

Total Variation Mixing of Weighted Quadratic Gibbs Samplers on a Cube

Alper Ferudun*

11 October 2026

Abstract

Consider uniform random-scan coordinate Gibbs sampling on $[0, 1]^d$ for a density proportional to $\exp\{-a \sum_{i < j} c_{ij}(x_i - x_j)^2\}$, where $c_{ij} = c_{ji} \geq 0$. For every fixed nonzero weight matrix, we prove a worst-start total-variation mixing bound of order a as $a \rightarrow \infty$, with an explicit finite-parameter bound and logarithmic dependence on the requested accuracy. Together with the Wasserstein lower bound of Gerencsér and Ottolini, this gives the sharp order $\Theta(a)$ at total-variation tolerance $1/4$, including disconnected graphs and isolated vertices. The proof adapts the non-strongly-convex entropy interpolation of Ascolani, Lavenant and Zanella to the compact cube. A coordinate-coverage decomposition handles arbitrary deterministic starts without assigning finite entropy to the chain's remaining singular part. The result supplies the total-variation upper bound discussed in the weighted model's earlier analysis. It concerns exact conditional Gibbs updates, not arbitrary Metropolis proposals, and does not optimize the dependence on dimension or graph parameters.

1 Model and main bound

Let $Q = [0, 1]^d$, $d \geq 2$, and let $C = (c_{ij})$ be symmetric with $c_{ii} = 0$ and $c_{ij} \geq 0$. For $a \geq 0$, define

$$V(x) = a \sum_{i < j} c_{ij}(x_i - x_j)^2, \quad \pi(dx) = Z^{-1} e^{-V(x)} dx \quad (x \in Q). \quad (1)$$

Write P_i for exact resampling of coordinate i from its conditional distribution under π , and put $P = d^{-1} \sum_{i=1}^d P_i$. One step is one randomly chosen coordinate update, not one sweep. For probability measures, use $\text{TV}(\mu, \nu) = \sup_B |\mu(B) - \nu(B)|$, and set

$$d_P(t) = \sup_{x \in Q} \text{TV}(P^t(x, \cdot), \pi), \quad t_{\text{mix}}(\epsilon) = \min\{t \in \mathbb{N}_0 : d_P(t) \leq \epsilon\}.$$

All logarithms without a subscript are natural.

Gerencsér [2] proved the sharp order on the square. Gerencsér and Ottolini [3] established the corresponding fixed-graph order in Wasserstein distance in arbitrary dimension and discussed the total-variation upper bound as a remaining question in Section 6.1 of their preprint. Their parameter A satisfies $a = A^2$. We obtain the upper bound by applying the entropy-interpolation mechanism of Ascolani, Lavenant and Zanella [4, Section 6] directly on the cube and first smoothing a high-probability branch of every point start. The method and the matching lower bound are prior results; the purpose here is their explicit compact-support, worst-start application.

*Mercury Software GmbH. Email: alper@mercurycodelab.com. GitHub: [AlperTheKing](https://github.com/AlperTheKing).

Put

$$W = \sum_{i < j} c_{ij}, \quad c_i = \sum_{j \neq i} c_{ij}, \quad c_{\max} = \max_i c_i.$$

Theorem 1.1. *For the kernel P associated with (1), define*

$$m = \lceil d \log(8d) \rceil, \quad R^2 = a \max\{W + mc_{\max}, 4W\}, \quad T = m + \lceil 64dR^2 \rceil. \quad (2)$$

Then $d_P(T) \leq 1/4$. For every $0 < \epsilon < 1$,

$$t_{\text{mix}}(\epsilon) \leq T \lceil \log_2(1/\epsilon) \rceil. \quad (3)$$

No connectivity assumption on C is needed.

For fixed C , this is $O_C((1+a)\log(1/\epsilon))$. In particular, the exponent coefficient, not its square, determines the asymptotic time scale. If $aW = 0$, the target is uniform and the bound uses only coordinate coverage.

Corollary 1.2. *For every fixed C with at least one positive weight,*

$$t_{\text{mix}}(1/4) = \Theta_C(a) \quad (a \rightarrow \infty).$$

The upper bound is proved in Sections 2–4; the deduction of the corollary from the existing lower bound is given in Section 5. The constants are deliberately elementary. Neither their growing- d dependence nor their graph dependence is claimed optimal.

2 A smooth branch from every starting point

The conditional density used by P_i is

$$g_i(t \mid x_{-i}) = \frac{\exp\{-a \sum_{j \neq i} c_{ij}(t - x_j)^2\}}{\int_0^1 \exp\{-a \sum_{j \neq i} c_{ij}(u - x_j)^2\} du}. \quad (4)$$

It is smooth and strictly positive on the closed cube, and

$$e^{-ac_i} \leq g_i(t \mid x_{-i}) \leq e^{ac_i}. \quad (5)$$

Indeed, the sum in the exponent lies between 0 and c_i .

Lemma 2.1. *For every $x \in Q$, the law at time m in (2) admits*

$$P^m(x, \cdot) = (1 - \eta)\mu + \eta\nu, \quad 0 \leq \eta \leq 1/8,$$

where ν is a probability measure and μ has a smooth, strictly positive density on Q satisfying

$$\text{KL}(\mu \mid \pi) \leq a(W + mc_{\max}). \quad (6)$$

Every subsequent μP^n also has a smooth, strictly positive density.

Proof. Let E be the event that each coordinate is chosen at least once among the first m independent uniform coordinate choices. Then

$$\eta := \mathbb{P}(E^c) \leq d(1 - 1/d)^m \leq de^{-m/d} \leq 1/8.$$

Take μ to be the law at time m conditional on E . For a fixed covering word $i_1, \dots, i_m$, the joint density of its m update values is the product of their sequential conditional densities. By (5), it is at most $\exp(amc_{\max})$ on $[0, 1]^m$.

Each final coordinate is the update value at its last occurrence in the word. These are d distinct variables. Marginalizing over the other $m - d$ variables, on a domain of volume one, shows that the final density is at most $\exp(amc_{\max})$. This argument does not assume independence of the final coordinates. Products of the smooth positive conditional densities, and their compact-domain integrals, are smooth and positive. The finite mixture over covering words has the same properties and the same upper bound.

Since $0 \leq V \leq aW$ and $Z \leq 1$, the density of π is at least e^{-aW} . The pointwise likelihood-ratio bound proves (6). If a law has smooth positive density f , its image under P_i has density $f_{-i}(x_{-i})g_i(x_i | x_{-i})$. This verifies the last assertion. $\square$

The conditioning in Lemma 2.1 is essential. For any fixed number of random-scan updates there can be a positive probability that a coordinate has not been updated. From a deterministic start, that branch is singular relative to π , and the entropy of the *unconditional* law can still be infinite.

3 Entropy interpolation on the cube

We give the compact-cube form of the argument underlying [4, Theorem 6.1]. Its proof avoids silently placing a hard-wall potential under a whole-space finite-potential assumption. Write $H(f dx) = \int_Q f \log f dx$.

Lemma 3.1. *Let U be a smooth convex function on a neighborhood of Q with $\partial_{ii}U \leq L_i$, where $L_i \geq 0$, and let $\rho(dx) \propto e^{-U(x)} dx$ on Q . Let K be its uniform random-scan conditional Gibbs kernel. For every smooth, strictly positive density σ on Q and every $0 \leq s \leq 1$,*

$$\text{KL}(\sigma K | \rho) \leq (1 - s/d) \text{KL}(\sigma | \rho) + \frac{s^2}{2d} \sum_i L_i. \quad (7)$$

Proof. Let $F : Q \rightarrow Q$ be the increasing triangular Knothe–Rosenblatt map with $F_{\#}\sigma = \rho$. Successive conditional cumulative distribution functions construct F . Their densities are smooth and uniformly positive on the compact cube; solving each conditional CDF equation gives a smooth triangular diffeomorphism, with $\partial_i F_i > 0$ and $F_i(x_{<i}, 0) = 0$, $F_i(x_{<i}, 1) = 1$. This is the regularity case in [4, Proposition 5.9].

Define S_i by leaving the other coordinates fixed and putting $(S_i x)_i = (1 - s)x_i + sF_i(x)$. Each S_i is a triangular diffeomorphism of Q , and

$$\det DS_i(x) = 1 - s + s\partial_i F_i(x).$$

It preserves the complementary marginal σ_{-i} . The KL chain rule implies that conditional resampling minimizes relative entropy among laws with that marginal. Thus, if K_i is the i th coordinate kernel,

$$\text{KL}(\sigma K_i | \rho) \leq \text{KL}((S_i)_{\#}\sigma | \rho). \quad (8)$$

Change of variables and the concavity of the logarithm give

$$H((S_i)_\# \sigma) - H(\sigma) = - \int_Q \log(1 - s + s \partial_i F_i) d\sigma \leq -s \int_Q \log(\partial_i F_i) d\sigma.$$

Since $\det DF = \prod_i \partial_i F_i$ and $F_\# \sigma = \rho$, summing yields

$$\sum_i (H((S_i)_\# \sigma) - H(\sigma)) \leq s(H(\rho) - H(\sigma)). \quad (9)$$

Coordinate Taylor bounds and convexity of U imply

$$\begin{aligned} \sum_i (U(S_i x) - U(x)) &\leq s \nabla U(x) \cdot (F(x) - x) + \frac{s^2}{2} \sum_i L_i (F_i(x) - x_i)^2 \\ &\leq s(U(F(x)) - U(x)) + \frac{s^2}{2} \sum_i L_i, \end{aligned}$$

because both $F_i(x)$ and x_i lie in $[0, 1]$. Integrate, use $F_\# \sigma = \rho$, and add (9). The normalization constants cancel, giving

$$\frac{1}{d} \sum_i \text{KL}((S_i)_\# \sigma \mid \rho) \leq (1 - s/d) \text{KL}(\sigma \mid \rho) + \frac{s^2}{2d} \sum_i L_i.$$

Finally combine (8) with convexity of KL in its first argument and $\sigma K = d^{-1} \sum_i \sigma K_i$. $\square$

For the potential (1), convexity follows from the nonnegative weights, $L_i = 2ac_i$, and

$$B := \sum_i L_i = 4aW. \quad (10)$$

Zero values of L_i are harmless; no inverse curvature matrix or positive spectral gap of the graph is used in the upper bound.

4 Recursion and arbitrary accuracy

Fix a starting point and apply Lemma 2.1. Write $h_n = \text{KL}(\mu P^n \mid \pi)$. Data processing gives $h_n \leq h_0 \leq R^2$, and all these laws meet the regularity assumptions of Lemma 3.1. If $R^2 > 0$, choose $s = h_n/R^2$ there and use $B \leq R^2$ to obtain

$$h_{n+1} \leq h_n - \frac{h_n^2}{2dR^2}. \quad (11)$$

If an entropy is zero, all subsequent ones vanish. Otherwise,

$$\frac{1}{h_{n+1}} - \frac{1}{h_n} \geq \frac{1}{2dR^2}, \quad h_n \leq \frac{2dR^2}{n + 2dR^2/h_0} \leq \frac{2dR^2}{n + 2d}.$$

Pinsker's inequality then gives

$$\text{TV}(\mu P^n, \pi) \leq \sqrt{\frac{dR^2}{n + 2d}}. \quad (12)$$

Future scan choices are independent of the coverage event, so

$$P^{m+n}(x, \cdot) = (1 - \eta)\mu P^n + \eta\nu P^n.$$

For $n = \lceil 64dR^2 \rceil$, (12) contributes at most $1/8$, and the arbitrary uncovered branch contributes at most another $1/8$. This proves $d_P(T) \leq 1/4$ uniformly in x . If $R^2 = 0$, then $aW = 0$ and every covering word already gives the uniform target exactly, proving the same conclusion with $T = m$.

For the accuracy statement, define the contraction coefficient

$$\bar{d}_P(t) = \sup_{x, y \in Q} \text{TV}(P^t(x, \cdot), P^t(y, \cdot)).$$

Stationarity and the triangle inequality give $d_P(t) \leq \bar{d}_P(t) \leq 2d_P(t)$. The standard contraction-coefficient inequality gives $\bar{d}_P(t+u) \leq \bar{d}_P(t)\bar{d}_P(u)$. Consequently $d_P(\ell T) \leq \bar{d}_P(\ell T) \leq 2^{-\ell}$, which proves (3).

5 Matching order and scope

Proof of Corollary 1.2. The upper bound follows from Theorem 1.1. For the lower bound use [3, Theorem 2.4] with its accuracy parameter $\delta = 1/16$. For fixed nonzero C , there is $b_C > 0$ such that, for all sufficiently large a and all integer $t \leq b_C a$, the worst-start Wasserstein distance with cost $\|x - y\|_\infty$ is at least $3/16$. This is an expected transport cost, not an essential-supremum transport metric. Since $\|x - y\|_\infty \leq 1$ on Q , maximal coupling gives $W_1 \leq \text{TV}$. Hence $t_{\text{mix}}(1/8) > \lfloor b_C a \rfloor$. The contraction-coefficient argument from Section 4 also gives $t_{\text{mix}}(1/8) \leq 3t_{\text{mix}}(1/4)$. Therefore $t_{\text{mix}}(1/4) > \lfloor b_C a \rfloor / 3$, as required. $\square$

The original AIM question [1] uses the word Metropolis but does not fix its proposal. The official workshop page links the square Gibbs analysis [2]. Theorem 1.1 treats this exact conditional-resampling interpretation and the weighted higher-dimensional model of [3]. A stationary density alone cannot give a mixing bound uniform over Metropolis proposals: replacing a proposal by an arbitrarily lazy version can make its chain arbitrarily slow. No proposal-independent assertion is intended.

The general interpolation method is due to [4]; the lower bound is due to [3]. The compact-cube argument and coverage decomposition make their combination explicit for point starts. The note does not claim absolute priority, sharp graph constants, or a conclusion uniform in growing dimension. It is an unrefereed, self-audited preprint. AI tools assisted the derivation, source review, writing and computational checks; this is not independent peer review or formal verification. The accompanying exact finite checks support the algebra and constants but do not replace the analytic proof.

References

- [1] American Institute of Mathematics, *Markov chain mixing times*, Problem 6.2 (P. Diaconis), <http://aimpl.org/markovmixing/6/>; workshop publications: <https://aimath.org/pastworkshops/markovmixing.html>.
- [2] B. Gerencsér, *Mixing time of an unaligned Gibbs sampler on the square*, Stochastic Processes and their Applications **129** (2019), 3570–3584. doi:10.1016/j.spa.2018.10.004. arXiv:1706.08694v2.
- [3] B. Gerencsér and A. Ottolini, *Rates of convergence for Gibbs sampling in the analysis of almost exchangeable data*, Stochastic Processes and their Applications **165** (2023), 440–464. doi:10.1016/j.spa.2023.08.008. Theorem and section numbering follows arXiv:2010.15539v2.
- [4] F. Ascolani, H. Lavenant and G. Zanella, *Entropy contraction of the Gibbs sampler under log-concavity*, arXiv:2410.00858v2, 2024, revised 20 January 2026.