Supplementary Materials: Online Decentralized Leverage Score Sampling for Streaming Multidimensional Time Series

1 Proofs

Proof of Theorem 5.1.

Let $\mathbf{U}_t = 1_{\{\mathbf{X}_t \in \mathcal{E}_r\}} \mathbf{X}_t$. By equation (2), we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{n} \sum_{t=1}^{n} \mathbf{U}_t \mathbf{U}_t' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{U}_t \mathbf{e}_t' \right), \tag{1}$$

which is understood as $-\sqrt{n}\boldsymbol{\beta}$ if the invertibility fails. Note that

$$\mathbb{E}[\operatorname{vec}(\mathbf{U}_t \mathbf{e}_t') \operatorname{vec}((\mathbf{U}_t \mathbf{e}_t')'] = \Omega \otimes \Gamma(r). \tag{2}$$

For any column vector $\mathbf{a} \in \mathbb{R}^{K^2 p}$, the linear combination $\mathbf{a}' \operatorname{vec}(\mathbf{U}_t)\mathbf{e}_t$ forms a stationary martingale difference in t with respect to the filtration $\mathcal{F}_t = \sigma(\mathbf{e}_i,\ i \le t)$ since $\mathbf{U}_t$ is $\mathcal{F}_{t-1}$-measurable and $\mathbf{e}_t$ is centered and independent of $\mathcal{F}_{t-1}$. By (2) and the Martingale Central Limit Theorem (Theorem 35.12 of [Billingsley 1995 Probability and Measure 3rd ed.]), as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{a}' \operatorname{vec}(\mathbf{U}_t \mathbf{e}_t') \xrightarrow{d} N(0, \mathbf{a}'\Omega \otimes \Gamma(r)\mathbf{a}).$$

In view of the Cramer-Wold Device, we have thus shown that as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \operatorname{vec}(\mathbf{U}_t \mathbf{e}_t') \xrightarrow{d} N(\mathbf{0}, \Omega \otimes \Gamma(r)). \tag{3}$$

On the other hand, each component of the $\mathbf{U}_t$ is a causal linear filter of i.i.d. (thus ergodic) $\mathbf{e}_t$, and is hence an ergodic sequence by Lemma 10.5 of [Kallenberg 2002 Foundations of Modern Probability 2nd ed]. Therefore, by the Birkhoff Ergodic Theorem (Theorem 10.6 of [Kallenberg]) applied to each entry, one has almost surely as $n \to \infty$ that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{U}_t \mathbf{U}_t' \to \Gamma(r). \tag{4}$$

At last, notice that the invertible matrices of a fixed size form an open subset under the product topology. Hence $\frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_t \mathbf{U}_t'$ is invertible with probability tending to one as $n \to \infty$. Combining (1), (3) and (4) yields (12). □

Proof of Theorem 5.2.

The case of consecutive sampling can be directly deduced from Theorem 5.1 by letting $E = \mathbb{R}^m$ and substituting n by nq. For the Bernoulli sampling, the proof can be carried out similarly as the proof of Theorem 5.1. In particular, the indicator $1_{\{\mathbf{X}_t \in E\}}$ is replaced by i.i.d. Bernoulli(q) variables independent of the time series $(\mathbf{Y}_t)$, which still retains the martingale property used in the proof of Theorem 5.1.

□

Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89.

Proof of Theorem 5.3.

(a) Since $\mathbf{e}_t$'s are Gaussian, for each $t \in \mathbb{Z}$, $\mathbf{X}_t \sim N(\mathbf{0}, \Gamma)$. Let $\mathbf{X} = (X_1, \dots, X_m) \overset{d}{:=} \mathbf{X}_t$, and let $\mathbf{Z} = \Gamma^{-1/2}\mathbf{X} \sim N(\mathbf{0}, I_m)$. Then

$$\Pr(\mathbf{X} \in \mathcal{E}_r) = \Pr(\mathbf{Z} \in \mathcal{D}_r) = \Pr(\chi^2_m > r^2) = Q(m, r).$$

(b) For any column vector $\mathbf{a} \in \mathbb{R}^m$ with $\|\mathbf{a}\| = 1$, define

$$\begin{aligned} F(\mathbf{a}; \mathcal{E}_r) &:= \mathbf{a}'(\Gamma(r) - Q(m,r)\Gamma)\mathbf{a} \\ &= \mathbb{E}\left[\left(\sum_{i=1}^m a_i X_i\right)^2 [1_{\{\mathbf{X} \in \mathcal{E}_r\}} - Q(m,r)]\right]. \end{aligned} \tag{5}$$

Let ϕ_Γ denote the density of $N(\mathbf{0}, \Gamma)$. Then by a change of variable $\mathbf{x} = \Gamma^{1/2}\mathbf{y}$,

$$\begin{aligned} F(\mathbf{a}; \mathcal{E}_r) &= \int (\mathbf{a}'\mathbf{x})^2 [1_{\mathcal{E}_r}(\mathbf{x}) - Q(m,r)] \phi_\Gamma(\mathbf{x}) d\mathbf{x} \\ &= \int (\mathbf{a}' P' \Lambda^{1/2} P \mathbf{y})^2 [1_{\mathcal{D}_r}(\mathbf{y}) - Q(m,r)] \phi_{I_m}(\mathbf{y}) d\mathbf{y}. \end{aligned}$$

Let $\mathbf{b} = (b_1, \dots, b_m)' = P\mathbf{a}$. By orthogonality of P, we have $\|\mathbf{b}\| = 1$ as well. By a change of variable $\mathbf{z} = (z_1, \dots, z_m)' = P\mathbf{y}$, and using the invariance of $d\mathbf{z}$, ϕ_I and $\mathcal{D}_r$ with respect to an orthogonal transform, we have

$$\begin{aligned} &F(\mathbf{a}(\mathbf{b}); \mathcal{E}_r) \\ &= \int \left(\sum_{i=1}^m b_i \lambda_i^{1/2} z_i\right)^2 [1_{\mathcal{D}_r}(\mathbf{z}) - Q(m,r)] \phi_{I_m}(\mathbf{z}) d\mathbf{z}. \end{aligned} \tag{6}$$

By the symmetry of $\mathcal{D}_r$ and ϕ_{I_m}, the "covariance"

$$\int z_i z_j [1_{\mathcal{D}_r}(\mathbf{z}) - Q(m,r)] \phi_{I_m}(\mathbf{z}) d\mathbf{z} = 0, \qquad \text{if } i \neq j.$$

Hence

$$\begin{aligned} F(\mathbf{a}(\mathbf{b}); \mathcal{E}_r) &= \int \sum_{i=1}^m b_i^2 \lambda_i z_i^2 [1_{\mathcal{D}_r}(\mathbf{z}) - Q(m,r)] \phi_{I_m}(\mathbf{z}) d\mathbf{z} \\ &= \left(\int z_1^2 [1_{\mathcal{D}_r}(\mathbf{z}) - Q(m,r)] \phi_{I_m}(\mathbf{z}) d\mathbf{z}\right) \left(\sum_{i=1}^m b_i^2 \lambda_i\right). \end{aligned}$$

Note that

$$\min_{\|\mathbf{b}\|=1} \left(\sum_{i=1}^m b_i^2 \lambda_i\right) = \lambda_{min},$$

which is positive since Γ is non-singular by assumption. On the other hand, we have

$$\begin{aligned} \int_{\mathcal{D}_r} z_1^2 \phi_{I_m}(\mathbf{z}) d\mathbf{z} &= \frac{1}{m} \int_{\mathcal{D}_r} \|\mathbf{z}\|^2 \\ \phi_{I_m}(\mathbf{z}) d\mathbf{z} &= \frac{1}{m} \mathbb{E}[\chi^2_m 1_{\{\chi^2_m > r^2\}}] = T(m,r). \end{aligned}$$

and

$$\int z_1^2 \phi_{I_m}(\mathbf{z}) d\mathbf{z} = 1.$$

Hence

$$\min_{\|\mathbf{b}\|=1} F(\mathbf{a}(\mathbf{b}); \mathcal{E}_r) = \lambda_{\min} [T(m,r) - Q(m,r)].$$

□