Electronic Companion

Appendix A: Impact of Time Augmentation

First, recall that the time-augmented process of a d-dimensional continuous-time stochastic process $\mathbf{X}_t = (X_t^1, X_t^2, \ldots, X_t^d)^\top \in \mathbb{R}^d$, $0 \le t \le T$ is a $(d+1)$-dimensional process (Chevyrev and Kormilitzin 2016, Lyons and McLeod 2022)

$$\tilde{\mathbf{X}}_t = \left(t, \mathbf{X}_t^\top\right)^\top = \left(t, X_t^1, X_t^2, \ldots, X_t^d\right)^\top. \tag{A.1}$$

The time augmentation does not change the core block-diagonal structure between signature components.

In particular, for a d-dimensional Brownian motion $\mathbf{X}$ given by (7), if all signature components with the time dimension are grouped together, with other signature components arranged in recursive order (see Definition B.1 in Appendix B), the correlation matrix for Itô signature of $\tilde{\mathbf{X}}$ with orders truncated to K is given by

$$\begin{pmatrix} \Psi_{0,0} & \Psi_{0,1} & \Psi_{0,2} & \cdots & \Psi_{0,K} \\ \Psi_{1,0} & \Omega_1 & 0 & \cdots & 0 \\ \Psi_{2,0} & 0 & \Omega_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ \Psi_{K,0} & 0 & 0 & \cdots & \Omega_K \end{pmatrix},$$

with Ω_i defined by (10) and $\Psi_{0,m}$ the correlation matrix between all signature components with the time dimension and all m-th order signature components without the time dimension.

Similarly, for the Stratonovich signature of a $(d+1)$-dimensional time-augmented Brownian motion given by (7) and (A.1), or the Itô or Stratonovich signature of a $(d+1)$-dimensional time-augmented OU process given by (8) and (A.1), if we group all signature components with the time dimension together and other signature components together, the correlation matrix for the signature with orders truncated to K can be given by

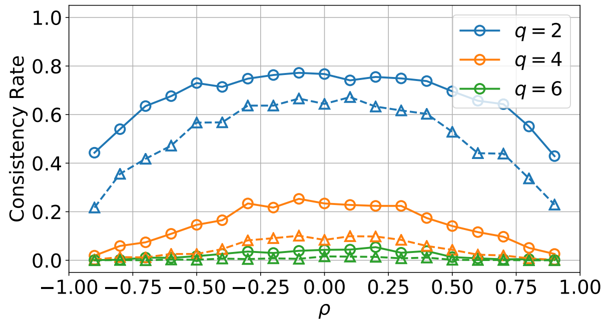
$$\begin{pmatrix} \Psi_{0,0} & \Psi_{0,\mathrm{odd}} & \Psi_{0,\mathrm{even}} \\ \Psi_{\mathrm{odd},0} & \Psi_{\mathrm{odd}} & 0 \\ \Psi_{\mathrm{even},0} & 0 & \Psi_{\mathrm{even}} \end{pmatrix}, \tag{A.2}$$

where Ψ_{odd} and Ψ_{even} are defined by (13), $\Psi_{0,0}$ is the correlation matrix between all signature components with the time dimension, and $\Psi_{0,\mathrm{odd}}$ ($\Psi_{0,\mathrm{even}}$) is the correlation matrix between all signature components with the time dimension and all odd (even) order signature components without the time dimension.

Simulation. Now we perform simulations to study the consistency of signature using Lasso regression for the time-augmented Brownian motion. We consider $\tilde{\mathbf{X}}$, the time-augmentation of a 2-dimensional Brownian motion with an inter-dimensional correlation of ρ. The simulation setups are the same as in Section 4.

Figure A.1 shows the consistency rates for different values of inter-dimensional correlation ρ, and different numbers of true predictors q. The time augmentation generally increases the correlation between signature components and, therefore, leads to a lower consistency rate for Lasso compared to the case without time augmentation (Figure 1(a)). However, the main relationships of the consistency rate with respect to ρ and q remain the same.

Figure A.1 Consistency rates for the time-augmented Brownian motion with different values of inter-dimensional correlation ρ and different numbers of true predictors q. Solid (dashed) lines correspond to the Itô (Stratonovich) signature.



Learning option payoffs. We also demonstrate the ability of signature to learn option payoffs when incorporating time augmentation. Following our framework in Section 5.1.1, we consider two underlying assets, eight different option payoff functions, and three different types of predictors (Sig, RSam, and USam). The only difference in this section is that we also include the time dimension when calculating these three types of predictors.

Figure A.2 shows R^2 as a function of the penalization parameter of the Lasso regression λ, when using different types of predictors with time augmentation. Similar to our observations without time augmentation (Figure 4), both in-sample and out-of-sample R^2 values for Lasso regression with signature components as predictors consistently outperform those for Lasso regression with random sampling and equidistant sampling as predictors.

By comparing Figure A.2 and Figure 4, we also find that R^2 values using signature with time augmentation outperform those without time augmentation, particularly for path-dependent options. This demonstrates that, although the signature of time-augmentation paths has a lower consistency rate due to the inclusion of more predictors, it is more effective in approximating various nonlinear payoff functions, thanks to the universal nonlinearity (Theorem 1).

Appendix B: Technical Details and Examples for the Calculation of Correlation Structures

This appendix provides details and examples for calculating the correlation structures of signature. Appendices B.1 and B.2 discuss the Brownian motion and the OU process, respectively.

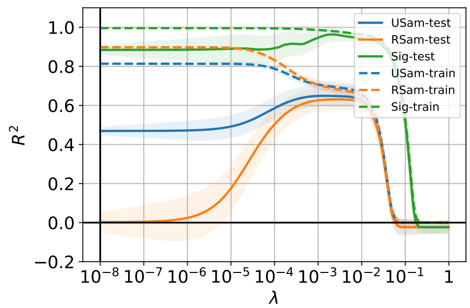
B.1. Brownian Motion

Itô Signature. Proposition 1 and Theorem 3 in the main paper give explicit formulas for calculating the correlation structure of the Itô signature for Brownian motion. The "recursive order" mentioned in Theorem 3 is defined as follows.

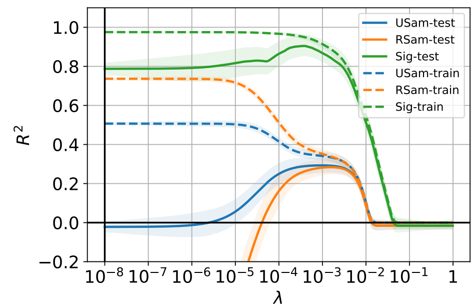
DEFINITION B.1 (RECURSIVE ORDER). Consider a d-dimensional process $\mathbf{X}$. We order the indices of all of its 1st order signature components as

$$1 \quad 2 \quad \cdots \quad d.$$

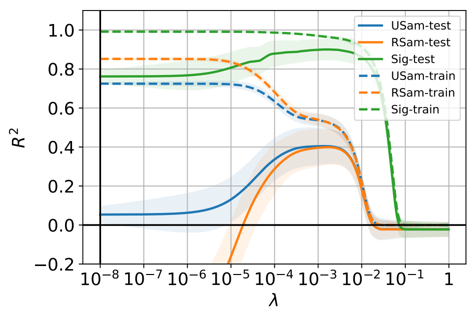
Figure A.2 In-sample and out-of-sample R^2 for learning option payoffs using different types of predictors with time augmentation.



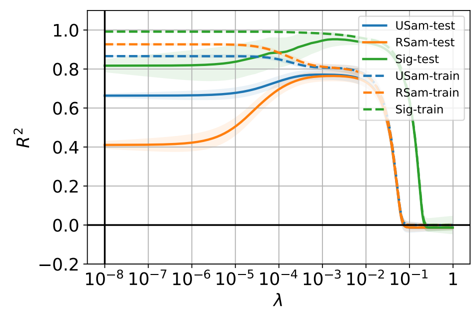
(a) Call option.



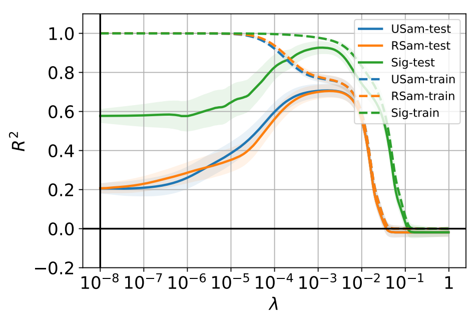
(b) Put option.



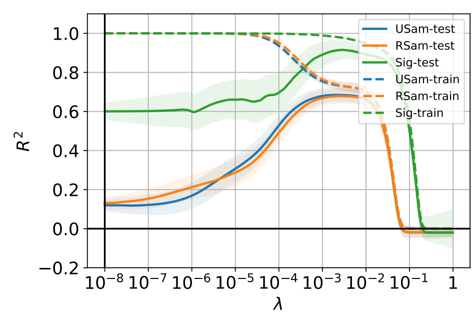
(c) Asian option.



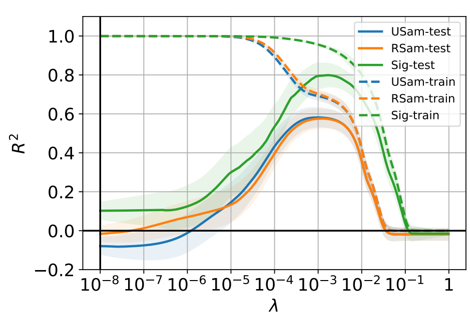
(d) Lookback option.



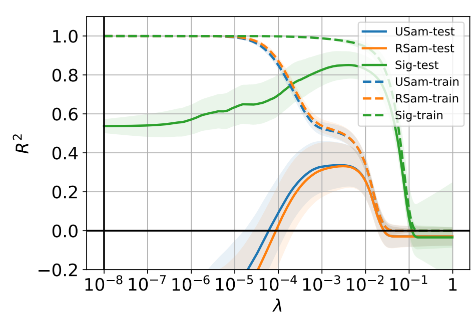
(e) Rainbow option I.



(f) Rainbow option II.



(g) Rainbow option III.



(h) Rainbow option IV.

Then, if all k-th order signature components are ordered as

$$r_1 \quad r_2 \quad \cdots \quad r_{d^k},$$

we define the orders of all $(k+1)$-th order signature components as

$$r_1,1 \quad r_2,1 \quad \cdots \quad r_{d^k},1 \quad r_1,2 \quad r_2,2 \quad \cdots \quad r_{d^k},2 \quad \cdots\cdots\cdots \quad r_1,d \quad r_2,d \quad \cdots \quad r_{d^k},d.$$

For example, for a $d=3$-dimensional process, the recursive order of its signature is

- 1st order: $1 \quad 2 \quad 3$
- 2nd order: $1,1 \quad 2,1 \quad 3,1 \quad 1,2 \quad 2,2 \quad 3,2 \quad 1,3 \quad 2,3 \quad 3,3$
- 3rd order: $1,1,1 \quad 2,1,1 \quad 3,1,1 \quad 1,2,1 \quad 2,2,1 \quad 3,2,1 \quad 1,3,1 \quad 2,3,1 \quad 3,3,1$
  $1,1,2 \quad 2,1,2 \quad 3,1,2 \quad 1,2,2 \quad 2,2,2 \quad 3,2,2 \quad 1,3,2 \quad 2,3,2 \quad 3,3,2$
  $1,1,3 \quad 2,1,3 \quad 3,1,3 \quad 1,2,3 \quad 2,2,3 \quad 3,2,3 \quad 1,3,3 \quad 2,3,3 \quad 3,3,3$
- ...

To provide intuition for Proposition 1 and Theorem 3 in the main paper, the following two examples show the correlation structures of Itô signatures for 2-dimensional Brownian motions with inter-dimensional correlations $\rho = 0.6$ and $\rho = 0$, respectively.

EXAMPLE B.1. Consider a 2-dimensional Brownian motion given by (7) with an inter-dimensional correlation of $\rho = 0.6$. Figure B.1(a) shows the correlation matrix of its Itô signature calculated using Proposition 1. The figure illustrates Theorem 3—the correlation matrix has a block diagonal structure, and each block of the matrix is the Kronecker product of the inter-dimensional correlation matrix $\begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}$.

EXAMPLE B.2. Consider a 2-dimensional Brownian motion given by (7) with an inter-dimensional correlation of $\rho = 0$. Figure B.1(b) shows the correlation matrix of its Itô signature calculated using Proposition 1. When $\rho = 0$, the block diagonal correlation matrix reduces to an identity matrix, indicating that all of its Itô signature components are mutually uncorrelated.

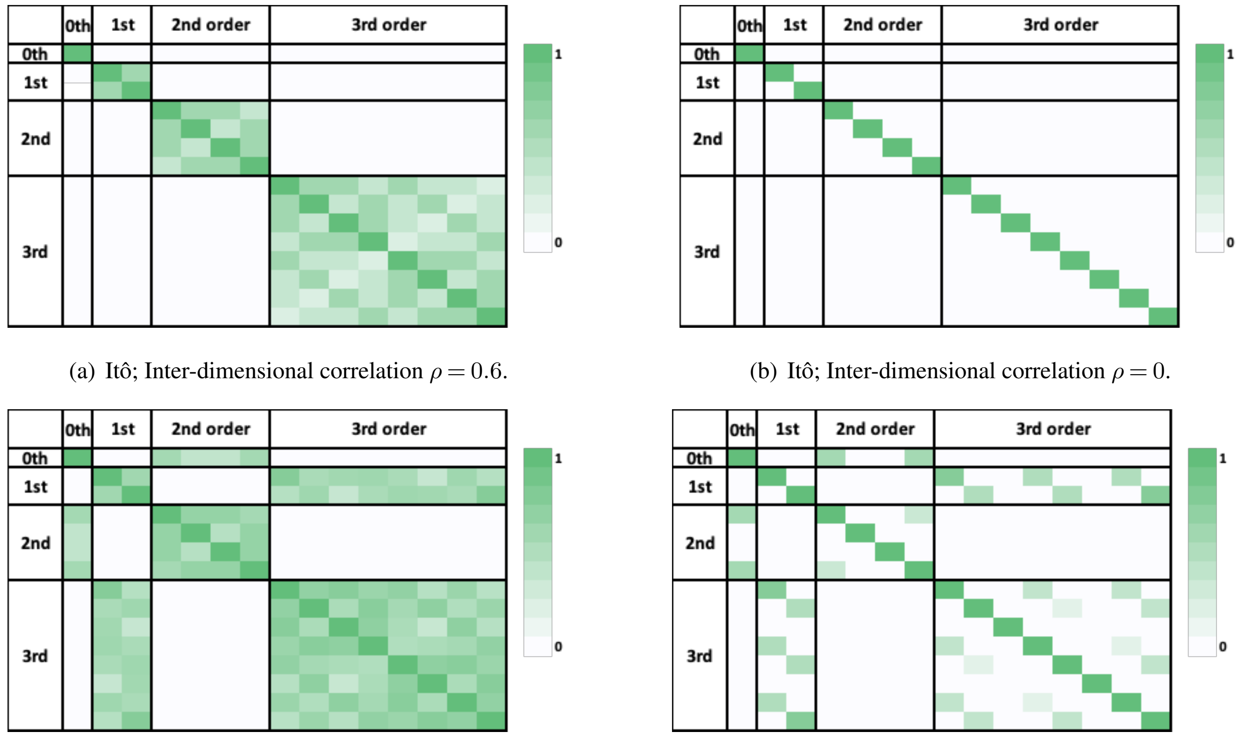
Stratonovich Signature. Proposition 2 and Theorem 4 in the main paper provide formulas for calculating the correlation structure of the Stratonovich signature for a Brownian motion. The following proposition gives the concrete recursive formulas for calculating $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_{2n},S} S(\mathbf{X})_t^{j_1,\ldots,j_{2m},S}\right]$ and $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_{2n-1},S} S(\mathbf{X})_t^{j_1,\ldots,j_{2m-1},S}\right]$, which extends Proposition 2 in the main paper.

PROPOSITION B.1. *Let $\mathbf{X}$ be a d-dimensional Brownian motion given by* (7). *For any $l,t \geq 0$ and $m,n \in \mathbb{N}^+$, define $f_{2n,2m}(l,t) := \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_{2n},S} S(\mathbf{X})_t^{j_1,\ldots,j_{2m},S}\right]$, we have*

$$f_{2n,2m}(l,t) = g_{2n,2m}(l,t) + \frac{1}{2}\rho_{j_{2m-1}j_{2m}}\sigma_{j_{2m-1}}\sigma_{j_{2m}} \int_0^t f_{2n,2m-2}(l,s)\mathrm{d}s, \tag{B.1}$$

$$g_{2n,2m}(l,t) = \rho_{i_{2n}j_{2m}}\sigma_{i_{2n}}\sigma_{j_{2m}} \int_0^{l\wedge t} f_{2n-1,2m-1}(s,s)\mathrm{d}s + \frac{1}{2}\rho_{i_{2n-1}i_{2n}}\sigma_{i_{2n-1}}\sigma_{i_{2n}} \int_0^l g_{2n-2,2m}(s,t)\mathrm{d}s, \tag{B.2}$$

Figure B.1 Correlation matrices of signatures for 2-dimensional Brownian motions.



(a) Itô; Inter-dimensional correlation $\rho = 0.6$.

(b) Itô; Inter-dimensional correlation $\rho = 0$.

(c) Stratonovich; Inter-dimensional correlation $\rho = 0.6$.

(d) Stratonovich; Inter-dimensional correlation $\rho = 0$.

with initial conditions

$$f_{0,0}(l,t) = 1, \tag{B.3}$$

$$g_{0,2m}(l,t) = 0. \tag{B.4}$$

In addition, define $f_{2n-1,2m-1}(l,t) := \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_{2n-1},S} S(\mathbf{X})_t^{j_1,\ldots,j_{2m-1},S}\right]$*, we have*

$$f_{2n-1,2m-1}(l,t) = g_{2n-1,2m-1}(l,t) + \frac{1}{2}\rho_{j_{2m-2}j_{2m-1}}\sigma_{j_{2m-2}}\sigma_{j_{2m-1}} \int_0^t f_{2n-1,2m-3}(l,s)\mathrm{d}s, \tag{B.5}$$

$$\begin{aligned} g_{2n-1,2m-1}(l,t) = {} & \rho_{i_{2n-1}j_{2m-1}}\sigma_{i_{2n-1}}\sigma_{j_{2m-1}} \int_0^{l\wedge t} f_{2n-2,2m-2}(s,s)\mathrm{d}s \\ & + \frac{1}{2}\rho_{i_{2n-2}i_{2n-1}}\sigma_{i_{2n-2}}\sigma_{i_{2n-1}} \int_0^l g_{2n-3,2m-1}(s,t)\mathrm{d}s, \end{aligned} \tag{B.6}$$

with initial conditions

$$f_{1,1}(l,t) = \rho_{i_1 j_1}\sigma_{i_1}\sigma_{j_1}(l\wedge t), \tag{B.7}$$

$$g_{1,2m-1}(l,t) = \rho_{i_1 j_{2m-1}} \frac{1}{2^{m-1}} \frac{(l\wedge t)^{m-1}}{(m-1)!} \sigma_{i_1} \prod_{k=1}^{2m-1} \sigma_{j_k} \prod_{k=1}^{m-1} \rho_{j_{2k-1}j_{2k}}. \tag{B.8}$$

Here, $x\wedge y$ *represents the smaller value between* x *and* y*.*

The following two examples show the correlation structures of Stratonovich signatures for 2-dimensional Brownian motions with inter-dimensional correlations $\rho = 0.6$ and $\rho = 0$, respectively, calculated using Proposition 2 and Theorem 4 in the main paper and Proposition B.1.

EXAMPLE B.3. Consider a 2-dimensional Brownian motion given by (7) with an inter-dimensional correlation of $\rho = 0.6$. Figure B.1(c) shows the correlation matrix of its Stratonovich signature calculated using Propositions 2 and B.1. The figure illustrates that the correlation matrix has an odd–even alternating structure.

EXAMPLE B.4. Consider a 2-dimensional Brownian motion given by (7) with an inter-dimensional correlation of $\rho = 0$. Figure B.1(d) shows the correlation matrix of its Stratonovich signature calculated using Propositions 2 and B.1. The figure demonstrates that the correlation matrix has an odd–even alternating structure, even though different dimensions of the Brownian motion are mutually independent ($\rho = 0$). This is different from the result for Itô signature shown in Example B.2, where all Itô signature are mutually uncorrelated.

In this case, assume that one includes all Stratonovich signature components of orders up to $K = 4$ in the Lasso regression given by (4), and the true model given by (2) has beta coefficients $\beta_0 = 0$, $\beta_1 > 0$, $\beta_2 > 0$, $\beta_{1,1} > 0$, $\beta_{1,2} > 0$, $\beta_{2,1} > 0$, $\beta_{2,2} < 0$, and $\beta_{i_1,i_2,i_3} = \beta_{i_1,i_2,i_3,i_4} = 0$. Let Δ^2 be the correlation matrix between all predictors given by Theorem 4. Then, by Proposition 2,

$$\begin{aligned}\Delta^2_{A^{*c},A^*}(\Delta^2_{A^*,A^*})^{-1}\mathrm{sign}(\boldsymbol{\beta}_{A^*}) = (&0, 0.77, 0.5, 0, 0.5, 0.5, 0, 0.5, 0.77, 1.01, 0.73, 0.47, 0,\\ &0.47, 0, 0.58, 0.73, 0.73, -0.58, 0, 0.47, 0, 0.47, 0.73, -1.01)^\top,\end{aligned}$$

which does not satisfy the irrepresentable conditions I and II defined in Definition 4 because $|-1.01| > 1$.

B.2. OU Process

Deriving explicit formulas for calculating the exact correlation between signature components of OU processes (both Itô and Stratonovich) is complicated. Here we provide an example to show the general approach for calculating the correlation. The proof of this example is given in Appendix E, and one can use a similar routine to compute the correlation for other setups of OU processes.

EXAMPLE B.5. Consider a 1-dimensional OU process $\mathbf{X}_t = Y_t$ with a mean reversion speed $\kappa > 0$, which is driven by

$$\mathrm{d}Y_t = -\kappa Y_t \mathrm{d}t + \mathrm{d}W_t, \quad Y_0 = 0. \tag{B.9}$$

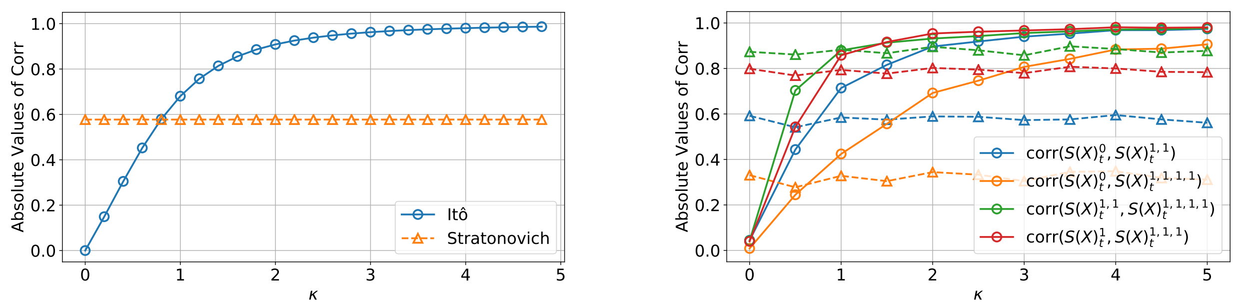
The correlation coefficients between its 0-th order and 2nd order of signature are

$$\begin{aligned}\frac{\mathbb{E}\left[S(\mathbf{X})_T^{0,I} S(\mathbf{X})_T^{1,1,I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_T^{0,I}\right]^2 \mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right]^2}} &= \frac{-2\kappa T - e^{-2\kappa T} + 1}{\sqrt{4\kappa T e^{-2\kappa T} + 3e^{-4\kappa T} - 6e^{-2\kappa T} - 4\kappa T + 3 + 4\kappa^2 T^2}},\\ \frac{\mathbb{E}\left[S(\mathbf{X})_T^{0,S} S(\mathbf{X})_T^{1,1,S}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_T^{0,S}\right]^2 \mathbb{E}\left[S(\mathbf{X})_T^{1,1,S}\right]^2}} &= \frac{\sqrt{3}}{3},\end{aligned}$$

for Itô and Stratonovich signature, respectively. The proof is provided in Appendix E.

Figure B.2(a) shows the absolute values of correlation coefficients between the 0-th order and 2nd order signature components calculated using the formulas above under different values of κ with $T = 1$. Notably, the correlation for Itô signature increases with respect to κ, while the correlation for Stratonovich signature remains fixed at $\sqrt{3}/3$.

Figure B.2 Absolute values of correlation coefficients between signature components of the 1-dimensional OU process. Solid (dashed) lines correspond to the Itô (Stratonovich) signature.



(a) Correlation between the 0-th and the 2nd order signature components.

(b) Correlation between the first four order signature components.

We further perform simulations to estimate the correlation coefficients for higher-order signature components of the OU process. We generate 10,000 sample paths of the OU process using the methods discussed in Appendix D. For each path, we calculate the corresponding signature components and then estimate the sample correlation matrix based on the 10,000 simulated samples. Figure B.2(b) shows the simulation results for the absolute values of correlation coefficients between the first four order signature components under different values of κ. Consistent with the observation in Figure B.2(a), the correlations for Itô signature increase with respect to κ, while the correlations for Stratonovich signature remain relatively stable. Notably, the correlations for Itô signature are zero when $\kappa = 0$, which reduces to the results for a Brownian motion. In addition, when κ is sufficiently large, the absolute values of correlation coefficients for Itô signature exceed those for Stratonovich signature.

Recall that the irrepresentable condition, as defined in Definition 4, illustrates that a higher correlation generally leads to poorer consistency. Therefore, based on Example B.5, we can expect that the Lasso is more consistent when using Itô signature for small values of κ (weaker mean reversion), and more consistent when using Stratonovich signature for large values of κ (stronger mean reversion). This provides a theoretical explanation for our observations in Section 4.3 of the main paper—When processes are sufficiently rough or mean reverting (El Euch et al. 2018, Gatheral et al. 2018), using Lasso with Stratonovich signature will likely lead to higher statistical consistency compared to Itô signature.

Appendix C: Technical Details for Consistency of Signature

C.1. Tightness of the Sufficient Condition for Consistency

In this appendix, we investigate the irrepresentable condition for Itô signature of a multi-dimensional Brownian motion with constant inter-dimensional correlation. This analysis not only provides further insights into the irrepresentable condition but also demonstrates the tightness of the sufficient condition presented in Theorem 5 in our main paper.

The following proposition characterizes the irrepresentable condition for a Brownian motion with constant inter-dimensional correlation when using Itô signature. For mathematical simplicity, we assume that only the first order signature components are included in the regression model.

PROPOSITION C.1. *For a multi-dimensional Brownian motion given by* (7) *with equal inter-dimensional correlation* $\rho = \rho_{ij}$*, assume that only its first order Itô signature components are included in* (2)*, and that all true beta coefficients are positive. Then, the irrepresentable conditions I and II hold if* $\rho \in (-\frac{1}{2\#A_1^*}, 1)$*, and do not hold if* $\rho \in (-\frac{1}{\#A_1^*}, -\frac{1}{2\#A_1^*}]$.

REMARK C.1. Proposition C.1 only discusses the results for $\rho \in (-\frac{1}{\#A_1^*}, 1)$. If $\rho \leq -\frac{1}{\#A_1^*}$, then the inter-dimensional correlation matrix for the Brownian motion is not positive definite.

Proposition C.1 demonstrates that the sufficient condition (14) is tight when the inter-dimensional correlation ρ is constant and negative. Meanwhile, for $\rho > 0$, the irrepresentable conditions always hold but may not satisfy (14).

C.2. Consistency of Lasso with General Predictors in Finite Sample

In this appendix, we present additional results on the consistency of Lasso with general predictors (not necessarily signature components) in finite sample.

Consider a linear regression model with N samples and p predictors $X_1, \ldots, X_p$ given by

$$y = X\boldsymbol{\beta} + \varepsilon, \tag{C.1}$$

where $\varepsilon \in \mathbb{R}^N$ is a vector of independent and normally distributed white noise with mean zero and variance σ^2, $X \in \mathbb{R}^{N\times p}$ is the random design matrix with each row represents a random sample of $(X_1, \ldots, X_p)^\top$, $y \in \mathbb{R}^N$ is the target to predict, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of beta coefficients. Assume that X has full column rank. Given a tuning parameter $\lambda > 0$, we adopt the Lasso estimator given by

$$\hat{\boldsymbol{\beta}}^N(\lambda) = \arg\min_{\hat{\boldsymbol{\beta}}} \left\{ \left\| y - \tilde{X}\hat{\boldsymbol{\beta}} \right\|_2^2 + \lambda \left\| \hat{\boldsymbol{\beta}} \right\|_1 \right\} \tag{C.2}$$

to identify the true predictors, where $\tilde{X}$ represents the standardized version of X across N samples by the l_2-norm, whose (n, j)-entry is defined by

$$\tilde{X}_{n,j} = \frac{X_{n,j}}{\sqrt{\sum_{m=1}^N (X_{m,j})^2 \Big/ N}}, \quad n = 1, 2, \ldots, N;\ j = 1, 2, \ldots, p.$$

Therefore, the sample covariance matrix calculated using $\tilde{X}$ is the same as the sample correlation matrix of X.

Denote by $\hat{\Delta}$ and Δ the sample correlation matrix and the population correlation matrix of all predictors in the Lasso regression, respectively. Because the number of samples, N, is finite, $\hat{\Delta}$ may deviate from Δ. Therefore, when studying the consistency of Lasso, $\hat{\Delta}$ may not satisfy the irrepresentable condition even if Δ does. Nevertheless, in this appendix, we show that $\hat{\Delta}$ will satisfy the irrepresentable condition with high probability.

Our analysis aligns with existing works in high-dimensional statistics. For example, Wainwright (2009) assumes that the predictors are normally distributed, while Cai et al. (2022) and Wüthrich and Zhu (2023) assume a sub-Gaussian distribution. However, these results cannot be directly applied in our setup, because the predictors in our paper are signature components, which are neither Gaussian nor sub-Gaussian.

We first introduce some notations. Denote the set of true predictors by A^*, false predictors by A^{*c}, the number of true predictors by $q = \#A^*$, the population covariance matrix of all predictors in the Lasso regression by Σ, the population correlation matrix of all predictors in the Lasso regression by Δ, the correlation matrix between predictors in sets A and B by Δ_{AB}, and the volatility of components of ε in (C.1) by σ. We also let $\tilde{\boldsymbol{\beta}}$ be the vector containing all standardized beta coefficients of the true model whose j-th component is given by

$$\tilde{\beta}_j = \beta_j \cdot \sqrt{\frac{1}{N}\sum_{m=1}^{N}(X_{m,j})^2}.$$

The following result shows the consistency of Lasso under the assumption that all predictors have finite fourth moments.

THEOREM C.1. *For the Lasso regression given by* (C.1) *and* (C.2)*, assume that the following two conditions hold:*

(i) The irrepresentable condition II in Definition 4 holds for the population correlation matrix, i.e., there exists some $\gamma \in (0,1]$ *such that* $\left\|\Delta_{A^{*c}A^*}\Delta_{A^*A^*}^{-1}\right\|_\infty \leq 1-\gamma$*;*

(ii) The predictors have finite fourth moments, i.e., there exists $K < \infty$ *such that* $\mathbb{E}[X_i^4] \leq K$ *for all* $i = 1, \dots, p$*.*

In addition, we assume that the sequence of regularization parameters $\{\lambda_N\}$ *satisfies* $\lambda_N > \frac{4\sigma}{\gamma}\sqrt{\frac{2\ln p}{N}}$*. Then, the following properties hold with probability greater than*

$$\left(1 - \frac{8p^4\sigma_{\max}^4(\sigma_{\min}^4 + K)}{N\xi^2\sigma_{\min}^4}\right)\left(1 - 4e^{-cN\lambda_N^2}\right)$$

for some constant $c > 0$*.*

(a) The Lasso has a unique solution $\hat{\boldsymbol{\beta}}^N(\lambda_N) \in \mathbb{R}^p$ *with its support contained within the true support, and satisfies*

$$\left\|\hat{\boldsymbol{\beta}}^N(\lambda_N) - \tilde{\boldsymbol{\beta}}\right\|_\infty \leq \lambda_N \left[\frac{\zeta(2+2\alpha\zeta+\gamma)}{2+2\alpha\zeta} + \frac{4\sigma}{\sqrt{\frac{1}{2}C_{\min}}}\right] =: h(\lambda_N);$$

(b) If in addition $\min_{i\in A^*}|\tilde{\beta}_i| > h(\lambda_N)$*, then* $\operatorname{sign}(\hat{\boldsymbol{\beta}}^N(\lambda_N)) = \operatorname{sign}(\tilde{\boldsymbol{\beta}})$.

Here, $\sigma_{\min} = \min_{1\le i\le p}\sqrt{\Sigma_{ii}}$*,* $\sigma_{\max} = \max_{1\le i\le p}\sqrt{\Sigma_{ii}}$*,* $\alpha = \|\Delta_{A^{*c}A^*}\|_\infty$*,* $\zeta = \left\|\Delta^{-1}_{A^*A^*}\right\|_\infty$*,* $C_{\min} = \Lambda_{\min}(\Delta_{A^*A^*}) = \frac{1}{\left\|\Delta^{-1}_{A^*A^*}\right\|_2} > 0$*, and* $\xi = \min\left\{g_\Sigma^{-1}\left(\frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}\right), g_\Sigma^{-1}\left(\frac{C_{\min}}{2\sqrt{p}}\right)\right\} > 0$*, where the definition of* $g_\Sigma(\cdot)$ *is given by* (16).

A detailed proof of Theorem C.1 can be found in Appendix E.

Appendix D: Additional Details for Simulation

This appendix provides technical details, computational cost, more numerical experiments, and robustness checks for the simulations conducted in this paper.

D.1. More Technical Details

Throughout our simulations in the paper, we set the time index $0 = t_0 < t_1 < \cdots < t_n = T$ with $t_{k+1} - t_k = \Delta t = T/n$ for any $k \in \{0, 1, \ldots, n-1\}$ and $n = 100$.

Simulation of Processes. We simulate the i-th dimension of the Brownian motion W_t^i, and OU process Y_t^i, by discretizing the stochastic differential equations of the processes using the Euler–Maruyama schemes given by

- Brownian motion: $W^i_{t_{k+1}} = W^i_{t_k} + \sqrt{\Delta t}\varepsilon^i_k$, $W^i_0 = 0$;
- OU process: $Y^i_{t_{k+1}} = Y^i_{t_k} - \kappa_i Y^i_{t_k}\Delta t + \sqrt{\Delta t}\varepsilon^i_k$, $Y^i_0 = 0$,

with ε^i_k randomly drawn from the standard normal distribution.

The i-th dimension of the random walk and AR(1) model, both denoted by Z_t^i, are simulated using the following formulas.

- Random walk: $Z^i_{t_{k+1}} = Z^i_{t_k} + e^i_k$, $Z^i_0 = 0$;
- AR(1) model: $Z^i_{t_{k+1}} = \phi_i Z^i_{t_k} + \varepsilon^i_k$, $Z^i_0 = 0$,

with e^i_k randomly drawn from

$$\mathbb{P}(e^i_k = +1) = \mathbb{P}(e^i_k = -1) = 0.5,$$

and ε^i_k randomly drawn from the standard normal distribution.

After simulating each dimension of the processes, we simulate the inter-dimensional correlation between different dimensions of the processes using the Cholesky decomposition. Finally, we generate $\mathbf{X}$ using (7) or (8).

In all the simulations, we set the length of the processes $T = 1$, and the initial values of the processes to zero. These choices have no impact on the results because the signature of a path $\mathbf{X}$ is invariant under a time reparametrization and a shift of the starting point of $\mathbf{X}$ (Chevyrev and Kormilitzin 2016).

Calculation of Integrals. The calculation of Itô and Stratonovich signatures requires the calculation of Itô and Stratonovich integrals. By definition, they are computed using the following schemes.

- Itô integral: $\int_0^T A_t \mathrm{d}B_t \approx \sum_{k=0}^{n-1} A_{t_k}(B_{t_{k+1}} - B_{t_k})$;
- Stratonovich integral: $\int_0^T A_t \circ \mathrm{d}B_t \approx \sum_{k=0}^{n-1} \frac{1}{2}(A_{t_k} + A_{t_{k+1}})(B_{t_{k+1}} - B_{t_k})$.

D.2. Computational Details

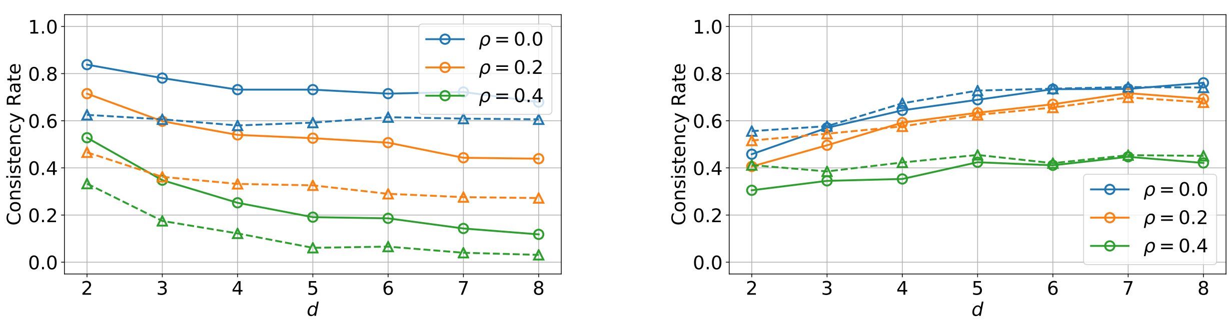
- The simulations are implemented using Python 3.7.
- The simulations are run on a laptop with an Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz.
- The random seed is set to 0 for reproducibility.
- The Lasso regressions are performed using the `sklearn.linear_model.lars_path` package.
- Each individual experiment, including generating 100 paths, calculating their signatures, and performing the Lasso regression, can be completed within one second.

D.3. Impact of the Dimension of the Process and the Number of Samples

Most simulations in Section 4 of our main paper consider the case of $d = 2$ (dimension of the process) and $N = 100$ (number of samples).

Figure D.1 shows how the consistency of Lasso varies with the dimension of the process d, with Figure D.1(a) for the Brownian motion and Figure D.1(b) for the OU process with $\kappa = 2$. We set the number of true predictors to be three. Other simulation setups remain the same as in Section 4.1 of the main paper.

Figure D.1 Consistency rates for the Brownian motion and the OU process with different numbers of dimensions d and different values of inter-dimensional correlation ρ. Solid (dashed) lines correspond to the Itô (Stratonovich) signature.

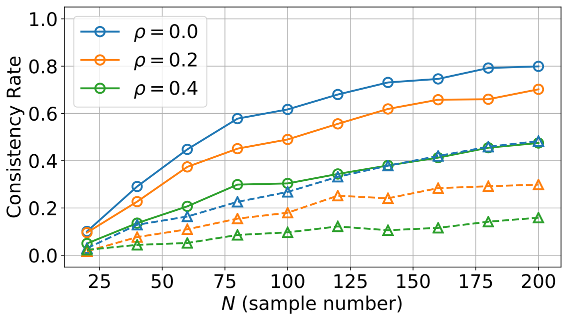


(a) Brownian motion. (b) OU process.

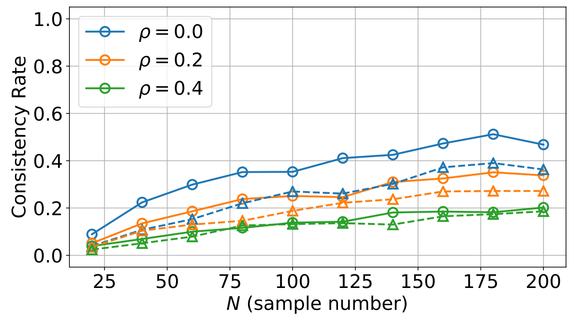
First, for the Brownian motion, the consistency rate decreases with d. This can be attributed to the fact that the inter-dimensional correlation of the process leads to stronger correlations between signature components as more dimensions are included. Second, for the OU process, the consistency rate increases with d because the inter-dimensional correlation of the process is weaker than the correlation between the increments of the OU process itself.

Figure D.2 shows the relationship between the consistency rate and the number of samples. In general, we find that the consistency rate increases as the number of samples increases.

Figure D.2 Consistency rates for the Brownian motion and the OU process with different numbers of samples N and different values of inter-dimensional correlation ρ. Solid (dashed) lines correspond to the Itô (Stratonovich) signature.



(a) Brownian motion.



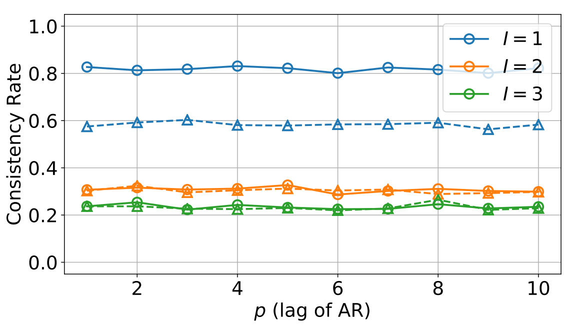
(b) OU process.

D.4. The ARIMA Process

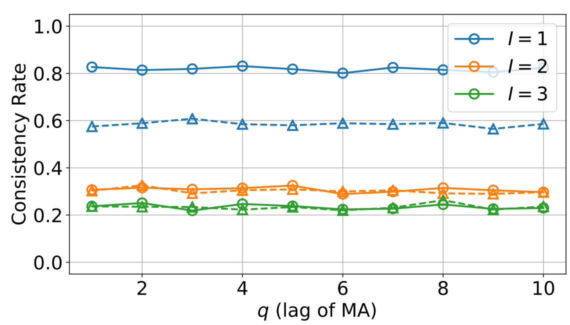
This appendix examines the consistency of signature for the ARIMA(p, I, q) model, where p is the lag of AR, I is the degree of differencing, and q is the lag of MA.

Figure D.3 shows how the consistency rate varies with p, q, and I. We find that the consistency rate does not exhibit any apparent dependence on p and q, but does highly rely on I. Specifically, the consistency rate generally decreases as I increases due to the stronger correlation between the increments of the ARIMA processes introduced by I.

Figure D.3 Consistency rates for the ARIMA(p, I, q) with different lags of AR, p, lags of MA, q, and degrees of differencing, I. Solid (dashed) lines correspond to the Itô (Stratonovich) signature.



(a) Consistency rates for different p and I.



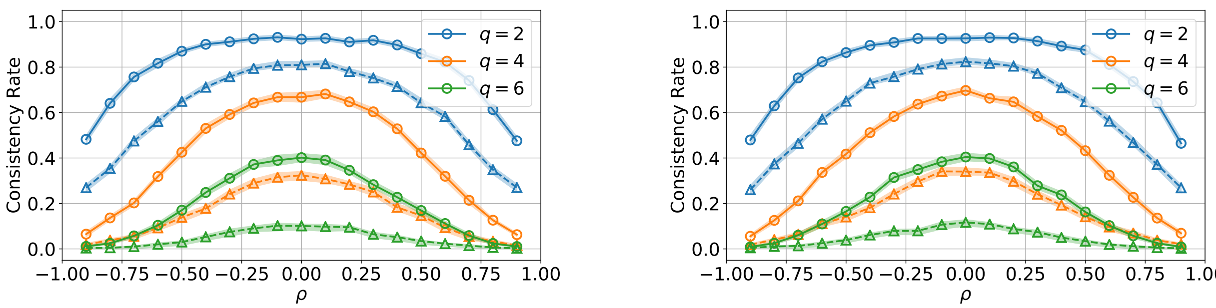
(b) Consistency rates for different q and I.

D.5. Robustness Checks

To show the robustness of our simulations shown in Section 4 of the main paper, we present Figures D.4 and D.5, which include confidence intervals (shaded regions) for the estimated consistency rates of the Brownian motion/random walk and OU process/AR(1) model, respectively.

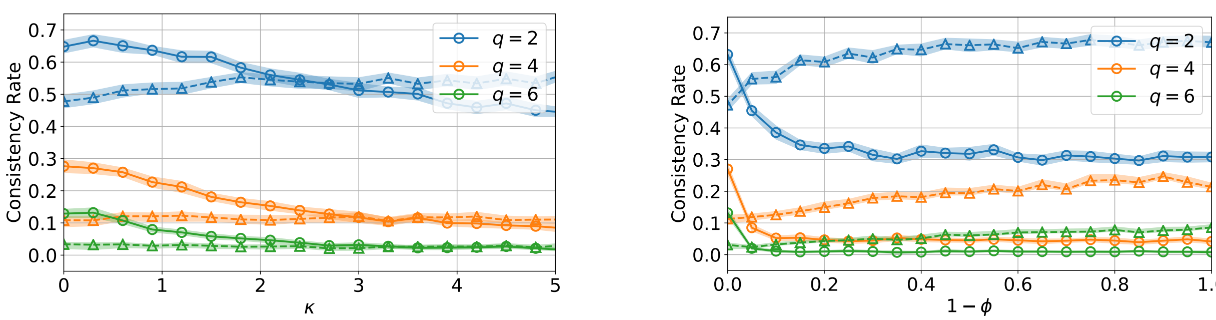
In Figures D.4 and D.5, we estimate the consistency rate by repeating the procedure described in Section 4.1 100 times, and this process is repeated 30 times to obtain the confidence interval for the estimation. Thus, these confidence intervals are based on 30 estimations of the consistency rate, with each estimation calculated using 100 experiments.

Figure D.4 Consistency rates for the Brownian motion and the random walk with different values of inter-dimensional correlation ρ and different numbers of true predictors q. Solid (dashed) lines correspond to the Itô (Stratonovich) signature. Shaded regions are confidence intervals of the experiments.



(a) Brownian motion. (b) Random walk.

Figure D.5 Consistency rates for the OU process and the AR(1) model with different parameters (κ and $1-\phi$) and different numbers of true predictors q. Solid (dashed) lines correspond to the Itô (Stratonovich) signature. Shaded regions are confidence intervals of the experiments.



(a) OU process. (b) AR(1) model.

We observe that the confidence intervals of the consistency rates shown in Figures D.4 and D.5 are narrow. Moreover, the observations made in Section 4 are consistent with the results presented here, further confirming the robustness of our findings.

Appendix E: Lemmas and Proofs

This appendix provides the proofs of all theoretical results in this article and lemmas used in the proofs.

E.1. Lemmas

LEMMA E.1. *Assume that $\hat{\Sigma}$ and Σ are $p\times p$ positive definite matrices with diagonal entries $\{\hat{\sigma}_i^2\}_{i=1}^p$ and $\{\sigma_i^2\}_{i=1}^p$, respectively, with $\sigma_{\min}=\min_{1\le i\le p}\sigma_i$ and $\sigma_{\max}=\max_{1\le i\le p}\sigma_i$. Let $\hat{\Delta}$ and Δ be $p\times p$ matrices with (i,j)-entries $\hat{\Delta}_{ij}=\hat{\Sigma}_{ij}/(\hat{\sigma}_i\hat{\sigma}_j)$ and $\Delta_{ij}=\Sigma_{ij}/(\sigma_i\sigma_j)$ for $i,j=1,2,\ldots,p$, respectively. For $\epsilon<\sigma_{\min}^2$, if $\|\hat{\Sigma}-\Sigma\|_\infty\le\epsilon$, we have $\|\hat{\Delta}-\Delta\|_\infty\le g_\Sigma(\epsilon)$, where $g_\Sigma(\cdot)$ is given by* (16)*.*

Proof of Lemma E.1. For any $i,j=1,2,\ldots,p$,

$$|\hat{\Sigma}_{ij}-\Sigma_{ij}|\le\|\hat{\Sigma}-\Sigma\|_\infty\le\epsilon,$$

which implies that

$$\hat{\Sigma}_{ij}\in\left(\Sigma_{ij}-\epsilon,\Sigma_{ij}+\epsilon\right).$$

Hence, for $\epsilon<\sigma_{\min}^2$,

$$\hat{\sigma}_i\in\left(\sqrt{\sigma_i^2-\epsilon},\sqrt{\sigma_i^2+\epsilon}\right).$$

Now we estimate the difference between Δ_{ij} and $\hat{\Delta}_{ij}$. If $\Delta_{ij}>0$,

$$\begin{aligned}\Delta_{ij}-\hat{\Delta}_{ij}=\frac{\Sigma_{ij}}{\sigma_i\sigma_j}-\frac{\hat{\Sigma}_{ij}}{\hat{\sigma}_i\hat{\sigma}_j}&\le\frac{\Sigma_{ij}}{\sigma_i\sigma_j}-\frac{\Sigma_{ij}-\epsilon}{\sqrt{\sigma_i^2+\epsilon}\cdot\sqrt{\sigma_j^2+\epsilon}}=\frac{\Sigma_{ij}\sqrt{\sigma_i^2+\epsilon}\sqrt{\sigma_j^2+\epsilon}-\left(\Sigma_{ij}-\epsilon\right)\sigma_i\sigma_j}{\sigma_i\sigma_j\sqrt{\sigma_i^2+\epsilon}\sqrt{\sigma_j^2+\epsilon}}\\&\le\frac{\Sigma_{ij}\left(\sqrt{\sigma_i^2+\epsilon}\sqrt{\sigma_j^2+\epsilon}-\sigma_i\sigma_j\right)+\epsilon\sigma_i\sigma_j}{\sigma_i^2\sigma_j^2}=\frac{\Sigma_{ij}\cdot\frac{\epsilon(\sigma_i^2+\sigma_j^2)+\epsilon^2}{\sqrt{\sigma_i^2+\epsilon}\sqrt{\sigma_j^2+\epsilon}+\sigma_i\sigma_j}+\epsilon\sigma_i\sigma_j}{\sigma_i^2\sigma_j^2}\\&\le\frac{\Sigma_{ij}\cdot\frac{\epsilon(\sigma_i^2+\sigma_j^2)+\epsilon^2}{2\sigma_i\sigma_j}+\epsilon\sigma_i\sigma_j}{\sigma_i^2\sigma_j^2}=\Delta_{ij}\cdot\frac{\epsilon(\sigma_i^2+\sigma_j^2)+\epsilon^2}{2\sigma_i^2\sigma_j^2}+\frac{\epsilon}{\sigma_i\sigma_j}\le\Delta_{ij}\cdot\frac{2\epsilon\sigma_{\min}^2+\epsilon^2}{2\sigma_{\min}^4}+\frac{\epsilon}{\sigma_{\min}^2}.\end{aligned}\tag{E.1}$$

Meanwhile,

$$\begin{aligned}\hat{\Delta}_{ij}-\Delta_{ij}=&\frac{\hat{\Sigma}_{ij}}{\hat{\sigma}_i\hat{\sigma}_j}-\frac{\Sigma_{ij}}{\sigma_i\sigma_j}\le\frac{\Sigma_{ij}+\epsilon}{\sqrt{\sigma_i^2-\epsilon}\cdot\sqrt{\sigma_j^2-\epsilon}}-\frac{\Sigma_{ij}}{\sigma_i\sigma_j}=\frac{\left(\Sigma_{ij}+\epsilon\right)\sigma_i\sigma_j-\Sigma_{ij}\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}}{\sigma_i\sigma_j\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}}\\=&\frac{\Sigma_{ij}\cdot\left(\sigma_i\sigma_j-\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}\right)+\epsilon\sigma_i\sigma_j}{\sigma_i\sigma_j\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}}\\=&\Delta_{ij}\cdot\frac{\epsilon(\sigma_i^2+\sigma_j^2)-\epsilon^2}{\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}\left(\sigma_i\sigma_j+\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}\right)}+\frac{\epsilon}{\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}}\\\le&\Delta_{ij}\cdot\frac{\epsilon(\sigma_i^2+\sigma_j^2)}{\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}\left(\sigma_i\sigma_j+\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}\right)}+\frac{\epsilon}{\sqrt{\sigma_i^2-\epsilon}\sqrt{\sigma_j^2-\epsilon}}\\\le&\Delta_{ij}\cdot\frac{2\epsilon\sigma_{\min}^2}{\sqrt{\sigma_{\min}^2-\epsilon}\sqrt{\sigma_{\min}^2-\epsilon}\left(\sigma_{\min}^2+\sqrt{\sigma_{\min}^2-\epsilon}\sqrt{\sigma_{\min}^2-\epsilon}\right)}+\frac{\epsilon}{\sqrt{\sigma_{\min}^2-\epsilon}\sqrt{\sigma_{\min}^2-\epsilon}}\\=&\Delta_{ij}\cdot\frac{2\epsilon\sigma_{\min}^2}{\left(\sigma_{\min}^2-\epsilon\right)\left(2\sigma_{\min}^2-\epsilon\right)}+\frac{\epsilon}{\sigma_{\min}^2-\epsilon}.\end{aligned}\tag{E.2}$$

Combining (E.1) and (E.2), we see

$$|\hat{\Delta}_{ij}-\Delta_{ij}| \le \Delta_{ij}\cdot \frac{2\epsilon\sigma_{\min}^2}{(\sigma_{\min}^2-\epsilon)\,(2\sigma_{\min}^2-\epsilon)}+\frac{\epsilon}{\sigma_{\min}^2-\epsilon}.$$

For the case of $\Delta_{ij}\le 0$, one can similarly establish

$$|\hat{\Delta}_{ij}-\Delta_{ij}| \le |\Delta_{ij}|\cdot \frac{2\epsilon\sigma_{\min}^2}{(\sigma_{\min}^2-\epsilon)\,(2\sigma_{\min}^2-\epsilon)}+\frac{\epsilon}{\sigma_{\min}^2-\epsilon}.$$

Hence,

$$\begin{aligned}\sum_{1\le j\le p}|\hat{\Delta}_{ij}-\Delta_{ij}| = \sum_{1\le j\le p, j\ne i}|\hat{\Delta}_{ij}-\Delta_{ij}| &\le \frac{2\epsilon\sigma_{\min}^2}{(\sigma_{\min}^2-\epsilon)\,(2\sigma_{\min}^2-\epsilon)}\cdot\sum_{1\le j\le p, j\ne i}|\Delta_{ij}|+\frac{(p-1)\epsilon}{\sigma_{\min}^2-\epsilon}\\ &\le \frac{2\epsilon\sigma_{\min}^2(p-1)\rho}{(\sigma_{\min}^2-\epsilon)\,(2\sigma_{\min}^2-\epsilon)}+\frac{(p-1)\epsilon}{\sigma_{\min}^2-\epsilon}.\end{aligned}$$

Finally,

$$\|\hat{\Delta}-\Delta\|_\infty \le \frac{2\epsilon\sigma_{\min}^2(p-1)\rho}{(\sigma_{\min}^2-\epsilon)\,(2\sigma_{\min}^2-\epsilon)}+\frac{(p-1)\epsilon}{\sigma_{\min}^2-\epsilon}=g_\Sigma(\epsilon). \qquad \square$$

LEMMA E.2. *Let A and B be invertible $p\times p$ matrices satisfying $\|I-A^{-1}B\|<1$, where I is an $p\times p$ identity matrix. Then,*

$$\|A^{-1}-B^{-1}\| \le \frac{\|A^{-1}\|^2\|A-B\|}{1-\|A^{-1}\|\|A-B\|}.$$

Here, $\|\cdot\|$ is any specific sub-multiplicative matrix norm.

Proof of Lemma E.2. Since $\|I-A^{-1}B\|<1$, we have $B^{-1}A=(A^{-1}B)^{-1}=\sum_{n=0}^\infty(I-A^{-1}B)^n$. Thus, $B^{-1}=\sum_{n=0}^\infty(I-A^{-1}B)^nA^{-1}$. Hence,

$$\begin{aligned}\|A^{-1}-B^{-1}\| &= \left\|I-\sum_{n=0}^\infty(I-A^{-1}B)^nA^{-1}\right\| = \left\|\sum_{n=1}^\infty(I-A^{-1}B)^nA^{-1}\right\|\\ &\le \|A^{-1}\|\cdot\sum_{n=1}^\infty\|I-A^{-1}B\|^n = \|A^{-1}\|\cdot\frac{\|I-A^{-1}B\|}{1-\|I-A^{-1}B\|} = \|A^{-1}\|\cdot\frac{\|A^{-1}(A-B)\|}{1-\|A^{-1}(A-B)\|}\\ &\le \|A^{-1}\|\cdot\frac{\|A^{-1}\|\cdot\|A-B\|}{1-\|A^{-1}\|\cdot\|A-B\|} = \frac{\|A^{-1}\|^2\|A-B\|}{1-\|A^{-1}\|\|A-B\|}. \qquad \square\end{aligned}$$

LEMMA E.3. *Let $\mathbf{X}$ be a d-dimensional Brownian motion given by* (7) *or an OU process given by* (8)*. For any $k=1,2,\ldots$, there exists a constant $\lambda_k<\infty$ such that for all $0\le t\le T$ and $i_1,\ldots,i_k\in\{1,2,\ldots,d\}$,*

$$\mathbb{E}\left(S(\mathbf{X})_t^{i_1,\ldots,i_k,I}\right)^4 \le \lambda_k, \tag{E.3}$$

and

$$\mathbb{E}\left(S(\mathbf{X})_t^{i_1,\ldots,i_k,S}\right)^4 \le \lambda_k. \tag{E.4}$$

Proof of Lemma E.3. The proofs for OU process and Brownian motion are similar, and we will focus on the case of the Brownian motion. We first prove (E.3) by induction. Let $\sigma_{\max} = \max_{j=1,\dots,d}\sigma_j$. When $k=1$,

$$\mathbb{E}\left(S(\mathbf{X})_t^{i_1,I}\right)^4 = \mathbb{E}(X_t^{i_1})^4 = 3\sigma_{i_1}^4 t^2 \le 3\sigma_{\max}^4 T^2 =: \lambda_1 < \infty.$$

Now, for $n>1$, assume that (E.3) holds for all $k<n$. Then, for $k=n$, the quadratic variation of the Itô signature component satisfies

$$\begin{aligned}
\mathbb{E}\left(\left[S(\mathbf{X})^{i_1,\dots,i_n,I}\right]_t\right)^2 &= \mathbb{E}\left(\int_0^t \left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},I}\right)^2 \sigma_{i_n}^2 \mathrm{d}s\right)^2 \\
&= \sigma_{i_n}^4 \cdot \mathbb{E}\int_0^t\int_0^t \left(S(\mathbf{X})_w^{i_1,\dots,i_{n-1},I}\right)^2 \left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},I}\right)^2 \mathrm{d}w\mathrm{d}s \\
&= \sigma_{i_n}^4 \cdot \int_0^t\int_0^t \mathbb{E}\left(\left(S(\mathbf{X})_w^{i_1,\dots,i_{n-1},I}\right)^2 \left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},I}\right)^2\right) \mathrm{d}w\mathrm{d}s \\
&\le \sigma_{i_n}^4 \cdot \int_0^t\int_0^t \sqrt{\mathbb{E}\left(S(\mathbf{X})_w^{i_1,\dots,i_{n-1},I}\right)^4 \cdot \mathbb{E}\left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},I}\right)^4}\mathrm{d}w\mathrm{d}s \\
&\le \sigma_{i_n}^4 \cdot \int_0^t\int_0^t \sqrt{\lambda_{n-1}\cdot\lambda_{n-1}}\mathrm{d}w\mathrm{d}s = \lambda_{n-1}\sigma_{i_n}^4 t^2.
\end{aligned}$$

Thus, by the Burkholder–Davis–Gundy inequality, there exists a constant $c<\infty$ such that for all $0\le t\le T$ and $i_1,\dots,i_n \in \{1,2,\dots,d\}$,

$$\mathbb{E}\left(S(\mathbf{X})_t^{i_1,\dots,i_n,I}\right)^4 \le c\cdot\mathbb{E}\left(\left[S(\mathbf{X})^{i_1,\dots,i_n,I}\right]_t\right)^2 \le c\lambda_{n-1}\sigma_{i_n}^4 t^2 \le c\lambda_{n-1}\sigma_{\max}^4 T^2 =: \lambda_n < \infty.$$

This implies that (E.3) holds when $k=n$, which completes the proof of (E.3).

Now we prove (E.4). By the relationship between the Stratonovich integral and the Itô integral, we have

$$\begin{aligned}
S(\mathbf{X})_t^{i_1,\dots,i_k,S} &= \int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{k-1},S} \circ \mathrm{d}X_s^{i_k} \\
&= \int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{k-1},S}\mathrm{d}X_s^{i_k} + \frac{1}{2}\left[S(\mathbf{X})^{i_1,\dots,i_{k-1},S}, X^{i_k}\right]_t,
\end{aligned}$$

where $[A,B]_t$ represents the quadratic covariation between processes A and B from time 0 to t. Furthermore, by properties of the quadratic covariation,

$$\begin{aligned}
\left[S(\mathbf{X})^{i_1,\dots,i_{k-1},S}, X^{i_k}\right]_t &= \int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{k-2},S}\mathrm{d}\left[X^{i_{k-1}}, X^{i_k}\right]_s \\
&= \rho_{i_{k-1}i_k}\sigma_{i_{k-1}}\sigma_{i_k}\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{k-2},S}\mathrm{d}s.
\end{aligned}$$

Therefore,

$$S(\mathbf{X})_t^{i_1,\dots,i_k,S} = \int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{k-1},S}\mathrm{d}X_s^{i_k} + \frac{1}{2}\rho_{i_{k-1}i_k}\sigma_{i_{k-1}}\sigma_{i_k}\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{k-2},S}\mathrm{d}s. \tag{E.5}$$

We prove (E.4) by induction. Let $\sigma_{\max} = \max_{j=1,\dots,d}\sigma_j$. When $k=1$, we have

$$\mathbb{E}\left(S(\mathbf{X})_t^{i_1,S}\right)^4 = \mathbb{E}\left(S(\mathbf{X})_t^{i_1,I}\right)^4 = \mathbb{E}(X_t^{i_1})^4 = 3\sigma_{i_1}^4 t^2 \le 3\sigma_{\max}^4 T^2 =: \lambda_1 < \infty.$$

When $k=2$, by (E.3) and (E.5), there exists a constant C such that

$$\begin{aligned}\mathbb{E}\left(S(\mathbf{X})_t^{i_1,i_2,S}\right)^4 = \mathbb{E}\left(S(\mathbf{X})_t^{i_1,i_2,I} + \frac{1}{2}\rho_{i_1 i_2}\sigma_{i_1}\sigma_{i_2}t\right)^4 &\le 8\mathbb{E}\left(S(\mathbf{X})_t^{i_1,i_2,I}\right)^4 + \frac{1}{2}\rho_{i_1 i_2}^4\sigma_{i_1}^4\sigma_{i_2}^4 t^4\\ &\le 8C + \frac{1}{2}\sigma_{\max}^8 T^4 =: \lambda_2 < \infty.\end{aligned}$$

For $n>2$, assume that (E.4) holds for all $k<n$. Thus, for $k=n$, we have

$$\begin{aligned}\mathbb{E}\left(\left[\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\mathrm{d}X_s^{i_n}\right]_t\right)^2 &= \mathbb{E}\left(\int_0^t \left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\right)^2\sigma_{i_n}^2\mathrm{d}s\right)^2\\ &= \sigma_{i_n}^4\cdot\mathbb{E}\int_0^t\int_0^t \left(S(\mathbf{X})_w^{i_1,\dots,i_{n-1},S}\right)^2\left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\right)^2\mathrm{d}w\mathrm{d}s\\ &= \sigma_{i_n}^4\cdot\int_0^t\int_0^t \mathbb{E}\left(\left(S(\mathbf{X})_w^{i_1,\dots,i_{n-1},S}\right)^2\left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\right)^2\right)\mathrm{d}w\mathrm{d}s\\ &\le \sigma_{i_n}^4\cdot\int_0^t\int_0^t\sqrt{\mathbb{E}\left(S(\mathbf{X})_w^{i_1,\dots,i_{n-1},S}\right)^4\cdot\mathbb{E}\left(S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\right)^4}\mathrm{d}w\mathrm{d}s\\ &\le \sigma_{i_n}^4\cdot\int_0^t\int_0^t\sqrt{\lambda_{n-1}\cdot\lambda_{n-1}}\mathrm{d}w\mathrm{d}s = \lambda_{n-1}\sigma_{i_n}^4 t^2.\end{aligned}$$

Hence, by the Burkholder–Davis–Gundy inequality, there exists a constant $c<\infty$ such that for all $0\le t\le T$ and $i_1,\dots,i_n\in\{1,2,\dots,d\}$,

$$\mathbb{E}\left(\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\mathrm{d}X_s^{i_n}\right)^4 \le c\cdot\mathbb{E}\left(\left[\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\mathrm{d}X_s^{i_n}\right]_t\right)^2 \le c\lambda_{n-1}\sigma_{i_n}^4 t^2.$$

In addition, we have

$$\begin{aligned}&\mathbb{E}\left(\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}\mathrm{d}s\right)^4\\ =&\mathbb{E}\int_0^t\int_0^t\int_0^t\int_0^t S(\mathbf{X})_w^{i_1,\dots,i_{n-2},S}S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}S(\mathbf{X})_u^{i_1,\dots,i_{n-2},S}S(\mathbf{X})_v^{i_1,\dots,i_{n-2},S}\mathrm{d}w\mathrm{d}s\mathrm{d}u\mathrm{d}v\\ =&\int_0^t\int_0^t\int_0^t\int_0^t \mathbb{E}\left(S(\mathbf{X})_w^{i_1,\dots,i_{n-2},S}S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}S(\mathbf{X})_u^{i_1,\dots,i_{n-2},S}S(\mathbf{X})_v^{i_1,\dots,i_{n-2},S}\right)\mathrm{d}w\mathrm{d}s\mathrm{d}u\mathrm{d}v\\ \le&\int_0^t\int_0^t\int_0^t\int_0^t \frac{1}{4}\Big(\mathbb{E}\left(S(\mathbf{X})_w^{i_1,\dots,i_{n-2},S}\right)^4 + \mathbb{E}\left(S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}\right)^4\\ &\qquad + \mathbb{E}\left(S(\mathbf{X})_u^{i_1,\dots,i_{n-2},S}\right)^4 + \mathbb{E}\left(S(\mathbf{X})_v^{i_1,\dots,i_{n-2},S}\right)^4\Big)\mathrm{d}w\mathrm{d}s\mathrm{d}u\mathrm{d}v\\ \le&\int_0^t\int_0^t\int_0^t\int_0^t \frac{1}{4}\cdot 4\lambda_{n-2}\mathrm{d}w\mathrm{d}s\mathrm{d}u\mathrm{d}v = \lambda_{n-2}t^4.\end{aligned}$$

Therefore, by (E.5),

$$\mathbb{E}\left(S(\mathbf{X})_t^{i_1,\dots,i_n,S}\right)^4 \le 8\mathbb{E}\left(\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\mathrm{d}X_s^{i_n}\right)^4 + \frac{1}{2}\rho_{i_{n-1}i_n}^4\sigma_{i_{n-1}}^4\sigma_{i_n}^4\,\mathbb{E}\left(\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}\mathrm{d}s\right)^4$$

$$\leq 8c\lambda_{n-1}\sigma_{i_n}^4 t^2 + \frac{1}{2}\rho_{i_{n-1}i_n}^4\sigma_{i_{n-1}}^4\sigma_{i_n}^4\lambda_{n-2}t^4$$
$$\leq 8c\lambda_{n-1}\sigma_{\max}^4 T^2 + \frac{1}{2}\sigma_{\max}^8\lambda_{n-2}T^4 =: \lambda_n < \infty.$$

This implies that (E.4) holds when $k = n$, which completes the proof. □

LEMMA E.4. *For the Lasso regression given by* (C.1) *and* (C.2), *assume that conditions (i) and (ii) in Theorem C.1 hold. Then,*

$$\mathbb{P}\left(\Lambda_{\min}(\hat{\Delta}_{A^*A^*}) \geq \frac{1}{2}C_{\min}\right) \geq 1 - \frac{4p^4\sigma_{\max}^4(\sigma_{\min}^4 + K)}{N\xi^2\sigma_{\min}^4}$$

holds with $\xi = g_\Sigma^{-1}\left(\frac{C_{\min}}{2\sqrt{p}}\right) > 0$, *and the definition of* $g_\Sigma(\cdot)$ *and other notations the same as in Theorem C.1.*

Proof of Lemma E.4. Condition (ii) implies that

$$\mathbb{E}\left[\frac{X_i^4}{\Sigma_{ii}^2}\right] \leq \mathbb{E}\left[\frac{X_i^4}{\sigma_{\min}^4}\right] \leq \frac{K}{\sigma_{\min}^4}.$$

Hence, by Ravikumar et al. (2011, Lemma 2), for any $i, j \in \{1, \ldots, p\}$,

$$\mathbb{P}\left(\left|\hat{\Sigma}_{ij} - \Sigma_{ij}\right| > \frac{\xi}{p}\right) \leq \frac{4p^2\sigma_{\max}^4(1 + \frac{K}{\sigma_{\min}^4})}{N\xi^2}.$$

Thus,

$$\mathbb{P}\left(\sum_{j=1}^p \left|\hat{\Sigma}_{ij} - \Sigma_{ij}\right| \leq \xi\right) \geq 1 - \frac{4p^3\sigma_{\max}^4(1 + \frac{K}{\sigma_{\min}^4})}{N\xi^2},$$

which further implies that

$$\mathbb{P}\left(\|\hat{\Sigma} - \Sigma\|_\infty \leq \xi\right) \geq 1 - \frac{4p^4\sigma_{\max}^4(1 + \frac{K}{\sigma_{\min}^4})}{N\xi^2} = 1 - \frac{4p^4\sigma_{\max}^4(\sigma_{\min}^4 + K)}{N\xi^2\sigma_{\min}^4}. \tag{E.6}$$

Therefore, by Lemma E.1, we have

$$\mathbb{P}\left(\|\hat{\Delta} - \Delta\|_2 \leq \frac{C_{\min}}{2}\right) \geq \mathbb{P}\left(\|\hat{\Delta} - \Delta\|_\infty \leq \frac{C_{\min}}{2\sqrt{p}}\right) \geq \mathbb{P}\left(\|\hat{\Sigma} - \Sigma\|_\infty \leq g_\Sigma^{-1}\left(\frac{C_{\min}}{2\sqrt{p}}\right)\right)$$
$$\geq \mathbb{P}\left(\|\hat{\Sigma} - \Sigma\|_\infty \leq \xi\right) \geq 1 - \frac{4p^4\sigma_{\max}^4(\sigma_{\min}^4 + K)}{N\xi^2\sigma_{\min}^4}.$$

Now, whenever $\|\hat{\Delta} - \Delta\|_2 \leq \frac{C_{\min}}{2}$ holds, we have

$$\|I - \Delta_{A^*A^*}^{-1}\hat{\Delta}_{A^*A^*}\|_2 \leq \|\Delta_{A^*A^*}^{-1}\|_2 \cdot \|\Delta_{A^*A^*} - \hat{\Delta}_{A^*A^*}\|_2$$
$$= \frac{1}{C_{\min}} \cdot \|\Delta_{A^*A^*} - \hat{\Delta}_{A^*A^*}\|_2 \leq \frac{1}{C_{\min}} \cdot \|\hat{\Delta} - \Delta\|_2 \leq \frac{1}{2} < 1,$$

which implies that

$$\begin{aligned}\|\hat{\Delta}_{A^*A^*}^{-1}\|_2 &\le \|\hat{\Delta}_{A^*A^*}^{-1} - \Delta_{A^*A^*}^{-1}\|_2 + \|\Delta_{A^*A^*}^{-1}\|_2 = \|\hat{\Delta}_{A^*A^*}^{-1} - \Delta_{A^*A^*}^{-1}\|_2 + \frac{1}{C_{\min}} \\ &\le \frac{\frac{1}{C_{\min}^2}\cdot\frac{C_{\min}}{2}}{1-\frac{1}{C_{\min}}\cdot\frac{C_{\min}}{2}} + \frac{1}{C_{\min}} = \frac{2}{C_{\min}},\end{aligned}$$

where the second inequality holds because of Lemma E.2. Therefore, $\|\hat{\Delta}-\Delta\|_2 \le \frac{C_{\min}}{2}$ implies

$$\Lambda_{\min}(\hat{\Delta}_{A^*A^*}) = \frac{1}{\|\hat{\Delta}_{A^*A^*}^{-1}\|_2} \ge \frac{1}{2}C_{\min}.$$

Thus,

$$\mathbb{P}\left(\Lambda_{\min}(\hat{\Delta}_{A^*A^*}) \ge \frac{1}{2}C_{\min}\right) \ge 1 - \frac{4p^4\sigma_{\max}^4(\sigma_{\min}^4+K)}{N\xi^2\sigma_{\min}^4},$$

which completes the proof. □

LEMMA E.5. *For the Lasso regression given by* (C.1) *and* (C.2)*, assume that conditions (i) and (ii) in Theorem C.1 hold. Then,*

$$\mathbb{P}\left(\left\|\hat{\Delta}_{A^{*c}A^*}\hat{\Delta}_{A^*A^*}^{-1}\right\|_\infty \le 1-\frac{\gamma}{2}\right) \ge 1 - \frac{4p^4\sigma_{\max}^4(\sigma_{\min}^4+K)}{N\xi^2\sigma_{\min}^4}$$

holds with $\xi = g_\Sigma^{-1}\left(\frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}\right) > 0$*, and the definition of* $g_\Sigma(\cdot)$ *and other notations the same as in Theorem C.1.*

Proof of Lemma E.5. By Lemma E.1,

$$\begin{aligned}\mathbb{P}\left(\|\hat{\Delta}-\Delta\|_\infty \le \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}\right) &\ge \mathbb{P}\left(\|\hat{\Sigma}-\Sigma\|_\infty \le g_\Sigma^{-1}\left(\frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}\right)\right) \\ &\ge \mathbb{P}\left(\|\hat{\Sigma}-\Sigma\|_\infty \le \xi\right) \ge 1 - \frac{4p^4\sigma_{\max}^4(\sigma_{\min}^4+K)}{N\xi^2\sigma_{\min}^4},\end{aligned}$$

where the last inequality holds by (E.6). Whenever $\|\hat{\Delta}-\Delta\|_\infty \le \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}$ holds, we have

$$\begin{aligned}\|\hat{\Delta}_{A^{*c}A^*} - \Delta_{A^{*c}A^*}\|_\infty &\le \|\hat{\Delta}-\Delta\|_\infty \le \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}, \\ \|\hat{\Delta}_{A^*A^*} - \Delta_{A^*A^*}\|_\infty &\le \|\hat{\Delta}-\Delta\|_\infty \le \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}, \\ \|\hat{\Delta}_{A^{*c}A^*}\|_\infty &\le \|\Delta_{A^{*c}A^*}\|_\infty + \|\hat{\Delta}_{A^{*c}A^*} - \Delta_{A^{*c}A^*}\|_\infty \le \alpha + \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)},\end{aligned}$$

and

$$\begin{aligned}\|I - \Delta_{A^*A^*}^{-1}\hat{\Delta}_{A^*A^*}\|_\infty &\le \|\Delta_{A^*A^*}^{-1}\|_\infty \cdot \|\Delta_{A^*A^*} - \hat{\Delta}_{A^*A^*}\|_\infty = \zeta\cdot\|\Delta_{A^*A^*} - \hat{\Delta}_{A^*A^*}\|_\infty \\ &\le \zeta\cdot\|\hat{\Delta}-\Delta\|_\infty \le \frac{\gamma}{2+2\alpha\zeta+\gamma} < 1.\end{aligned}$$

Therefore, applying Lemma E.2 yields

$$\|\hat{\Delta}_{A^*A^*}^{-1} - \Delta_{A^*A^*}^{-1}\|_\infty \le \frac{\|\Delta_{A^*A^*}^{-1}\|_\infty^2 \cdot \|\hat{\Delta}_{A^*A^*} - \Delta_{A^*A^*}\|_\infty}{1 - \|\Delta_{A^*A^*}^{-1}\|_\infty \cdot \|\hat{\Delta}_{A^*A^*} - \Delta_{A^*A^*}\|_\infty} \le \frac{\zeta^2 \cdot \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}}{1 - \zeta \cdot \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}} = \frac{\gamma\zeta}{2+2\alpha\zeta},$$

which further implies that

$$\|\hat{\Delta}_{A^*A^*}^{-1}\|_\infty \le \|\Delta_{A^*A^*}^{-1}\|_\infty + \|\hat{\Delta}_{A^*A^*}^{-1} - \Delta_{A^*A^*}^{-1}\|_\infty \le \zeta + \frac{\gamma\zeta}{2+2\alpha\zeta} = \frac{\zeta(2+2\alpha\zeta+\gamma)}{2+2\alpha\zeta}.$$

Hence, $\|\hat{\Delta} - \Delta\|_\infty \le \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}$ implies that

$$\begin{aligned}
\left\|\hat{\Delta}_{A^{*c}A^*}\hat{\Delta}_{A^*A^*}^{-1}\right\|_\infty \le & \left\|\hat{\Delta}_{A^{*c}A^*}\hat{\Delta}_{A^*A^*}^{-1} - \hat{\Delta}_{A^{*c}A^*}\Delta_{A^*A^*}^{-1}\right\|_\infty + \\
& \left\|\hat{\Delta}_{A^{*c}A^*}\Delta_{A^*A^*}^{-1} - \Delta_{A^{*c}A^*}\Delta_{A^*A^*}^{-1}\right\|_\infty + \left\|\Delta_{A^{*c}A^*}\Delta_{A^*A^*}^{-1}\right\|_\infty \\
\le & \|\hat{\Delta}_{A^{*c}A^*}\|_\infty \cdot \|\hat{\Delta}_{A^*A^*}^{-1} - \Delta_{A^*A^*}^{-1}\|_\infty + \\
& \|\Delta_{A^*A^*}^{-1}\|_\infty \cdot \|\hat{\Delta}_{A^{*c}A^*} - \Delta_{A^{*c}A^*}\|_\infty + 1 - \gamma \\
\le & \left(\alpha + \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}\right) \cdot \frac{\gamma\zeta}{2+2\alpha\zeta} + \zeta \cdot \frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)} + 1 - \gamma \\
= & 1 - \frac{\gamma}{2}.
\end{aligned}$$

Therefore,

$$\mathbb{P}\left(\left\|\hat{\Delta}_{A^{*c}A^*}\hat{\Delta}_{A^*A^*}^{-1}\right\|_\infty \le 1 - \frac{\gamma}{2}\right) \ge 1 - \frac{4p^4\sigma_{\max}^4(\sigma_{\min}^4 + K)}{N\xi^2\sigma_{\min}^4}. \qquad \square$$

E.2. Proofs

Proof of Theorem 2. For any $\theta > 0$,

$$\begin{aligned}
\mathbb{P}(|L_a - L_b| > \eta) \ge & \mathbb{P}\left(\left|\sum_{i=1}^p c_i S_i\right| > \eta, \|S\|_2 < \theta\sqrt{p\|\Sigma\|_2}\right) \\
= & \mathbb{P}\left(\left|\sum_{i=1}^p c_i S_i\right| > \eta \middle| \|S\|_2 < \theta\sqrt{p\|\Sigma\|_2}\right) \cdot \mathbb{P}\left(\|S\|_2 < \theta\sqrt{p\|\Sigma\|_2}\right). \qquad \text{(E.7)}
\end{aligned}$$

By Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left(\|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right) \ge 1 - \mathbb{P}\left(\|S\|_2 > \theta\sqrt{p\|\Sigma\|_2}\right) \ge 1 - \frac{\mathbb{E}\|S\|_2}{\theta\sqrt{p\|\Sigma\|_2}} \\
\ge 1 - \frac{\sqrt{\mathbb{E}\|S\|_2^2}}{\theta\sqrt{p\|\Sigma\|_2}} = 1 - \frac{\sqrt{\mathrm{tr}(\Sigma)}}{\theta\sqrt{p\|\Sigma\|_2}} \ge 1 - \frac{\sqrt{p\|\Sigma\|_2}}{\theta\sqrt{p\|\Sigma\|_2}} = 1 - \frac{1}{\theta}. \qquad \text{(E.8)}
\end{aligned}$$

In addition, applying Markov's inequality to $X = \theta\|C\|_\infty p\sqrt{\|\Sigma\|_2} - |\sum_{i=1}^p c_i S_i|$ for a sufficiently small $\eta > 0$ yields

$$\begin{aligned}
\mathbb{P}\left(\left|\sum_{i=1}^p c_i S_i\right| \le \eta \middle| \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right) = \mathbb{P}\left(X \ge \theta\|C\|_\infty p\sqrt{\|\Sigma\|_2} - \eta \middle| \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right) \\
\le \frac{\mathbb{E}\left[X \middle| \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right]}{\theta\|C\|_\infty p\sqrt{\|\Sigma\|_2} - \eta} = \frac{\theta\|C\|_\infty p\sqrt{\|\Sigma\|_2} - \mathbb{E}\left[|\sum_{i=1}^p c_i S_i| \middle| \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right]}{\theta\|C\|_\infty p\sqrt{\|\Sigma\|_2} - \eta}.
\end{aligned}$$

Hence,

$$\mathbb{P}\left(\left|\sum_{i=1}^{p} c_i S_i\right| > \eta \,\middle|\, \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right) \ge \frac{\mathbb{E}\left[|\sum_{i=1}^{p} c_i S_i| \,\middle|\, \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right] - \eta}{\theta\|C\|_\infty p\sqrt{\|\Sigma\|_2} - \eta}. \tag{E.9}$$

Under the condition of $\|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}$,

$$\left|\sum_{i=1}^{p} c_i S_i\right| \le \|C\|_\infty \cdot \|S\|_1 \le \|C\|_\infty \cdot \sqrt{p}\|S\|_2 \le \theta\|C\|_\infty p\sqrt{\|\Sigma\|_2},$$

and by multiplying both sides of the above inequality by $|\sum_{i=1}^{p} c_i S_i|$ and taking expectations, we obtain

$$\mathbb{E}\left[\left|\sum_{i=1}^{p} c_i S_i\right| \,\middle|\, \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right] \ge \frac{\mathbb{E}\left[(\sum_{i=1}^{p} c_i S_i)^2 \,\middle|\, \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right]}{\theta\|C\|_\infty p\sqrt{\|\Sigma\|_2}}. \tag{E.10}$$

Thus, by combining (E.7), (E.8), (E.9), and (E.10), we obtain

$$\mathbb{P}(|L_a - L_b| > \eta) \ge \left(1 - \frac{1}{\theta}\right) \cdot \frac{\frac{\mathbb{E}\left[(\sum_{i=1}^{p} c_i S_i)^2 \middle| \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right]}{\theta\|C\|_\infty p\sqrt{\|\Sigma\|_2}} - \eta}{\theta\|C\|_\infty p\sqrt{\|\Sigma\|_2} - \eta}.$$

When θ is sufficiently large, $\mathbb{E}\left[(\sum_{i=1}^{p} c_i S_i)^2 \middle| \|S\|_2 \le \theta\sqrt{p\|\Sigma\|_2}\right] > 0$ because the distribution of S is non-degenerate. Therefore, (5) holds. Furthermore, (6) is a direct result of the triangle inequality, which completes the proof. □

Proof of Proposition 1. First, we have

$$\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_n,I}\right] = \mathbb{E}\left[\int_0^t S(\mathbf{X})_s^{i_1,\dots,i_{n-1}} \mathrm{d}X_s^{i_n}\right] = 0 \tag{E.11}$$

Next we prove $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_n,I} S(\mathbf{X})_t^{j_1,\dots,j_m,I}\right] = 0$ for $m \ne n$ by induction. Without loss of generality, we assume that $m > n$. When $n = 1$, for any $m > 1$, we have

$$\begin{aligned}\mathbb{E}\left[S(\mathbf{X})_t^{i_1,I} S(\mathbf{X})_t^{j_1,\dots,j_m,I}\right] &= \mathbb{E}\left[\left(\int_0^t \mathrm{d}X_s^{i_1}\right)\left(\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{m-1}} \mathrm{d}X_s^{j_m}\right)\right] \\ &= \int_0^t \mathbb{E}\left[S(\mathbf{X})_s^{j_1,\dots,j_{m-1},I}\right] \rho_{i_1 j_m} \sigma_{i_1} \sigma_{j_m} \mathrm{d}s = 0,\end{aligned}$$

where the second equality uses the Itô isometry and the third equality uses (E.11). Now assume $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_n,I} S(\mathbf{X})_t^{j_1,\dots,j_m,I}\right] = 0$ for any $m > n$. Then,

$$\begin{aligned}&\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_{n+1},I} S(\mathbf{X})_t^{j_1,\dots,j_{m+1},I}\right] \\ =&\mathbb{E}\left[\left(\int_0^t S(\mathbf{X})_t^{i_1,\dots,i_n,I} \mathrm{d}X_s^{i_{n+1}}\right)\left(\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_m,I} \mathrm{d}X_s^{j_{m+1}}\right)\right] \\ =&\int_0^t \mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_k,I} S(\mathbf{X})_s^{j_1,\dots,j_m,I}\right] \rho_{i_{n+1} j_{m+1}} \sigma_{i_{n+1}} \sigma_{j_{m+1}} \mathrm{d}s = 0.\end{aligned}$$

This proves $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}S(\mathbf{X})_t^{j_1,\ldots,j_m,I}\right]=0$.

We finally prove $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}S(\mathbf{X})_t^{j_1,\ldots,j_n,I}\right]=\frac{t^n}{n!}\prod_{k=1}^n \rho_{i_kj_k}\sigma_{i_k}\sigma_{j_k}$ by induction. When $n=1$,

$$\begin{aligned}&\mathbb{E}\left[S(\mathbf{X})_t^{i_1,I}S(\mathbf{X})_t^{j_1,I}\right]\\ =&\mathbb{E}\left[\left(\int_0^t \mathrm{d}X_s^{i_1}\right)\left(\int_0^t \mathrm{d}X_s^{j_1}\right)\right]=\int_0^t \rho_{i_1j_1}\sigma_{i_1}\sigma_{j_1}\mathrm{d}s=t\rho_{i_1j_1}\sigma_{i_1}\sigma_{j_1}.\end{aligned}$$

Now, assume $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}S(\mathbf{X})_t^{j_1,\ldots,j_n,I}\right]=\frac{t^n}{n!}\prod_{k=1}^n \rho_{i_kj_k}\sigma_{i_k}\sigma_{j_k}$, then

$$\begin{aligned}&\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_{n+1},I}S(\mathbf{X})_t^{j_1,\ldots,j_{n+1},I}\right]\\ =&\mathbb{E}\left[\left(\int_0^t S(\mathbf{X})_s^{i_1,\ldots,i_n,I}\mathrm{d}X_s^{i_{n+1}}\right)\left(\int_0^t S(\mathbf{X})_s^{j_1,\ldots,j_n,I}\mathrm{d}X_s^{j_{n+1}}\right)\right]\\ =&\int_0^t \mathbb{E}\left[S(\mathbf{X})_s^{i_1,\ldots,i_n,I}S(\mathbf{X})_s^{j_1,\ldots,j_n,I}\right]\rho_{i_{n+1}j_{n+1}}\sigma_{i_{n+1}}\sigma_{j_{n+1}}\mathrm{d}s\\ =&\int_0^t \left(\frac{s^n}{n!}\prod_{k=1}^n \rho_{i_kj_k}\sigma_{i_k}\sigma_{j_k}\right)\rho_{i_{n+1}j_{n+1}}\sigma_{i_{n+1}}\sigma_{j_{n+1}}\mathrm{d}s=\frac{t^{n+1}}{(n+1)!}\prod_{k=1}^{n+1}\rho_{i_kj_k}\sigma_{i_k}\sigma_{j_k}.\end{aligned}$$

Therefore, $\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}S(\mathbf{X})_t^{j_1,\ldots,j_n,I}\right]=\frac{t^n}{n!}\prod_{k=1}^n \rho_{i_kj_k}\sigma_{i_k}\sigma_{j_k}$. □

Proof of Theorem 3. By Proposition 1, for any n,

$$\frac{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}S(\mathbf{X})_t^{j_1,\ldots,j_n,I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}\right]^2\mathbb{E}\left[S(\mathbf{X})_t^{j_1,\ldots,j_n,I}\right]^2}}=\frac{\frac{t^n}{n!}\prod_{k=1}^n \rho_{i_kj_k}\sigma_{i_k}\sigma_{j_k}}{\sqrt{\frac{t^n}{n!}\prod_{k=1}^n \sigma_{i_k}\sigma_{i_k}\cdot\frac{t^n}{n!}\prod_{k=1}^n \sigma_{j_k}\sigma_{j_k}}}=\prod_{k=1}^n \rho_{i_kj_k},$$

implying

$$\frac{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,I}S(\mathbf{X})_t^{j_1,I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,I}\right]^2\mathbb{E}\left[S(\mathbf{X})_t^{j_1,I}\right]^2}}=\rho_{i_1j_1}$$

and

$$\frac{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}S(\mathbf{X})_t^{j_1,\ldots,j_n,I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}\right]^2\mathbb{E}\left[S(\mathbf{X})_t^{j_1,\ldots,j_n,I}\right]^2}}=\rho_{i_nj_n}\cdot\frac{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_{n-1},I}S(\mathbf{X})_t^{j_1,\ldots,j_{n-1},I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_{n-1},I}\right]^2\mathbb{E}\left[S(\mathbf{X})_t^{j_1,\ldots,j_{n-1},I}\right]^2}}.$$

This proves the Kronecker product structure given by (10).

Proposition 1 also implies that, for any $m\neq n$,

$$\frac{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}S(\mathbf{X})_t^{j_1,\ldots,j_m,I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_n,I}\right]^2\mathbb{E}\left[S(\mathbf{X})_t^{j_1,\ldots,j_m,I}\right]^2}}=0.$$

This proves that Itô signatures of different orders are uncorrelated and, therefore, the correlation matrix is block diagonal. □

Proof of Proposition 2. Equations

$$\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_{2n-1},S}\right]=0$$

and

$$\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_{2n},S}S(\mathbf{X})_t^{j_1,\dots,j_{2m-1},S}\right]=0$$

can be proven using a similar approach to the proof of Theorem 4. Now we prove

$$\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_{2n},S}\right]=\frac{1}{2^n}\frac{t^n}{n!}\prod_{k=1}^{n}\rho_{i_{2k-1}i_{2k}}\prod_{k=1}^{2n}\sigma_{i_k} \tag{E.12}$$

by induction. If $n=0$, (E.12) holds because of (B.3) in Proposition B.1. Now we assume that (E.12) holds for $n=j$. Then, when $n=j+1$, by Proposition B.1,

$$\begin{aligned}\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_{2(j+1)},S}\right]=&\frac{1}{2}\rho_{i_{2j+1}i_{2j+2}}\sigma_{i_{2j+1}}\sigma_{i_{2j+2}}\int_0^t\frac{1}{2^j}\frac{s^j}{j!}\prod_{k=1}^{j}\rho_{i_{2k-1}i_{2k}}\prod_{k=1}^{2j}\sigma_{i_k}\mathrm{d}s\\=&\frac{1}{2^{j+1}}\frac{t^{j+1}}{(j+1)!}\prod_{k=1}^{j+1}\rho_{i_{2k-1}i_{2k}}\prod_{k=1}^{2(j+1)}\sigma_{i_k}.\end{aligned}$$

Therefore, (E.12) holds. □

Proof of Theorem 4. For the Stratonovich signature of a Brownian motion, this is a direct corollary of Proposition 2. For both the Itô and Stratonovich signatures of an OU process, we only need to prove that, for an odd number m and an even number n, we have

$$\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_m}S(\mathbf{X})_t^{j_1,\dots,j_n}\right]=0$$

for any $i_1,\dots,i_m$ and $j_1,\dots,j_n$ taking values in $\{1,2,\dots,d\}$. Here the signatures can be defined in the sense of either Itô or Stratonovich.

Consider the reflected OU process, $\check{\mathbf{X}}_t=-\mathbf{X}_t$. By definition, $\check{\mathbf{X}}_t$ is also an OU process with the same mean reversion parameter. Therefore, the signatures of $\check{\mathbf{X}}_t$ and $\mathbf{X}_t$ should have the same distribution. In particular, we have

$$\mathbb{E}\left[S(\check{\mathbf{X}})_t^{i_1,\dots,i_m}S(\check{\mathbf{X}})_t^{j_1,\dots,j_n}\right]=\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\dots,i_m}S(\mathbf{X})_t^{j_1,\dots,j_n}\right]. \tag{E.13}$$

Now we consider the definition of the signature

$$S(\mathbf{X})_t^{i_1,\dots,i_m}=\int_{0<t_1<\dots<t_m<t}\mathrm{d}X_{t_1}^{i_1}\cdots\mathrm{d}X_{t_k}^{i_m},$$

where the integral can be defined in the sense of either Itô or Stratonovich. We therefore have

$$\begin{aligned}S(\check{\mathbf{X}})_t^{i_1,\dots,i_m}=S(-\mathbf{X})_t^{i_1,\dots,i_m}&=\int_{0<t_1<\dots<t_m<t}\mathrm{d}(-X_{t_1}^{i_1})\cdots\mathrm{d}(-X_{t_k}^{i_m})\\&=(-1)^m\int_{0<t_1<\dots<t_m<t}\mathrm{d}X_{t_1}^{i_1}\cdots\mathrm{d}X_{t_k}^{i_m}=(-1)^mS(\mathbf{X})_t^{i_1,\dots,i_m}.\end{aligned}$$

Similarly, we have

$$S(\check{\mathbf{X}})_t^{j_1,\dots,j_n}=(-1)^nS(\mathbf{X})_t^{j_1,\dots,j_n}.$$

Therefore,

$$= -\mathbb{E}\left[S(\mathbf{X})_t^{i_1,\ldots,i_m} S(\mathbf{X})_t^{j_1,\ldots,j_n}\right],$$

and combining this with (E.13) leads to the result. $\square$

Proof of Theorem 5. Note that, for a block diagonal correlation matrix Δ, the irrepresentable conditions given by Definition 4 hold if and only if they hold for each block. Thus, the first necessary and sufficient condition for the irrepresentable conditions holds due to Theorem 3. The second sufficient condition holds due to (E.18) in the proof of Theorem 7 because $\rho < \frac{1}{2q_{\max}-1}$ implies $\gamma > 0$. This completes the proof. $\square$

Proof of Theorem 6. Note that, for a block diagonal correlation matrix Δ, the irrepresentable conditions given by Definition 4 hold if and only if they hold for each block. Thus, this result holds because of Theorem 4. $\square$

Proof of Theorem 7. We use Theorem C.1 to obtain the result. Lemma E.3 implies that the finite fourth-moment condition for the Itô signature of Brownian motion holds. By Theorem 3, the correlation matrix of the Itô signature of Brownian motion exhibits a block-diagonal structure

$$\Delta^1 = \operatorname{diag}\{\Omega_0, \Omega_1, \Omega_2, \ldots, \Omega_K\},$$

whose diagonal blocks Ω_k are given by

$$\Omega_k = \underbrace{\Omega \otimes \Omega \otimes \cdots \otimes \Omega}_{k}, \quad k = 1, 2, \ldots, K,$$

and $\Omega_0 = 1$. Because $\rho = \max_{i \neq j} |\rho_{ij}|$, for any $k = 1, 2, \ldots, K$, we have $\max_{i \neq j} \{|\Omega_{k,ij}|\} \leq \rho$, where $\Omega_{k,ij}$ is the (i,j)-entry of Ω_k. Hence,

$$\|\Omega_{k,A_k^{*c}A_k^*}\|_\infty \leq \#A_k^* \cdot \rho \leq q_{\max}\rho \tag{E.14}$$

and

$$\|\Omega_{k,A_k^*A_k^*}\|_2 \leq \sqrt{\#A_k^*} \cdot \|\Omega_{k,A_k^*A_k^*}\|_\infty \leq \sqrt{\#A_k^*} \cdot (1 + (\#A_k^* - 1)\rho) \leq \sqrt{q_{\max}}\,(1 + (q_{\max} - 1)\rho). \tag{E.15}$$

Let $X = (X_1, \ldots, X_{\#A_k^*})^\top \in \mathbb{R}^{\#A_k^*}$ be any vector of constants satisfying $\|X\|_\infty = 1$. Without loss of generality, we assume $X_1 = 1$. Therefore,

$$\begin{aligned}
\|\Omega_{k,A^*A^*}X\|_\infty &\geq |(\Omega_{k,A^*A^*})_{1,1}X_1 + \cdots + (\Omega_{k,A^*A^*})_{1,\#A_k^*}X_{\#A_k^*}| \\
&= |1 + (\Omega_{k,A^*A^*})_{1,2}X_2 + \cdots + (\Omega_{k,A^*A^*})_{1,\#A_k^*}X_{\#A_k^*}| \\
&\geq 1 - |(\Omega_{k,A^*A^*})_{1,2}X_2| - \cdots - |(\Omega_{k,A^*A^*})_{1,\#A_k^*}X_{\#A_k^*}| \\
&\geq 1 - (\#A_k^* - 1)\rho \geq 1 - (q_{\max} - 1)\rho,
\end{aligned}$$

which implies that

$$\|\Omega_{k,A^*A^*}^{-1}\|_\infty = \frac{1}{\min_{\|X\|_\infty=1}\|\Omega_{k,A^*A^*}X\|_\infty} \le \frac{1}{1-(q_{\max}-1)\rho}, \tag{E.16}$$

$$\|\Omega_{k,A^{*c}A^*}\Omega_{k,A^*A^*}^{-1}\|_\infty \le \|\Omega_{k,A^{*c}A^*}\|_\infty \cdot \|\Omega_{k,A^*A^*}^{-1}\|_\infty \le \frac{q_{\max}\rho}{1-(q_{\max}-1)\rho}. \tag{E.17}$$

Equations (E.14), (E.15), (E.16), and (E.17) lead to the parameters for Theorem C.1 given by

$$\begin{aligned}
\alpha &= \|\Delta_{A^{*c}A^*}\|_\infty = \max_{1\le k\le K}\|\Omega_{k,A^{*c}A^*}\|_\infty \le q_{\max}\rho,\\
\zeta &= \left\|\Delta_{A^*A^*}^{-1}\right\|_\infty = \max_{1\le k\le K}\left\|\Omega_{k,A^*A^*}^{-1}\right\|_\infty \le \frac{1}{1-(q_{\max}-1)\rho},\\
C_{\min} &= \Lambda_{\min}(\Delta_{A^*A^*}) = \frac{1}{\|\Delta_{A^*A^*}^{-1}\|_2} = \frac{1}{\max_{1\le k\le K}\|\Omega_{k,A^*A^*}^{-1}\|_2}\\
&\qquad \ge \frac{1}{\max_{1\le k\le K}\sqrt{q_{\max}}\|\Omega_{k,A^*A^*}^{-1}\|_\infty} \ge \frac{1-(q_{\max}-1)\rho}{\sqrt{q_{\max}}},\\
\gamma &= \min_{1\le k\le K}\left\{1-\left\|\Omega_{k,A^{*c}A^*}\Omega_{k,A^*A^*}^{-1}\right\|_\infty\right\} \ge \frac{1-(2q_{\max}-1)\rho}{1-(q_{\max}-1)\rho}.
\end{aligned} \tag{E.18}$$

Plugging these into Theorem C.1 leads to the result. □

Proof of Theorem 8. Theorem 4 implies that the correlation structure can be represented by $\mathrm{diag}\{\Psi_{\mathrm{odd}}, \Psi_{\mathrm{even}}\}$. Lemma E.3 implies that the finite fourth-moment condition for Stratonovich signature of Brownian motion holds, while Lemma E.3 implies that the finite fourth-moment condition for both Itô and Stratonovich signature of OU process holds. Combining these with Theorem C.1 leads to the result. □

Proof of Proposition B.1. For any $l, t \ge 0$ and $m, n = 0, 1, \ldots$, define

$$\begin{aligned}
f_{n,m}(l,t) &:= \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_n,S} S(\mathbf{X})_t^{j_1,\ldots,j_m,S}\right],\\
g_{n,m}(l,t) &:= \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_n,S}\int_0^t S(\mathbf{X})_s^{j_1,\ldots,j_{m-1},S}\mathrm{d}X_s^{j_m}\right].
\end{aligned}$$

Then, by (E.5) in the proof of Lemma E.3 and Fubini's theorem,

$$\begin{aligned}
f_{n,m}(l,t) &= \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_n,S} S(\mathbf{X})_t^{j_1,\ldots,j_m,S}\right]\\
&= \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_n,S}\left(\int_0^t S(\mathbf{X})_s^{j_1,\ldots,j_{m-1},S}\mathrm{d}X_s^{j_m} + \frac{1}{2}\rho_{j_{m-1}j_m}\sigma_{j_{m-1}}\sigma_{j_m}\int_0^t S(\mathbf{X})_s^{j_1,\ldots,j_{m-2},S}\mathrm{d}s\right)\right]\\
&= g_{n,m}(l,t) + \frac{1}{2}\rho_{j_{m-1}j_m}\sigma_{j_{m-1}}\sigma_{j_m}\mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_n,S}\int_0^t S(\mathbf{X})_s^{j_1,\ldots,j_{m-2},S}\mathrm{d}s\right]\\
&= g_{n,m}(l,t) + \frac{1}{2}\rho_{j_{m-1}j_m}\sigma_{j_{m-1}}\sigma_{j_m}\int_0^t \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\ldots,i_n,S} S(\mathbf{X})_s^{j_1,\ldots,j_{m-2},S}\right]\mathrm{d}s\\
&= g_{n,m}(l,t) + \frac{1}{2}\rho_{j_{m-1}j_m}\sigma_{j_{m-1}}\sigma_{j_m}\int_0^t f_{n,m-2}(l,s)\mathrm{d}s.
\end{aligned}$$

This proves (B.1) and (B.5). In addition, by Itô isometry and Fubini's theorem,

$$\begin{aligned}
g_{n,m}(l,t) =& \mathbb{E}\left[S(\mathbf{X})_l^{i_1,\dots,i_n,S}\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{m-1},S}\mathrm{d}X_s^{j_m}\right]\\
=&\mathbb{E}\Bigg[\left(\int_0^l S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\mathrm{d}X_s^{i_n}+\frac{1}{2}\rho_{i_{n-1}i_n}\sigma_{i_{n-1}}\sigma_{i_n}\int_0^l S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}\mathrm{d}s\right)\\
&\qquad\qquad\cdot\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{m-1},S}\mathrm{d}X_s^{j_m}\Bigg]\\
=&\mathbb{E}\left[\int_0^l S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}\mathrm{d}X_s^{i_n}\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{m-1},S}\mathrm{d}X_s^{j_m}\right]\\
&\qquad+\frac{1}{2}\rho_{i_{n-1}i_n}\sigma_{i_{n-1}}\sigma_{i_n}\mathbb{E}\left[\int_0^l S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}\mathrm{d}s\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{m-1},S}\mathrm{d}X_s^{j_m}\right]\\
=&\rho_{i_nj_m}\sigma_{i_n}\sigma_{j_m}\int_0^{l\wedge t}\mathbb{E}\left[S(\mathbf{X})_s^{i_1,\dots,i_{n-1},S}S(\mathbf{X})_s^{j_1,\dots,j_{m-1},S}\right]\mathrm{d}s\\
&\qquad+\frac{1}{2}\rho_{i_{n-1}i_n}\sigma_{i_{n-1}}\sigma_{i_n}\int_0^l\mathbb{E}\left[S(\mathbf{X})_s^{i_1,\dots,i_{n-2},S}\int_0^t S(\mathbf{X})_u^{j_1,\dots,j_{m-1},S}\mathrm{d}X_u^{j_m}\right]\mathrm{d}s\\
=&\rho_{i_nj_m}\sigma_{i_n}\sigma_{j_m}\int_0^{l\wedge t}f_{n-1,m-1}(s,s)\mathrm{d}s+\frac{1}{2}\rho_{i_{n-1}i_n}\sigma_{i_{n-1}}\sigma_{i_n}\int_0^l g_{n-2,m}(s,t)\mathrm{d}s.
\end{aligned}$$

This proves (B.2) and (B.6).

Now we prove the initial conditions. First, (B.3) follows from the definition of 0-th order of signature. Second, (B.4) follows from the property of Itô integral

$$g_{0,2m}(l,t)=\mathbb{E}\left[\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{2m-1},S}\mathrm{d}X_s^{j_{2m}}\right]=0.$$

Third,

$$\begin{aligned}
f_{1,1}(l,t)&=\mathbb{E}\left[S(\mathbf{X})_l^{i_1,S}S(\mathbf{X})_t^{j_1,S}\right]=\mathbb{E}\left[\int_0^l 1\circ\mathrm{d}X_s^{i_1}\int_0^t 1\circ\mathrm{d}X_s^{j_1}\right]\\
&=\mathbb{E}\left[X_l^{i_1}X_t^{j_1}\right]=\rho_{i_1j_1}\sigma_{i_1}\sigma_{j_1}(l\wedge t),
\end{aligned}$$

which proves (B.7). Fourth, by Itô isometry,

$$\begin{aligned}
g_{1,2m-1}(l,t)&=\mathbb{E}\left[S(\mathbf{X})_l^{i_1,S}\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{2m-2},S}\mathrm{d}X_s^{j_{2m-1}}\right]\\
&=\mathbb{E}\left[\int_0^l 1\circ\mathrm{d}X_s^{i_1}\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{2m-2},S}\mathrm{d}X_s^{j_{2m-1}}\right]\\
&=\mathbb{E}\left[\int_0^l \mathrm{d}X_s^{i_1}\int_0^t S(\mathbf{X})_s^{j_1,\dots,j_{2m-2},S}\mathrm{d}X_s^{j_{2m-1}}\right]\\
&=\int_0^{l\wedge t}\mathbb{E}\left[S(\mathbf{X})_s^{j_1,\dots,j_{2m-2},S}\right]\rho_{i_1j_{2m-1}}\sigma_{i_1}\sigma_{j_{2m-1}}\mathrm{d}s\\
&=\rho_{i_1j_{2m-1}}\sigma_{i_1}\sigma_{j_{2m-1}}\int_0^{l\wedge t}f_{0,2m-2}(s,s)\mathrm{d}s.
\end{aligned}$$

In addition, by using (B.1) recursively, we can obtain that

$$f_{0,2m-2}(s,s) = \frac{1}{2^{m-1}} \frac{s^{m-1}}{(m-1)!} \prod_{k=1}^{m-1} \rho_{j_{2k-1}j_{2k}} \prod_{k=1}^{2m-2} \sigma_{j_k}.$$

Therefore,

$$g_{1,2m-1}(l,t) = \rho_{i_1 j_{2m-1}} \frac{1}{2^{m-1}} \frac{(l \wedge t)^{m-1}}{(m-1)!} \sigma_{i_1} \prod_{k=1}^{2m-1} \sigma_{j_k} \prod_{k=1}^{m-1} \rho_{j_{2k-1}j_{2k}},$$

which proves (B.8). □

Proof of Example B.5. The solution to stochastic differential equation (B.9) can be explicitly expressed as

$$Y_t = \int_0^t e^{-\kappa(t-s)} \mathrm{d}W_s, \quad t \geq 0,$$

where W_t is a standard Brownian motion. Therefore, by Itô isometry, Y_t is a Gaussian random variable with zero mean and

$$\mathrm{Var}(Y_t) = \mathbb{E}\left[Y_t^2\right] = \mathbb{E}\left[\int_0^t e^{-\kappa(t-s)} \mathrm{d}W_s\right]^2 = \int_0^t \left[e^{-\kappa(t-s)}\right]^2 \mathrm{d}s = \frac{1-e^{-2\kappa t}}{2\kappa}.$$

Now we calculate the correlation coefficient for its Itô and Stratonovich signature, respectively.

Itô Signature. By the definition of signature and (B.9),

$$\begin{aligned} \mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right] = \mathbb{E}\left[\int_0^T Y_t \mathrm{d}Y_t\right] &= -\kappa \mathbb{E}\left[\int_0^T Y_t^2 \mathrm{d}t\right] + \mathbb{E}\left[\int_0^T Y_t \mathrm{d}W_t\right] = -\kappa \int_0^T \mathbb{E}\left[Y_t^2\right] \mathrm{d}t \\ &= -\kappa \int_0^T \frac{1-e^{-2\kappa t}}{2\kappa} \mathrm{d}t = -\frac{T}{2} + \frac{1-e^{-2\kappa T}}{4\kappa}. \end{aligned} \tag{E.19}$$

For the second moment, by Itô isometry,

$$\begin{aligned} \mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right]^2 = \mathbb{E}\left[\int_0^T Y_t \mathrm{d}Y_t\right]^2 &= \mathbb{E}\left[-\kappa \int_0^T Y_t^2 \mathrm{d}t + \int_0^T Y_t \mathrm{d}W_t\right]^2 \\ &= \kappa^2 \int_0^T \int_0^T \mathbb{E}\left[Y_t^2 Y_s^2\right] \mathrm{d}t \mathrm{d}s - 2\kappa \mathbb{E}\left[\int_0^T Y_t^2 \mathrm{d}t \int_0^T Y_t \mathrm{d}W_t\right] + \int_0^T \mathbb{E}\left[Y_t^2\right] \mathrm{d}t \\ &=: \text{(a)} - \text{(b)} + \text{(c)}. \end{aligned}$$

It is easy to calculate Term (c):

$$\text{(c)} = \int_0^T \mathbb{E}\left[Y_t^2\right] \mathrm{d}t = \int_0^T \frac{1-e^{-2\kappa t}}{2\kappa} \mathrm{d}t = \frac{T}{2\kappa} + \frac{e^{-2\kappa T}-1}{4\kappa^2}. \tag{E.20}$$

To derive Term (a), we need to calculate $\mathbb{E}[Y_t^2 Y_s^2]$. Assume that $s < t$ and denote $M_t = \int_0^t e^{\kappa u} \mathrm{d}W_u$, we have $Y_t = e^{-\kappa t} M_t$, and therefore

$$\begin{aligned} \mathbb{E}\left[Y_t^2 Y_s^2\right] = e^{-2\kappa(t+s)} \mathbb{E}\left[M_t^2 M_s^2\right] &= e^{-2\kappa(t+s)} \mathbb{E}\left[(M_t - M_s + M_s)^2 M_s^2\right] \\ &= e^{-2\kappa(t+s)} \left[\mathbb{E}\left[(M_t - M_s)^2 M_s^2\right] + 2\mathbb{E}\left[(M_t - M_s) M_s^3\right] + \mathbb{E}\left[M_s^4\right]\right]. \end{aligned}$$

Because $M_t - M_s = \int_s^t e^{\kappa u} \mathrm{d}W_u$ is a Gaussian random variable with mean 0 and variance

$$\mathrm{Var}(M_t - M_s) = \mathbb{E}\left[(M_t - M_s)^2\right] = \mathbb{E}\left[\int_s^t e^{\kappa u} \mathrm{d}W_u\right]^2 = \int_s^t \left[e^{\kappa u}\right]^2 \mathrm{d}u = \frac{e^{2\kappa t} - e^{2\kappa s}}{2\kappa},$$

and M_t has independent increments, we have

$$\begin{aligned}
\mathbb{E}\left[Y_t^2 Y_s^2\right] &= e^{-2\kappa(t+s)}\left[\mathbb{E}\left[(M_t - M_s)^2\right]\mathbb{E}\left[M_s^2\right] + 2\mathbb{E}\left[M_t - M_s\right]\mathbb{E}\left[M_s^3\right] + \mathbb{E}\left[M_s^4\right]\right] \\
&= e^{-2\kappa(t+s)}\left[\frac{e^{2\kappa t} - e^{2\kappa s}}{2\kappa} \cdot \frac{e^{2\kappa s} - 1}{2\kappa} + 0 + 3\left(\frac{e^{2\kappa s} - 1}{2\kappa}\right)^2\right] \\
&= \frac{1 + 2e^{-2\kappa t + 2\kappa s} - e^{-2\kappa s} - 5e^{-2\kappa t} + 3e^{-2\kappa t - 2\kappa s}}{4\kappa^2}
\end{aligned}$$

when $s < t$. One can similarly write the corresponding formula for the case of $s > t$ and therefore

$$\begin{aligned}
\text{(a)} &= \kappa^2 \int_0^T \int_0^T \mathbb{E}\left[Y_t^2 Y_s^2\right] \mathrm{d}t \mathrm{d}s \\
&= \frac{1}{4}\left(T^2 + \frac{T}{\kappa} + \frac{10Te^{-2\kappa T}}{2\kappa} + \frac{3e^{-4\kappa T}}{4\kappa^2} - \frac{9}{4\kappa^2} + \frac{3e^{-2\kappa T}}{2\kappa^2}\right).
\end{aligned}$$

For Term (b), note that

$$2\kappa\mathbb{E}\left[\int_0^T Y_t^2 \mathrm{d}t \int_0^T Y_t \mathrm{d}W_t\right] = 2\kappa \int_0^T \mathbb{E}\left[Y_s^2 \int_0^T Y_t \mathrm{d}W_t\right] \mathrm{d}s,$$

By Itô's lemma,

$$\mathrm{d}Y_s^2 = 2Y_s \mathrm{d}Y_s + \mathrm{d}[Y, Y]_s = -2\kappa Y_s^2 \mathrm{d}s + 2Y_s \mathrm{d}W_s + \mathrm{d}s,$$

which implies that

$$Y_s^2 = -2\kappa \int_0^s Y_u^2 \mathrm{d}u + 2\int_0^s Y_u \mathrm{d}W_u + \int_0^s \mathrm{d}u.$$

Therefore, for $s < T$, with the help of Itô isometry and (E.20), we have

$$\begin{aligned}
f(s) &= \mathbb{E}\left[Y_s^2 \int_0^T Y_t \mathrm{d}W_t\right] \\
&= \mathbb{E}\left[\left(-2\kappa \int_0^s Y_u^2 \mathrm{d}u + 2\int_0^s Y_u \mathrm{d}W_u + \int_0^s \mathrm{d}u\right) \int_0^T Y_t \mathrm{d}W_t\right] \\
&= -2\kappa \int_0^s \mathbb{E}\left(Y_u^2 \int_0^T Y_t \mathrm{d}W_t\right) \mathrm{d}u + 2\int_0^s \mathbb{E}\left[Y_t^2\right] \mathrm{d}t + 0 \\
&= -2\kappa \int_0^s f(u) \mathrm{d}u + \frac{s}{\kappa} + \frac{e^{-2\kappa s} - 1}{2\kappa^2},
\end{aligned}$$

and taking derivatives of both sides leads to

$$\frac{\mathrm{d}f}{\mathrm{d}s} = -2\kappa f(s) + \frac{1}{\kappa} - \frac{e^{-2\kappa s}}{\kappa}.$$

Solving this ordinary differential equation with respect to f with initial condition $f(0) = 0$, we obtain that

$$f(s) = \frac{1}{2\kappa^2} - \frac{se^{-2\kappa s}}{\kappa} - \frac{e^{-2\kappa s}}{2\kappa^2}.$$

Therefore,

$$(\mathrm{b}) = 2\kappa\int_0^T f(s)\mathrm{d}s = \frac{T}{\kappa} + \frac{Te^{-2\kappa T}}{\kappa} + \frac{e^{-2\kappa T}-1}{\kappa^2}.$$

Finally,

$$\mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right]^2 = (\mathrm{a}) - (\mathrm{b}) + (\mathrm{c}) = \frac{Te^{-2\kappa T}}{4\kappa} + \frac{3e^{-4\kappa T}}{16\kappa^2} - \frac{3e^{-2\kappa T}}{8\kappa^2} - \frac{T}{4\kappa} + \frac{3}{16\kappa^2} + \frac{T^2}{4}. \tag{E.21}$$

Therefore,

$$\frac{\mathbb{E}\left[S(\mathbf{X})_T^{0,I}S(\mathbf{X})_T^{1,1,I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_T^{0,I}\right]^2\mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right]^2}} = \frac{\mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right]^2}}$$
$$= \frac{-2\kappa T - e^{-2\kappa T} + 1}{\sqrt{4\kappa Te^{-2\kappa T} + 3e^{-4\kappa T} - 6e^{-2\kappa T} - 4\kappa T + 3 + 4\kappa^2T^2}},$$

where the 0-th order of signature is defined as 1.

Stratonovich Signature. The Stratonovich integral and the Itô integral are related by

$$\int_0^t A_s \circ \mathrm{d}B_s = \int_0^t A_s\mathrm{d}B_s + \frac{1}{2}[A,B]_t.$$

Therefore,

$$S(\mathbf{X})_T^{1,S} = \int_0^T 1\circ\mathrm{d}Y_t = \int_0^T 1\mathrm{d}Y_t + \frac{1}{2}[1,Y]_T = \int_0^T 1\mathrm{d}Y_t = S(\mathbf{X})_T^{1,I} = Y_T,$$

and

$$S(\mathbf{X})_T^{1,1,S} = \int_0^T S(\mathbf{X})_T^{1,S}\circ\mathrm{d}Y_t = \int_0^T Y_t\circ\mathrm{d}Y_t = \int_0^T Y_t\mathrm{d}Y_t + \frac{1}{2}[Y,Y]_T = S(\mathbf{X})_T^{1,1,I} + \frac{T}{2},$$

where we use the fact that $[1,Y]_T = 0$ and $[Y,Y]_T = T$. Now by (E.19) and (E.21), we have

$$\mathbb{E}\left[S(\mathbf{X})_T^{1,1,S}\right] = \mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right] + \frac{T}{2} = \frac{1-e^{-2\kappa T}}{4\kappa},$$

and

$$\mathbb{E}\left[S(\mathbf{X})_T^{1,1,S}\right]^2 = \mathbb{E}\left[S(\mathbf{X})_T^{1,1,I} + \frac{T}{2}\right]^2$$
$$= \mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right]^2 + T\mathbb{E}\left[S(\mathbf{X})_T^{1,1,I}\right] + \frac{T^2}{4} = \frac{3(1-e^{-2\kappa T})^2}{16\kappa^2}.$$

Therefore,

$$\frac{\mathbb{E}\left[S(\mathbf{X})_T^{0,S}S(\mathbf{X})_T^{1,1,S}\right]}{\sqrt{\mathbb{E}\left[S(\mathbf{X})_T^{0,S}\right]^2\mathbb{E}\left[S(\mathbf{X})_T^{1,1,S}\right]^2}} = \frac{\sqrt{3}}{3}.$$

□

Proof of Proposition C.1. Let $a = \#A_1^*$ and $b = \#A_1^{*c}$. Under the equal inter-dimensional correlation assumption, we have $\Sigma_{A^*,A^*} = (1-\rho)I_a + \rho \mathbf{1}_a \mathbf{1}_a^\top$, where I_a is an $a \times a$ identity matrix and $\mathbf{1}_a$ is an a-dimensional all-one vector. In addition, $\Sigma_{A^{*c},A^*} = \rho \mathbf{1}_b \mathbf{1}_a^\top$, where $\mathbf{1}_b$ is a b-dimensional all-one vector. By the Sherman–Morrison formula,

$$\Sigma_{A^*,A^*}^{-1} = \frac{1}{1-\rho} I_a - \frac{\rho}{(1-\rho)(1+(a-1)\rho)} \mathbf{1}_a \mathbf{1}_a^\top.$$

Therefore, since all true beta coefficients are positive, we have

$$\Sigma_{A^{*c},A^*} \Sigma_{A^*,A^*}^{-1} \operatorname{sign}(\boldsymbol{\beta}_{A^*}) = \frac{a\rho}{1+(a-1)\rho} \mathbf{1}_a.$$

Hence, the irrepresentable condition

$$\left\| \Sigma_{A^{*c},A^*} \Sigma_{A^*,A^*}^{-1} \operatorname{sign}(\boldsymbol{\beta}_{A^*}) \right\| = \frac{a|\rho|}{1+(a-1)\rho} < 1$$

holds if and only if $\frac{a|\rho|}{1+(a-1)\rho} < 1$. One can easily verify that this holds if $\rho \in (-\frac{1}{2\#A_1^*}, 1)$, and does not hold if $\rho \in (-\frac{1}{\#A_1^*}, -\frac{1}{2\#A_1^*}]$. This completes the proof. □

Proof of Theorem C.1. For $\xi = \min\left\{ g_\Sigma^{-1}\left(\frac{\gamma}{\zeta(2+2\alpha\zeta+\gamma)}\right), g_\Sigma^{-1}\left(\frac{C_{\min}}{2\sqrt{p}}\right) \right\} > 0$, Lemmas E.4 and E.5 imply that

$$\mathbb{P}\left(\Lambda_{\min}(\hat{\Delta}_{A^*A^*}) \geq \frac{1}{2} C_{min}\right) \geq 1 - \frac{4p^4 \sigma_{\max}^4 (\sigma_{\min}^4 + K)}{N\xi^2 \sigma_{\min}^4},$$
$$\mathbb{P}\left(\left\| \hat{\Delta}_{A^{*c}A^*} \hat{\Delta}_{A^*A^*}^{-1} \right\|_\infty \leq 1 - \frac{\gamma}{2}\right) \geq 1 - \frac{4p^4 \sigma_{\max}^4 (\sigma_{\min}^4 + K)}{N\xi^2 \sigma_{\min}^4}.$$

Hence,

$$\mathbb{P}\left(\Lambda_{\min}(\hat{\Delta}_{A^*A^*}) \geq \frac{1}{2} C_{\min}, \left\| \hat{\Delta}_{A^{*c}A^*} \hat{\Delta}_{A^*A^*}^{-1} \right\|_\infty \leq 1 - \frac{\gamma}{2}\right) \geq 1 - \frac{8p^4 \sigma_{\max}^4 (\sigma_{\min}^4 + K)}{N\xi^2 \sigma_{\min}^4}. \tag{E.22}$$

Equation (E.22) gives the probability that the conditions of Wainwright (2009, Theorem 1) hold. Therefore, applying Wainwright (2009, Theorem 1) yields the result. □

References

Cai TT, Zhang AR, Zhou Y (2022) Sparse group Lasso: Optimal sample complexity, convergence rate, and statistical inference. *IEEE Transactions on Information Theory* 68(9):5975–6002.

Chevyrev I, Kormilitzin A (2016) A primer on the signature method in machine learning, arXiv preprint arXiv:1603.03788.

El Euch O, Fukasawa M, Rosenbaum M (2018) The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics* 22:241–280.

Gatheral J, Jaisson T, Rosenbaum M (2018) Volatility is rough. *Quantitative Finance* 18(6):933–949.

Lyons T, McLeod AD (2022) Signature methods in machine learning, arXiv preprint arXiv:2206.14674.

Ravikumar P, Wainwright MJ, Raskutti G, Yu B (2011) High-dimensional covariance estimation by minimizing l_1-penalized log-determinant divergence. *Electronic Journal of Statistics* 5:935–980.

Wainwright MJ (2009) Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1-constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory* 55(5):2183–2202.

Wüthrich K, Zhu Y (2023) Omitted variable bias of Lasso-based inference methods: A finite sample analysis. *The Review of Economics and Statistics* 105(4):982–997.