frac{x^2}{2}\left(1 - \frac{1}{(2.5)^2}\right)\right),$$

$$\frac{q(y)}{\hat{q}(y)} = 1.25 \exp\left(-\frac{y^2}{2}\left(\frac{1}{4} - \frac{1}{(2.5)^2}\right)\right).$$

À $x = 0$ et $y = 0$, l'exponentielle vaut 1, donc :

$$M_p = 2.5, \quad M_q = 1.25.$$

Ainsi, la domination est formellement vérifiée :

$$p(x) \leq 2.5\, \hat{p}(x), \quad q(y) \leq 1.25\, \hat{q}(y).$$