IMPROVED RATES OF LEARNING GAUSSIAN GRAPHICAL MODELS FROM GLAUBER DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

We give a polynomial-time algorithm for learning Gaussian graphical models from a single Glauber trajectory, reducing the observation time from $\widetilde{O}(k^3/\alpha^5)$ (Shen et al., 2026) to $\widetilde{O}(k^2/\alpha^2)$. Here k is the maximum degree, α is the minimum normalized edge strength, and each coordinate updates at Poisson rate one. The guarantee is uniform over arbitrary initial states and independent of mixing time and condition number. Our method tests for edges using the three-update patterns i, j, i and i, i, j. Changes in other variables can distort the test, so we combine results across randomly chosen interval lengths, using weights that reduce this error. This permits longer windows and a higher rate of informative observations. The resulting bound has optimal inverse-square dependence on edge strength, up to logarithmic factors.

1 INTRODUCTION

In high-dimensional statistics, we often want to identify dependencies between variables despite random variation in the data. Gaussian graphical models (GGMs) are a common way to represent these conditional dependencies among continuous random variables. Their applications include studying brain connectivity (Huang et al., 2010) and financial systemic risk (Cerchiello & Giudici, 2016).

Write $[d] = \{1, \dots, d\}$. A GGM on d variables is a Gaussian random vector $x \sim \mathcal{N}(0, \Sigma)$, with its precision matrix $\Theta = \Sigma^{-1}$ encoding conditional independence between coordinates:

$$x_i \perp x_j \mid x_{[d]\setminus\{i,j\}} \quad \Longleftrightarrow \quad \Theta_{ij} = 0, \qquad i \neq j.$$

The associated graph of a GGM has an edge $\{i, j\}$ exactly when $\Theta_{ij} \neq 0$. A natural modeling assumption is that each variable depends directly on only a small number of others, even when the total number d of variables is large. We therefore consider k-sparse GGMs, whose graphs have maximum degree at most k.

> *Problem* 1 (Structure learning for sparse GGMs). Given observations from an unknown k-sparse GGM on d variables, recover its conditional-independence graph with high probability. How many observations are needed, and can recovery be performed in polynomial time?

Assume each edge has normalized strength at least α: $|\Theta_{ij}|/\sqrt{\Theta_{ii}\Theta_{jj}} \geq \alpha$. Under just sparsity and this signal condition, Misra et al. (2020) show that $O(k\alpha^{-2} \log d)$ independent samples suffice information-theoretically. Conversely, Wang et al. (2010) give a lower bound of $\Omega(\alpha^{-2} \log d)$ for worst-case recovery with success probability at least $3/4$, when $0 < \alpha \leq 1/4$. Statistical and computational efficiency need not come together: the algorithm of Misra et al. (2020) searches exhaustively over candidate neighborhoods, with up to d^k choices per vertex and $d^{O(k)}$ total computation. For general k, a polynomial-time algorithm with the same sample guarantee is not known under these assumptions (Kelner et al., 2024, Appendix D).

There is evidence for a computational-statistical gap in this problem. Kelner et al. (2024, Remark 43) show that detecting a sparse, near-critical negative spike in a Wishart model reduces to independent-sample GGM structure learning. They also establish a low-degree barrier for this detection problem (Kelner et al., 2024, Theorem 41). Recently, there has also been encouraging progress

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

in related settings: Chandrasekaran et al. (2026) obtain efficient prediction guarantees for sparse linear regression with uniformly random supports, even with highly ill-conditioned designs.

The observation process offers another way to approach efficient structure learning. In an interacting system, local changes reveal how one variable responds to the current values of the others. Learning from such updates has been studied for Ising models and other discrete graphical models (Bresler et al., 2018; Gaitonde et al., 2025). We study the Gaussian analogue: a single trajectory of Glauber dynamics. Each coordinate updates according to an independent rate-one Poisson process. At an update of coordinate i, its value is resampled from the conditional Gaussian distribution given the other coordinates, which remain unchanged. Write $x(t)$ for the observed state and $z = x(0)$ for its initial value. We observe z and the times, labels, and resulting states of all updates. The initial state may be arbitrary; the Gaussian distribution is stationary for the dynamics. Over observation time T, the trajectory length is the update count $n \sim \mathrm{Poisson}(dT)$, so $\mathbb{E}n = dT$.

Converting this trajectory into approximately independent observations by waiting for mixing can be costly: sparsity and minimum edge strength allow arbitrarily slow mixing; see Shen et al. (2026, Section 1 and Appendix D). Under the same observation model, Shen et al. (2026) learn directly from the updates, with guarantees independent of mixing time, conditioning, and the initial state. Their polynomial-time algorithm recovers the graph from a trajectory observed for time $O(k^3\alpha^{-5}\log(d/\delta))$, where δ is the failure probability. They also prove an observation-time lower bound of $\Omega(\alpha^{-2}\log d)$. Our upper bound matches the inverse-square dependence on edge strength required by this lower bound, up to logarithmic factors, and improves the dependence on degree.

1.1 Results

Let $\Theta \succ 0$ be a precision matrix on d variables, with maximum graph degree at most k and normalized edge strengths at least α. Let δ be the desired failure probability. The algorithms receive k, α, δ and the observed trajectory, including its initial state and update labels. The initial state is arbitrary and independent of future clocks and update noises. Each theorem guarantees exact graph recovery with probability at least $1 - \delta$, uniformly over Θ and the initial state. The notation $\widetilde{O}$ suppresses only logarithmic factors in $d, \alpha^{-1}, \delta^{-1}$.

Theorem 1.1 (Local structure learning)**.** *There is a randomized polynomial-time algorithm that recovers the graph from a trajectory observed for time*

$$T = \widetilde{O}(k^2\alpha^{-2}).$$

More precisely, $T \geq Ck^2\alpha^{-2}\log^7(C/\alpha)\log(d/\delta)$ suffices for a universal constant C.

For larger degrees, ordinary least squares gives a complementary guarantee. At each vertex's updates, we regress its new value on the other coordinates immediately before the update, and recover edges by thresholding products of the regression coefficients in the two directions.

Theorem 1.2 (Least-squares structure learning)**.** *Nodewise ordinary least squares recovers the graph in polynomial time from a trajectory observed for time*

$$T = O\left(d\alpha^{-2}\log\frac{Cd}{\alpha\delta}\right),$$

where C is a universal constant.

Together, these algorithms give a polynomial-time observation bound of $\widetilde{O}(\min\{k^2, d\}\alpha^{-2})$. If exhaustive search is allowed, we can replace the dimension dependence of OLS by the sparsity.

Theorem 1.3 (Subset selection)**.** *Selecting the least-squares fit with minimum residual over supports of size at most k at each vertex, followed by the same product threshold, recovers the graph from a trajectory observed for time*

$$T = O\left(k\alpha^{-2}\log\frac{Cd}{\alpha\delta}\right).$$

Its runtime is $d^{O(k)}\,\mathrm{poly}(T)$.

The lower bound of Shen et al. (2026) already holds for degree-one graphs with stationary initialization, $\alpha < 1/4$, and constant failure probability. The optimal dependence on degree remains open.

Ideas behind the proofs. Shen et al. (2026) use the pattern (i, i, j, i), where the extra i-update makes the additive numerator noise independent of the denominator when neighboring coordinates do not update. We use the shorter patterns (i, i, j) and (i, j, i) and look at the signs of the coordinate increments. When the other coordinates are fixed, (i, i, j) gives a correlation signal when the mean j-increment is small. When this mean is large, (i, j, i) gives a signal through the effect of the j-increment on the next i-update. The two signals have the same sign, so averaging the patterns works for every starting state. Longer intervals give more observations, but neighboring updates introduce bias. We combine several lengths with weights that cancel the first few bias terms (linear, quadratic, and so on) while preserving the edge signal.

For the regression results, each coordinate update gives a linear equation with fresh Gaussian noise. We analyze the noise accumulated between updates to lower-bound the empirical Gram matrix, and use a martingale inequality to bound the response-noise term. These bounds control the coefficient errors through the normal equations for OLS and a comparison with the true model's residual for sparse OLS.

1.2 Related work

Learning from dynamics. Bresler et al. (2018) initiated the study of learning graphical models from Glauber dynamics for Ising models, obtaining recovery guarantees without a mixing assumption. For higher-order Markov random fields, Gaitonde et al. (2025) show that access to dynamics can bypass the noisy-parity barrier associated with independent-sample learning. These results illustrate how the observation process can change the computational problem. Simchowitz et al. (2018, Section 2.4 and Theorem 2.4) similarly analyze ordinary least squares from a single linear-system trajectory without requiring mixing. Their proof uses conditional anti-concentration to lower-bound the empirical Gram matrix, an upper bound to make this control uniform over directions, and martingale bounds for the response-noise term. Our regression proof follows this strategy. Here fresh noise enters one coordinate at a time; we analyze the noise accumulated between successive updates of each target to obtain coefficientwise guarantees for graph recovery.

Comparison of structure-learning guarantees. Recent work has studied Problem 1 from a single Glauber trajectory (Tirukkonda et al., 2026; Shen et al., 2026; Tirukkonda & Dasarathy, 2026). Our results improve the polynomial parameter dependence of uniform observation bounds in both the polynomial-time and exhaustive-search regimes. Earlier local and mixing-based guarantees require additional contraction and variance assumptions (Tirukkonda et al., 2026).[1]

Table 1: Structure learning from a single Gaussian Glauber trajectory. All rows allow arbitrary initial states, and T uses rate-one coordinate updates.

Method	Observation time T	Computation
Shen et al.	$\widetilde{O}(k^3\alpha^{-5})$	Polynomial
LS-DuNeS	$O(k^2\alpha^{-2}H)$	$d^{O(k)}\,\mathrm{poly}(T)$
AU-DuNeS	$\widetilde{O}(k^4\alpha^{-2})$	$d^{O(k)}\,\mathrm{poly}(T)$
This work: local	$\widetilde{O}(k^2\alpha^{-2})$	Polynomial
This work: OLS	$\widetilde{O}(d\alpha^{-2})$	Polynomial
This work: subset	$\widetilde{O}(k\alpha^{-2})$	$d^{O(k)}\,\mathrm{poly}(T)$

Here $H = \log\big(Cdk\,\rho_{2k}(\Theta)(1+z^\top\Theta z)/(\alpha\delta)\big)$, where $\rho_{2k}(\Theta)$ is the maximal normalized covariance eigenvalue over $2k$-coordinate subsets, defined precisely in Appendix E.

The local algorithm retains the polynomial computation and uniformity of Shen et al. (2026), while improving the observation bound by a factor of k/α^3 up to logarithms. The sparse algorithms of Tirukkonda & Dasarathy (2026) test possible neighborhoods for each vertex, with $O(d^{2k+1})$ tests in total. LS-DuNeS permits arbitrary starts, but its required horizon depends on covariance geometry

[1] In the updated arXiv version, the local and DICE-based mixing bounds have polynomial factors $k^3(\beta_{\min}\sigma_{\min})^{-5}$ and $k\alpha^{-2}/(1-r)$, respectively; Appendix E records the additional assumptions and horizon restriction. Here $\beta_{\min} = \min_{i\neq j:\Theta_{ij}\neq 0}|\Theta_{ij}|/\Theta_{ii}$, $\sigma_{\min} = \min_i \Theta_{ii}^{-1/2}$, and $r = \max_i \sum_{j\neq i}|\Theta_{ij}|/\Theta_{ii} < 1$.

and initial energy through H; these quantities are unbounded in the model class. Our local algorithm retains its quadratic degree dependence with a uniform budget and polynomial computation. Relative to the uniform AU-DuNeS guarantee, the local result improves the degree dependence from quartic to quadratic and removes exhaustive search.

When exhaustive computation is allowed on both sides, subset selection improves the degree dependence to linear, by factors of k and k^3 relative to LS-DuNeS and AU-DuNeS, respectively, while retaining a uniform budget. These comparisons concern polynomial factors; the local bound has additional logarithmic factors.

For dense graphs, fixing the predictor set to contain all other vertices gives polynomial-time consequences of the DuNeS analyses: quadratic dimension dependence for LS-DuNeS, with its additional instance-dependent logarithm, and cubic dimension dependence for the uniform AU-DuNeS bound. Our OLS result reduces both to linear dimension dependence with a uniform budget. Appendix E gives these bounds and shows how to obtain them for dense graphs.

Tirukkonda & Dasarathy (2026) first obtained the optimal α^{-2} dependence without a mixing assumption, even from an arbitrary initial state.

2 Technical overview

2.1 Local algorithm

Our local algorithm tests each pair of vertices by examining short intervals of the trajectory. We construct a statistic from two three-update patterns whose ideal average expectation is zero on nonedges and has magnitude at least a constant times α on edges. Updates at other coordinates introduce a bias. By expressing this bias as a power series in the interval length, we can combine observations at several lengths to cancel its leading terms.

For the analysis, rescale coordinate x_i by $\sqrt{(\Theta_{\mathrm{raw}})_{ii}}$, where Θ_{raw} is the original precision matrix. The new precision matrix satisfies

$$\Theta_{ij} = \frac{(\Theta_{\mathrm{raw}})_{ij}}{\sqrt{(\Theta_{\mathrm{raw}})_{ii}(\Theta_{\mathrm{raw}})_{jj}}}, \qquad \Theta_{ii} = 1.$$

This preserves the graph and normalized edge strengths; Lemma A.6 gives the precise rescaling statement and proof. Thus, for an estimator invariant under positive coordinate scaling, we may analyze the rescaled coordinates with $\Theta_{ii} = 1$.

An update of coordinate i then has the form

$$x_i \leftarrow -\sum_{\ell \neq i} \Theta_{i\ell} x_\ell + \xi, \qquad \xi \sim \mathcal{N}(0,1),$$

with a fresh independent noise at each update. Coordinates update according to independent rate-one Poisson processes, and the trajectory includes the update labels. Recall that each vertex has at most k neighbors and every edge satisfies $|\Theta_{ij}| \geq \alpha$, where $0 < \alpha < 1$. We seek recovery with failure probability at most $\delta \in (0,1)$. The initial state may be arbitrary and is independent of future update times and noises.

Two complementary update patterns. Fix $i \neq j$ and write $\theta = \Theta_{ij}$. We accept an interval if its complete sequence of i- and j-updates is either (i,i,j) or (i,j,i); updates at other coordinates are unrestricted. Let Δx_i be the second resulting i-value minus the first, and Δx_j the change at the sole j-update. Our statistic is

$$g_{ij} = \mathbf{1}\{\Delta x_i \, \Delta x_j < 0\} - \frac{1}{2}. \tag{1}$$

Positive coordinate rescaling multiplies $\Delta x_i \, \Delta x_j$ by a positive factor, so it leaves g_{ij} unchanged. This justifies analyzing the statistic with $\Theta_{ii} = 1$, while the algorithm computes it directly in the observed coordinates.

First consider an ideal experiment starting from x, in which all other coordinates remain fixed and the two patterns are chosen with equal probability. Denote the expected statistic by $F_0(x)$. Let b be

the mean j-increment in this experiment, which is the same for both patterns, and let ξ_1, ξ_2, ξ_3 be the independent standard Gaussian noises at the designated updates. The increments have the following representations:

pattern	$U := \Delta x_i$	$V := \Delta x_j$
(i,i,j)	$\xi_2 - \xi_1$	$b - \theta\xi_2 + \xi_3$
(i,j,i)	$-\theta V + \xi_3 - \xi_1$	$b - \theta\xi_1 + \xi_2$.

If $\theta = 0$, then U is centered and independent of V in both rows, so $F_0(x) = 0$. For an edge, the two rows provide complementary signals. For example, the first row gives

$$\mathrm{Cov}(U,V) = -\theta, \qquad \mathrm{Corr}(U,V) = -\frac{\theta}{\sqrt{2(1+\theta^2)}}.$$

When $|b| \le 1$, the shift in V is bounded, and this correlation yields a sign bias of magnitude at least a constant times $|\theta|$. However, a large $|b|$ can make the sign of V almost deterministic and weaken this signal. The second pattern handles this regime: its means are $\mathbb{E}U = -\theta b$ and $\mathbb{E}V = b$, reflecting the effect of the j-increment on the second i-update. For $|b| \ge 1$, these shifts yield a sign bias of magnitude at least a constant times $|\theta|$. In both patterns the bias has the sign of θ, so they cannot cancel each other. The Gaussian sign calculation in Lemma C.1 therefore gives, for every x,

$$F_0(x) = 0 \quad \text{if } \theta = 0, \qquad \mathrm{sign}(\theta) F_0(x) \ge \frac{|\theta|}{50} \quad \text{if } \theta \ne 0. \tag{2}$$

The cost of short intervals. In an actual interval, neighboring coordinates may update and change the conditional means. For $u > 0$, let $\mathcal{A}_{ij}((0,u])$ denote acceptance on $(0,u]$ and define

$$F_u(x) = \mathbb{E}_x[g_{ij} \mid \mathcal{A}_{ij}((0,u])].$$

The two accepted patterns have equal probability. Thus $F_0(x)$ is the ideal quantity that $F_u(x)$ approaches as the interval shrinks.

The probability of an intervening neighbor update is $O(ku)$ for small u. Treating every such observation as arbitrary contamination requires this probability to be small compared with the signal α. This is the reason for the interval length $u \asymp \alpha/k$ in the analysis of Shen et al. (2026, Remark 5.10). The paper's four-update pattern (i,i,j,i) occurs with probability $\Theta(u^4)$ per interval, giving $\Theta(u^3)$ observations per unit time. Collecting $O(\alpha^{-2}\log(d/\delta))$ observations therefore requires time of order

$$\frac{\alpha^{-2}\log(d/\delta)}{u^3} \asymp k^3 \alpha^{-5} \log(d/\delta).$$

The three-update patterns above are more frequent, but shrinking their intervals to α/k would still be expensive. We instead use the expansion of the bias $F_u(x) - F_0(x)$ to control it at longer interval lengths.

How sparsity controls the expansion. For a bounded function f of the state, let $Q_v f(x)$ be its expectation after a single update of coordinate v, starting from x, and put $\mathcal{D}_v f = Q_v f - f$. If f depends only on coordinates in a set S, then $\mathcal{D}_v f = 0$ whenever $v \notin S$. However, when $v \in S$, averaging over its new value can introduce dependence on its neighbors, since those neighbors determine its conditional mean.

To see this for our statistic, consider the final i-update in the pattern (i,j,i). Let r_i be the recorded first i-value and r_j the initial j-value. With these two values held fixed, the terminal score is the function

$$f(x) = \mathbf{1}\{(x_i - r_i)(x_j - r_j) < 0\} - \frac{1}{2}.$$

Write $\mu_v(x) = -\sum_{\ell \ne v} \Theta_{v\ell} x_\ell$ for the conditional mean at coordinate v. Just before the final i-update, put $\chi = \mathrm{sign}(x_j - r_j)$; the preceding Gaussian j-update makes $x_j \ne r_j$ almost surely. Integrating over the new i-value gives

$$Q_i f(x) = \Phi(\chi[r_i - \mu_i(x)]) - \frac{1}{2},$$

where Φ is the standard Gaussian distribution function. The score itself reads only i and j, but this conditional expectation also reads the neighbors of i.

For a background coordinate $v \notin \{i, j\}$, we have $Q_v f = f$, and hence

$$\begin{aligned}\mathcal{D}_v \mathcal{D}_i f(x) &= Q_v Q_i f(x) - Q_i f(x) \\ &= \Phi\left(\frac{\chi[r_i - \mu_i(x) + \Theta_{iv}(\mu_v(x) - x_v)]}{\sqrt{1+\Theta_{iv}^2}}\right) - \Phi(\chi[r_i - \mu_i(x)])\,.\end{aligned}$$

Write $N(v) = \{j \neq v : \Theta_{vj} \neq 0\}$ for the neighborhood of coordinate v. If $v \notin N(i)$, then $\Theta_{iv} = 0$ and this difference vanishes. Otherwise, $\mu_v(x)$ can introduce dependence on $N(v)$. Thus a v-update can affect the score through the subsequent i-update.

Between designated updates, only background coordinates evolve. Their generator is $\mathcal{L}_0 = \sum_{v \notin \{i,j\}} \mathcal{D}_v$. For example, a background gap of length t followed by the final i-update gives

$$e^{t\mathcal{L}_0} Q_i f = Q_i f + t \sum_v \mathcal{D}_v Q_i f + \frac{t^2}{2} \sum_{v,w} \mathcal{D}_v \mathcal{D}_w Q_i f + \cdots,$$

where both sums run over background coordinates. Expanding one second-order term gives

$$\mathcal{D}_v \mathcal{D}_w Q_i f = Q_v Q_w Q_i f - Q_v Q_i f - Q_w Q_i f + Q_i f.$$

The term vanishes unless $w \in N(i)$ and v belongs to the dependence set created after averaging the w-update, which is contained in $N(i) \cup N(w) \cup \{j\}$. After applying ℓ background difference operators, whose coordinate indices $v_1, \ldots, v_\ell$ may repeat, the resulting function $\mathcal{D}_{v_1} \cdots \mathcal{D}_{v_\ell} Q_i f$ depends only on coordinates in $N(i) \cup N(v_1) \cup \cdots \cup N(v_\ell) \cup \{j\}$, a set of at most $(\ell+1)k + 1$ coordinates.

Conditional on an accepted pattern, its three update times divided by u are uniform order statistics in $(0, 1)$. The four gaps before, between, and after the designated updates are therefore u times random fractions whose distribution does not depend on u. We expand the background evolution in these gaps and compose it with the three designated updates. The statistic also remembers the initial j-value and the first resulting i-value; these recorded values remain fixed during background updates.

Each nonzero difference operator can add at most k coordinates to the dependence set. Since Q_v averages the function, $\|\mathcal{D}_v f\|_\infty \leq 2\|f\|_\infty$. Expanding over the four gaps gives (Lemma C.4)

$$F_u(x) = F_0(x) + \sum_{\ell \geq 1} c_\ell(x) u^\ell, \qquad |c_\ell(x)| \leq \frac{1}{2}(2k)^\ell \binom{\ell+4}{4}, \tag{3}$$

for $0 \leq u < 1/(2k)$. The series converges uniformly in x on every smaller closed interval. At $u \leq 1/(8k)$, its terms are bounded by a polynomial in ℓ times $4^{-\ell}$. The coefficients may depend on the unknown graph and on x, but this common bound lets us control the tail without estimating them using an interpolation technique below.

Canceling the leading bias terms. For $r \geq 2$ distinct positive lengths $u_0, \ldots, u_{r-1}$, the Lagrange interpolation weights

$$w_a = \prod_{\ell \neq a} \frac{-u_\ell}{u_a - u_\ell} \quad \text{satisfy} \quad \sum_a w_a = 1, \qquad \sum_a w_a u_a^\ell = 0 \quad (1 \leq \ell < r).$$

Consequently,

$$\sum_a w_a F_{u_a}(x) = F_0(x) + \sum_{\ell \geq r} c_\ell(x) \sum_a w_a u_a^\ell. \tag{4}$$

The first $r - 1$ bias terms disappear without estimating their coefficients.

The choice of lengths must also keep the weights small. Set $u_{\max} = 1/(8k)$ and $u_{\min} = u_{\max}/r^2$, and take the r Chebyshev extrema on $[u_{\min}, u_{\max}]$:

$$u_a = u_{\min} + \frac{u_{\max} - u_{\min}}{2}\left(1 - \cos\frac{a\pi}{r-1}\right), \qquad 0 \leq a < r.$$

For these lengths, the interpolation weights satisfy $\sum_a |w_a| < 4$ (Lemma C.6). Combining this bound with (3) gives

$$\left|\sum_a w_a F_{u_a}(x) - F_0(x)\right| \leq 2 \sum_{\ell \geq r} 4^{-\ell} \binom{\ell+4}{4}.$$

The tail decreases exponentially in r. Choosing $r = \lceil 10 + \log_2(1/\alpha) \rceil$ makes it at most $\varepsilon := \alpha/400$, for every x. Thus we can use lengths between $1/(8kr^2)$ and $1/(8k)$ and still keep the bias below the edge signal. The longer intervals make the accepted patterns more frequent: their probability is of order u^3 per interval, hence their rate per unit time is of order u^2. Lengths of order $1/k$, up to logarithmic factors, therefore cost k^2 time per accepted observation, again up to logarithmic factors.

Combining observations along one trajectory. At the beginning of window b, independently choose K_b uniformly from $\{0, \dots, r-1\}$ and observe an interval of length u_{K_b}. Let x_b be the state at that boundary, and write $g_{ij,b}$ for the statistic g_{ij} in window b when accepted.

The acceptance probability is $p_{\mathrm{acc}}(u) = e^{-2u}u^3/3$: exactly three updates of the pair must occur, in one of the two accepted orders. An accepted observation contributes $\widetilde{g}_{ij,b} = r w_{K_b} g_{ij,b}/p_{\mathrm{acc}}(u_{K_b})$, and a rejected observation contributes zero. Writing $\mathcal{F}_{b-1}$ for the history before the length is drawn, we have

$$\mathbb{E}[\widetilde{g}_{ij,b} \mid \mathcal{F}_{b-1}] = \sum_{a=0}^{r-1} \frac{1}{r}\, p_{\mathrm{acc}}(u_a)\, \frac{r w_a}{p_{\mathrm{acc}}(u_a)} F_{u_a}(x_b) = \sum_a w_a F_{u_a}(x_b). \tag{5}$$

Let $M_b^{ij} = \widetilde{g}_{ij,b} - \mathbb{E}[\widetilde{g}_{ij,b} \mid \mathcal{F}_{b-1}]$ be the centered score. To bound its fluctuations, put $p_{\min} = p_{\mathrm{acc}}(u_{\min}) \asymp k^{-3}r^{-6}$. Boundedness of g_{ij} and $\sum_a |w_a| < 4$ give

$$\mathbb{E}[\widetilde{g}_{ij,b}^2 \mid \mathcal{F}_{b-1}] = \frac{r}{4} \sum_a \frac{w_a^2}{p_{\mathrm{acc}}(u_a)} \le \frac{4r}{p_{\min}}, \qquad |\widetilde{g}_{ij,b}| \le \frac{2r}{p_{\min}}.$$

Thus the conditional variance of M_b^{ij} is at most $4r/p_{\min}$, and $|M_b^{ij}| \le 4r/p_{\min}$ deterministically. Freedman's inequality (Fact A.9) combines these two controls to give, for $0 < \varepsilon \le 1$,

$$\Pr\left(\left| \frac{1}{N_{\mathrm{win}}} \sum_{b=1}^{N_{\mathrm{win}}} M_b^{ij} \right| > \varepsilon \right) \le 2\exp\left(-\frac{c N_{\mathrm{win}} \varepsilon^2 p_{\min}}{r} \right).$$

Taking $\varepsilon = \alpha/400$ and a union bound over the vertex pairs, it is therefore enough to use

$$N_{\mathrm{win}} = O\left(\frac{r}{p_{\min}\alpha^2} \log \frac{d}{\delta} \right) = O\big(k^3 r^7 \alpha^{-2} \log(d/\delta)\big)$$

windows. Combining this fluctuation bound with the interpolation bias bound gives, simultaneously for all pairs with probability at least $1 - \delta$,

$$\left| \frac{1}{N_{\mathrm{win}}} \sum_{b=1}^{N_{\mathrm{win}}} \widetilde{g}_{ij,b} - \frac{1}{N_{\mathrm{win}}} \sum_{b=1}^{N_{\mathrm{win}}} F_0(x_b) \right| \le 2\varepsilon.$$

On a nonedge, every $F_0(x_b)$ is zero; on an edge, all have the sign of θ and magnitude at least $\alpha/50$. Thresholding the absolute average score at $\alpha/100$ therefore recovers the graph.

Each window has length at most $u_{\max} = 1/(8k)$, so the total observation time is at most $N_{\mathrm{win}} u_{\max}$. The resulting sufficient observation time is

$$T = O\big(k^2 \alpha^{-2} \log^7(C/\alpha) \log(d/\delta)\big).$$

2.2 Least-squares algorithms

When a coordinate updates, its new value is a linear combination of the other coordinates plus independent Gaussian noise. The observations are dependent and need not be near stationarity.

The estimators and their invariances. For observation time $T \ge 8$, retain the first $q = 2\lfloor T/4 \rfloor$ updates of each coordinate, so $q \asymp T$. Declare failure if any coordinate has fewer than q updates by time T, or if a tested design is singular. The total trajectory update count n is random, with $\mathbb{E}n = dT$.

The update equation has no intercept, so Algorithms 2 and 3 fit without one. We include an intercept here to simplify the presentation and analysis; Section D.1 shows that the same bounds hold without it. Let $\tau_{i,\ell}$ be the time of the ℓth update of coordinate i, and write $y_{i,\ell} = x_i(\tau_{i,\ell})$ for its response.

Write $\mathcal{R}_i(b_0, b)$ for the residual sum of squares over these q updates, predicting the new i-value from the other coordinates immediately before each update, with intercept b_0 and slopes $b = (b_j)_{j\neq i}$. The dense and subset-selection estimators are, respectively,

$$(\widehat{b}_{i,0}^{\mathrm{OLS}}, \widehat{\beta}_i^{\mathrm{OLS}}) \in \operatorname*{arg\,min}_{b_0\in\mathbb{R},\, b\in\mathbb{R}^{d-1}} \mathcal{R}_i(b_0, b),$$
$$(\widehat{b}_{i,0}^{\mathrm{subset}}, \widehat{\beta}_i^{\mathrm{subset}}) \in \operatorname*{arg\,min}_{\substack{b_0\in\mathbb{R},\, b\in\mathbb{R}^{d-1}\\ \|b\|_0\le k}} \mathcal{R}_i(b_0, b), \tag{6}$$

where $\|b\|_0$ counts the nonzero slopes; ties between supports are resolved by a fixed ordering. For either estimator, write $\widehat{\beta}_{ij} = (\widehat{\beta}_i)_j$, with zero entries outside the selected support, and define the estimated graph by

$$\widehat{E} = \Big\{\{i, j\} : i < j, \quad \widehat{\beta}_{ij}\widehat{\beta}_{ji} > \alpha^2/4\Big\}. \tag{7}$$

Suppose each coordinate x_j is rescaled to $a_j x_j$, with $a_j > 0$. For target i, rescaling b_0 by a_i and each b_j by a_i/a_j multiplies every residual by a_i. Thus every candidate support's minimum residual loss is multiplied by the same factor a_i^2, preserving the selected support.

$\widehat{\beta}_{ij}$ and $\widehat{\beta}_{ji}$ are multiplied by a_i/a_j and a_j/a_i, respectively. These factors cancel in $\widehat{\beta}_{ij}\widehat{\beta}_{ji}$, so $\widehat{E}$ is unchanged. We may therefore use the normalization from Section 2.1, writing Θ for the rescaled precision matrix, with $\Theta_{jj} = 1$. At the ℓth update of target i,

$$y_{i,\ell} = -\sum_{j\neq i} \Theta_{ij} x_j(\tau_{i,\ell}-) + \xi_{i,\ell}, \qquad \xi_{i,\ell} \sim \mathcal{N}(0, 1), \tag{8}$$

where $\xi_{i,\ell}$ is independent of the preceding history. The true slopes are $\beta_{ij} = -\Theta_{ij}$ and the true intercept is zero.

With an intercept, $\widehat{\beta}_i$ is unchanged by the initial-state translation $\widetilde{x} = x - z$, where $z = x(0)$. In those translated coordinates the true intercept is $-(\Theta z)_i$.

Uniform design and noise control. Fix a support-size bound $1 \le s \le d-1$, a target i, and a predictor support $S \subseteq [d] \setminus \{i\}$ with $p = |S| \le s$. Write $x_{S,\ell} = x_S(\tau_{i,\ell}-)$ and $\bar{x}_S = \frac{1}{q}\sum_{\ell=1}^{q} x_{S,\ell}$. Let $X_{i,S} \in \mathbb{R}^{q\times(p+1)}$ have rows $(1, x_{S,\ell}^\top)$, let $y_i = (y_{i,1}, \dots, y_{i,q})^\top$, and let $\xi_i = (\xi_{i,1}, \dots, \xi_{i,q})^\top$ collect the true update innovations. Define

$$G_{i,S} = \frac{1}{q} X_{i,S}^\top X_{i,S}, \qquad h_{f,i,S} = X_{i,S}^\top \xi_i, \qquad G_{i,S}^c = \frac{1}{q}\sum_{\ell=1}^{q} (x_{S,\ell} - \bar{x}_S)(x_{S,\ell} - \bar{x}_S)^\top. \tag{9}$$

$h_{f,i,S}$ is a sum over the q updates, while $G_{i,S}$ and $G_{i,S}^c$ are averages. The translation $\widetilde{x} = x - z$ leaves $G_{i,S}^c$ unchanged and, whenever $G_{i,S}$ is invertible, also leaves $h_{f,i,S}^\top G_{i,S}^{-1} h_{f,i,S}$ unchanged. Set

$$B_i = \int_0^1 e^{-2t\Theta_{-i,-i}}\, \mathrm{d}t, \qquad B_{i,S} = (B_i)_{SS}, \qquad L_s = s\log(CdT/\delta) + \log(d/\delta).$$

The matrix B_i lower-bounds the covariance of fresh noise introduced during a unit interval in which target i is held fixed, after averaging over the other coordinates' clocks.

The uniform design and score bounds (Lemmas D.1 and D.2) show that, if $T \gtrsim L_s$, then with probability at least $1 - \delta$ all retained updates occur by T and, simultaneously for every target i and nonempty support $1 \le |S| \le s$,

$$G_{i,S}^c \succeq cB_{i,S}, \qquad h_{f,i,S}^\top G_{i,S}^{-1} h_{f,i,S} \lesssim qL_s. \tag{10}$$

Moreover, deterministically, $(B_{i,S}^{-1})_{jj} \le 3$ for $j \in S$ (Fact D.7). The bounds hide only absolute constants. Here $h_{f,i,S} = X_{i,S}^\top \xi_i$ uses ξ_i from (8).

From prediction error to coefficient error. Work on the event where (10) holds simultaneously for all targets and supports, and suppress these indices. For an augmented coefficient error $\Delta_f = (\Delta_0, \Delta^\top)^\top$, where $\Delta = \widehat{\beta} - \beta_{\mathrm{true}}$ contains the slope errors and $\beta_{\mathrm{true}} = (-\Theta_{ij})_{j\in S}$,

$$\Delta_f^\top G \Delta_f = (\Delta_0 + \bar{x}_S^\top \Delta)^2 + \Delta^\top G^c \Delta.$$

The inequality $G^c \succeq cB$ in (10) gives $((G^c)^{-1})_{jj} \le c^{-1}(B^{-1})_{jj} \le 3/c$. Consequently, Cauchy–Schwarz yields

$$|\Delta_j|^2 \le \left((G^c)^{-1}\right)_{jj} \Delta^\top G^c \Delta \le \frac{3}{c}\, \Delta_f^\top G \Delta_f. \tag{11}$$

Dense ordinary least squares. Apply $G^c \succeq cB$ and $h_f^\top G^{-1} h_f \lesssim qL_s$ from (10) with $s = d-1$. Since the regression includes every true predictor, the normal equations give

$$\Delta_f = \frac{1}{q} G^{-1} h_f, \qquad \Delta_f^\top G \Delta_f = \frac{1}{q^2} h_f^\top G^{-1} h_f \lesssim \frac{L_{d-1}}{T}.$$

Equation (11) therefore bounds every squared slope error by $Cd\log(CdT/\delta)/T$.

Exhaustive subset selection. Apply $G^c \succeq cB$ and $h_f^\top G^{-1} h_f \lesssim qL_s$ from (10) with $s = \min\{2k, d-1\}$. Let S_{true} be the true support and S_{selected} the selected support, and set $S = S_{\mathrm{true}} \cup S_{\mathrm{selected}}$. Pad both augmented coefficient vectors with zeros on the union, and let Δ_f be their difference. The true coefficients are feasible in the search, so the selected fit has no larger residual sum of squares:

$$\|\xi_i - X_{i,S}\Delta_f\|^2 \le \|\xi_i\|^2.$$

Expanding, substituting $X_{i,S}^\top \xi_i = h_f$ and $X_{i,S}^\top X_{i,S} = qG$, and applying Cauchy–Schwarz gives

$$\Delta_f^\top G \Delta_f \le \frac{2}{q} \Delta_f^\top h_f \le \frac{2}{q} \sqrt{\Delta_f^\top G \Delta_f}\, \sqrt{h_f^\top G^{-1} h_f}.$$

Thus

$$\Delta_f^\top G \Delta_f \le \frac{4}{q^2} h_f^\top G^{-1} h_f \lesssim \frac{L_s}{T}, \qquad |\Delta_j|^2 \lesssim \frac{k\log(CdT/\delta)}{T}. \tag{12}$$

Once the upper bound on $|\Delta_j|^2$ in (12) is below α^2, a support S_{selected} omitting any true edge cannot win, as the squared coefficient error for that omitted edge is already at least α^2. Because (10) holds simultaneously for all supports with $1 \le |S| \le s$, it applies to any nonempty $S_{\mathrm{true}} \cup S_{\mathrm{selected}}$, even though S_{selected} depends on the observed trajectory.

Graph recovery and observation cost. The OLS inequality $|\Delta_j|^2 \lesssim d\log(CdT/\delta)/T$ and the subset-selection inequality (12) give $|\widehat{\beta}_{ij} - \beta_{ij}| \le \alpha/4$ at the respective observation times

$$T_{\mathrm{dense}} \gtrsim d\alpha^{-2} \log\frac{Cd}{\alpha\delta}, \qquad T_{\mathrm{subset}} \gtrsim k\alpha^{-2}\log\frac{Cd}{\alpha\delta}.$$

These choices also satisfy $T \gtrsim L_s$ in (10). For a nonedge, $\beta_{ij} = \beta_{ji} = 0$, so $|\widehat{\beta}_{ij}\widehat{\beta}_{ji}| \le \alpha^2/16$. For an edge, $\beta_{ij} = \beta_{ji} = -\Theta_{ij}$ with $|\Theta_{ij}| \ge \alpha$. Both $\widehat{\beta}_{ij}$ and $\widehat{\beta}_{ji}$ have the sign of $-\Theta_{ij}$ and satisfy $|\widehat{\beta}_{ij}|, |\widehat{\beta}_{ji}| \ge 3\alpha/4$, so $\widehat{\beta}_{ij}\widehat{\beta}_{ji} \ge 9\alpha^2/16$. Thresholding $\widehat{\beta}_{ij}\widehat{\beta}_{ji}$ at $\alpha^2/4$ as in (7) therefore recovers the graph.

Poisson concentration converts these time bounds into total-update budgets of orders $\widetilde{O}(d^2\alpha^{-2})$ and $\widetilde{O}(dk\alpha^{-2})$, respectively. Dense fitting takes $O(d^3T + d^4)$ arithmetic operations; exhaustive subset selection takes $d^{O(k)}\,\mathrm{poly}(T)$. Both statistical guarantees remain uniform over the arbitrary initial state.

AI USE STATEMENT

The paper was written by the human authors in collaboration with ChatGPT 6 Astra. AI was used throughout the theoretical development and writing, and the results were developed primarily by ChatGPT 6 Astra with human guidance and assistance. The authors have fully verified the proofs and take full responsibility for the accuracy, originality, and integrity of the paper.

We found the ideas behind these results surprising and elegant. Working through them changed our understanding of what can be learned from Glauber dynamics. We chose to share the results because we believe they provide new insight into the problem, beyond the improvement in rates.

REPRODUCIBILITY STATEMENT

This is a theoretical work. All assumptions are stated in the main text, and complete proofs of all results are given in the appendix.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, volume 24, pp. 2312–2320, 2011.

Noga Alon and Edward R. Scheinerman. Degrees of freedom versus dimension for containment orders. *Order*, 5:11–16, 1988. doi: 10.1007/BF00143892. URL https://web.math.princeton.edu/~nalon/PDFS/warren4.pdf.

Mathias Beiglböck and Pietro Siorpaes. Pathwise versions of the Burkholder–Davis–Gundy inequality. *Bernoulli*, 21(1):360–373, 2015. doi: 10.3150/13-BEJ570.

Guy Bresler, David Gamarnik, and Devavrat Shah. Learning graphical models from the Glauber dynamics. *IEEE Transactions on Information Theory*, 64(6):4072–4080, 2018. doi: 10.1109/TIT.2017.2713828.

Paola Cerchiello and Paolo Giudici. Conditional graphical models for systemic risk estimation. *Expert Systems with Applications*, 43:165–174, 2016. doi: 10.1016/j.eswa.2015.08.047.

Gautam Chandrasekaran, Raghu Meka, and Konstantinos Stavropoulos. Sparse linear regression is easy on random supports. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pp. 2242–2253. Association for Computing Machinery, 2026. doi: 10.1145/3798129.3800926.

Luc Devroye, László Györfi, and Gábor Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996. doi: 10.1007/978-1-4612-0711-5.

Klaus-Jochen Engel and Rainer Nagel. *One-Parameter Semigroups for Linear Evolution Equations*, volume 194 of *Graduate Texts in Mathematics*. Springer, 2000. doi: 10.1007/b97696.

Jason Gaitonde, Ankur Moitra, and Elchanan Mossel. Bypassing the noisy parity barrier: Learning higher-order Markov random fields from dynamics. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pp. 348–359, 2025.

Shuai Huang, Jing Li, Liang Sun, Jieping Ye, Adam Fleisher, Teresa Wu, Kewei Chen, Eric Reiman, and Alzheimer's Disease NeuroImaging Initiative. Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50(3):935–949, 2010. doi: 10.1016/j.neuroimage.2009.12.120.

Jonathan Kelner, Frederic Koehler, Raghu Meka, and Dhruv Rohatgi. Lasso with latents: Efficient estimation, covariate rescaling, and computational-statistical gaps. In *Proceedings of the 37th Conference on Learning Theory*, volume 247 of *Proceedings of Machine Learning Research*, pp. 2840–2886, 2024. URL https://proceedings.mlr.press/v247/kelner24a.html.

David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, 2 edition, 2017. doi: 10.1090/mbk/107. URL https://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf. With contributions by Elizabeth L. Wilmer.

Sidhant Misra, Marc Vuffray, and Andrey Y. Lokhov. Information theoretic optimal learning of Gaussian graphical models. In *Proceedings of the 33rd Conference on Learning Theory*, pp. 2888–2909, 2020.

Eric Shen, Tony Wu, Mahbod Majid, and Ankur Moitra. Learning Gaussian graphical models from a Glauber trajectory without mixing. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2606.31230v1. Cited version: arXiv:2606.31230v1.

Max Simchowitz, Horia Mania, Stephen Tu, Michael I. Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Proceedings of the 31st Conference on Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 439–473, 2018.

Vignesh Tirukkonda and Gautam Dasarathy. Mixing-free and signal-optimal learning of Gaussian graphical models from Glauber dynamics, 2026. URL https://arxiv.org/abs/2607.18559v2.

Vignesh Tirukkonda, Anirudh Rayas, and Gautam Dasarathy. Local and mixing-based algorithms for Gaussian graphical model selection from Glauber dynamics, 2026. URL https://arxiv.org/abs/2412.18594v3.

Wei Wang, Martin J. Wainwright, and Kannan Ramchandran. Information-theoretic bounds on model selection for Gaussian Markov random fields. In *IEEE International Symposium on Information Theory*, pp. 1373–1377, 2010.

Hugh E. Warren. Lower bounds for approximation by nonlinear manifolds. *Transactions of the American Mathematical Society*, 133(1):167–178, 1968. doi: 10.1090/S0002-9947-1968-0226281-1.

A PRELIMINARIES

Notation. The ambient dimension is d, the maximum-degree bound is k, and n denotes the total number of observed coordinate updates. We write $[d] = \{1, \dots, d\}$, and let e_i denote the ith standard basis vector. For $S \subseteq [d]$, x_S denotes the restriction of x to S, and M_{SS} the corresponding principal submatrix of a matrix M. The subscript $-i$ means that coordinate i is omitted. We use $\|x\|$ for the Euclidean norm and $\|M\|_{\mathrm{op}}$ for the induced matrix norm. For symmetric matrices, $M \succeq N$ means that $M - N$ is positive semidefinite. We write $\mathbf{1}\{\cdot\}$ for an indicator and $\|f\|_\infty = \sup_x |f(x)|$ for the supremum norm of a bounded function. The notation $a \lesssim b$ means $a \le Cb$ for a universal constant C; $a \asymp b$ means that both $a \lesssim b$ and $b \lesssim a$ hold. Unless a base is specified, logarithms are natural, and $\widetilde{O}(\cdot)$ suppresses factors polylogarithmic in d, $1/\alpha$, and $1/\delta$.

A.1 GAUSSIAN GRAPHICAL MODELS

A Gaussian graphical model (GGM) is a Gaussian distribution $\mathcal{N}(0, \Sigma)$ with $\Sigma \succ 0$. Its precision matrix is $\Theta = \Sigma^{-1}$. The associated graph $\mathcal{G} = ([d], E)$ has an edge $\{i, j\}$ whenever $i \neq j$ and $\Theta_{ij} \neq 0$. We write $N(i) = \{j \neq i : \Theta_{ij} \neq 0\}$.

Definition A.1 ((α, k)-sparse GGM)**.** Let $0 < \alpha < 1$ and $k \ge 1$. A GGM is (α, k)-sparse if its graph has maximum degree at most k and

$$\frac{|\Theta_{ij}|}{\sqrt{\Theta_{ii}\Theta_{jj}}} \ge \alpha \qquad \text{for every } \{i, j\} \in E.$$

Given the other coordinate values x_{-i}, a conditional draw at vertex i has law

$$x_i^{\mathrm{new}} \sim \mathcal{N}\left(-\sum_{j \neq i} \frac{\Theta_{ij}}{\Theta_{ii}} x_j,\ \frac{1}{\Theta_{ii}} \right). \tag{13}$$

A.2 CONTINUOUS-TIME GLAUBER DYNAMICS

We use the same observation model as Shen et al. (2026). Each coordinate updates according to an independent rate-1 Poisson clock. When coordinate i updates, we resample its value according to (13), leaving every other coordinate unchanged. So if the state just before an update is x, the new value at i is

$$x_i^{\mathrm{new}} = -\sum_{j \neq i} \frac{\Theta_{ij}}{\Theta_{ii}} x_j + \Theta_{ii}^{-1/2} \xi, \qquad \xi \sim \mathcal{N}(0, 1). \tag{14}$$

The Gaussian noises are independent across updates and independent of the clocks and the initial state. We denote the resulting right-continuous process by $(x(t))_{t \ge 0}$ and its fixed initial state by $z = x(0)$. This is the continuous-time random Gibbs sampler, also known as heat-bath or Glauber dynamics (Levin & Peres, 2017).

The learner observes z and all update times, updated indices, and new coordinate values up to time T. It is given d, k, α and a failure probability $0 < \delta < 1$, but does not know Θ. The trajectory length $n = n(T)$ is the total number of updates observed by time T; it is Poisson with mean dT. The Poisson Chernoff bound gives

$$\mathbb{P}\big(n \notin [dT/2, 3dT/2]\big) \le 2e^{-dT/12}.$$

Thus $dT/2 \le n \le 3dT/2$ with probability at least $1 - \delta$ whenever $dT \ge 12 \log(2/\delta)$.

Definition A.2 (Uniform structure learning)**.** An algorithm learns (α, k)-sparse GGMs uniformly in time $T = T(d, k, \alpha, \delta)$ if, for every such GGM and every deterministic initial state $z \in \mathbb{R}^d$, it recovers E with probability at least $1 - \delta$ from the trajectory up to time T.

Poisson clocks. For a set $S \subseteq [d]$ and an interval I, let $N_S(I)$ be the number of updates in I with indices in S. We use the following standard properties of Poisson processes.

Fact A.3 (Poisson counts and update times)**.** *Let $N(I)$ count the updates of a rate-r Poisson process in a deterministic interval I of length u. Then*

$$\mathbb{P}(N(I) = \ell) = e^{-ru}\frac{(ru)^\ell}{\ell!}, \qquad \ell = 0, 1, 2, \dots.$$

Conditional on $N(I) = \ell$, the update times are the order statistics of ℓ independent uniform points in I.

Fact A.4 (Updating multiple coordinate clocks)**.** *For a nonempty set $S \subseteq [d]$, the combined updates of the coordinates in S form a Poisson process of rate $|S|$. Their labels are independent and uniform on S, independently of the update times. The clocks outside S are independent of these times and labels.*

Fact A.5 (Restarting the clocks)**.** *Let τ be an almost surely finite stopping time for the history of the trajectory, clocks, and any independent algorithmic randomness revealed so far. Conditional on this history through τ, the clock increments after τ are independent rate-1 Poisson processes, and subsequent update noises are independent standard Gaussians, independent of those clocks. Under this conditioning, the future trajectory has the law of a fresh trajectory started at $x(\tau)$. The same conclusion holds after conditioning on a window length drawn at τ independently of the history and future clocks and noises.*

This is the strong Markov property of the independent marked Poisson processes driving the dynamics.

A.3 Diagonal rescaling

Lemma A.6. *Let $D_a = \mathrm{diag}(a_1, \dots, a_d)$ with $a_i > 0$. If $x(t)$ follows Glauber dynamics with precision matrix Θ, then $D_a x(t)$, with the same update times and labels, follows Glauber dynamics with precision matrix $D_a^{-1}\Theta D_a^{-1}$. These two precision matrices have the same normalized precision matrix and the same graph.*

Proof. Multiplying (14) by a_i gives

$$a_i x_i^{\mathrm{new}} = -\sum_{j\neq i}\frac{a_i\Theta_{ij}}{a_j\Theta_{ii}}(a_j x_j) + \frac{a_i}{\sqrt{\Theta_{ii}}}\xi.$$

These are precisely the conditional mean and noise standard deviation for $D_a^{-1}\Theta D_a^{-1}$. Also,

$$\frac{\Theta_{ij}/(a_i a_j)}{\sqrt{(\Theta_{ii}/a_i^2)(\Theta_{jj}/a_j^2)}} = \frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}}. \qquad \square$$

For normalization, write Θ_{raw} for the original precision and set

$$D_{\mathrm{raw}} = \mathrm{diag}\big((\Theta_{\mathrm{raw}})_{11}, \dots, (\Theta_{\mathrm{raw}})_{dd}\big).$$

Applying $D_a = D_{\mathrm{raw}}^{1/2}$ to both the trajectory and its initial state gives precision

$$\Theta = D_{\mathrm{raw}}^{-1/2}\Theta_{\mathrm{raw}}D_{\mathrm{raw}}^{-1/2}, \qquad \Theta_{ii} = 1.$$

In these normalized coordinates, we retain $x(t)$ for the trajectory and $z = x(0)$ for its initial state. An estimator invariant under positive coordinate rescaling can therefore be analyzed with $\Theta_{ii} = 1$ while using the original observations. The edge-strength condition becomes $|\Theta_{ij}| \geq \alpha$ on every edge. Positive definiteness gives $|\Theta_{ij}| < 1$ for $i \neq j$, since the corresponding principal minor has determinant $1 - \Theta_{ij}^2 > 0$. Sparsity gives $\|\Theta\|_{\mathrm{op}} \leq k + 1$ by the maximum absolute row sum bound. We impose no positive lower bound on the smallest eigenvalue of this normalized precision.

A.4 Update operators and the transition semigroup

For a precision matrix Θ, not necessarily normalized, a bounded measurable function $f : \mathbb{R}^d \to \mathbb{R}$, and a vertex v, define

$$(Q_v f)(x) = \mathbb{E}\left[f\left(x + e_v\left[-x_v - \sum_{w\neq v}\frac{\Theta_{vw}}{\Theta_{vv}}x_w + \Theta_{vv}^{-1/2}\xi\right]\right)\right], \qquad \xi \sim \mathcal{N}(0,1). \tag{15}$$

Thus $(Q_v f)(x)$ is the expected value of f after one update at v from state x, and $\|Q_v f\|_\infty \le \|f\|_\infty$. When Θ has diagonal one, this simplifies to

$$(Q_v f)(x) = \mathbb{E}\left[f\left(x + e_v\left[-(\Theta x)_v + \xi\right]\right)\right], \qquad \xi \sim \mathcal{N}(0,1).$$

Write $\mathcal{D}_v = Q_v - \mathrm{Id}$ for the expected change from that update.

For $\mathcal{B} \subseteq [d]$, let $x^{\mathcal{B}}(t)$ be the dynamics in which only the clocks in $\mathcal{B}$ run. Its transition semigroup and infinitesimal generator are, respectively,

$$(P_t^{\mathcal{B}} f)(x) = \mathbb{E}_x[f(x^{\mathcal{B}}(t))], \qquad \mathcal{L}_{\mathcal{B}} = \sum_{v \in \mathcal{B}} \mathcal{D}_v. \tag{16}$$

Here $\mathbb{E}_x$ denotes expectation for the process started at x. The next lemma is the bounded-generator semigroup formula of Engel & Nagel (2000, Theorem I.3.7 and Proposition I.3.5).

Lemma A.7. *For every bounded measurable f,*

$$P_t^{\mathcal{B}} f = e^{t\mathcal{L}_{\mathcal{B}}} f = \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} \mathcal{L}_{\mathcal{B}}^\ell f.$$

The series converges in supremum norm for every fixed t. Moreover, $\frac{\mathrm{d}}{\mathrm{d}t} P_t^{\mathcal{B}} f = \mathcal{L}_{\mathcal{B}} P_t^{\mathcal{B}} f = P_t^{\mathcal{B}} \mathcal{L}_{\mathcal{B}} f$, where the derivative is taken in supremum norm.

Proof. If $\mathcal{B} = \varnothing$, the transition operator is the identity and $\mathcal{L}_{\mathcal{B}} = 0$. Otherwise, conditional on ℓ updates in $\mathcal{B}$ by time t, their labels are independent and uniform on $\mathcal{B}$ by Fact A.4. Averaging one update therefore applies $|\mathcal{B}|^{-1} \sum_{v \in \mathcal{B}} Q_v$, and averaging ℓ successive updates applies its ℓth power. Averaging over the Poisson count from Fact A.3 gives

$$P_t^{\mathcal{B}} = e^{-|\mathcal{B}|t} \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} \Big(\sum_{v \in \mathcal{B}} Q_v \Big)^\ell = e^{t\mathcal{L}_{\mathcal{B}}}.$$

Since $\|\mathcal{L}_{\mathcal{B}}\| \le 2|\mathcal{B}|$ in the operator norm induced by the supremum norm, the exponential series and its derivative converge in operator norm on every bounded time interval. □

Definition A.8 (Coordinate support)**.** A function f is supported on $S \subseteq [d]$ if $f(x) = f(x')$ whenever $x_S = x'_S$. Thus support refers to the coordinates on which f depends, rather than the set on which f is nonzero.

Suppose f is supported on S. If $v \notin S$, updating coordinate v leaves f unchanged, so $Q_v f = f$ and $\mathcal{D}_v f = 0$. If $v \in S$, the new value at v depends on the current state only through the coordinates in $N(v)$. Hence $Q_v f$ depends only on $(S \setminus \{v\}) \cup N(v)$, and subtracting f shows that $\mathcal{D}_v f = Q_v f - f$ is supported on $S \cup N(v)$. Finally, the triangle inequality and $\|Q_v f\|_\infty \le \|f\|_\infty$ give

$$\|\mathcal{D}_v f\|_\infty \le \|Q_v f\|_\infty + \|f\|_\infty \le 2\|f\|_\infty. \tag{17}$$

A.5 Martingales

We use the following consequence of Freedman's inequality.

Fact A.9 (Freedman's inequality)**.** *Let $(M_a, \mathcal{F}_a)_{a=0}^K$ be a martingale with $M_0 = 0$. Suppose its increments satisfy $|M_a - M_{a-1}| \le b$ and*

$$\sum_{a=1}^{K} \mathbb{E}[(M_a - M_{a-1})^2 \mid \mathcal{F}_{a-1}] \le v \qquad \textit{almost surely}.$$

Then, for every $\lambda > 0$,

$$\mathbb{P}(|M_K| \ge \lambda) \le 2\exp\left(-\frac{\lambda^2}{2(v + b\lambda/3)}\right).$$

Fact A.10 (Fourth moment of a jump martingale)**.** *Let M be a real, locally square-integrable càdlàg martingale starting at zero, with no continuous martingale part and with continuous predictable quadratic variation $\langle M\rangle$. Write $\Delta M_r = M_r - M_{r-}$ for its jump at time r. Then*

$$\mathbb{E}|M_t|^4 \lesssim \mathbb{E}\langle M\rangle_t^2 + \mathbb{E}\sum_{0<r\le t} |\Delta M_r|^4. \tag{18}$$

The assertion applies after localization whenever the right side is finite.

Proof. The Burkholder–Davis–Gundy inequality with exponent 4 gives $\mathbb{E}|M_t|^4 \lesssim \mathbb{E}[M]_t^2$; see Beiglböck & Siorpaes (2015, Theorem 1.1 and its continuous-time extension). Here $[M]_t = \sum_{0<r\le t} |\Delta M_r|^2$. Since $\langle M\rangle$ is continuous, the martingale $[M] - \langle M\rangle$ has jumps $|\Delta M_r|^2$. Martingale isometry and the inequality $(a+b)^2 \le 2a^2 + 2b^2$ yield

$$\mathbb{E}[M]_t^2 \le 2\mathbb{E}\langle M\rangle_t^2 + 2\mathbb{E}([M]_t - \langle M\rangle_t)^2 = 2\mathbb{E}\langle M\rangle_t^2 + 2\mathbb{E}\sum_{0<r\le t} |\Delta M_r|^4.$$

Localization and Fatou's lemma give the stated version. □

Fact A.11 (Gaussian self-normalized score)**.** *Let $(\mathcal{F}_a)_{a=0}^q$ be a filtration. Let $\psi_a \in \mathbb{R}^r$ be $\mathcal{F}_{a-1}$-measurable and let ε_a be $\mathcal{F}_a$-measurable with $\varepsilon_a \mid \mathcal{F}_{a-1} \sim \mathcal{N}(0,1)$, for $a = 1, \dots, q$. Define*

$$V = \sum_{a=1}^q \psi_a \psi_a^\top, \qquad g = \sum_{a=1}^q \psi_a \varepsilon_a.$$

For any fixed positive-definite Γ_0 and $0 < \epsilon < 1$, with probability at least $1-\epsilon$,

$$g^\top (V+\Gamma_0)^{-1} g \le \log\det(I + \Gamma_0^{-1/2} V \Gamma_0^{-1/2}) + 2\log(1/\epsilon). \tag{19}$$

The result also holds conditional on auxiliary randomness, provided the stated Gaussian law and predictability hold under that conditioning and Γ_0 is fixed under it.

Proof. We give the Gaussian-mixture argument for this specialization of the self-normalized inequality of Abbasi-Yadkori et al. (2011, Theorem 1). When V is nonsingular, completing the square gives

$$\sup_v \{v^\top g - \tfrac{1}{2} v^\top V v\} = \tfrac{1}{2} g^\top V^{-1} g. \tag{20}$$

For a direction v fixed before the innovations, put $b_a = v^\top \psi_a$. Completing the square in the scalar Gaussian density gives

$$\mathbb{E}[e^{b\varepsilon - b^2/2}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/2 + bt - b^2/2}\, \mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(t-b)^2/2}\, \mathrm{d}t = 1.$$

Because b_a is known before ε_a, the product $\prod_{a=1}^\ell \exp\{b_a\varepsilon_a - b_a^2/2\}$ is a nonnegative martingale of expectation one. The tower property, without any independence between its factors, therefore gives

$$\mathbb{E}\exp\{v^\top g - \tfrac{1}{2} v^\top V v\} = 1.$$

The maximizing direction $V^{-1}g$ in (20) depends on the data, so it cannot be substituted into this fixed-direction expectation identity.

Instead average the exponential against an independent $v \sim \mathcal{N}(0, \Gamma_0^{-1})$, and call the resulting scalar $\mathcal{Z}$. Tonelli's theorem gives $\mathbb{E}\mathcal{Z} = 1$. To evaluate the integral, put $\Lambda = V + \Gamma_0$ and $b = \Lambda^{-1} g$. The combined exponent from the Gaussian density and the martingale is

$$-\tfrac{1}{2} v^\top \Lambda v + v^\top g = -\tfrac{1}{2}(v-b)^\top \Lambda (v-b) + \tfrac{1}{2} g^\top \Lambda^{-1} g.$$

Integrating the translated Gaussian density yields

$$\mathcal{Z} = \frac{\det(\Gamma_0)^{1/2}}{\det(V+\Gamma_0)^{1/2}} \exp\{\tfrac{1}{2} g^\top (V+\Gamma_0)^{-1} g\}. \tag{21}$$

This calculation is valid even when V is singular because $V + \Gamma_0$ is positive definite. Applying Markov's inequality to $\mathcal{Z}$ and taking logarithms proves (19). □

B ALGORITHM PSEUDOCODE

All three algorithms use the labeled trajectory $(x(t))_{0\le t\le T}$, including $z = x(0)$, in the original coordinate scales. Algorithmic randomness is independent of the trajectory's clocks and update noises. Assume an (α, k)-sparse GGM as in Definition A.1, with $d \ge 2$, $1 \le k \le d-1$, and failure probability $\delta \in (0, 1/4)$.

Local observations and parameters. For $i < j$ and an interval I, let $\mathcal{A}_{ij}(I)$ be the event that the complete subsequence of i- and j-update labels in I is (i, i, j) or (i, j, i); other coordinates may update freely. Let Δx_i be the second resulting i-value minus the first, and Δx_j the change at the sole j-update. On this event, define $g_{ij}(I) = \mathbf{1}\{\Delta x_i\, \Delta x_j < 0\} - 1/2$. For the bth window I_b, write $g_{ij,b} = g_{ij}(I_b)$ when it is accepted.

Set $r = \lceil 10 + \log_2(1/\alpha)\rceil$, $u_{\min} = 1/(8kr^2)$, and $u_{\max} = 1/(8k)$. For $0 \le a < r$, define $u_a = u_{\min} + (u_{\max} - u_{\min})(1 - \cos(a\pi/(r-1)))/2$. The weight is $w_a = \prod_{\ell\ne a} \frac{-u_\ell}{u_a - u_\ell}$, where the product ranges over $0 \le \ell < r$; the acceptance probability is $p_a = e^{-2u_a} u_a^3/3$. Use $N_{\mathrm{win}} = \lceil 48re^{2u_{\min}} u_{\min}^{-3}(400/\alpha)^2 \log(4d^2/\delta)\rceil$ windows.

Algorithm 1 Local structure learning

Require: Trajectory up to time $T \ge N_{\mathrm{win}} u_{\max}$; parameters d, k, α, δ.
1: Set u_a, w_a, p_a and N_{win} as defined above.
2: $t_0 \leftarrow 0$; $S_{ij} \leftarrow 0$ for all $i < j$.
3: **for** $b = 1, \dots, N_{\mathrm{win}}$ **do**
4: Draw K_b uniformly from $\{0, \dots, r-1\}$, independently of the past.
5: $t_b \leftarrow t_{b-1} + u_{K_b}$; $I_b \leftarrow (t_{b-1}, t_b]$.
6: **for** $1 \le i < j \le d$ **do**
7: **if** $\mathcal{A}_{ij}(I_b)$ occurs **then**
8: Compute $g_{ij,b} = \mathbf{1}\{\Delta x_i\, \Delta x_j < 0\} - 1/2$.
9: $S_{ij} \leftarrow S_{ij} + r w_{K_b} g_{ij,b}/p_{K_b}$.
10: **end if**
11: **end for**
12: **end for**
13: **return** $\{\{i, j\} : i < j,\ |S_{ij}/N_{\mathrm{win}}| > \alpha/100\}$.

Computational complexity. For an event-list input containing n coordinate updates, the arithmetic cost is $O(n + d^2 N_{\mathrm{win}} + r^2)$: one pass through the updates, $O(d^2)$ pair tests per window, and $O(r^2)$ operations to form the interpolation weights.

Ordinary least squares. Theorem 1.2 guarantees recovery with probability at least $1-\delta$ when $T \geq Cd\alpha^{-2}\log(Cd/(\alpha\delta))$, for a sufficiently large universal constant C. Each target contributes $q = 2\lfloor T/4 \rfloor$ retained regression observations. The fits below omit the intercept; Section D proves the guarantee for this choice and for the version with an intercept.

Algorithm 2 Structure learning by ordinary least squares

Require: Labeled trajectory through time $T \geq 8$; d, α.
1: **return** FAILURE if the event-list scan reads more than $\lceil 2dT \rceil$ updates.
2: $q \leftarrow 2\lfloor T/4 \rfloor$.
3: **return** FAILURE if any coordinate has fewer than q updates by T.
4: **for** $i = 1, \ldots, d$ **do**
5: Using the first q updates of i, regress its new value on the other coordinates immediately before each update, without an intercept.
6: **return** FAILURE if the least-squares solution is not unique.
7: Let $\widehat{\beta}_{ij}$ be the fitted coefficient of coordinate j, for $j \neq i$.
8: **end for**
9: **return** $\big\{\{i,j\} : i < j,\ \widehat{\beta}_{ij}\widehat{\beta}_{ji} > \alpha^2/4\big\}$.

Computational complexity. Standard dense least squares uses $O(qd^2)$ arithmetic operations per vertex. Including the event-list scan, the total cost is $O(n + d^3 q) = O(n + d^3 T)$.

Sparse least squares. Theorem 1.3 guarantees recovery with probability at least $1-\delta$ when $T \geq Ck\alpha^{-2}\log(Cd/(\alpha\delta))$, for a sufficiently large universal constant C.

Algorithm 3 Sparse least squares by exhaustive subset selection

Require: Inputs of Algorithm 2, together with k.
1: Run Algorithm 2, replacing its per-vertex fit by the following rule:
2: Fit the same regression on every predictor support of size at most k.
3: **return** FAILURE if any candidate least-squares solution is not unique.
4: Keep the fit with the smallest residual sum of squares; break ties by a fixed order.
5: Use its coefficients as $\widehat{\beta}_{ij}$, setting those outside the selected support to zero.

Computational complexity. There are $\sum_{p=0}^{k}\binom{d-1}{p}$ candidate supports per vertex, including the empty support. The arithmetic cost is $O\Big(n + dqk^2\sum_{p=0}^{k}\binom{d-1}{p}\Big)$, using $O(qk^2)$ operations per candidate fit.

C Proof of the local structure-learning guarantee

We prove Theorem 1.1 for Algorithm 1, using the accepted-pattern event $\mathcal{A}_{ij}(I)$ and observer g_{ij} defined in Section B. For an accepted interval $I = (t, t+u]$, write $x^0 = x(t)$ and let x^1, x^2, x^3 be the states immediately after its three designated updates. The increments defining g_{ij} are then

$$\Delta x_i = x_i^3 - x_i^1, \qquad \Delta x_j = x_j^3 - x_j^0.$$

As noted in Section 2.1, $|g_{ij}| = 1/2$ and the observer is invariant under positive coordinate rescaling. Lemma A.6 therefore allows us to work with precision matrix Θ with $\Theta_{jj} = 1$ throughout this proof.

Fix a pair $i < j$ and write $\theta = \Theta_{ij}$. Positive definiteness gives $|\theta| < 1$. We suppress the fixed pair (i,j) in the notation for expected scores.

C.1 The ideal sign gap

Recall the ideal experiment from Section 2.1: start at x, choose (i,i,j) or (i,j,i) with equal probability, perform those three updates, and hold every other coordinate fixed. Its expected observer is $F_0(x)$.

Lemma C.1 (Ideal sign gap). *For every $x \in \mathbb{R}^d$, if $\theta = 0$, then $F_0(x) = 0$. If $\theta \neq 0$, then*

$$\operatorname{sign}(\theta) F_0(x) \geq \frac{|\theta|}{50}.$$

Proof. Put

$$\mu_i = -\sum_{v \neq i} \Theta_{iv} x_v, \qquad b = -\sum_{v \notin \{i,j\}} \Theta_{jv} x_v - \theta \mu_i - x_j.$$

Let ξ_1, ξ_2, ξ_3 be the independent standard Gaussian noises at the three updates. Substituting into the update rule gives

pattern	$U := \Delta x_i$	$V := \Delta x_j$
(i,i,j)	$\xi_2 - \xi_1$	$b - \theta\xi_2 + \xi_3$
(i,j,i)	$-\theta V + \xi_3 - \xi_1$	$b - \theta\xi_1 + \xi_2$.

(22)

If $\theta = 0$, then U is centered Gaussian and independent of V in both cases, so $\mathbb{P}(UV < 0) = 1/2$ and $F_0(x) = 0$.

We may assume $\theta > 0$. Indeed, for fixed b, replacing θ by $-\theta$ and negating (ξ_1, ξ_2) in the first row, or (ξ_1, ξ_3) in the second, sends (U, V) to $(-U, V)$. These changes preserve the laws of the noises and negate $g = \mathbf{1}\{UV < 0\} - 1/2$ almost surely. Thus a bound valid for every b and $\theta > 0$ gives the required bound for $\theta < 0$ as well. Write ϕ and Φ for the standard normal density and distribution function.

For the (i,i,j) pattern, $U \sim \mathcal{N}(0, 2)$ and

$$V \mid U = u \sim \mathcal{N}\left(b - \frac{\theta}{2}u, \varsigma^2\right), \qquad \varsigma^2 = 1 + \frac{\theta^2}{2}.$$

Conditioning on U and pairing its positive and negative values gives

$$\begin{aligned}\mathbb{E}[g \mid (i,i,j)] &= \int_0^\infty \frac{\phi(u/\sqrt{2})}{\sqrt{2}} \left[\mathbb{P}(V < 0 \mid U = u) + \mathbb{P}(V > 0 \mid U = -u) - 1\right] \mathrm{d}u \\ &= \int_0^\infty \frac{\phi(u/\sqrt{2})}{\sqrt{2}} \left[\Phi\left(\frac{b + \theta u/2}{\varsigma}\right) - \Phi\left(\frac{b - \theta u/2}{\varsigma}\right)\right] \mathrm{d}u \geq 0.\end{aligned}$$

To lower-bound this integral, for any $\mu \in \mathbb{R}$ and $a \geq 0$,

$$\begin{aligned}\Phi(\mu + a) - \Phi(\mu - a) &= \phi(\mu) \int_0^a e^{-t^2/2}(e^{\mu t} + e^{-\mu t})\, \mathrm{d}t \\ &\geq 2a\phi(\mu) e^{-a^2/2},\end{aligned}$$

since $e^{\mu t} + e^{-\mu t} \geq 2$ and $e^{-t^2/2} \geq e^{-a^2/2}$ on $[0, a]$. Applying this with $\mu = b/\varsigma$ and $a = \theta u/(2\varsigma)$ yields

$$\begin{aligned}\mathbb{E}[g \mid (i,i,j)] &\geq \frac{\theta\phi(b/\varsigma)}{\varsigma} \int_0^\infty \frac{u}{2\sqrt{\pi}} \exp\left(-\frac{1 + \theta^2}{4\varsigma^2}u^2\right) \mathrm{d}u \\ &= \frac{\theta\varsigma\phi(b/\varsigma)}{\sqrt{\pi}(1 + \theta^2)}.\end{aligned}$$

The last equality uses $\int_0^\infty u e^{-cu^2}\, \mathrm{d}u = 1/(2c)$. If $|b| \leq 1$, then $\varsigma \geq 1$, $\phi(b/\varsigma) \geq \phi(1)$, and $1 + \theta^2 \leq 2$ because $\theta < 1$. Therefore

$$\mathbb{E}[g \mid (i,i,j)] \geq \frac{\theta e^{-1/2}}{2\pi\sqrt{2}} > \frac{\theta}{16}. \tag{23}$$

For the (i,j,i) pattern, the Gaussian pair (U, V) has

$$\mathbb{E}U = -\theta b, \qquad \mathbb{E}V = b, \qquad \operatorname{Var}(U) = v := 2 - \theta^2 + \theta^4 \leq 2, \qquad \operatorname{Var}(V) = w := 1 + \theta^2 \leq 2,$$

and $\operatorname{Cov}(U, V) = -\theta^3 < 0$. In particular, $\psi(t) := \mathbb{P}(U < 0 \mid V = t)$ is increasing in t. If V' is an independent copy of V, then

$$2\operatorname{Cov}(\psi(V), \mathbf{1}\{V > 0\}) = \mathbb{E}\left[(\psi(V) - \psi(V'))(\mathbf{1}\{V > 0\} - \mathbf{1}\{V' > 0\})\right] \geq 0.$$

Hence $\mathbb{P}(U<0,V>0)\geq\mathbb{P}(U<0)\mathbb{P}(V>0)$. The same argument with two decreasing functions gives the corresponding bound for $\mathbb{P}(U>0,V<0)$. Adding the two bounds,

$$\begin{aligned}\mathbb{E}[g\mid(i,j,i)]&\geq\Phi\left(\frac{\theta b}{\sqrt{v}}\right)\Phi\left(\frac{b}{\sqrt{w}}\right)+\Phi\left(-\frac{\theta b}{\sqrt{v}}\right)\Phi\left(-\frac{b}{\sqrt{w}}\right)-\frac{1}{2}\\&=2\left[\Phi\left(\frac{\theta|b|}{\sqrt{v}}\right)-\frac{1}{2}\right]\left[\Phi\left(\frac{|b|}{\sqrt{w}}\right)-\frac{1}{2}\right]\geq 0.\end{aligned}$$

If $|b|\geq 1$, monotonicity of Φ and $v,w\leq 2$ give

$$\mathbb{E}[g\mid(i,j,i)]\geq 2\left[\Phi\left(\frac{\theta}{\sqrt{2}}\right)-\frac{1}{2}\right]\left[\Phi\left(\frac{1}{\sqrt{2}}\right)-\frac{1}{2}\right]\geq\frac{\theta}{25}. \tag{24}$$

For the last inequality, use

$$\Phi(t)-\frac{1}{2}=\int_0^t\phi(\tau)\,\mathrm{d}\tau\geq t\phi(1)>\frac{t}{5},\qquad 0<t\leq 1.$$

Both patterns have nonnegative expected score. If $|b|\leq 1$, their average is at least $\theta/32$ by (23); if $|b|\geq 1$, it is at least $\theta/50$ by (24). Thus $F_0(x)\geq\theta/50$ for $\theta>0$. The sign symmetry above proves the claim for $\theta<0$. □

C.2 The effect of background updates

Definition C.2 (Mean observer at interval length u)**.** For $u>0$ and an initial state x, define

$$F_u(x)=\mathbb{E}_x[g_{ij}\mid\mathcal{A}_{ij}((0,u])],$$

where $x=x^0$ is the state at the start of the interval, before all three designated updates, and all coordinates are allowed to update. The first resulting i-value x_i^1 is used to compute Δx_i, and is averaged over in this expectation.

The value at zero refers to the ideal experiment.

Lemma C.3 (Acceptance probability)**.** *For every $u>0$, the acceptance probability is*

$$p_{\mathrm{acc}}(u):=\mathbb{P}(\mathcal{A}_{ij}((0,u]))=\frac{e^{-2u}u^3}{3}. \tag{25}$$

Proof. The combined clock of i and j has rate 2, and each label is independently i or j with equal probability. Thus the probability of either specified order is

$$\mathbb{P}(N_{\{i,j\}}((0,u])=3)\,2^{-3}=e^{-2u}\frac{(2u)^3}{3!}\,2^{-3}=\frac{e^{-2u}u^3}{6}.$$

The two orders have equal probability, proving (25). □

Lemma C.4 (Power-series bound for the mean observer)**.** *There are coefficients $c_\ell(x)$ such that, for $0\leq u<1/(2k)$,*

$$F_u(x)=\sum_{\ell=0}^{\infty}c_\ell(x)u^\ell,\qquad c_0(x)=F_0(x),\qquad |c_\ell(x)|\leq\frac{1}{2}(2k)^\ell\binom{\ell+4}{4}. \tag{26}$$

The series converges uniformly in x and $u\in[0,\rho]$ for every $\rho<1/(2k)$.

Proof. By Lemma C.3, the two accepted orders are equally likely. Conditional on either order, its three update times divided by u are uniform order statistics in $(0,1)$. All other clocks remain independent rate-1 clocks.

Augment the current coordinate vector x' with two registers: r_j stores the initial j-value, and r_i stores the resulting value at the first designated i-update. The terminal function is

$$f(x',r_j,r_i)=\mathbf{1}\{(x_i'-r_i)(x_j'-r_j)<0\}-\frac{1}{2}.$$

It has supremum norm $1/2$. With r_j, r_i held fixed, its physical coordinate support is $\{i, j\}$. We keep the registers fixed when counting support below. Background updates leave both registers fixed. Initialize them to $(r_j, r_i) = (x_j, 0)$; the initial value of r_i is immaterial because the first designated update overwrites it.

Fix an order and its designated times. Let s_0, s_1, s_2, s_3 be the four intervening gap lengths divided by u, so that $s_a \geq 0$ and $\sum_{a=0}^{3} s_a = 1$. On this augmented state space, let $\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3$ be the designated update operators. The operator $\mathcal{Q}_1$ also saves the first resulting i-value in r_i. Each $\mathcal{Q}_a$ is a contraction in supremum norm. Let

$$\mathcal{L}_0 = \sum_{v \notin \{i,j\}} \mathcal{D}_v$$

be the background generator, acting only on physical coordinates. During each gap only the background clocks run, so Lemma A.7 gives the transition operator $e^{us_a\mathcal{L}_0}$ for gap a. Composing these transitions with the three designated updates, the expected observer conditional on the gaps and order is the evaluation at $(x, x_j, 0)$ of

$$e^{us_0\mathcal{L}_0}\,\mathcal{Q}_1 e^{us_1\mathcal{L}_0}\,\mathcal{Q}_2 e^{us_2\mathcal{L}_0}\,\mathcal{Q}_3 e^{us_3\mathcal{L}_0} f. \tag{27}$$

The factors are listed in chronological order, but act on the terminal function from right to left. For fixed finite d, $\mathcal{L}_0$ is bounded on bounded measurable functions, so Lemma A.7 lets us expand all four exponentials and collect coefficients.

Consider a term with a total of ℓ background difference operators, and apply its operators backward from f. A background operator $\mathcal{D}_v$ gives zero unless v is in the current physical support. If it is nonzero, it adds at most k coordinates to that support and multiplies the norm bound by at most two. Each designated operator also adds at most k physical coordinates: its new value depends only on the neighbors of the updated vertex and a fresh noise. Saving that value in a register does not introduce any further physical dependence. Hence, before the $(h+1)$st background operator, the support has size at most

$$2 + 3k + hk \leq k(h+5).$$

There are therefore at most

$$k^\ell \prod_{h=0}^{\ell-1} (h+5)$$

choices of nonzero background labels, for any fixed allocation among the four gaps. The norm of each such term is at most $2^\ell \|f\|_\infty = 2^{\ell-1}$.

If the four exponential orders are $\ell_0, \ldots, \ell_3$, their scalar coefficient is $\prod_{a=0}^{3} s_a^{\ell_a}/\ell_a!$. Summing over all allocations with total ℓ, the multinomial theorem gives

$$\sum_{\ell_0+\cdots+\ell_3=\ell} \prod_{a=0}^{3} \frac{s_a^{\ell_a}}{\ell_a!} = \frac{(s_0+s_1+s_2+s_3)^\ell}{\ell!} = \frac{1}{\ell!}.$$

Thus the coefficient of u^ℓ in (27) has absolute value at most

$$\frac{(2k)^\ell}{2\ell!} \prod_{h=0}^{\ell-1} (h+5) = \frac{1}{2}(2k)^\ell \binom{\ell+4}{4}.$$

Averaging this bound over the gap fractions and the two equally likely orders proves the coefficient bound. Its majorant is summable for $2ku < 1$, which also proves the asserted uniform convergence. Finally, setting $u = 0$ replaces every background semigroup by the identity and leaves exactly the frozen-coordinate experiment. Thus $c_0(x) = F_0(x)$. □

C.3 Cancellation across interval lengths

For a fixed starting state x, a weighted combination of interval lengths gives

$$\sum_a w_a F_{u_a}(x) = F_0(x) \sum_a w_a + \sum_{\ell \geq 1} c_\ell(x) \sum_a w_a u_a^\ell.$$

Lemma C.5 (Cancellation of powers)**.** *Let $r \geq 1$ be an integer, let $u_0, \dots, u_{r-1}$ be distinct positive numbers, and define*

$$w_a = \prod_{\substack{0 \leq \ell < r \\ \ell \neq a}} \frac{-u_\ell}{u_a - u_\ell}. \tag{28}$$

Then, for every polynomial p of degree at most $r-1$,

$$\sum_{a=0}^{r-1} w_a p(u_a) = p(0).$$

In particular, $\sum_a w_a = 1$ and $\sum_a w_a u_a^\ell = 0$ for $1 \leq \ell < r$.

Proof. The Lagrange basis polynomials

$$\ell_a(u) = \prod_{\ell \neq a} \frac{u - u_\ell}{u_a - u_\ell}$$

satisfy $\ell_a(u_j) = \mathbf{1}\{a = j\}$. Thus $p(u) - \sum_a p(u_a)\ell_a(u)$ has degree at most $r-1$ and vanishes at all r nodes, so it is identically zero. Evaluating at $u = 0$ proves the claim, since $\ell_a(0) = w_a$. □

Use the window parameters and Chebyshev nodes specified before Algorithm 1. In particular, $u_{\max} = 1/(8k)$, $u_{\min} = u_{\max}/r^2$, and $u_0 = u_{\min} < u_1 < \cdots < u_{r-1} = u_{\max}$. Their weights are exactly (28).

Lemma C.6 (Bound on the weights)**.** *For these nodes,*

$$W := \sum_{a=0}^{r-1} |w_a| = \cosh\left(2(r-1)\operatorname{artanh}(1/r)\right) < 4.$$

Proof. Map $[u_{\min}, u_{\max}]$ to $[-1, 1]$ by $\zeta(u) = (2u - u_{\min} - u_{\max})/(u_{\max} - u_{\min})$. Then

$$\zeta(0) = -\frac{r^2+1}{r^2-1}, \qquad \zeta(u_a) = -\cos\frac{a\pi}{r-1}.$$

Let T_{r-1} be the Chebyshev polynomial, characterized by $T_{r-1}(\cos t) = \cos((r-1)t)$. At the chosen nodes, $T_{r-1}(\zeta(u_a)) = (-1)^{r-1-a}$. Since $u_0 < \cdots < u_{r-1}$ are positive, the product formula gives $\operatorname{sign}(w_a) = (-1)^a$. The summands $w_a T_{r-1}(\zeta(u_a))$ therefore all have sign $(-1)^{r-1}$. Applying Lemma C.5 to the degree-$(r-1)$ polynomial $T_{r-1}(\zeta(u))$ yields

$$W = |T_{r-1}(\zeta(0))| = \cosh\left((r-1)\operatorname{arcosh}\frac{r^2+1}{r^2-1}\right).$$

The identity $\operatorname{arcosh}((r^2+1)/(r^2-1)) = 2\operatorname{artanh}(1/r)$ gives the claimed expression. Finally,

$$\operatorname{artanh}(1/r) = \int_0^{1/r} \frac{\mathrm{d}t}{1-t^2} < \frac{r}{r^2-1} < \frac{1}{r-1},$$

so $W < \cosh(2) < 4$. □

Lemma C.7 (Interpolation error)**.** *For every state x,*

$$\left| \sum_{a=0}^{r-1} w_a F_{u_a}(x) - F_0(x) \right| \leq \frac{\alpha}{400}.$$

Proof. By Lemmas C.5 and C.4,

$$\begin{aligned} \left| \sum_a w_a F_{u_a}(x) - F_0(x) \right| &\leq W \sup_{0 \leq u \leq u_{\max}} \sum_{\ell \geq r} |c_\ell(x)| u^\ell \\ &\leq \frac{W}{2} \sum_{\ell \geq r} 4^{-\ell} \binom{\ell+4}{4}. \end{aligned}$$

Here $2ku_{\max} = 1/4$. Since $0 < \alpha < 1$, our choice gives $r \geq 11$. For $\ell \geq 11$, consecutive terms in the final sum have ratio

$$\frac{4^{-(\ell+1)}\binom{\ell+5}{4}}{4^{-\ell}\binom{\ell+4}{4}} = \frac{\ell+5}{4(\ell+1)} \leq \frac{1}{3}.$$

Also $\binom{\ell+4}{4} \leq (2/3)2^\ell$ for $\ell \geq 11$: it holds at $\ell = 11$, and the ratio of the left side to 2^ℓ decreases thereafter. Thus

$$\sum_{\ell \geq r} 4^{-\ell}\binom{\ell+4}{4} \leq \frac{3}{2}\,4^{-r}\binom{r+4}{4} \leq 2^{-r}.$$

Using $W < 4$ and $2^{-r} \leq \alpha/2^{10}$, the interpolation error is less than $2^{1-r} \leq \alpha/512 < \alpha/400$. □

C.4 Concentration and graph recovery

Proof of Theorem 1.1. Use Algorithm 1 with its stated parameters. For this proof, put $p_{\min} = e^{-2u_{\min}}u_{\min}^3/3$, $\varepsilon = \alpha/400$, $L = \log(4d^2/\delta)$, and $\overline{V} = 4r/p_{\min}$. Its explicit window budget then equals $N_{\text{win}} = \lceil 4\overline{V}\varepsilon^{-2}L \rceil$. Since $e^{-2u}u^3/3$ increases for $0 < u < 3/2$ and $u_{\max} \leq 1/8$, every p_a is at least $p_{\min}$. Recall that $t_0 = 0$, $t_b = t_{b-1} + u_{K_b}$, and $I_b = (t_{b-1}, t_b]$, where K_b is drawn uniformly and independently at the start of window b. Let $\mathcal{F}_{b-1}$ be the history through time t_{b-1}, including the previous length choices but before drawing K_b. Write $x_b = x(t_{b-1})$ for the current state. For a fixed pair $i < j$, let its contribution in window b be

$$\widetilde{g}_{ij,b} = \begin{cases} \dfrac{rw_{K_b}}{p_{K_b}} g_{ij,b}, & \mathcal{A}_{ij}(I_b) \text{ occurs}, \\ 0, & \text{otherwise}. \end{cases}$$

Thus the algorithm's accumulated score is $S_{ij} = \sum_{b=1}^{N_{\text{win}}} \widetilde{g}_{ij,b}$. The fresh choice of K_b and the restart property in Fact A.5 give

$$\mathbb{E}[\widetilde{g}_{ij,b} \mid \mathcal{F}_{b-1}] = \sum_{a=0}^{r-1} \frac{1}{r}\,\frac{rw_a}{p_a}\,p_a F_{u_a}(x_b) = \sum_{a=0}^{r-1} w_a F_{u_a}(x_b). \tag{29}$$

Every term on the right starts from the same state x_b. Lemma C.7 therefore implies

$$|\mathbb{E}[\widetilde{g}_{ij,b} \mid \mathcal{F}_{b-1}] - F_0(x_b)| \leq \varepsilon. \tag{30}$$

Separate averages of the accepted observations at each length would involve different starting states, so cancellation would not follow directly from the interpolation identity.

Next, $g_{ij,b}^2 = 1/4$ on every accepted observation. Consequently,

$$\begin{aligned} \mathbb{E}[(\widetilde{g}_{ij,b})^2 \mid \mathcal{F}_{b-1}] &= \frac{r}{4}\sum_{a=0}^{r-1}\frac{w_a^2}{p_a} \\ &\leq \frac{r}{4p_{\min}}\left(\sum_a |w_a|\right)^2 \leq \frac{4r}{p_{\min}} = \overline{V}. \end{aligned}$$

Also $|\widetilde{g}_{ij,b}| \leq rW/(2p_{\min}) \leq 2r/p_{\min}$. The centered increments $M_b^{ij} = \widetilde{g}_{ij,b} - \mathbb{E}[\widetilde{g}_{ij,b} \mid \mathcal{F}_{b-1}]$ thus satisfy $|M_b^{ij}| \leq \overline{V}$, and their total conditional variance is at most $N_{\text{win}}\overline{V}$. Freedman's inequality (Fact A.9) gives

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{1}{N_{\text{win}}}\sum_{b=1}^{N_{\text{win}}} M_b^{ij}\right| \geq \varepsilon\right) \\ &\qquad \leq 2\exp\left(-\frac{N_{\text{win}}\varepsilon^2}{2\overline{V}(1+\varepsilon/3)}\right) \leq 2e^{-L}. \end{aligned}$$

The last inequality uses $N_{\text{win}} \geq 4\overline{V}\varepsilon^{-2}L$ and $\varepsilon < 1$. Taking a union bound over the fewer than $d^2/2$ pairs, with probability at least $1 - \delta/4$ we have simultaneously

$$\left|\frac{S_{ij}}{N_{\text{win}}} - \frac{1}{N_{\text{win}}}\sum_{b=1}^{N_{\text{win}}} F_0(x_b)\right| \leq 2\varepsilon = \frac{\alpha}{200}.$$

For a nonedge, each $F_0(x_b)$ is zero, so $|S_{ij}/N_{\mathrm{win}}| \le \alpha/200$. For an edge, Lemma C.1 gives the same sign in every window and

$$\operatorname{sign}(\Theta_{ij})\frac{1}{N_{\mathrm{win}}}\sum_{b=1}^{N_{\mathrm{win}}} F_0(x_b) \ge \frac{|\Theta_{ij}|}{50} \ge \frac{\alpha}{50}.$$

It follows that $|S_{ij}/N_{\mathrm{win}}| \ge 3\alpha/200$. The threshold $\alpha/100$ therefore distinguishes all edges from nonedges. Every window has length at most $u_{\max}$, so N_{win} windows finish by the deterministic time $N_{\mathrm{win}}u_{\max}$. Since $u_{\min} = 1/(8kr^2)$ and $u_{\min} \le 1/8$,

$$p_{\min}^{-1} = 3e^{2u_{\min}}u_{\min}^{-3} = O(k^3r^6).$$

Thus

$$N_{\mathrm{win}} = O(k^3r^7\alpha^{-2}L), \qquad N_{\mathrm{win}}u_{\max} = O(k^2r^7\alpha^{-2}L),$$

which gives the stated bounds because $r = O(\log(C/\alpha))$.

Finally, suppose the event-list input contains n updates. One pass through this list suffices to process the windows. Within each window, keep the starting value, the update count, and the times and resulting values of the first two updates of each coordinate. These determine acceptance and the observer for every pair. Checking all pairs costs $O(d^2)$ per window, and forming the weights by their product formulas costs $O(r^2)$. The total arithmetic cost is therefore $O(n + d^2N_{\mathrm{win}} + r^2)$. □

D ORDINARY LEAST SQUARES ALONG THE UPDATE TRAJECTORY

Theorems D.4 and D.5 give the detailed OLS and sparse OLS guarantees, respectively, and imply Theorems 1.2 and 1.3. Each allows fits with or without an intercept. Their proofs follow the shared regression bounds; the remaining subsections prove those bounds.

D.1 SETUP AND UNIFORM REGRESSION BOUNDS

We use the observation model of Section A.2 and the normalization in Lemma A.6. In this section $x(t)$ denotes the normalized observed trajectory, Θ its precision, and $z = x(0)$ its fixed initial state. Thus $\Theta_{jj} = 1$, and (14) becomes $x_j^{\mathrm{new}} = x_j^{\mathrm{old}} - (\Theta x^{\mathrm{old}})_j + \xi$. All bounds are uniform over fixed z, so conditioning on an independent random initial state gives the same guarantees.

Let $d \geq 2$, $1 \leq s \leq d-1$, $0 < \delta < 1/4$, and $T \geq 8$. Set

$$q = 2\lfloor T/4 \rfloor, \qquad L_s = L_s(\delta), \qquad L_s(\varepsilon) = s\log(CdT/\varepsilon) + \log(d/\varepsilon). \tag{31}$$

For T divisible by four, $q = T/2$. For each target i, retain its first q updates, at times $\tau_{i,1}, \dots, \tau_{i,q}$. We analyze these updates on an indefinitely continued trajectory and later intersect with the event that they all occur by T.

Fix $S \subseteq [d] \setminus \{i\}$ with $p = |S| \leq s$. The observed predictors and their empirical mean are

$$x_{S,\ell} = x_S(\tau_{i,\ell}-), \qquad \bar{x}_S = \frac{1}{q}\sum_{\ell=1}^{q} x_{S,\ell}.$$

The responses are $y_{i,\ell} = x_i(\tau_{i,\ell})$, collected in $y_i = (y_{i,1}, \dots, y_{i,q})^\top \in \mathbb{R}^q$. Observations are rows of the $q \times (p+1)$ augmented design:

$$X = \begin{bmatrix} 1 & x_{S,1}^\top \\ \vdots & \vdots \\ 1 & x_{S,q}^\top \end{bmatrix}, \qquad G = \frac{1}{q}X^\top X, \qquad G^c = \frac{1}{q}\sum_{\ell=1}^{q}(x_{S,\ell} - \bar{x}_S)(x_{S,\ell} - \bar{x}_S)^\top. \tag{32}$$

Let $\xi = (\xi_1, \dots, \xi_q)^\top$ collect the innovations at the retained updates of i, and define the unnormalized response score $h_f = X^\top \xi$. The row of X at update ℓ is known before ξ_ℓ is drawn, although later rows can depend on ξ_ℓ.

Let σ be the complete clock schedule, with z still fixed. Write

$$\bar{X} = \mathbb{E}[X \mid \sigma], \qquad Z = X - \bar{X}, \qquad D = \frac{1}{q}\bar{X}^\top \bar{X}. \tag{33}$$

The matrix $\bar{X}$ is the design generated by setting all innovations to zero. It is distinct from the empirical predictor mean $\bar{x}_S$. Its intercept column consists of ones, and the intercept column of Z is zero. In particular, D is the Gram matrix of the conditional mean; $\mathbb{E}[G \mid \sigma] = D + q^{-1}\mathbb{E}[Z^\top Z \mid \sigma]$.

The deterministic comparison matrices are

$$B_i = \int_0^1 e^{-2t\Theta_{-i,-i}}\,\mathrm{d}t, \qquad B_{i,S} = (B_i)_{SS}. \tag{34}$$

The restriction to S is taken after computing the full $(d-1)$-dimensional matrix B_i. It generally differs from $\int_0^1 e^{-2t\Theta_{SS}}\,\mathrm{d}t$. Suppressing i, S, define

$$B = B_{i,S}, \quad J = \mathrm{diag}(0, B), \quad R = \max\{1, C_R s d^2 T/\delta\}, \quad F = J + D/R, \tag{35}$$

where C_R is a sufficiently large universal constant. The parameter R is predetermined and F is fixed conditional on the clocks. For $S = \varnothing$, set $G = D = 1$, $J = 0$, and $F = 1/R$. Subscripts i, S will be restored when simultaneity matters.

Lemma D.1 (Centered design excitation)**.** *There are universal constants $c, C > 0$ such that, if $T \geq CL_s$, then with probability at least $1-\delta$ all retained updates occur by T and, simultaneously for every target i and every nonempty $S \subseteq [d] \setminus \{i\}$ with $|S| \leq s$,*

$$G_{i,S}^c \succeq cB_{i,S}. \tag{36}$$

The proof also establishes $c_F F_{i,S} \preceq G_{i,S} \preceq 2RF_{i,S}$ for a universal $c_F > 0$, including empty supports.

Lemma D.2 (Self-normalized response score)**.** *Under the same assumptions, with probability at least $1-\delta$ all retained updates occur by T, all the augmented Grams are nonsingular, and simultaneously for every target i and every support $|S| \le s$,*

$$h_{f,i,S}^{\top} G_{i,S}^{-1} h_{f,i,S} \le CqL_s. \tag{37}$$

Both lemmas hold on a common event of probability at least $1-\delta$.

Regression models and estimators. For the recovery proofs, assume $1 \le k \le d-1$, $0 < \alpha < 1$, maximum graph degree at most k, and $|\Theta_{ij}| \ge \alpha$ on every edge in the normalized model. At the ℓth retained update of target i, the normalized dynamics give

$$y_{i,\ell} = \sum_{j \ne i} \beta_{ij} x_j(\tau_{i,\ell}-) + \xi_{i,\ell}, \qquad \beta_{ij} = -\Theta_{ij}. \tag{38}$$

The innovation $\xi_{i,\ell}$ is standard Gaussian and independent of the history before that update. Write $\xi_i = (\xi_{i,1}, \dots, \xi_{i,q})^{\top}$ when displaying the target index. The true intercept is zero, and the true support $S_{\mathrm{true}} = \{j \ne i : \beta_{ij} \ne 0\}$ has size at most k. For $b_0 \in \mathbb{R}$ and $b \in \mathbb{R}^{d-1}$, define

$$\mathcal{R}_i(b_0, b) = \sum_{\ell=1}^{q} \left(y_{i,\ell} - b_0 - \sum_{j \ne i} b_j x_j(\tau_{i,\ell}-) \right)^2. \tag{39}$$

The estimators in Algorithms 2 and 3 are

$$\begin{aligned} \widehat{\beta}_i^{\mathrm{OLS}} &\in \operatorname*{arg\,min}_{b \in \mathbb{R}^{d-1}} \mathcal{R}_i(0, b), \\ \widehat{\beta}_i^{\mathrm{subset}} &\in \operatorname*{arg\,min}_{\substack{b \in \mathbb{R}^{d-1} \\ \|b\|_0 \le k}} \mathcal{R}_i(0, b). \end{aligned} \tag{40}$$

The variants with an intercept also minimize over $b_0 \in \mathbb{R}$. Subset selection fits every support of size at most k, including the empty support, and breaks ties between supports by a fixed ordering. An empty support has the unique empty slope vector; its intercept is zero for a fit through the origin and the empirical response mean for a fit with an intercept. Either algorithm declares failure if a target has fewer than q updates by T or a tested least-squares fit is not unique. Both algorithms also stop reading the event list and declare failure if its update count exceeds $\lceil 2dT \rceil$. Pad the selected slope vector by zeros outside its support and return

$$\widehat{E} = \left\{ \{i, j\} : i < j, \quad \widehat{\beta}_{ij} \widehat{\beta}_{ji} > \alpha^2/4 \right\}. \tag{41}$$

Apply Lemmas D.1 and D.2 with failure probability $\delta/2$ and support bound s. Denote their common event by $\mathcal{E}_s$. It has probability at least $1 - \delta/2$ whenever $T \ge CL_s(\delta/2)$. On this event all retained observations are available, every relevant augmented Gram is nonsingular, and

$$G_{i,S}^{c} \succeq cB_{i,S} \quad (S \ne \varnothing), \qquad h_{f,i,S}^{\top} G_{i,S}^{-1} h_{f,i,S} \le CqL_s(\delta/2). \tag{42}$$

Lemma D.3 (Regression reduction)**.** *For either intercept convention, the fitted reciprocal products and selected supports are invariant under positive coordinate rescaling. On $\mathcal{E}_s$, every nonempty support S of size at most s and augmented coefficient error $\Delta_f = (\Delta_0, \Delta^{\top})^{\top}$ satisfy $|\Delta_j|^2 \le (3/c)\Delta_f^{\top} G_{i,S} \Delta_f$ for $j \in S$. All fits through the origin on supports of size at most s are unique on this event.*

Suppose $1 \le s_0 \le d$, $s \le 2s_0$, and, for every target, the union of the true and fitted supports has size at most s. If on $\mathcal{E}_s$ the prediction errors satisfy $\Delta_f^{\top} G_{i,S} \Delta_f \le K_0 L_s(\delta/2)/q$ for a universal constant K_0, then a sufficiently large universal C_0 makes (41) *recover the graph with probability at least $1-\delta$ whenever $T \ge C_0 s_0 \alpha^{-2} \log(C_0 d/(\alpha\delta))$. On this success event all retained observations are available and the update count is at most $3dT/2$, so the event-list cap is not reached.*

Proof. Rescaling. Let Θ_{raw} be the precision in the original coordinates. For positive factors a_j, replace each coordinate x_j by $a_j x_j$. The coefficient transformation $(b_0, b_j) \mapsto (a_i b_0, (a_i/a_j) b_j)$ is a bijection on the fits of every fixed support and multiplies each residual for target i by a_i. It preserves

the zero-intercept constraint, uniqueness, and the ordering of all candidate residual losses, since every minimum loss is multiplied by a_i^2. The fixed rule for ties therefore selects the same support. The fitted reciprocal product is unchanged because the factors a_i/a_j and a_j/a_i cancel. Taking $a_j = \sqrt{(\Theta_{\mathrm{raw}})_{jj}}$ and using Lemma A.6 justifies the normalization above while the algorithms use the original observations. In the normalized model, each edge has $|\beta_{ij}| = |\Theta_{ij}| \geq \alpha$ and $\beta_{ij} = \beta_{ji}$.

Coordinatewise error. Fact D.7, proved below, gives the deterministic bound $(B_{i,S}^{-1})_{jj} \leq 3$ for every $j \in S$. Fix a nonempty S and suppress i, S. For any augmented coefficient error $\Delta_f = (\Delta_0, \Delta^\top)^\top$, where $\beta_{\mathrm{true}} = (\beta_{ij})_{j \in S}$ and $\Delta = \widehat{\beta} - \beta_{\mathrm{true}}$, expanding around the empirical predictor mean gives

$$\Delta_f^\top G \Delta_f = (\Delta_0 + \bar{x}_S^\top \Delta)^2 + \Delta^\top G^c \Delta. \tag{43}$$

Inversion of the design bound yields $((G^c)^{-1})_{jj} \leq 3/c$. Writing

$$\Delta_j = ((G^c)^{-1/2} e_j)^\top (G^c)^{1/2} \Delta,$$

and applying Cauchy–Schwarz gives

$$|\Delta_j|^2 \leq ((G^c)^{-1})_{jj}\, \Delta^\top G^c \Delta \leq \frac{3}{c}\, \Delta_f^\top G \Delta_f. \tag{44}$$

This inequality applies to any two coefficient vectors on S, including the selected and true fits on their union support.

Restriction to fits without an intercept. Let

$$G^0 = \frac{1}{q} \sum_{\ell=1}^{q} x_{S,\ell} x_{S,\ell}^\top, \qquad h = \sum_{\ell=1}^{q} x_{S,\ell} \xi_\ell. \tag{45}$$

These are the slope principal block of G and the slope portion of h_f, respectively. For nonempty S, on $\mathcal{E}_s$, $G^0 = G^c + \bar{x}_S \bar{x}_S^\top \succeq cB$, so every tested fit through the origin is unique. Completing the square in the variational formula for a positive quadratic form gives

$$\begin{aligned} h^\top (G^0)^{-1} h &= \sup_u \{2u^\top h - u^\top G^0 u\} \\ &\leq \sup_{a,u} \{2(a,u)^\top h_f - (a,u)^\top G (a,u)\} = h_f^\top G^{-1} h_f. \end{aligned} \tag{46}$$

The inequality restricts the second supremum to $a = 0$. For a zero-intercept coefficient error, take $\Delta_0 = 0$ in (43)–(44).

Observation time and the edge threshold. The prediction-error assumption and (44) bound every squared slope error by $KL_s(\delta/2)/q$, where $K = 3K_0/c$. Outside the union support both slopes are zero; this also covers an empty union. To make this bound at most $\alpha^2/16$, set

$$L_0 = \log \frac{C_0 d}{\alpha \delta}, \qquad T_0 = C_0 s_0 \alpha^{-2} L_0. \tag{47}$$

Here C_0 is a sufficiently large universal constant. Since $s \leq 2s_0$ and $s_0 \leq d$, taking $C_0 \geq 2C$ gives, at $T = T_0$,

$$\log(2CdT_0/\delta) \leq 2L_0 + \log L_0 \leq 3L_0, \qquad \log(2d/\delta) \leq L_0,$$
$$L_s(\delta/2) \leq (3s+1)L_0 \leq 7 s_0 L_0.$$

Since $q \geq T/4$ for $T \geq 8$, choosing $C_0 \geq 448K$ makes $KL_s(\delta/2)/q \leq \alpha^2/16$ at T_0; increasing C_0 also ensures $T_0 \geq 8$ and $T_0 \geq CL_s(\delta/2)$. The function $L_s(\delta/2)/T$ decreases for $T \geq T_0$, so the same conclusions hold for every $T \geq T_0$.

If all normalized slope errors are at most $\alpha/4$, a nonedge has $|\widehat{\beta}_{ij}\widehat{\beta}_{ji}| \leq \alpha^2/16$. On an edge both estimates have the sign of $\beta_{ij} = \beta_{ji}$ and magnitude at least $3\alpha/4$, so their product is at least $9\alpha^2/16$. Thus (41) returns exactly the true graph, also in the original coordinate scales.

The total number n of coordinate updates by T is $\mathrm{Poisson}(dT)$. The Poisson bound in Section A.2 gives

$$\Pr\{n \notin [dT/2, 3dT/2]\} \leq 2e^{-dT/12} \leq \delta/2 \tag{48}$$

when $dT \geq 12 \log(4/\delta)$, which follows from (47) after enlarging C_0. Intersecting this count event with $\mathcal{E}_s$ costs at most δ in total; no independence is needed. On the count event the event-list cap is never reached. □

D.2 ORDINARY LEAST SQUARES

Theorem D.4 (OLS graph recovery). *Let $d \geq 2$, $0 < \alpha < 1$, and $0 < \delta < 1/4$. Let $\Theta_{\mathrm{raw}} \succ 0$ be a precision matrix whose edges satisfy $|(\Theta_{\mathrm{raw}})_{ij}|/\sqrt{(\Theta_{\mathrm{raw}})_{ii}(\Theta_{\mathrm{raw}})_{jj}} \geq \alpha$. Observe its rate-one coordinate Glauber trajectory from an arbitrary initial state independent of future clocks and innovations. There is a universal constant C_0 such that, for every*

$$T \geq T_{\mathrm{OLS}} := C_0 d\alpha^{-2} \log \frac{C_0 d}{\alpha\delta},$$

Algorithm 2, using $q = 2\lfloor T/4 \rfloor$ retained updates per target, returns the exact graph and the update count is at most $3dT/2$ with probability at least $1-\delta$. The same guarantee holds when each regression includes an intercept. It is uniform over the precision matrix and initial state. The runtime is $O(d^3T)$ on every trajectory. At $T = T_{\mathrm{OLS}}$, the update bound is $O(d^2\alpha^{-2}\log(C_0 d/(\alpha\delta)))$.

Proof. By Lemma D.3, it suffices to analyze the normalized model. Take $s = d-1$ and work on $\mathcal{E}_{d-1}$. For every target i, the support $S = [d] \setminus \{i\}$ contains the true neighborhood, so (38) is the fitted linear model. Set $\beta_{\mathrm{true}} = (-\Theta_{ij})_{j\in S}$, so $y_i = X(0, \beta_{\mathrm{true}}^\top)^\top + \xi_i$. For a fit with an intercept, let $\Delta_f = (\widehat{b}_{i,0}, (\widehat{\beta}_i - \beta_{\mathrm{true}})^\top)^\top$. The residual vector is $\xi_i - X\Delta_f$. Differentiating its squared norm gives $X^\top(\xi_i - X\Delta_f) = 0$, hence

$$\begin{aligned} qG\Delta_f = h_f, \qquad \Delta_f = \frac{1}{q} G^{-1} h_f, \\ \Delta_f^\top G \Delta_f = \frac{1}{q^2} h_f^\top G^{-1} h_f \leq \frac{C L_{d-1}(\delta/2)}{q}. \end{aligned} \tag{49}$$

For the zero-intercept fit used by Algorithm 2, the slope normal equations instead give $qG^0\Delta = h$, where $\Delta = \widehat{\beta}_i - \beta_{\mathrm{true}}$. With $\Delta_f = (0, \Delta^\top)^\top$, equation (46) yields

$$\begin{aligned} \Delta_f^\top G \Delta_f = \Delta^\top G^0 \Delta = \frac{1}{q^2} h^\top (G^0)^{-1} h \\ \leq \frac{1}{q^2} h_f^\top G^{-1} h_f \leq \frac{C L_{d-1}(\delta/2)}{q}. \end{aligned}$$

Equation (44) therefore bounds every squared slope error by $KL_{d-1}(\delta/2)/q$ for either estimator, simultaneously over all ordered pairs. The common event also ensures that no availability or uniqueness check returns failure.

For every $T \geq T_{\mathrm{OLS}}$, Lemma D.3 with $s_0 = d$ makes every slope error at most $\alpha/4$ and proves graph recovery. The intersection with (48) has probability at least $1-\delta$ and uses at most $3dT/2 = O(d^2\alpha^{-2}\log(Cd/(\alpha\delta)))$ updates at the stated horizon. This proves the claimed time and update bounds, uniformly over each fixed initial state.

For an event-list input with n updates, maintain the current state in one capped pass and collect the first q regression observations per target. Forming a Gram and response cross-product costs $O(qd^2)$ per target, and solving each dense system costs $O(d^3)$. The total is $O(\min\{n, dT\} + d^3q + d^4) = O(d^3T)$, because the capped scan reads at most $\lceil 2dT \rceil + 1$ updates and the prescribed horizon gives $q \geq d$. This runtime bound holds even on trajectories that return failure. □

D.3 SPARSE ORDINARY LEAST SQUARES

Theorem D.5 (Sparse OLS graph recovery). *Let $d \geq 2$, $1 \leq k \leq d-1$, $0 < \alpha < 1$, and $0 < \delta < 1/4$. Let $\Theta_{\mathrm{raw}} \succ 0$ have graph degree at most k and $|(\Theta_{\mathrm{raw}})_{ij}|/\sqrt{(\Theta_{\mathrm{raw}})_{ii}(\Theta_{\mathrm{raw}})_{jj}} \geq \alpha$ on every edge. Observe its rate-one coordinate Glauber trajectory from an arbitrary initial state independent of future clocks and innovations. There is a universal constant C_0 such that, for every*

$$T \geq T_{\mathrm{subset}} := C_0 k \alpha^{-2} \log \frac{C_0 d}{\alpha \delta},$$

Algorithm 3, using $q = 2\lfloor T/4 \rfloor$ retained updates per target and searching over supports of size at most k, returns the exact graph and the update count is at most $3dT/2$ with probability at least $1 - \delta$. The same guarantee holds when each regression includes an intercept. It is uniform over the precision matrix and initial state. On every trajectory the runtime is

$$O\left(dTk^2 \sum_{p=0}^{k} \binom{d-1}{p} \right) = d^{O(k)} \operatorname{poly}(T).$$

At $T = T_{\mathrm{subset}}$, the update bound is $O(dk\alpha^{-2} \log(C_0 d/(\alpha\delta)))$.

Proof. By Lemma D.3, it suffices to analyze the normalized model. Take $s = \min\{2k, d-1\}$ and work on $\mathcal{E}_s$. Every tested support has size at most $k \leq s$, so all its fits are unique, with the empty-support convention given above. Fix a target i and write S_{selected} for the chosen support. Both S_{true} and S_{selected} have size at most k; their union

$$S = S_{\mathrm{true}} \cup S_{\mathrm{selected}}$$

has size at most s. This union is used in the proof; the search still tests only supports of size at most k.

Pad the true and selected slopes by zeros on S, and put $\Delta = \widehat{\beta} - \beta_{\mathrm{true}}$ on that union. Let Δ_0 be the fitted intercept, or zero when the intercept is omitted, and let $\Delta_f = (\Delta_0, \Delta^\top)^\top$. The true fit, with zero intercept, is feasible in either search. Its residual vector is exactly ξ_i by (38). Since the selected loss is no larger,

$$\|\xi_i - X_{i,S}\Delta_f\|^2 \leq \|\xi_i\|^2. \tag{50}$$

Expand the left side and cancel the common noise norm:

$$\|\xi_i\|^2 - 2\Delta_f^\top X_{i,S}^\top \xi_i + \Delta_f^\top X_{i,S}^\top X_{i,S} \Delta_f \leq \|\xi_i\|^2,$$
$$q\Delta_f^\top G \Delta_f \leq 2\Delta_f^\top h_f.$$

On $\mathcal{E}_s$, G is positive definite. Cauchy–Schwarz in its quadratic form gives

$$\Delta_f^\top G \Delta_f \leq \frac{2}{q}\Delta_f^\top h_f \leq \frac{2}{q}\sqrt{\Delta_f^\top G \Delta_f}\sqrt{h_f^\top G^{-1} h_f}.$$

If the left side is zero the next bound is immediate; otherwise divide by its square root and square. In either case,

$$\Delta_f^\top G \Delta_f \leq \frac{4}{q^2} h_f^\top G^{-1} h_f \leq \frac{C L_s(\delta/2)}{q}. \tag{51}$$

For nonempty S, equation (44) gives $|\Delta_j|^2 \leq K L_s(\delta/2)/q$ for each $j \in S$. Outside S both slopes are zero. If S is empty, all slope errors are zero and no slope matrix needs to be inverted; the augmented matrix is simply $G = 1$. Thus the same coordinatewise conclusion holds in every case. The bounds are simultaneous over all supports, so they apply to the data-selected union without conditioning on that union.

For every $T \geq T_{\mathrm{subset}}$, Lemma D.3 with $s_0 = k$ bounds all slope errors by $\alpha/4$, so the product rule recovers every edge and nonedge. Availability and uniqueness hold on $\mathcal{E}_s$, and intersecting with (48) gives probability at least $1 - \delta$. At the stated horizon the update count is at most $O(dk\alpha^{-2} \log(Cd/(\alpha\delta)))$. All these statements are uniform over each fixed initial state.

There are $\sum_{p=0}^{k} \binom{d-1}{p}$ candidate supports per target. For a support of size p, forming and solving its least-squares system uses $O(q(p+1)^2 + (p+1)^3)$ arithmetic operations, including an intercept if

desired. The prescribed horizon gives $q \geq k+1$, so this is $O(qk^2)$, also covering the empty candidate. Including the event-list scan with its update cap, the total cost is

$$O\left(\min\{n, dT\} + dqk^2 \sum_{p=0}^{k} \binom{d-1}{p}\right).$$

Since $\sum_{p=0}^{k} \binom{d-1}{p} \leq (k+1)d^k$ and $q \leq T/2$, this is $d^{O(k)}\,\mathrm{poly}(T)$ for every trajectory. □

D.4 Fresh noise in the coordinate dynamics

With the initial state z fixed, define the full-trajectory noise by

$$\eta^{\mathrm{full}}(t) = x(t) - \mathbb{E}[x(t) \mid \sigma].$$

It starts at zero and obeys the same linear update rule as the observed process, with precision Θ. To isolate noise within a gap between consecutive updates of target i, let t_0 be the first update time. Fix the state $x(t_0)$ and the clocks in the gap, and write $\widetilde{x}(t) = x(t) - z$ for the initial-state shift. Coordinate $\widetilde{x}_i$ is constant there, and, for $j \neq i$,

$$\widetilde{x}_j^{\mathrm{new}} = \widetilde{x}_j^{\mathrm{old}} - (\Theta_{-i,-i}\widetilde{x}_{-i}^{\mathrm{old}})_j - \Theta_{ji}\widetilde{x}_i - (\Theta z)_j + \xi.$$

Let $\mu(r)$ be the conditional mean of $\widetilde{x}_{-i}(t_0 + r)$ over the innovations introduced during the gap, and put $\eta(r) = \widetilde{x}_{-i}(t_0 + r) - \mu(r)$. Subtracting the mean update cancels both fixed terms, so $\eta(0) = 0$ and η follows the linear update rule with conditional precision $\Theta_{-i,-i}$. Subsystem coordinates retain their original labels below.

Let $\Omega \in \mathbb{R}^{r_\Omega \times r_\Omega}$ be positive definite with $\Omega_{jj} = 1$, and let η be its zero-start Glauber process as in Section A.2. Its normalized update rule is

$$\eta^{\mathrm{new}} = \eta^{\mathrm{old}} + e_j\big(-(\Omega\eta^{\mathrm{old}})_j + \xi\big). \tag{52}$$

The choices are $r_\Omega = d$, $\Omega = \Theta$ for the full trajectory, and $r_\Omega = d-1$, $\Omega = \Theta_{-i,-i}$ for a target gap. In this subsection, σ denotes the clock schedule of this generic process. For $t > 0$, define

$$B_t(\Omega) = \int_0^t e^{-2r\Omega}\,\mathrm{d}r, \qquad C_t = \mathbb{E}[\eta_t\eta_t^\top], \qquad Q_\sigma = \mathrm{Cov}(\eta_t \mid \sigma). \tag{53}$$

The terminal time in Q_σ will always be fixed when it is used.

Lemma D.6 (Fresh-noise covariance and fourth moment)**.** *For the process* (52)*,*

$$B_t(\Omega) \preceq C_t \preceq 2B_t(\Omega) \preceq 2tI_{r_\Omega}, \qquad B_t(\Omega)^{-1} \preceq 2\Omega + t^{-1}I_{r_\Omega} \quad (t > 0). \tag{54}$$

There is a universal constant C_4 such that, for every $u_0 \in \mathbb{R}^{r_\Omega}$ and $t \geq 1$,

$$\mathbb{E}(u_0^\top\eta_t)^4 \leq C_4\big(u_0^\top B_t(\Omega)u_0\big)^2. \tag{55}$$

Consequently, there are universal constants $a, p_0 > 0$ such that, for $u_0 \neq 0$ and $1 \leq t \leq 2$,

$$\mathbb{P}_\sigma\big\{u_0^\top Q_\sigma u_0 > a\,u_0^\top B_1(\Omega)u_0\big\} \geq p_0. \tag{56}$$

All expectations in the first two displays average over both clocks and innovations. The last probability averages over clocks alone.

Proof. For a coordinate j, write $A_j = I_{r_\Omega} - e_je_j^\top\Omega$ for the linear part of its update. A direct calculation, using $\Omega_{jj} = 1$, gives

$$A_j\Omega^{-1}A_j^\top + e_je_j^\top = \Omega^{-1}.$$

For a fixed clock schedule with successive update labels $j_1, \ldots, j_{N_0}$ up to time t, put $K_\sigma = A_{j_{N_0}} \cdots A_{j_1}$, with $K_\sigma = I_{r_\Omega}$ when $N_0 = 0$. The covariance starts at zero and each update maps a covariance Q to $A_jQA_j^\top + e_je_j^\top$. Induction using the preceding identity yields

$$Q_\sigma = \Omega^{-1} - K_\sigma\Omega^{-1}K_\sigma^\top, \qquad 0 \preceq Q_\sigma \preceq \Omega^{-1}. \tag{57}$$

Conditional on the clocks, η_t is centered Gaussian because it is a linear function of independent centered Gaussian innovations; without this conditioning, its law is generally a Gaussian mixture with clock-dependent covariance.

For an elapsed time r, let $r_j(r) = (\Omega\eta_{r-})_j$. Conditional on the clocks up to r, this residual is centered Gaussian. Its conditional variance is at most

$$e_j^\top \Omega\Omega^{-1}\Omega e_j = \Omega_{jj} = 1$$

by (57). It follows that

$$\mathbb{E}r_j(r)^2 \le 1, \qquad \mathbb{E}r_j(r)^4 \le 3, \qquad \mathbb{E}\big(-r_j(r)+\xi\big)^4 = \mathbb{E}r_j(r)^4 + 6\mathbb{E}r_j(r)^2 + 3 \le 12, \tag{58}$$

where ξ is a fresh independent standard Gaussian.

Use the joint clock-and-innovation filtration. The rate-one generator in (16), evaluated on each coordinate function using (52), has drift $-\Omega\eta$. The compensated process

$$\mathcal{M}_t = \eta_t + \int_0^t \Omega\eta_r \,\mathrm{d}r, \qquad \mathrm{d}\eta_r = -\Omega\eta_r \,\mathrm{d}r + \mathrm{d}\mathcal{M}_r \tag{59}$$

is a vector martingale. At such an update, $\Delta\mathcal{M}_r = e_j(-r_j(r)+\xi)$. Distinct coordinates do not update simultaneously almost surely, and the conditional second moment of this jump at coordinate j is $1 + r_j(r)^2$. Therefore its predictable quadratic variation is

$$\mathrm{d}\langle\mathcal{M}\rangle_r = \mathrm{Diag}\big(1 + r_1(r)^2, \dots, 1 + r_{r_\Omega}(r)^2\big)\,\mathrm{d}r. \tag{60}$$

This martingale statement uses the joint clock-and-innovation filtration; the full future clock schedule has not been conditioned on.

For fixed Ω, the covariance domination in (57) also gives finite coordinate fourth moments. Stopping therefore justifies these generator calculations for the unbounded coordinate and quadratic functions used here. For the second-moment equation, conditional on the pre-update state, the expected change in $\eta\eta^\top$ at coordinate j is

$$-e_j r_j \eta^\top - \eta r_j e_j^\top + e_j e_j^\top (1 + r_j^2).$$

Summing the rate-one contributions and taking expectations gives

$$C_t' = -\Omega C_t - C_t\Omega + I_{r_\Omega} + \mathrm{Diag}\big((\Omega C_t\Omega)_{11}, \dots, (\Omega C_t\Omega)_{r_\Omega r_\Omega}\big), \qquad C_0 = 0. \tag{61}$$

Each diagonal entry in the last term lies in $[0,1]$ by (58). Variation of constants expresses C_t as

$$\int_0^t e^{-(t-r)\Omega}\left[I_{r_\Omega} + \mathrm{Diag}\big((\Omega C_r\Omega)_{11}, \dots, (\Omega C_r\Omega)_{r_\Omega r_\Omega}\big)\right] e^{-(t-r)\Omega}\,\mathrm{d}r.$$

Bounding the middle matrix between I_{r_Ω} and $2I_{r_\Omega}$ proves the covariance bounds. The factor 2 in the exponent defining B_t comes from the two covariance propagation factors: $e^{-(t-r)\Omega} I_{r_\Omega} e^{-(t-r)\Omega} = e^{-2(t-r)\Omega}$. To prove the inverse bound, diagonalize Ω. For an eigenvalue $\lambda > 0$, the corresponding inverse eigenvalue of B_t is

$$\frac{2\lambda}{1 - e^{-2t\lambda}} \le 2\lambda + t^{-1}.$$

The scalar inequality follows from $e^x \ge 1 + x$, which implies $1 - e^{-x} \ge x/(1+x)$ for $x > 0$.

Fix the terminal time $t \ge 1$ and a direction u_0. Set

$$\begin{aligned} w_r &= e^{-\Omega(t-r)}u_0, \qquad N_\rho = \int_0^\rho w_r^\top \,\mathrm{d}\mathcal{M}_r \quad (0 \le \rho \le t), \\ b &= \int_0^t \|w_r\|_2^2 \,\mathrm{d}r = u_0^\top B_t(\Omega) u_0. \end{aligned} \tag{62}$$

Multiplying (59) by $e^{r\Omega}$ and integrating gives $N_t = u_0^\top \eta_t$. The predictable quadratic variation of this scalar martingale is

$$\langle N\rangle_t = \int_0^t \sum_{j=1}^{r_\Omega} w_{r,j}^2 (1 + r_j(r)^2)\,\mathrm{d}r.$$

If $u_0 = 0$ the desired bound is immediate; assume $b > 0$. Weighted Cauchy–Schwarz and (58) give

$$\mathbb{E}\langle N\rangle_t^2 \le b \int_0^t \sum_{j=1}^{r_\Omega} w_{r,j}^2 \mathbb{E}(1 + r_j(r)^2)^2 \,\mathrm{d}r$$
$$\le 6b \int_0^t \sum_{j=1}^{r_\Omega} w_{r,j}^2 \,\mathrm{d}r = 6b^2. \tag{63}$$

Here the factor 6 is $1 + 2\mathbb{E}r_j(r)^2 + \mathbb{E}r_j(r)^4 \le 6$; no independence between residuals at different coordinates or times is used.

For the jump contribution, rate-one Poisson compensation and the fresh innovation moment in (58) give

$$\mathbb{E} \sum_{0<r\le t} |\Delta N_r|^4 \le 12 \int_0^t \sum_{j=1}^{r_\Omega} w_{r,j}^4 \,\mathrm{d}r. \tag{64}$$

To bound the deterministic integral, Cauchy–Schwarz in the $B_t(\Omega)$ metric gives, for every r, j,

$$w_{r,j}^2 \le b\, e_j^\top e^{-\Omega(t-r)} B_t(\Omega)^{-1} e^{-\Omega(t-r)} e_j$$
$$\le b\, (B_t(\Omega)^{-1})_{jj} \le (2 + t^{-1})b. \tag{65}$$

For the middle inequality, the exponential and $B_t(\Omega)^{-1}$ are commuting positive functions of Ω, and the exponential has operator norm at most one. For the last inequality, use (54) and $\Omega_{jj} = 1$. Consequently,

$$\mathbb{E} \sum_{0<r\le t} |\Delta N_r|^4 \le 12(2 + t^{-1})b \int_0^t \sum_j w_{r,j}^2 \,\mathrm{d}r \le 36b^2 \qquad (t \ge 1). \tag{66}$$

Fact A.10 applies because N is a compensated Poisson integral with continuous predictable quadratic variation. By (63) and (66),

$$\mathbb{E}|N_t|^4 \lesssim \mathbb{E}\langle N\rangle_t^2 + \mathbb{E} \sum_{0<r\le t} |\Delta N_r|^4 \lesssim b^2.$$

Since $N_t = u_0^\top \eta_t$, this proves (55).

Conditional Gaussianity gives

$$\mathbb{E}_\sigma[u_0^\top Q_\sigma u_0] = u_0^\top C_t u_0 \ge b, \qquad \mathbb{E}_\sigma[(u_0^\top Q_\sigma u_0)^2] = \frac{1}{3}\mathbb{E}(u_0^\top \eta_t)^4 \le \frac{C_4}{3} b^2. \tag{67}$$

Paley–Zygmund applied to $V = u_0^\top Q_\sigma u_0$ gives

$$\mathbb{P}_\sigma\{V > b/2\} \ge \mathbb{P}_\sigma\{V > \mathbb{E}_\sigma V/2\} \ge \frac{(\mathbb{E}_\sigma V)^2}{4\mathbb{E}_\sigma V^2} \ge \frac{3}{4C_4}.$$

For $1 \le t \le 2$, spectral calculus gives $B_1(\Omega) \preceq B_t(\Omega) \preceq 2B_1(\Omega)$. In particular $b \ge u_0^\top B_1(\Omega) u_0$, which proves (56) with $a = 1/2$ and a universal $p_0 > 0$. □

Fact D.7 (Inverse bounds after restriction)**.** *For the matrices in* (34) *and every nonempty predictor set* $S \subseteq [d] \setminus \{i\}$*,*

$$(B_{i,S}^{-1})_{jj} \le 3 \quad (j \in S), \qquad \|B_{i,S}^{-1}\|_{\mathrm{op}} \le \mathrm{tr}(B_{i,S}^{-1}) \le 3|S|. \tag{68}$$

Proof. Lemma D.6 gives $B_i^{-1} \preceq 2\Theta_{-i,-i} + I$, whose diagonal entries are 3. To pass to a principal restriction, write $B_i = \left(\begin{smallmatrix} B_{i,S} & E \\ E^\top & H \end{smallmatrix}\right)$ when the complementary block is nonempty. The corresponding block of the inverse is

$$(B_i^{-1})_{SS} = (B_{i,S} - EH^{-1}E^\top)^{-1} \succeq B_{i,S}^{-1}.$$

If the complement is empty the same comparison is equality. Thus each diagonal entry of $B_{i,S}^{-1}$ is at most 3. A positive-definite matrix has operator norm at most its trace, proving the remaining bounds. In particular, $B_{i,S} \succeq (3|S|)^{-1} I_{|S|}$. □

D.5 Clock intervals and uniform excitation

For each target i, consider its first q updates on the indefinitely continued trajectory and the $q/2$ disjoint intervals

$$I_{i,\ell} = (\tau_{i,2\ell-1}, \tau_{i,2\ell}), \qquad 1 \le \ell \le q/2. \tag{69}$$

Let $Q_{i,\ell}$ be the clock-conditioned covariance of the fresh noise in $I_{i,\ell}$. The gap dynamics in Section D.4 use precision $\Theta_{-i,-i}$, and $Q_{i,\ell}$ depends only on the interval's clock schedule, not its starting state.

By the Poisson-clock properties in Section A.2, the interval lengths for a fixed i are iid $\mathrm{Exp}(1)$, independent of the other coordinate clocks. Conditional on the entire schedule of coordinate i, the other clocks have independent increments on these disjoint intervals; their relative schedules depend in law only on the respective interval lengths. Integrating over the independent lengths proves that the matrices $Q_{i,1}, \dots, Q_{i,q/2}$ are iid. This argument is made before intersecting with the event that the first q updates occur by time T.

Fix a nonempty support S and write $B = B_{i,S}$ as in (34). Embed $u \in \mathbb{R}^{|S|}$ into the full subsystem by setting its coordinates outside S to zero. The fixed-time clock estimate (56), together with $\Pr(|I_{i,\ell}| \in [1,2]) = e^{-1} - e^{-2}$, gives universal constants $a_*, \pi_* > 0$ such that

$$\Pr_{\sigma}\{u^\top (Q_{i,\ell})_{SS} u > a_* u^\top B u\} \ge \pi_* \qquad (u \ne 0). \tag{70}$$

Here and below, $\Pr_\sigma$ averages over clocks. We need a positive fraction of successful intervals for every direction at once, since the directions used to cover the design will depend on the clocks.

Lemma D.8 (Uniform counts for quadratic inequalities)**.** *Let $p, N \ge 1$ be integers, let B be positive definite on $\mathbb{R}^p$, and let $Q_1, \dots, Q_N$ be iid symmetric matrices. Suppose that, for fixed $a_*, \pi_* > 0$,*

$$\Pr\{u^\top (Q_1 - a_* B) u > 0\} \ge \pi_* \qquad \textit{for every } u \ne 0.$$

Then, with constants depending only on π_*,

$$\Pr\big\{\exists u \ne 0 : \quad \#\{\ell : u^\top (Q_\ell - a_* B) u > 0\} < \pi_* N/2\big\} \le (CN)^{Cp} e^{-cN}. \tag{71}$$

Proof. Write $f_u(Q) = \mathbf{1}\{u^\top (Q - a_* B) u > 0\}$. On any N_0 fixed matrices, these labels are determined by the signs of N_0 quadratic polynomials in the p coordinates of u. Warren's sign-pattern bound, including zero signs as in Alon & Scheinerman (1988, Theorem 2), gives at most $(16eN_0/p)^p$ patterns when $N_0 \ge p$; see also Warren (1968). For $N_0 < p$ there are at most 2^{N_0} binary patterns. Consequently the growth function satisfies

$$\mathcal{S}(N_0) := \sup_{A_1, \dots, A_{N_0}} \#\{(f_u(A_1), \dots, f_u(A_{N_0})) : u \ne 0\} \le (CN_0)^{Cp}. \tag{72}$$

Put $P_N f = N^{-1} \sum_{\ell=1}^N f(Q_\ell)$ and $Pf = \mathbb{E} f(Q_1)$. The Vapnik–Chervonenkis deviation inequality (Devroye et al., 1996, Theorem 12.5) gives

$$\Pr\left\{\sup_{u \ne 0} |P_N f_u - P f_u| > \pi_*/2\right\} \le 8 \mathcal{S}(N) e^{-N\pi_*^2/128} \le (CN)^{Cp} e^{-cN}.$$

Since $Pf_u \ge \pi_*$ for every nonzero u, any count below $\pi_* N/2$ belongs to this deviation event, proving the claim. □

Apply Lemma D.8 with $N = q/2$ to each target and nonempty support. Define $\rho = \pi_*/4$ and the event $\mathcal{E}_{\mathrm{clock}}$ by

$$\#\left\{\ell : u^\top (Q_{i,\ell})_{SS} u > a_* u^\top B_{i,S} u\right\} \ge \rho q \quad \text{for all } i,\ 1 \le |S| \le s,\ u \ne 0. \tag{73}$$

Let K_s count all target/support pairs, including the empty supports:

$$K_s = d \sum_{p=0}^{s} \binom{d-1}{p} \le d(Cd)^s. \tag{74}$$

The event is determined only by the clocks, and

$$\Pr(\mathcal{E}_{\mathrm{clock}}^c) \le d \sum_{p=1}^{s} \binom{d-1}{p} (Cq)^{Cp} e^{-cq} \le K_s (Cq)^{Cs} e^{-cq}. \tag{75}$$

The successful intervals can vary with the direction. The union bound does not require independence between different targets or supports.

D.6 A SCALAR LOWER TAIL FOR THE OBSERVED DESIGN

Fix a complete clock schedule $\sigma \in \mathcal{E}_{\mathrm{clock}}$, a target i, a support S, and an augmented direction $v = (a, u)$ that is fixed conditional on these clocks. Let $\mathcal{H}_{i,\ell}$ contain all innovations up to and including the odd update $\tau_{i,2\ell-1}$. Conditional on σ and $\mathcal{H}_{i,\ell}$, the gap-start state is fixed. The fresh randomness in $x_{S,2\ell}$ then comes from the innovations inside $I_{i,\ell}$. Therefore

$$a + u^\top x_{S,2\ell} \,\big|\, (\sigma, \mathcal{H}_{i,\ell}) \sim N(\mu_\ell(v), \nu_\ell(u)), \qquad \nu_\ell(u) = u^\top (Q_{i,\ell})_{SS} u. \tag{76}$$

The mean $\mu_\ell(v)$ can depend on the initial state and all earlier innovations. The variance is fixed by the clocks.

For $W \sim N(\mu, \nu)$ and $\lambda > 0$, completing the square in the Gaussian density yields

$$\mathbb{E}e^{-\lambda W^2} = \frac{1}{\sqrt{1+2\lambda\nu}} \exp\left(-\frac{\lambda\mu^2}{1+2\lambda\nu}\right) \le (1+2\lambda\nu)^{-1/2}. \tag{77}$$

The formula includes $\nu = 0$ by continuity, and the bound is uniform in μ.

Suppose $u \neq 0$, and put $b = u^\top B u > 0$ and $\lambda = b^{-1}$. For each successful interval in (73), (76)–(77) bound its conditional Laplace factor by $(1+2a_*)^{-1/2}$; the factor for every other interval is at most one. The observations from earlier intervals are measurable with respect to $\mathcal{H}_{i,\ell}$. Iterating conditional expectation from the last interval to the first therefore gives

$$\begin{aligned}\mathbb{E}_\xi\left[\exp\left\{-\frac{1}{b}\sum_{\ell=1}^{q/2}(a + u^\top x_{S,2\ell})^2\right\}\,\middle|\,\sigma\right] &\le \prod_{\ell=1}^{q/2}(1+2\nu_\ell(u)/b)^{-1/2} \\ &\le (1+2a_*)^{-\rho q/2} = e^{-\kappa q},\end{aligned} \tag{78}$$

where $\kappa = (\rho/2)\log(1+2a_*) > 0$ and $\mathbb{E}_\xi$ averages over the innovations only. The tower property also integrates out the intervening updates between the displayed intervals. The observed values themselves need not be independent.

Choose $c_0 = c_1 = \kappa/2$. Markov's inequality applied to (78) gives

$$\Pr_\xi\left\{\frac{1}{q}\sum_{\ell=1}^{q/2}(a + u^\top x_{S,2\ell})^2 < c_0 b \,\middle|\, \sigma\right\} \le e^{-c_1 q}.$$

The normalized Gram matrix contains these terms among its q nonnegative summands:

$$v^\top G v = \frac{1}{q}\sum_{\ell=1}^{q}(a + u^\top x_{S,\ell})^2 \ge \frac{1}{q}\sum_{\ell=1}^{q/2}(a + u^\top x_{S,2\ell})^2.$$

Consequently, with $J = \mathrm{diag}(0, B)$,

$$\Pr_\xi\{v^\top G v < c_0 v^\top J v \mid \sigma\} \le e^{-c_1 q}. \tag{79}$$

If $u = 0$, then $v^\top J v = 0$ and the strict failure event is empty, so the same bound holds; this also covers the empty support. Thus (79) holds for any direction fixed after conditioning on the clocks, including the centers of a clock-dependent net.

D.7 A NOISE ENVELOPE AND THE DESIGN COVERING ARGUMENT

One noise event for all supports. Recall the full-trajectory noise η^{full} from Section D.4. Its value at t uses only clocks and innovations up to t, so its left limits are predictable. The covariance estimate in Lemma D.6 gives $\mathbb{E}[\eta^{\mathrm{full}}(t)\eta^{\mathrm{full}}(t)^\top] \preceq 2tI$. Define its total energy sampled just before updates by

$$A_T = \sum_{i=1}^{d}\sum_{\ell:\,\tau_{i,\ell}\le T} \|\eta_{-i}^{\mathrm{full}}(\tau_{i,\ell}-)\|^2.$$

Here the inner sum includes all updates of i by T. Compensating the rate-one Poisson counts gives

$$\mathbb{E}A_T = \sum_{i=1}^{d} \int_0^T \mathbb{E}\|\eta_{-i}^{\mathrm{full}}(t)\|^2 \,\mathrm{d}t \le d(d-1)T^2 \le d^2T^2. \tag{80}$$

Set

$$\mathcal{E}_U = \{A_T \le 5d^2T^2/\delta\}, \qquad \mathcal{E}_{\mathrm{av}} = \{\tau_{i,q} \le T \text{ for all } i\}.$$

Markov's inequality gives $\Pr(\mathcal{E}_U^c) \le \delta/5$. On $\mathcal{E}_{\mathrm{av}}$, all rows of each design-noise matrix Z occur in the sum defining A_T. For $v = (a, u)$,

$$\frac{1}{q}\|Zv\|^2 \le \frac{A_T}{q}\|u\|^2 \le \frac{3sA_T}{q} u^\top Bu.$$

The last inequality uses the deterministic estimate $\|B^{-1}\|_{\mathrm{op}} \le 3|S| \le 3s$ proved above. Since $q \ge T/4$, choosing $C_R \ge 60$ in (35) yields, simultaneously for all targets and supports,

$$\frac{1}{q} Z^\top Z \preceq RJ \quad \text{on } \mathcal{E}_{\mathrm{av}} \cap \mathcal{E}_U. \tag{81}$$

Let $\mathcal{E}_Z$ denote this simultaneous conclusion; the probability bound above is unconditional.

Comparison at each covering direction. For any nonzero $v = (a, u)$, $v^\top Fv > 0$: if $u \ne 0$, the J term is positive, and if $u = 0$, the intercept column of $\bar{X}$ gives $v^\top Dv = a^2$. Thus F is positive definite for every clock schedule. On $\mathcal{E}_Z$, the identities $\bar{X} = X - Z$ and $X = \bar{X} + Z$ imply

$$D \preceq 2G + 2RJ, \qquad F \preceq 3J + \frac{2}{R}G, \qquad G \preceq 2D + 2RJ = 2RF. \tag{82}$$

For example, the first inequality follows by applying $\|b - c\|^2 \le 2\|b\|^2 + 2\|c\|^2$ to Xv and Zv for each v.

On a good clock schedule, the scalar lower tail established above gives $v^\top Gv \ge c_0 v^\top Jv$ for each fixed clock-measurable direction v, except with probability $e^{-c_1 q}$. At any such successful direction, (82) gives

$$v^\top Fv \le \left(\frac{3}{c_0} + \frac{2}{R}\right) v^\top Gv \le \left(\frac{3}{c_0} + 2\right) v^\top Gv. \tag{83}$$

This implication uses only the scalar lower bound at that direction; it does not assume a spectral lower bound on all directions. Set $\kappa_0 = (3/c_0 + 2)^{-1}$.

Extension from a net. For each i, S and clock schedule, take a net $\mathcal{N}_{i,S}(\sigma)$ of the F-unit sphere with radius

$$\varepsilon_F = \frac{\sqrt{\kappa_0}}{2\sqrt{2R}} \quad \text{in the norm } \|v\|_F = (v^\top Fv)^{1/2}.$$

It can be obtained by applying $F^{-1/2}$ to a fixed Euclidean sphere net. A maximal separated set and the disjoint-ball volume comparison give

$$|\mathcal{N}_{i,S}(\sigma)| \le (1 + 2/\varepsilon_F)^{p+1} \le (C\sqrt{R})^{p+1}. \tag{84}$$

The chosen net is measurable in the clocks and does not depend on innovations.

Let $\mathcal{E}_{\mathrm{cover}}$ be the event that $v^\top Gv \ge c_0 v^\top Jv$ at every chosen center for every target and support. Recall the target/support count K_s in (74). Conditional on each schedule in $\mathcal{E}_{\mathrm{clock}}$, union the scalar failures over these centers and pairs, then integrate over the clocks. This gives

$$\begin{aligned}\Pr(\mathcal{E}_{\mathrm{clock}} \cap \mathcal{E}_{\mathrm{cover}}^c) &\le d \sum_{p=0}^{s} \binom{d-1}{p} (C\sqrt{R})^{p+1} e^{-c_1 q} \\ &\le K_s (C\sqrt{R})^{s+1} e^{-c_1 q}.\end{aligned} \tag{85}$$

This probability calculation precedes intersection with $\mathcal{E}_Z$; it does not condition on the design-noise bound.

On $\mathcal{E}_Z \cap \mathcal{E}_{\text{clock}} \cap \mathcal{E}_{\text{cover}}$, every center has G-norm at least $\sqrt{\kappa_0}$ by (83). For an arbitrary F-unit vector v, choose a center v_* within ε_F. The upper bound in (82) gives

$$\sqrt{v^\top G v} \geq \sqrt{v_*^\top G v_*} - \sqrt{(v-v_*)^\top G (v-v_*)} \geq \sqrt{\kappa_0} - \sqrt{2R}\,\varepsilon_F = \frac{\sqrt{\kappa_0}}{2}.$$

By homogeneity, with $c_F = \kappa_0/4$,

$$c_F F \preceq G \preceq 2RF \quad \text{simultaneously for all } i, S. \tag{86}$$

Finally, completing the square in the intercept gives

$$(a,u)^\top G(a,u) = (a + u^\top \bar{x}_S)^2 + u^\top G^c u, \qquad u^\top G^c u = \inf_{a \in \mathbb{R}} (a,u)^\top G(a,u). \tag{87}$$

Since $F \succeq J$, (86) implies $G \succeq c_F J$ and therefore $G^c \succeq c_F B$. Equation (87) also proves the converse equivalence between $G \succeq cJ$ and $G^c \succeq cB$.

D.8 The predictable response score

When X has full column rank,

$$h_f^\top G^{-1} h_f = q\,\xi^\top X (X^\top X)^{-1} X^\top \xi \leq q\|\xi\|^2.$$

The last quantity is of order q^2. Obtaining the smaller order qL_s requires the freshness of each response innovation. Conditioning on the entire realized design would lose this property, because later predictors depend on earlier response innovations. We condition on clocks and reveal innovations chronologically instead.

Apply the Gaussian score inequality conditional on clocks. Fix a complete clock schedule and a positive-definite Γ that is fixed under this conditioning. The chronological order of updates is then fixed. Take $\mathcal{F}_{\ell-1}$ to contain all innovations strictly before $\tau_{i,\ell}$, including intervening nontarget updates, and $\mathcal{F}_q$ through the last response. The ℓth row of X is $\mathcal{F}_{\ell-1}$-measurable and ξ_ℓ is fresh conditional on that history. Thus Fact A.11 applies with

$$\psi_\ell = (1, x_{S,\ell}^\top)^\top, \qquad V = X^\top X = qG, \qquad g = h_f, \qquad \Gamma_0 = q\Gamma.$$

Substituting these variables into (19) gives, with conditional failure probability at most ε,

$$\frac{1}{q} h_f^\top (G+\Gamma)^{-1} h_f \leq \log\det(I + \Gamma^{-1/2} G \Gamma^{-1/2}) + 2\log(1/\varepsilon). \tag{88}$$

The factor $1/q$ comes from $(qG + q\Gamma)^{-1} = q^{-1}(G+\Gamma)^{-1}$.

Removing the regularization. Set $\Gamma = \gamma F$ for a fixed $0 < \gamma \leq c_F$. It is positive definite and clock-measurable for every schedule, whether or not the design comparison succeeds. On (86),

$$\Gamma \preceq G \preceq (2R/\gamma)\Gamma. \tag{89}$$

The lower comparison gives $G + \Gamma \preceq 2G$. Inversion reverses the order of positive-definite matrices, so

$$(G+\Gamma)^{-1} \succeq \tfrac{1}{2} G^{-1}. \tag{90}$$

For the determinant, let $\lambda_1, \ldots, \lambda_{p+1}$ be the eigenvalues of $\Gamma^{-1/2} G \Gamma^{-1/2}$. Congruence in (89) bounds each by $2R/\gamma$, and hence

$$\log\det(I + \Gamma^{-1/2} G \Gamma^{-1/2}) = \sum_{j=1}^{p+1} \log(1+\lambda_j) \leq (p+1)\log(1 + 2R/\gamma). \tag{91}$$

Combining (88)–(91) gives

$$h_f^\top G^{-1} h_f \leq 2q\big[(p+1)\log(1+2R/\gamma) + 2\log(1/\varepsilon)\big]. \tag{92}$$

Choose $\varepsilon = \delta/(5K_s)$ and union the mixture failures over every target and support, including $S = \varnothing$. The resulting event $\mathcal{E}_{\text{mix}}$ has $\Pr(\mathcal{E}_{\text{mix}}^c) \leq \delta/5$, by integration over the clocks. Only then do we intersect with the design comparison. Since $\log K_s \leq \log d + s\log(Cd)$ and $\log(CR) \leq C\log(CdT/\delta)$, the right side of (92) is at most CqL_s.

D.9 Completion of the regression bounds

By Fact A.3, the update count of each target by T is Poisson with mean T. Since $q \le T/2$, its elementary lower-tail bound gives $\Pr(\mathcal{E}_{\mathrm{av}}^c) \le de^{-T/8}$. All failures used in the proof are bounded by

$$\begin{aligned}&\Pr(\mathcal{E}_{\mathrm{av}}^c) + \Pr(\mathcal{E}_U^c) + \Pr(\mathcal{E}_{\mathrm{clock}}^c) + \Pr(\mathcal{E}_{\mathrm{clock}} \cap \mathcal{E}_{\mathrm{cover}}^c) + \Pr(\mathcal{E}_{\mathrm{mix}}^c) \\ &\quad \le de^{-T/8} + \frac{\delta}{5} + K_s(Cq)^{Cs}e^{-cq} + K_s(C\sqrt{R})^{s+1}e^{-c'q} + \frac{\delta}{5}.\end{aligned} \tag{93}$$

Because $q \ge T/4$, $s \le d-1$, and $R = \max\{1, C_R s d^2 T/\delta\}$, a sufficiently large universal constant in $T \ge CL_s$ makes each of the three exponential terms at most $\delta/5$. Thus the common event

$$\mathcal{E}_{\mathrm{av}} \cap \mathcal{E}_U \cap \mathcal{E}_{\mathrm{clock}} \cap \mathcal{E}_{\mathrm{cover}} \cap \mathcal{E}_{\mathrm{mix}}$$

has probability at least $1-\delta$. On it, (81), (86), and (92) prove both lemmas. No independence between these events, different targets, or different supports is required.

D.10 Centered regression identities and initial-state translation

Work on the common event just proved. With h as in (45), decompose the response score as

$$h_f = \begin{pmatrix} h_0 \\ h \end{pmatrix}, \qquad h_0 = \sum_{\ell=1}^{q} \xi_\ell, \qquad h' = \frac{h - \bar{x}_S h_0}{q}. \tag{94}$$

The centered score h' uses the empirical mean $\bar{x}_S$, so its summands need not be martingale differences.

For a fit with an intercept on a nonempty S containing the true neighborhood of i, (13) gives a correctly specified response model, with true slopes $\beta_{ij} = -\Theta_{ij}$ and true intercept zero. Write $\beta_{\mathrm{true}} = (\beta_{ij})_{j\in S}$ and $\Delta = \widehat{\beta} - \beta_{\mathrm{true}}$ for the slope error; let Δ_f also include the intercept error. The normal-equation formula $\Delta_f = q^{-1}G^{-1}h_f$ from (49) holds on any such support. Eliminating the intercept gives

$$\Delta = (G^c)^{-1}h', \qquad \frac{1}{q^2}h_f^\top G^{-1}h_f = \frac{h_0^2}{q^2} + (h')^\top (G^c)^{-1}h'. \tag{95}$$

To see these identities, the intercept equation is $\Delta_0 + \bar{x}_S^\top \Delta = h_0/q$; substituting it in the slope equation gives $G^c\Delta = h'$. The identity $\Delta_f^\top G \Delta_f = (\Delta_0 + \bar{x}_S^\top \Delta)^2 + \Delta^\top G^c \Delta$ then gives the displayed quadratic decomposition. In particular, the slope block of G^{-1} is $(G^c)^{-1}$. The initial-state translation maps the predictors to $\widetilde{x}_{S,\ell} = x_{S,\ell} - z_S$. With the invertible matrix $A_z = \left[\begin{smallmatrix} 1 & 0 \\ -z_S & I \end{smallmatrix}\right]$, the full design transforms as $X \mapsto XA_z^\top$ and the score as $h_f \mapsto A_z h_f$. Thus $G \mapsto A_z G A_z^\top$, preserving $h_f^\top G^{-1} h_f$. Also G^c and h' are unchanged. If the responses are shifted to $y_{i,\ell} - z_i$ as well, their true intercept becomes $-(\Theta z)_i$, while their slopes stay the same.

E COMPARISON WITH PRIOR TRAJECTORY BOUNDS

Write $z = x(0)$ for the initial state. For $1 \le s \le d$, define

$$D_U = \mathrm{diag}(\Theta_{jj} : j \in U), \qquad \rho_s(\Theta) = \max_{\substack{U \subseteq [d] \\ |U| = s}} \lambda_{\max}\Big(D_U^{1/2} (\Theta^{-1})_{UU} D_U^{1/2}\Big), \qquad E_0 = z^\top \Theta z.$$

This specializes to Tirukkonda & Dasarathy (2026, Definition 4) at $s = 2k$. Their Theorem 1 and Remark 3 give LS-DuNeS the time bound

$$T_{\mathrm{LS}} = O\left(k^2 \alpha^{-2} \log \frac{Cdk\rho_{2k}(\Theta)(1 + E_0)}{\alpha\delta}\right),$$

and their Theorem 2 gives $T_{\mathrm{AU}} = O(k^4 \alpha^{-2} \log(Cd/(\alpha\delta)))$. These sparse-search bounds assume $d \ge 2k + 1$. Their update budgets convert to rate-one-per-coordinate time by Poisson concentration, splitting the failure probability between recovery and collecting the required updates.

Dense adaptations. Fix $S_i = [d] \setminus \{i\}$, which contains every true predictor. For LS, the fixed-support arguments of Tirukkonda & Dasarathy (2026, Propositions 5 and 12 and their proofs) use tested dimension $d - 1$, active-clock size d, and $d(d - 1)$ ordered-pair tests. Each predictor refreshes immediately before a target update with probability $1/d$, giving a normalized Gram lower bound of order $1/d$. Writing q for retained rows per target, their potential bound gives an upper spectral envelope

$$\Lambda_+(q) \lesssim \rho_d(\Theta)\big(E_0 + dq + d\log(d/\delta)\big).$$

The covering and separation conditions therefore become

$$q \ge \frac{Cd}{\alpha^2}\left\{d \log\big(Cd\Lambda_+(q)\big) + \log\frac{d^2}{\delta}\right\}.$$

Solving this logarithmic condition gives the sufficient horizon

$$T_{\mathrm{LS,dense}} = O\left(d^2 \alpha^{-2} \log \frac{Cd^2 \rho_d(\Theta)(1 + E_0)}{\alpha\delta}\right).$$

Since $q \ge C \log(d/\delta)$, the Poisson lower-tail bound supplies all q rows per target by $T = 2q$ with failure probability at most $de^{-T/8}$. The d full regressions and $d(d - 1)$ coefficient tests take polynomial time.

For AU, Tirukkonda & Dasarathy (2026, Proposition 14 and Lemma 19) apply to each fixed triple with $S_i = [d] \setminus \{i\}$. Allocate failure probability $\delta/(4d(d - 1))$ to each test and its block count, reserving $\delta/2$ for Poisson conversion. The required number of blocks is $b = O(\alpha^{-2} \log(Cd/(\alpha\delta)))$ per ordered pair. The active clock now includes all d coordinates, and a disjoint four-update block has labels (i, i, j, i) with probability d^{-4}. Thus $O(d^4 b)$ total updates suffice simultaneously, or

$$T_{\mathrm{AU,dense}} = O\left(d^3 \alpha^{-2} \log \frac{Cd}{\alpha\delta}\right).$$

Computing the $d(d - 1)$ fixed-pair statistics takes polynomial time.

Earlier local and mixing bounds. The corrected results of Tirukkonda et al. (2026, Theorems 1–2 and Remark 1) give polynomial time dependences $k^3(\beta_{\min}\sigma_{\min})^{-5}$ and $k\alpha^{-2}/(1 - r)$, respectively. These expressions suppress logarithmic factors and constants that may depend on the initial state and model parameters. Both retain their Assumptions A1–A3 on edge coefficients, variances, and contraction; the local result also requires its stated horizon $T \le T_{\max}$.