Supplementary Material

Learning Gaussian Graphical Models via Multiplicative Weights

(Anamay Chaturvedi and Jonathan Scarlett, AISTATS 2020)

All citations below are to the reference list in the main document.

A Comparison of Runtimes

Recall that p denotes the number of nodes, d denotes the maximal degree, κ denotes the minimum normalized edge strength, and m denotes the number of samples. The runtimes of some existing algorithms in the literature for Gaussian graphical model selection (see Section 1.1 for an overview) are outlined as follows:

- The only algorithms with assumption-free sample complexity bounds depending only on (p, d, κ) have a high runtime of $p^{O(d)}$, namely, $O(p^{2d+1})$ in [Misra et al., 2017], and $O(p^{d+1})$ in [Kelner et al., 2019, Thm. 11].
- A greedy method in [Kelner et al., 2019, Thm. 7] has runtime $O\big(\big(d \log \frac{1}{\kappa}\big)^3 mp^2\big)$. The sample complexity for this algorithm is $O\big(\frac{d}{\kappa^2} \cdot \log \frac{1}{\kappa} \cdot \log n\big)$, but this result is restricted to attractive graphical models.
- To our knowledge, ℓ_1-based methods [Cai et al., 2011, 2016, d'Aspremont et al., 2008, Meinshausen et al., 2006, Ravikumar et al., 2011, Wang et al., 2016, Yuan and Lin, 2007] such as Graphical Lasso and CLIME do not have precise time complexities stated, perhaps because this depends strongly on the optimization algorithm used. We expect that a general-purpose solver would incur $O(p^3)$ time, and we note that [Kelner et al., 2019, Table 2] indeed suggests that these approaches are slower.
- In practice, we expect BigQUIC [Hsieh et al., 2013] to be one of the most competitive algorithms in terms of runtime, but no sample complexity bounds were given for this algorithm.
- Under the local separation condition and a walk-summability assumption, the algorithm of [Anandkumar et al., 2012] yields a runtime of $O(p^{2+\eta})$, where $\eta > 0$ is an integer specifying the local separation condition.

Hence, we see that our runtime of $O(mp^2)$ is competitive among the existing works – it is faster than other algorithms for which sample complexity bounds have been established.

B Proof of Lemma 2 (Properties of Multivariate Gaussians)

We restate the lemma for ease of reference.

Lemma 2. *Given a zero-mean multivariate Gaussian $X = (X_1, \dots, X_p)$ with inverse covariance matrix $\Theta = [\theta_{ij}]$, and given T independent samples $(X^1, \dots, X^T)$ with the same distribution as X, we have the following:*

1. *For any $i \in [p]$, we have $X_i = \eta_i + \sum_{j \neq i} \big(-\frac{\theta_{ij}}{\theta_{ii}}\big) X_j$, where η_i is a Gaussian random variable with variance $\frac{1}{\theta_{ii}}$, independent of all X_j for $j \neq i$.*
2. *$\mathbb{E}[X_i | X_{\bar{i}}] = \sum_{j \neq i} \big(\frac{-\theta_{ij}}{\theta_{ii}}\big) X_j = w^i \cdot X_{\bar{i}}$, where $w^i = \big(\frac{-\theta_{ij}}{\theta_{ii}}\big)_{j \neq i} \in \mathbb{R}^n$ (with $n = p - 1$).*
3. *Let λ and $\nu_{\max}$ be defined as in (4) and (6), set $B := \sqrt{2 \log \frac{2pT}{\delta}}$, and define $(\tilde{x}^t, \tilde{y}^t) := \frac{1}{B\sqrt{\nu_{\max}(\lambda+1)}}(x^t, y^t)$, where $(x^t, y^t) = (X^t_{\bar{i}}, X^t_i)$ for an arbitrary fixed coordinate i. Then, with probability at least $1 - \delta$, $\tilde{y}^t$ and all entries of $\tilde{x}^t$ $(t = 1, \dots, T)$ have absolute value at most $\frac{1}{\sqrt{\lambda+1}}$.*

Proof. The first claim is standard in the literature (e.g., see [Zhou et al., 2011, Eq. (4)]), and the second claim follows directly from the first.

For the third claim, let N be a Gaussian random variable with mean 0 and variance 1. We make use of the standard (Chernoff) tail bound

$$\mathbb{P}\left(|N| > x\right) \leq 2e^{-x^2/2}. \tag{41}$$

By scaling the standard Gaussian distribution, recalling the definition of $\nu_{\max}$ in (6), and using $B = \sqrt{2\log\frac{2pT}{\delta}}$, it follows that

$$\mathbb{P}(|x_i^t| > \sqrt{\nu_{\max}}B) \leq \mathbb{P}\left(|N| > \sqrt{2\log\frac{2pT}{\delta}}\right) \tag{42}$$

$$\leq 2\exp\left(-\log\frac{2pT}{\delta}\right) \tag{43}$$

$$\leq \frac{\delta}{pT}, \tag{44}$$

and hence

$$\mathbb{P}\left(|x_i^t| > \frac{1}{\sqrt{\lambda+1}}\right) \leq \frac{\delta}{pT}. \tag{45}$$

The same high probability bound holds similarly for $\tilde{y}^t$. By taking the union bound over these p events, and also over $t = 1, \ldots, T$, we obtain the desired result. □

C Establishing Lemma 4 (Martingale Concentration Bound)

Here we provide additional details on attaining Lemma 4 from a more general result in [van de Geer, 1995]. While the latter concerns continuous-time martingales, we first state some standard definitions for discrete-time martingales. Throughout the appendix, we distinguish between discrete time and continuous time by using notation such as $M_t, \mathcal{F}_t$ for the former, and $\tilde{M}_t, \tilde{\mathcal{F}}_t$ for the latter.

Definition 10. *Given a discrete-time martingale $\{M_t\}_{t=0,1,\ldots}$ with respect to a filtration $\{\mathcal{F}_t\}_{t=0,1,\ldots}$, we define the following:*

1. *The* compensator *of $\{M_t\}$ is defined to be*

$$V_t = \sum_{j=1}^{t} \mathbb{E}[M_j - M_{j-1} \,|\, \mathcal{F}_{j-1}]. \tag{46}$$

2. *A discrete-time process $\{W_t\}_{t=1,2,\ldots}$ defined on the same probability space as $\{M_t\}$ is said to be* predictable *if W_t is measurable with respect to $\mathcal{F}_{t-1}$.*

3. *We say that $\{M_t\}$ is* locally square integrable *if there exists a sequence of stopping times $\{\tau_k\}_{k=1}^{\infty}$ with $\tau_k \to \infty$ such that $\mathbb{E}[M_{\tau_k}^2] < \infty$ for all k.*

In the continuous-time setup of [van de Geer, 1995, Lemma 2.2], the preceding definitions are replaced by generalized notions, e.g., see [Liptser and Shiryayev, 1989]. Note that the notion of a compensator in the continuous-time setting is much more technical, in contrast with the explicit formula (46) for discrete time.

The setup of [van de Geer, 1995] is as follows: Let $\{\tilde{M}_t\}_{t\geq 0}$ be a locally square integrable continuous-time martingale with respect to to a filtration $\{\tilde{\mathcal{F}}_t\}_{t\geq 0}$ satisfying right-continuity ($\tilde{\mathcal{F}}_t = \cap_{s>t}\tilde{\mathcal{F}}_s$) and completeness ($\mathcal{F}_0$ includes all sets of null probability). For each $t > 0$, the martingale jump is defined as $\Delta\tilde{M}_t = \tilde{M}_t - \tilde{M}_{t-}$, where t_- represents an infinitesimal time instant prior to t. For each integer $m \geq 2$, a higher-order variation process $\{\sum_{s\leq t} |\Delta\tilde{M}_s|^m\}$ is considered, and its compensator is denoted by $\tilde{V}_{m,t}$. Then, we have the following.

Lemma 11. [van de Geer, 1995, Lemma 2.2] *Under the preceding setup for continuous-time martingales, suppose that for all $t \geq 0$ and some $0 < K < \infty$, it holds that*

$$\tilde{V}_{m,t} \leq \frac{m!}{2} K^{m-2} \tilde{R}_t, \qquad m = 2, 3, \dots, \tag{47}$$

for some predictable process $\tilde{R}_t$. Then, for any $a, b > 0$, we have

$$\mathbb{P}(\tilde{M}_t \geq a \text{ and } \tilde{R}_t \leq b^2 \text{ for some } t) \leq \exp\left(-\frac{a^2}{2aK + b^2}\right). \tag{48}$$

While Lemma 11 is stated for continuous-time martingales, we obtain the discrete-time version in Lemma 4 by considering the choice $\tilde{M}_t = M_{\lfloor t \rfloor}$, where $\{M_t\}_{t=0,1,\dots}$ is the discrete-time martingale. Due to the floor operation, the required right-continuity condition on the continuous-time martingale holds. Moreover, the definition of a compensator in (46) applied to the higher-order variation process with parameter m yields

$$V_{m,t} = \sum_{j=1}^{t} \mathbb{E}\big[|\Delta M_j|^m \,|\, \mathcal{F}_{j-1}\big] \tag{49}$$

with $\Delta M_t = M_t - M_{t-1}$, in agreement with the statement of Lemma 4. Finally, since we assumed that $\mathbb{E}[M_t^2] < \infty$ for all t in Lemma 4, the locally square integrable condition follows by choosing the trivial sequence of stopping times, $\tau_k = k$.

D Proof of Lemma 5 (Concentration of $\sum_j Z^j$)

Lemma 5 is restated as follows.

Lemma 5. $|\sum_{j=1}^T Z^j| = O\left(\sqrt{T \log \frac{1}{\delta}}\right)$ *with probability at least $1 - \delta$.*

Proof. Recall that $\mathbb{E}_{t-1}[\cdot]$ denotes expectation conditioned on the history up to index $t-1$. Using the notation of Lemma 4, we let $M_t = \sum_{j \leq t} Z^j$, which yields $\Delta M_t = Z^t$. The definition of Z^t in (15) ensures that $\mathbb{E}_{t-1}[Z^t] = 0$, so that M_t is a martingale. In addition, we have

$$V_{m,t} = \sum_{j=1}^{t} \mathbb{E}_{j-1}[|\Delta M_j|^m] = \sum_{j=1}^{t} \mathbb{E}_{j-1}[|Z^j|^m]. \tag{50}$$

To use Lemma 4, we need to bound $\sum_{j=1}^t \mathbb{E}_{j-1}[|Z^j|^m]$ for some appropriate choices of K and R_t in (12). The conditional moments of $|Z^j|$ are the central conditional moments of Q^j:

$$\begin{aligned} \mathbb{E}_{j-1}[|Z^j|^m] &= \mathbb{E}_{j-1}[|Q^j - \mathbb{E}_{j-1}[Q^j]|^m] && (51)\\ &\leq \mathbb{E}_{j-1}[2^m(|Q^j|^m + |\mathbb{E}_{j-1}[Q^j]|^m)] && (52)\\ &\leq 2^{m+1}\mathbb{E}_{j-1}[|Q^j|^m], && (53) \end{aligned}$$

where (51) follows from the definition of Z^j in (15), (52) uses $|a - b| \leq 2\max\{|a|, |b|\}$, and (53) follows from Jensen's inequality ($|\mathbb{E}[Q^j]|^m \leq \mathbb{E}[|Q^j|^m]$). Furthermore, we have that

$$\begin{aligned} \mathbb{E}_{j-1}[|Q^j|^m] &= \mathbb{E}_{j-1}[|(\lambda p^j \cdot \tilde{x}^j - \tilde{y}^j)(p^j - w/\lambda) \cdot \tilde{x}^j|^m] && (54)\\ &\leq \mathbb{E}_{j-1}[|(\lambda p^j \cdot \tilde{x}^j - \tilde{y}^j)|^{2m}]^{1/2}\mathbb{E}_{j-1}[|(p^j - w/\lambda) \cdot \tilde{x}^j|^{2m}]^{1/2}, && (55) \end{aligned}$$

where (54) uses the definition of Q^j in (14), and (55) follows from the Cauchy-Schwartz inequality. Both of the averages in (55) contain Gaussian random variables (with p^j fixed due to the conditioning); we proceed by establishing an upper bound on the variances. Since $(\tilde{x}^j, \tilde{y}^j) = \frac{1}{B\sqrt{\nu_{\max}(\lambda+1)}}(x^j, y^j)$, the definition of $\nu_{\max}$ (see (6)) implies that each coordinate has a variance of at most $\left(\frac{1}{B\sqrt{\lambda+1}}\right)^2$. Then, using that $\sum_i p_i^j = 1$, we have

$$\mathrm{Var}(\lambda p^j \cdot \tilde{x}^j - \tilde{y}^j) \leq (\lambda+1)^2 \max_{z \in \{\tilde{x}_1^j, \ldots, \tilde{x}_n^j \tilde{y}^j\}} \mathrm{Var}(z) \tag{56}$$

$$\leq \frac{\lambda+1}{B^2}, \tag{57}$$

and similarly, using $\sum_i p_i^j = 1$ and $\|w\| = \lambda$ (see Footnote 2),

$$\mathrm{Var}((p^j - w/\lambda) \cdot \tilde{x}^j) \leq \frac{4}{(\lambda+1)B^2}. \tag{58}$$

Next, we use the standard fact that if N is a Gaussian random variable with mean 0 and variance σ, then

$$\mathbb{E}[N^p] = \begin{cases} 0 & \text{if } p \text{ is odd} \\ \sigma^p (p-1)!! & \text{if } p \text{ is even.} \end{cases} \tag{59}$$

It then follows from (53) and (57)–(59) that

$$\mathbb{E}_{j-1}[|Z^j|^m] \leq 2^{m+1} \mathbb{E}_{j-1}[|(\lambda p^j \cdot \tilde{x}^j - \tilde{y}^j)|^{2m}]^{1/2} \mathbb{E}_{j-1}[|(p^j - w/\lambda) \cdot \tilde{x}^j|^{2m}]^{1/2} \tag{60}$$

$$\leq 2^{m+1} \left(\left(\frac{\lambda+1}{B^2}\right)^{2m} (2m-1)!! \left(\frac{4}{(\lambda+1)B^2}\right)^{2m} (2m-1)!! \right)^{1/2} \tag{61}$$

$$= 2^{m+1} \frac{4^m}{B^{4m}} (2m-1)!! \tag{62}$$

$$= 2^{m+1} \frac{4^m}{B^{4m}} (1 \cdot 3 \cdot \ldots \cdot (2m-1)) \tag{63}$$

$$\leq 2^{m+1} \frac{4^m}{B^{4m}} (2 \cdot 4 \cdot \ldots \cdot 2m) \tag{64}$$

$$= 2 \cdot 4^m \frac{4^m}{B^{4m}} m! \tag{65}$$

$$= \frac{m!}{2} \left(\frac{16}{B^4}\right)^{m-2} \frac{2^{10}}{B^8}, \tag{66}$$

and summing over $j = 1, \ldots, t$ gives

$$\sum_{j=1}^{t} \mathbb{E}_{k-1}[|Z^j|^m] \leq \frac{m!}{2} \left(\frac{16}{B^4}\right)^{m-2} \frac{2^{10} t}{B^8}. \tag{67}$$

Hence, using the notation of Lemma 4, it suffices to set $K = \frac{16}{B^4}$ and $R_t = \frac{2^{10} t}{B^8}$. Plugging everything in, we get

$$\mathbb{P}\left(\sum_{j=1}^{T} Z^j > a\right) < \exp\left(-\frac{a^2}{32a\frac{1}{B^4} + 2^{10}\frac{T}{B^8}}\right). \tag{68}$$

Let $a = 2^{10}\sqrt{T \log \frac{1}{\delta}}$. Then, since $B = \sqrt{2 \log \frac{2pT}{\delta}}$ is always greater then $\sqrt{\log \frac{1}{\delta}}$, we obtain

$$\mathbb{P}\left(\sum_{j=1}^{t} Z^j > 2^{10}\sqrt{T \log \frac{1}{\delta}}\right) \leq \frac{\delta}{2}. \tag{69}$$

By replacing Z^j by $-Z^j$ above, we get a symmetric lower bound on $\sum_j Z^j$, as all the moments used above remain the same. Applying the union bound, we get that $|\sum_{j=1}^{T} Z^j| = O\left(\sqrt{T \log \frac{1}{\delta}}\right)$ with probability at least $1 - \delta$. □

E Proof of Lemma 7 (Concentration of Empirical Risk)

Lemma 7 is restated as follows.

Lemma 7. *For $\gamma > 0$, $\rho \in (0,1]$, and fixed $v \in \mathbb{R}^n$ satisfying $\|v\|_1 \le \lambda$, there is some $M = O\big((\lambda+1)\frac{\log(1/\rho)}{\gamma}\big)$ such that*

$$\mathbb{P}\left(\left|\frac{1}{M}\sum_{j=1}^{M}\Big((v\cdot a^j - b^j)^2 - \Xi\Big) - \varepsilon(v)\right| \ge \gamma\right) \le \rho, \tag{32}$$

where $\{(a^j,b^j)\}_{j=1}^M$ are the normalized samples defined in Algorithm 2, and $\Xi = \mathbb{E}\big[\mathrm{Var}[b^j \mid a^j]\big]$.[7]

Proof. We first derive a simple equality:

$$\begin{aligned}\mathbb{E}[(v\cdot a^j - b^j)^2] &= \mathbb{E}\big[\mathbb{E}[(v\cdot a^j - b^j)^2 \mid a^j]\big] &(70)\\ &= \mathbb{E}\Big[\big(\mathbb{E}[v\cdot a^j - b^j \mid a^j]\big)^2 + \mathrm{Var}[b^j \mid a^j]\Big] &(71)\\ &= \mathbb{E}[(v\cdot a^j - w\cdot a^j)^2] + \mathbb{E}\big[\mathrm{Var}[b^j \mid a^j]\big] &(72)\\ &= \varepsilon(v) + \Xi, &(73)\end{aligned}$$

where (71) uses $\mathrm{Var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$, (72) uses the second part of Lemma 2, and (73) uses the definitions of $\varepsilon(v)$ and Ξ.

In the following, we recall Bernstein's inequality.

Lemma 12. [Boucheron et al., 2013, Corollary 2.11] *Let $Z_1,\dots,Z_n$ be independent real-valued random variables, and assume that there exist positive numbers ϑ and c such that*

$$\sum_{i=1}^{n}\mathbb{E}[(Z_i)_+^2] \le \vartheta \tag{74}$$

$$\sum_{i=1}^{n}\mathbb{E}[(Z_i)_+^q] \le \frac{q!}{2}\vartheta\cdot c^{q-2}, \tag{75}$$

where $(x)_+ = \max\{x,0\}$. Letting $S = \sum_{i=1}^n \big(Z_i - \mathbb{E}[Z_i]\big)$, we have for all $t>0$ that

$$\mathbb{P}(S \ge t) \le \exp\left(-\frac{t^2}{2(\vartheta + ct)}\right). \tag{76}$$

We would like to use Bernstein's inequality to bound the deviation of

$$\frac{1}{M}\sum_{j=1}^{M}\Big((v\cdot a^j - b^j)^2 - \Xi - \varepsilon(v)\Big) \tag{77}$$

from its mean value 0. To do so, we need to find constants ϑ and c as described in the statement of Bernstein's inequality above.

Recall that $\nu_{\max}$ upper bounds the variance of any marginal variable in each unnormalized sample, and that (a^j,b^j) are samples normalized by $B\sqrt{\nu_{\max}(\lambda+1)}$ with $B = \sqrt{2\log\frac{2pT}{\delta}} \ge 1$. Hence, the entries of (a^i,b^i) have variance at most $\frac{1}{\lambda+1}$, and since $\|v\|_1 \le \lambda$, this implies that $v\cdot a^j - b^j$ has variance at most $\lambda+1$.

Using the expression for the moments of a Gaussian distribution (see (59)), it follows that

$$\mathbb{E}[(v\cdot a^j - b^j)^4] \le 8(\lambda+1)^2, \tag{78}$$

[7] This quantity is the same for all values of j.

$$\mathbb{E}[(v \cdot a^j - b^j)^{2m}] \leq (2m-1)!!(\lambda+1)^m \tag{79}$$
$$\leq 2^m m!(\lambda+1)^m \tag{80}$$
$$= \frac{m!}{2}(8(\lambda+1)^2)(2(\lambda+1))^{m-2}, \tag{81}$$

where (80) is established in the same way as (65). Since $(v \cdot a^j - b^j)^2$ is a non-negative random variable, the non-central moments bound the central moments from above. Hence, it suffices to let $\vartheta = 8(\lambda+1)^2$ and $c = 2(\lambda+1)$, and we obtain from Bernstein's inequality that

$$\mathbb{P}\left(\left|\sum_{j=1}^{M}\left(\left(v \cdot a^j - b^j\right)^2 - \Xi - \varepsilon(v)\right)\right| \geq \gamma M\right) \leq \exp\left(\frac{-\gamma^2 M^2}{2(8(\lambda+1)^2 + 2(\lambda+1)\gamma M)}\right). \tag{82}$$

To simplify the notation, we let M_0 be such that $M = (\lambda+1)M_0$, which yields

$$\mathbb{P}\left(\left|\frac{1}{M}\sum_{j=1}^{M}\left(\left(v \cdot a^j - b^j\right)^2 - \Xi - \varepsilon(v)\right)\right| \geq \gamma\right) \leq \exp\left(\frac{-\gamma^2 M_0^2}{16 + 2\gamma M_0}\right). \tag{83}$$

If $\gamma M_0 \geq 1$, then the right hand side is less than or equal to $\exp\left(\frac{-\gamma M_0}{18}\right)$. Otherwise, if $\gamma M_0 < 1$, then the right hand side is less than $\exp\left(\frac{-\gamma^2 M_0^2}{18}\right)$. It follows that to have a deviation of γ with probability at most ρ, it suffices to set $M_0 = \frac{18\log(1/\rho)}{\gamma}$. Recalling that $M = (\lambda+1)M_0$, it follows that with $M = 18(\lambda+1)\frac{\log(1/\rho)}{\gamma}$, we attain the desired target probability ρ. $\square$

F Proof of Lemma 8 (Low Risk Implies an ℓ_∞ Bound)

Lemma 8 is restated as follows, and refers to the setup described in Section 4.

Lemma 8. *Under the preceding setup, if we have $\varepsilon(v) \leq \epsilon$, then we also have $\|v - w\|_\infty \leq \sqrt{\epsilon\theta_{\max}}$, where $\theta_{\max}$ is a uniform upper bound on the diagonal entries of Θ.*

Proof. Recall that $\varepsilon(v) = \mathbb{E}[((v-w) \cdot X_{\bar{i}})^2]$, where $w = \left(\frac{-\theta_{ij}}{\theta_{ii}}\right)_{j \neq i}$ is the neighborhood weight vector of the node i under consideration, and $X_{\bar{i}} = (X_j)_{j \neq i}$. To motivate the proof, note from Lemma 2 that $X_i = \eta_i + \sum_{j \neq i}(-\theta_{ij}/\theta_{ii})X_j$, where η_i is an $\mathcal{N}\left(0, \frac{1}{\theta_{ii}}\right)$ random variable independent of $\{X_j\}_{j \neq i}$, from which it follows that $\mathrm{Var}(X_i) \geq \mathrm{Var}(\eta_i) = 1/\theta_{ii}$. In the following, we apply similar ideas to $(v-w) \cdot X_{\bar{i}}$.

Specifically, for an arbitrary index $i^* \neq i$, we can lower bound the expected risk $\varepsilon(v)$ as follows:

$$\mathbb{E}[((v-w) \cdot X_{\bar{i}})^2]$$
$$= \mathrm{Var}((v-w) \cdot X_{\bar{i}}) \tag{84}$$
$$= \mathrm{Var}\left(\sum_{j \neq i}(v_j - w_j)X_j\right) \tag{85}$$
$$= \mathrm{Var}\left((v_{i^*} - w_{i^*})X_{i^*} + \sum_{j \notin \{i, i^*\}}(v_j - w_j)X_j\right) \tag{86}$$
$$= \mathrm{Var}\left((v_{i^*} - w_{i^*})\eta_{i^*} - (v_{i^*} - w_{i^*})\frac{\theta_{i^*i}}{\theta_{i^*i^*}}X_i + \sum_{j \notin \{i, i^*\}}\left((v_j - w_j) - (v_{i^*} - w_{i^*})\frac{\theta_{i^*j}}{\theta_{i^*i^*}}\right)X_j\right) \tag{87}$$
$$= \mathrm{Var}((v_{i^*} - w_{i^*})\eta_{i^*}) + \mathrm{Var}\left(-(v_{i^*} - w_{i^*})\frac{\theta_{i^*i}}{\theta_{i^*i^*}}X_i + \sum_{j \notin \{i, i^*\}}\left((v_j - w_j) - (v_{i^*} - w_{i^*})\frac{\theta_{i^*j}}{\theta_{i^*i^*}}\right)X_j\right) \tag{88}$$
$$\geq \mathrm{Var}((v_{i^*} - w_{i^*})\eta_{i^*}) \tag{89}$$
$$= |v_{i^*} - w_{i^*}|^2\mathrm{Var}(\eta_{i^*}), \tag{90}$$

where (84) follows since $\mathbb{E}[X_{\bar{i}}] = 0$, (87) follows from the first part of Lemma 2 applied to node i^*, and (88) uses the independence of η_{i^*} and $X_{\bar{i^*}}$. Since $\mathrm{Var}(\eta_{i^*}) = \frac{1}{\theta_{i^*i^*}}$ and $\varepsilon(v) \le \epsilon$, this gives $|v_{i^*} - w_{i^*}| \le \sqrt{\epsilon \theta_{i^*i^*}} \le \sqrt{\epsilon \theta_{\max}}$. Then, since this holds for all $i^* \neq i$, we deduce that $\|v - w\|_\infty \le \sqrt{\epsilon \theta_{\max}}$, as desired. □